РОССИЙСКИЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЦЕНТР ИНФОРМАЦИИ ПО СТАНДАРТИЗАЦИИ, МЕТРОЛОГИИ И ОЦЕНКЕ СООТВЕТСТВИЯ

На правах рукописи

МОРОЗОВ Владимир Петрович

МЕТОДЫ, МОДЕЛИ И АЛГОРИТМЫ СИНТЕЗА ИНФОРМАЦИОННЫХ СИСТЕМ ПОДДЕРЖКИ ПОРТФЕЛЬНОЙ ИНВЕСТИЦИОННОЙ ДЕЯТЕЛЬНОСТИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ОРГАНИЗАЦИЙ

Специальность 05.25.05 – Информационные системы и процессы

Диссертация
на соискание ученой степени
доктора технических наук

Научный консультант: доктор технических наук, доцент
Мистров Леонид Евгеньевич

Москва – 2017 г.

СОДЕРЖАНИЕ

Стр.

Введение …… 7

1. Анализ методического обеспечения информационных систем и процессов портфельной инвестиционной деятельности социально-экономической организации. Постановка проблемы…… 24

1.1. Информационный процесс принятия решений в портфельной инвестиционной деятельности социально-экономической организации…….. 24

1.2. Определение информационной системы поддержки портфельной инвестиционной деятельности социально-экономической организации……... 36

1.3. Анализ методического обеспечения информационных процессов и систем портфельной инвестиционной деятельности социально-экономической организации……... 39

1.4. Постановка проблемы…… 47

1.5. Вероятностная модель учета влияния неопределенности внешней среды при синтезе информационной системы поддержки портфельной инвестиционной деятельности……. 63

1.6. Требования, предъявляемые к информационной системе поддержки портфельной инвестиционной деятельности социально-экономической организации…… 68

1.7. Критерии оптимизации, постановка и решение задачи структурно-функционального синтеза систем поддержки портфельной инвестиционной деятельности социально-экономической организации…….. 71

1.8. Технологии проведения исследований и внешнесистемного синтеза системы…… 82

ВЫВОДЫ……... 94

2. Модели, методики и алгоритмы информационного процесса формирования инвестиционного портфеля социально-экономической организации……... 97
2.1. Основные модели портфельного инвестиционного анализа……. 97
2.2. Анализ информационных процессов модели Марковица…….. 99
2.2.1. Постановка задачи определения долей ценных бумаг в оптимальном портфеле Марковица……..... 106
2.2.2. Обобщенный алгоритм определения долей ценных бумаг в оптимальном портфеле Марковица…… 108
2.2.3. Программная реализация определения долей ценных бумаг в оптимальном портфеле Марковица……. 110
2.3. Анализ информационных процессов модели Шарпа……. 112
2.3.1. Постановка задачи определения долей ценных бумаг в оптимальном портфеле Шарпа…… 112
2.3.2. Алгоритм определения долей ценных бумаг в оптимальном портфеле Шарпа…… 116
2.3.3. Программная реализация определения долей ценных бумаг в оптимальном портфеле Шарпа…… 117
2.4. Информационные процессы нейромодифицированной модели Шарпа……. 119
2.5. Разработка нейрокомитетной модели Шарпа……. 122
2.6. Модифицированный генетический алгоритм распределения инвестиций……. 126
2.6.1. Разработка островной модели параллельных вычислений……. 137
2.6.2. Программная реализация модифицированного генетического алгоритма распределения инвестиций…… 141
ВЫВОДЫ……... 145
3. Методы, модели и алгоритмы прогнозирования значений активов в инвестиционном процессе формирования инвестиционного портфеля социально-экономической организации…… 147

3.1 Характеристика программного комплекса нейросетевого прогнозирования временных рядов……... 148
3.1.1 Теоретические аспекты, алгоритм и программная реализация многослойного персептрона……..... 149
3.1.2 Теоретические аспекты и программная реализация алгоритма обратного распространения ошибок……... 159
3.1.3 Теоретические аспекты и программная реализация алгоритма имитации отжига……... 171
3.1.4 Программная реализация генетического алгоритма…… 176
3.1.5 Теоретические аспекты и алгоритмы адаптации архитектуры персептрона……..... 176
3.1.6 Реализация эвристической оптимизации ИНС……..... 181
3.1.7 Реализация управления и визуализации многослойного персептрона ……..... 186
3.2 Порядок использования и результаты апробации программного комплекса нейросетевого прогнозирования временных рядов……... 189
ВЫВОДЫ……... 195
4. Методы, модели и алгоритмы терминологического информационного поиска прагматической информации в рамках портфельной инвестиционной деятельности социально-экономической организации…….. 197
4.1 Модель формализации текстовой информации……... 197
4.2 Математическая модель весовой распределенной информационной системы на тезаурусе……. 198
4.3 Математическая модель весовой обработки информации на терминологическом портрете……..... 205
4.4 Синтез структуры распределенной информационной системы…. 207
4.5 Алгоритм идентификации текстов…… 212
4.6 Характеристика иерархического терминопостроителя……... 216
4.7 Семантическая матрично-лексическая модель анализа текста….. 217

4.8 Индексный метод извлечения новых знаний……... 224
4.9. Программная реализация терминологического информационного поиска в информационной системе поддержки портфельной инвестиционной деятельности……. 227
ВЫВОДЫ……... 236
5. Модели и алгоритмы хранения и обработки данных портфельной инвестиционной деятельности социально-экономической организации …….……... 238
5.1 Состав и структура подсистемы ввода и хранения данных…… 238
5.2 Алгоритм эвристической оптимизации хранилища данных…….. 241
5.3 Особенности реализации загрузочной секции……. 245
5.4 Контроль качества и оптимизация хранения актуальных данных. 247
5.5 Реализация архивации данных……... 253
5.6 Реализация резервного копирования……. 257
5.7 Реализация ввода данных……... 272
ВЫВОДЫ……... 276
6. Методическое обеспечение внутрисистемного синтеза, оценка эффективности и развитие информационной системы поддержки портфельной инвестиционной деятельности социально-экономической организации…… 278
6.1 Характеристика прототипа информационной системы поддержки портфельной инвестиционной деятельности ……... 278
6.2 Методика интегральной оценки эффективности информационной системы поддержки портфельной инвестиционной деятельности в составе социально-экономической организации ……. 284
6.3 Методика и результаты оценки эффективности функционирования подсистемы формирования инвестиционного портфеля социально-экономической организации …… 287
6.4 Методика оценки эффективности функционирования подсистемы терминологического информационного поиска ……. 296

6.5 Предложения по содержанию внутрисистемного синтеза и перспективного развития информационных систем поддержки портфельной инвестиционной деятельности…… 302
ВЫВОДЫ…….. 316
Заключение …….. 321
Список сокращений……. 325
Список литературы…… 328
Приложение А. Копии актов внедрения, регистрации программ и сертификатов ……. 356
Приложение Б. Результаты сравнительной оценки основных моделей портфельного инвестиционного анализа…… 373
Приложение В. Характеристика современного методического обеспечения портфельной инвестиционной деятельности социально-экономической организации……... 375
Приложение Г. Характеристика частных информационных систем, программных решений и систем поддержки принятия решений в портфельной инвестиционной деятельности социально-экономической организации…… 381
Приложение Д. Характеристика зарубежных и отечественных информационных систем и программ для работы на рынке ценных бумаг.. 383
Приложение Е. Результаты сравнительного анализа основных моделей формализации текстовой информации…… 387

ВВЕДЕНИЕ

Актуальность проблемы.

Значимость применения информационных систем (ИС), включающих техническое, программное и организационное обеспечение, а также персонал для информационных потребителей – лиц, принимающих решения (ЛПР) в рамках портфельной инвестиционной деятельности (ПИД) социально-экономических организаций (СЭО) на рынке ценных бумаг (РЦБ), трудно переоценить. Поскольку, несмотря на то, что, во-первых, РЦБ является важным элементом в общей экономической системе развития государства, предназначенным для стабилизации денежных потоков в экономике, аккумулирования и перераспределения финансовых ресурсов; формирования и движения рыночных цен активов; развития конкуренции среди финансовых услуг; стимулирования финансовых вложений в развитие среднего и верхнего уровня функциональной деятельности СЭО – основы экономики, во-вторых, создает объективные предпосылки доступа СЭО к более дешевым (например, по сравнению с банковским кредитованием) финансовым ресурсам и способствует активизации инвестиционного процесса, стратегии реализации которого определены в решении Правительства России от 17.11.2008 № 1662-р (ред. от 08.08.2009) «О Концепции долгосрочного социально-экономического развития Российской Федерации до 2020 года» [26], тем не менее, процесс информационного взаимодействия с ним (РЦБ) для ведения успешной ПИД требует от ЛПР принятия точных и оперативных решений.

В виду сложности ПИД, содержанием которой является совокупность управленческих решений и действий органов управления СЭО для оптимизации вложения финансовых инвестиций в активы РЦБ в форме инвестиционного портфеля (ИП) с целью получения прибыли в виде процентов (дивидендов) и целесообразности ее оперативного проведения, применение соответствующей информационной технологии вообще, и ИС, как ее разновидности, в частности, необходимо. Усиление необходимости применения ИС обусловлено тем фактом, что одной из центральных проблем ПИД СЭО является

формирование такого ИП, состав активов которого, обеспечивал бы максимальный уровень его доходности при минимальном/заданном уровнях возможного риска. Формирование ИП представляет сложный и неоднозначный информационный процесс (ИФП), содержащий этапы: анализа РЦБ в интересах выбора актуальных и перспективных направлений вложения инвестиций, определения конкретных ценных бумаг (ЦБ) и долей вкладываемого в них капитала; микропрогнозирования (анализа динамики цен отдельных видов ЦБ) и макропрогнозирования (анализа изменения уровня цен акций относительно ЦБ с фиксированным доходом, например корпоративных облигаций), проведения диверсификации (выбор ЦБ с низкой взаимной корреляцией и др.), расчета значений доходности и риска сформированного ИП, корректировки состава и структуры ИП с учетом изменений факторов внешней и внутренней среды. Проблемный характер формирования ИП обусловлен не структурированным характером его представления как объекта информационного управления и связанной с этим сложностью аналитического представления, вследствие нестационарного вида ИФП его развития, нелинейной зависимости ИП от множества факторов, в том числе и стохастических (внешней среды [90], иррационального поведения инвесторов, непредсказуемых структурных изменений и др.), а также последействием принимаемых управленческих решений.

Анализ существующего опыта разработки ИС в интересах ПИД СЭО, представленного в работах Д. Галиева, Е. Герасимова, Ю. Куреленкововой, О. Рекундаль, А. Тихненко, А. Шапошниковой, Н. Шолоховой, В. Цибульниковой и др. выявил их следующие особенности: 1) ИС реализуют лишь фрагментарную информационную поддержку некоторых частных этапов ИФП формирования ИП (например, у Д Галиева – модели альтернативных мер риска и доходности, Ю. Куреленковой – модель учета комплексных мер риска, О. Рекундаль – методика весовой оценки инвестиционной привлекательности активов и др.) и не охватывают весь ИФП; 2) заявленные системы информационной поддержки принятия решений не охватывают полный спектр

вопросов формирования ИП; 3) вопросы предоставления инвесторам СЭО объективной прагматической информации, повышающей уровень их компетенции в них не нашли отражения; 4) ИС обеспечивают максимальную точность, достоверность и оперативность проведения расчетов на частных этапах ИФП для детерминированных исходных данных. Для стохастических же данных, обусловленных неопределенностью изменения политико-экономической обстановки эффективность таких ИС недостаточна для принятия оптимальных решений, вследствие необходимости детального учета на этапах ИФП более сложных механизмов влияния неопределенности на доходность и риск ИП [129]. Данный факт подтверждается исследованиями Е. Герасимова, утверждающего, что методы и сложность решения задач формирования оптимального ИП СЭО для детерминированных и стохастических исходных данных, а также статических и динамических постановок задач принципиально отличаются. В первом случае задача информационной поддержки выбора оптимальной структуры ИП основывается на решении задач нелинейного, стохастического или линейного программирования, как функции риска и способов учета информационной неопределенности. Во втором случае, когда динамика ИП описывается в агрегированном виде на основе уравнения капитала портфеля, а в качестве управляющих воздействий используются доли вложений в соответствующий актив, поиск стратегии оптимизации ИП приводит к NP-сложной проблеме численного решения системы уравнений на основе метода динамического программирования.

Проведенный анализ работ по проблеме оптимизации ИФП формирования ИП показывает, что они в основном направлены на поиск решения на одном из его этапов, не охватывая весь ИФП, для детерминированных исходных данных и могут использоваться для получения ответов только на некоторые частные вопросы. В целом же решение данной проблемы остается открытым, что обусловливает наличие объективно существующего ***противоречия*** между практической необходимостью реализации инвестором СЭО информационного процесса формирования оптимального ИП для детерми-

нированных и стохастических исходных данных и фактическим отсутствием методического обеспечения и инструментальных средств для его информационной поддержки.

Частичная информационная поддержка ИФП формирования ИП путем разработки методического обеспечения отдельных фрагментов некоторых его этапов, общую проблему не решает, поскольку не охватывает данный ИФП в целом, начиная от этапа сбора, обработки, хранения и представления стохастических данных об активах с целью выбора среди них (активов) наиболее эффективных для включения в ИП, и заканчивая расчетом характеристик ИП (доходности и риска) и их мониторинга в ходе информационного управления.

Необходимость учета стохастических данных об активах в формирования оптимального ИП и информационного управления им в условиях оперативного изменения обстановки на РЦБ, актуализирует проблему информационной поддержки данного ИФП, которую можно разрешить на основе разработки информационной системы поддержки портфельной инвестиционной деятельности (ИСППИД). Структурно, она представляет собой совокупность объединенных единством цели методов, моделей, алгоритмов и программ поиска, анализа, обобщения, обработки, хранения и предоставления инвестору СЭО прагматической информации о выборе целесообразных ЦБ, а также рациональном составе и структуре активов ИП, обеспечивающих максимизацию его доходности с учетом заданного риска, требуемой точности, полноты, достоверности и оперативности проведения расчетов.

Теоретическое обоснование целей, задач, состава, структуры, характеристик, особенностей функционирования и практического использования ИСППИД базируется на ее облике, включающем состав и структуру используемого методического обеспечения, формируемом в рамках соответствующей методологии проведения исследований [189].

Объект исследования – информационные системы и процессы формирования ИП СЭО применительно к стохастическим данным об активах.

Предмет исследования – методы, модели и алгоритмы информационных процессов сбора, анализа, обобщения, обработки, хранения и формирования оптимальных ИП СЭО применительно к стохастическим данным об активах.

Цель работы – разработка (развитие) методов, моделей и алгоритмов синтеза ИСППИД для повышения эффективности информационной поддержки портфельной инвестиционной деятельности СЭО при принятии решений формирования оптимального ИП в условиях стохастических данных об активах.

Достижение цели требует решения общей проблемы, декомпозируемой на ряд частных ***научных задач***:

- разработки технологии внешнесистемного синтеза структурно-функционального облика ИСППИД;

- разработки методического обеспечения автоматизации определения параметров исследуемых финансовых активов и формирования ИП СЭО;

- формирования методического обеспечения прогнозирования параметров активов ИП СЭО для восстановления неизвестных дискрет в стохастических данных, используемых финансовых инструментов;

- разработки методического обеспечения поиска прагматической информации о качественных и количественных характеристиках исследуемых активов, а также перспективном методическом обеспечении информационного управления ПИД для оптимизации ИП СЭО;

- разработки алгоритмического обеспечения сбора, обработки и хранения разнородных данных, необходимых для реализации вычислений на различных этапах ИФП формирования ИП в ПИД СЭО;

- разработки методического обеспечения внутрисистемного синтеза ИСППИД, оценки эффективности функционирования реализованного прототипа, включающего определение значений характеристик точности, полноты, достоверности и оперативности информации предоставляемой инвестору в процессе ПИД СЭО, а также обоснования решений по его перспективному

развитию с учетом особенностей функционирования.

В основе используемых **методов исследования** лежат положения теорий информационных систем и процессов, принятия решений [199], системного анализа [12, 61, 191, 192, 202, 205], графов, иерархических многоуровневых систем [76], а также аналитический аппарат булевой алгебры, информационного, математического, нейросетевого и эволюционного моделирования, логико-лингвистического анализа, управления проектами, дискретной и непрерывной оптимизации, объектно-ориентированного и дихотомического программирования.

Основными научными результатами являются:

- методы и модели синтеза структурно-функционального облика ИСППИД для обоснования оптимального состава используемого методического обеспечения подсистем формирования ИП, прогнозирования значений параметров ЦБ, терминологического информационного поиска, а также сбора, хранения и обработки информации, обеспечивающего функционирование системы с заданными показателями качества для стохастических данных об активах РЦБ в ПИД СЭО [144];

- методические основы информационной поддержки формирования ИП СЭО, содержащие классические (модель Г. Марковица и обобщенный алгоритм ее реализации, одноиндексная модель У. Шарпа) и модифицированные (нейромодифицированная и нейрокомитетная модели У. Шарпа, модифицированная эволюционная модель распределения инвестиций) модели, позволяющие реализовать оптимизацию ИП как для детерминированных, так и для стохастических данных об активах;

- система алгоритмов, обеспечивающая построение, структурно-функциональную оптимизацию и обучение искусственных нейронных сетей (ИНС) для прогнозирования неизвестных дискрет в стохастических данных об активах в процессе ПИД СЭО;

- методическое обеспечение поиска прагматической информации о качественно-количественных характеристиках исследуемых активов и перспек-

тивных методах информационного управления формированием и оптимизацией ИП в процессе ПИД СЭО для стохастических данных об активах, базирующихся на обработке понятий формируемых терминологических портретов текстовой информации [182];

- методы и алгоритмы сбора, хранения и обработки разнородных многомерных данных больших размерностей, основанных на концепциях построения хранилищ и витрин данных, учитывающих предысторию использования информации в ПИД СЭО;

- методики оценки эффективности функционирования основных подсистем (формирования ИП и терминологического поиска) и ИСППИД в ПИД СЭО в целом, базирующиеся на сравнительном численном анализе значений эксплуатационных характеристик (точности, полноты, достоверности и оперативности) для ручного и автоматизированного способов формирования ИП СЭО [180, 183];

- решения по внутрисистемному синтезу и развитию ИСППИД, основанные на точностных и временных показателях разработанного прототипа, учитывающие закономерности изменения эффективности формируемых ИП в ПИД СЭО от точности, полноты, достоверности и оперативности информации предоставляемой ИСППИД СЭО в условиях стохастических данных об активах.

Научная новизна результатов

1. Теоретические основы синтеза структурно-функционального облика ИСППИД, обеспечивающей информационную поддержку ПИД СЭО в составе подсистем формирования ИП, прогнозирования значений параметров ЦБ, терминологического информационного поиска, сбора, хранения и обработки информации, содержащие подходы, методы и модели, отличающиеся от ранее известных, обеспечением функционирования системы с заданными показателями качества применительно к стохастическим данным об активах.

2. Нейрокомитетная модель У. Шарпа и модифицированный генетический алгоритм распределения инвестиций в ИП в процессе ПИД СЭО, обес-

печивающих обработку активов со стохастическими данными, отличающиеся в первом случае, способом учета предыстории изменения доходностей активов при прогнозировании их значений в текущей ситуации в виде риск-эффектов с использованием аппарата искусственных нейронных сетей (ИНС), встроенной в одноиндексную оптимизационную модель У. Шарпа, во-втором, способом прогнозирования, реализованном в эволюционной модели оптимизации, позволяющем осуществить параллельную обработку информации.

3. Модифицированные методы и алгоритмы анализа, синтеза и структурно-функциональной оптимизации и обучения ИНС, впервые используемые для прогнозирования значений неизвестных дискрет в стохастических данных об активах в ПИД СЭО, отличающиеся многокритериальным управлением ИФП обучения и оптимизации отдельных его этапов, применением визуализации исследуемых структур ИНС, для повышения оперативности обучения ИНС и точности, получаемых на их основе, результатов, а также исследования их новых видов структур.

4. Впервые разработанные методы, модели и алгоритмы терминологического информационного поиска (ТИП) для получения прагматической конфиденциальной информации о качественно-количественных характеристиках исследуемых активов, перспективном методическом обеспечении информационного управления формированием эффективных ИП в процессе ПИД СЭО для стохастических данных об активах, учитывающие: а) неопределенности в описании формализуемых запросов и содержании найденных документов; б) перераспределение поисковой нагрузки в локальных информационных подсистемах в зависимости от значимости искомых терминов; в) весовую параметрическую оценку понятий, сравниваемых терминологических портретов предметной области и исходных текстов, обеспечивающую селекцию понятий семантически связанных в рассматриваемой области определения [158]; г) извлечение дополнительных понятий из семантических матрично-лексических сетей текстов на основе запоминания

терминологических текстовых индексов взаимосвязей в виде логических цепочек.

5. Методы и алгоритмы сбора, хранения и обработки информации, обеспечивающие реализацию вычислений на различных этапах ИФП формирования ИП в ПИД СЭО, отличающиеся от известных учетом разнородности, многомерности и предыстории использования данных, реализованных на концепциях построения хранилищ и витрин данных, а также основанные на механизме обмена данными между хранилищем и витриной в зависимости от их ранговой популярности и эмпирической оценки вероятности поступления запросов в виде потока событий, подчиняющегося закону Пуассона.

6. Методики оценки эффективности функционирования подсистем формирования ИП и терминологического информационного поиска, а также ИСППИД в целом, базирующиеся на сравнительном численном анализе значений эксплуатационных характеристик для ручного и автоматизированного способов формирования ИП в ПИД СЭО, которые в отличие от ранее известных учитывают интегральные характеристики эффективности функционирования подсистем и процедуры автоматизированного формирования тестовых баз данных, полученные на основе разработанного прототипа.

7. Решения по внутрисистемному синтезу ИСППИД, основанные на концептуальных положениях и технологических аспектах формирования структурно-функционального облика прототипа, отличающиеся реализацией мультиэтапного параллельного процесса итеративного согласования синтезируемых решений.

Исследования, проведенные в работе, соответствуют пп. 1, 3, 4 и 7 паспорта специальности 05.25.05.

Работа выполнена в рамках федеральной комплексной программы «Исследования и разработки по приоритетным направлениям науки и техники гражданского назначения» и проекта №10-07-00463 «Разработка математических моделей, синтез методов и алгоритмов при управлении бизнесс-процессами в системах организационного управления».

Ряд методов, моделей и алгоритмов, характеризующихся научной новизной, реализованы в программных продуктах, зарегистрированных в государственном фонде алгоритмов и программ, а также центре информационных технологий и систем органов исполнительной власти.

Теоретическая значимость полученных результатов заключается в:

- разработке методов и моделей синтеза структурно-функционального облика ИСППИД, обеспечивающей информационную поддержку ПИД СЭО, отличающихся от известных, обеспечением функционирования системы с заданными показателями качества в условиях стохастических данных об активах РЦБ;

- разработке методов, моделей и алгоритмов оптимизации ИФП формирования ИП в процессе ПИД СЭО - инвесторов РЦБ России в условиях стохастических данных об исследуемых активах;

- разработке способа системной информированности инвестора о состоянии фондового рынка и его активах на основе ИСППИД в процессе ПИД СЭО, отличающегося от известных (например, способа ассиметричной информированности, предложенного В.Н. Бурковым в рамках теории активных систем) отсутствием манипулирования;

- разработке нейрокомитетной модели У. Шарпа и модифицированного генетического алгоритма распределения инвестиций в ИП в процессе ПИД СЭО, позволяющих в отличие от существующих, обрабатывать активы РЦБ со стохастическими данными;

- разработке положений количественного учета влияния компетентности ЛПР на эффективность формирования ИП на основе модели снижения неопределенности внешней среды при формировании ИП с использованием ИСППИД;

- развитии теоретических аспектов ТИП в части определения и извлечения дополнительных знаний на основе индексного метода, обеспечивающего извлечение дополнительных понятий из семантических матрично-лексических сетей текстов путем запоминания терминологических текстовых

индексов взаимосвязей в виде логических цепочек;

- обосновании решений внутрисистемного синтеза ИСППИД в процессе ПИД СЭО, учитывающих мультиэтапный параллельный ИФП итеративного согласования синтезируемых решений.

Достоверность научных результатов

1. Достоверность развиваемых теоретических положений структурно-функционального синтеза ИСППИД СЭО основывается на применении классических исходных постановок задач, проверенных практикой, ясным физическим смыслом используемых трактовок, строгостью математического представления формализованных постановок задач в рамках единой структуры методов, моделей и алгоритмов их (модификаций) реализации, использованием системного подхода в интересах решения научной проблемы оптимизации ИФП формирования ИП [229, 231, 235], решения частных научных задач с использованием обобщенных положений известных теорий, учетом наиболее значимых факторов, влияющих на синтезируемые результаты и проведением вычислительных экспериментов, подтверждающих совпадение в частных случаях значений промежуточных и конечных результатов с известными.

2. Достоверность сформированного облика ИСППИД СЭО основывается на результатах опытной эксплуатации ее отдельных подсистем и системы в целом.

3. Достоверность результатов количественного оценивания эффективности функционирования ИСППИД в процессе ПИД СЭО подтверждается результатами ее апробации и настройки на исторических данных, используемых в качестве эталонных.

Положения, выносимые на защиту

1. Теоретические положения внешнесистемного структурно-функционального синтеза облика ИСППИД СЭО, содержащие: модель снижения неопределенности внешней среды при формировании ИП, отличающуюся учетом влияния уровня компетентности ЛПР на основные показатели

формируемого ИП; стохастическую модель учета влияния неопределенности внешней среды на состав ИСППИД, основанную на увеличении числа дополнительных операций для достижения поставленной цели и позволяющую обосновать оптимальный по Парето ее набор компонент по критериям выделяемого финансового ресурса и достижимой точности; модель определения состава ИСППИД, отличающуюся установлением Парето-оптимальных комплектов подсистем и модулей, путем многокритериального оценивания их исходного инвариантного множества на основе метода дихотомического программирования; метод оптимизации функциональной структуры ИСППИД, базирующийся на графо-аналитической оценке и весовой обработке взаимосвязей между структурными элементами дерева целей, отличающийся построением и обработкой матриц весовых коэффициентов «цели-средства-данные», в отличие от традиционных «цели-средства», что позволяет построить адаптивную структуру системы, функционирующую на основе реконфигурации используемых моделей, методов и алгоритмов в зависимости от вида обрабатываемых данных.

2. Система решений формирования ИП для стохастических данных об активах РЦБ, включающая: а) нейрокомитетную модель У. Шарпа аналитического учета предыстории изменения доходностей активов при прогнозировании их значений в текущей ситуации в виде риск-эффектов с использованием ИНС для получения более точных результатов прогноза; б) модифицированный генетический алгоритм, отличающийся: учетом построения перспективных поддеревьев решений, применением стратегии поиска решений «снизу-вверх» и использованием островной модели параллельных вычислений для повышения оперативности проведения расчетов за счет параллельной обработки информации [164].

3. Модифицированные методы и алгоритмы анализа, синтеза и структурно-функциональной оптимизации и обучения ИНС, используемые для прогнозирования значений активов, отличающиеся реализацией механизма запоминания «риск-эффектов» базирующегося на теореме об универсальной

аппроксимации, введением многокритериального управления ИФП обучения и оптимизации отдельных его этапов, применением визуализации исследуемых структур ИНС, позволяющие повысить оперативность обучения и корректность настройки ИНС и, в конечном счете, точность, получаемых на их основе, результатов, а также исследование новых видов ИНС.

4. Методы, модели и алгоритмы ТИП прагматической информации о качественных и количественных характеристиках исследуемых активов, а также перспективном методическом обеспечении (под методическим обеспечением понимается совокупность методов, моделей и алгоритмов, объединенных в рамках единого целевого назначения (например, решения задачи, обеспечения функционирования подсистемы и др.) информационного управления формированием и оптимизацией ИП в ПИД СЭО для стохастического изменения данных об активах, включающее математическую модель весовой обработки информации на терминологическом портрете, алгоритмы графематического, морфологического, синтаксического анализа и идентификации текстов, семантическую матрично-лексическую модель, иерархический терминопостроитель и метод извлечения новых знаний, учитывающее: а) неопределенности в описании формализуемых запросов и содержании найденных документов; б) перераспределение поисковой нагрузки в локальных информационных подсистемах в зависимости от значимости искомых терминов; в) весовую параметрическую оценку понятий, сравниваемых терминологических портретов предметной области и исходных текстов, обеспечивающую селекцию понятий семантически связанных в рассматриваемой предметной области; г) извлечение дополнительных понятий из семантических матрично-лексических сетей текстов на основе запоминания терминологических текстовых индексов взаимосвязей в виде логических цепочек.

5. Модифицированный алгоритм индексирования данных, содержащихся в хранилище и алгоритм оптимизации доступа, включенные в состав подсистемы сбора, хранения и обработки информации, отличающиеся использованием составных индексов, присваиваемых множеству близких

непересекающихся данных в реляционных таблицах представленных обобщенными индексными деревьями, определяемых функциями узловых делений, дополненных процедурами выделения областей пересечения записей и организацией обмена данными между хранилищем и витриной на основе их (данных) ранговой популярности и эмпирической оценки вероятности поступления запросов в виде Пуассоновского потока событий, учитывающего предысторию обрабатываемых данных, позволяющие реализовать более оперативную обработку запросов.

6. Методики оценки эффективности функционирования основных подсистем (формирования ИП, прогнозирования и терминологического поиска) и ИСППИД СЭО в целом, базирующиеся на сравнительном численном анализе значений эксплуатационных характеристик (точности, полноты, достоверности и оперативности) для ручного и автоматизированного способов формирования ИП СЭО, отличающиеся учетом интегральных характеристик оценки эффективности функционирования подсистем и автоматизированным формированием тестовых баз данных, позволяющие повысить полноту и объективность оценки качества функционирования системы.

7. Решения внутрисистемного синтеза системы, основанные на концептуальных положениях и технологических аспектах формирования структурно-функционального облика ИСППИД для информационной поддержки ПИД СЭО, отличающиеся реализацией мультиэтапного параллельного ИФП итеративного согласования принимаемых решений.

Практическая значимость работы заключается в внутрисистемном синтезе в соответствии с предложенным методическим обеспечением прототипа ИСППИД СЭО, реализующего модели, методы и алгоритмы информационного управления формированием ИП, прогнозирования значений характеристик ЦБ, ТИП, а также сбора, хранения и обработки информации, обеспечивающие повышение точности и оперативности принятия инвестиционных решений, а также поиск прагматической конфиденциальной информации для инвестора СЭО. Разработанный прототип ИСППИД СЭО на конкретных

примерах показал высокую эффективность распределения заданного объема финансовых инвестиций по исследуемым финансовым инструментам (погрешность не превышает 1,5 %), что подтверждено результатами численного исследования и соответствующими актами внедрения. Предложенные модели и алгоритмы прогнозирования и терминологического поиска носят универсальный характер и могут использоваться в информационных системах, различающихся предметной областью, степенью сложности и характером решаемых задач.

Реализация результатов работы. Методы, модели и алгоритмы синтеза информационных систем поддержки портфельной инвестиционной деятельности социально-экономических организаций реализованы:

1) в части использования математических моделей весовых распределенных информационных систем на тезаурусе и терминологическом портрете, а также алгоритма обмена данными на основе эмпирической оценки их ранговой популярности, при разработке федеральной государственной информационно-аналитической системы по управлению использованием радиочастотного спектра в ОАО «НВП «Протек»;

2) в части применения методов, моделей и алгоритмов подсистем формирования ИП и прогнозирования значений параметров ценных бумаг в инвестиционной деятельности организаций: ЗАО «Воронеж-Дом»; ООО УК «Жилпроект»; ООО «Воронежский фондовый интернет центр»; ООО «Декарт»;

3) в части использования методического обеспечения терминологического информационного поиска в документальной информационно-поисковой системе «Научные труды» в ФГБОУ ВО «ВГТУ», в интересах обучения молодых специалистов в ООО УК «Жилпроект» и в деятельности организаций ЗАО «Воронеж-Дом» и ООО «Декарт»;

4) в части применения методов и алгоритмов сбора, хранения и обработки информации в интересах совершенствования корпоративных баз данных в организациях: ЗАО «Воронеж-Дом»; ООО «Декарт»; ООО «Воронежский фондовый интернет центр»;

5) в части опытной эксплуатации прототипа ИСППИД СЭО в ООО УК «Жилпроект» и ФГБОУ ВО «ВГТУ»;

6) в части внедрения в образовательный процесс в ФГБОУ ВО «ВГТУ» в рамках дисциплин «Интеллектуальные технологии и представление знаний» и «Инвестиционный менеджмент» и при разработке материалов на регистрацию программных продуктов, а также в Центральном филиале ФГБОУ ВО «РГУП» при разработке учебно-методических материалов по дисциплинам: «Экономика», Информационные технологии в юридической деятельности», «Компьютерные технологии в экспертной деятельности».

Факты реализации подтверждены актами внедрения, приведенными в приложении А.

Апробация работы. Основные результаты исследований и научных разработок докладывались и получили положительную оценку на 27 международных и всероссийских научно-практических конференциях, научных сессиях, школах и форумах, в том числе: VI межвузовской научной конференции «Актуальные проблемы науки в студенческих исследованиях» (Воронеж, 2005); международной научно - практической конференции «Проблемы развития экономической системы России в XXI в.» (Воронеж, 2008); международной научно - практической конференции «Развитие экономической науки России в XXI в.» (Воронеж, 2009); материалах научной сессии преподавателей ВЭПИ-2010 (Воронеж, 2010); международной научно - практической конференции «Развитие инноваций в России» (Воронеж, 2010); всероссийской научно-практической конференции «Инновационные механизмы экономики и педагогики в современной России: теория и практика» (Воронеж, 2011); Всероссийской конференции с элементами научной школы «Математическое моделирование в технике и технологии» (Воронеж, 2011); международной молодежной конференции «Математические проблемы современной теории управления системами и процессами» (Воронеж, 2012); международной научно – практической конференции «Перспективы и проблемы инновационного развития социально-экономических систем» (Воронеж,

2012); международной молодежной научной школе «Летняя Суперкомпьютерная Академия» (Воронеж, 2012); международных научно - технических конференциях «Современные сложные системы управления. HTCS» (Ст. Оскол, 2012, Воронеж, 2013, 2014); IX Всероссийской школе-конференции молодых ученых «Управление большими системами» (Тамбов – Липецк, 2012); XXI и XXII всероссийских научных конференциях «Информатизация и информационная безопасность правоохранительных органов» (Москва, 2012, 2013); Proceedings of the XVIII-th International Open Science Conference «Modern informatization problems in economics and safety» (USA, 2013); международной научно-практической конференции «Теория активных систем - ТАС-2014» (Москва, ИПУ РАН, 2014); IV международной научно-практической конференции «Моделирование и наукоемкие информационные технологии в технических и социально-экономических системах, 2016» (Новокузнецк, СибГИУ, 2016); XXII международной заочной научно-практической конференции «Современные тенденции развития науки и технологий» (Белгород, АПНИ, 2017).

Публикации. По теме диссертации опубликована 71 научная работа, авторским объемом более 50 п.л, включая 3 монографии, 68 научных статей, в том числе 2 статьи в изданиях SCOPUS, 25 статей в изданиях, рекомендованных ВАК РФ для опубликования основных результатов диссертационных исследований на соискание ученой степени доктора наук и 6 программных продуктов зарегистрированных во Всероссийском научно-техническом информационном центре, государственном фонде алгоритмов и программ и ФГАНУ «Центр информационных технологий и систем органов исполнительной власти».

Структура и объем работы. Диссертация состоит из введения, шести глав, заключения, приложений, списка литературы из 263 наименований и содержит 98 рисунков и 28 таблиц.

1. АНАЛИЗ МЕТОДИЧЕСКОГО ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННЫХ СИСТЕМ И ПРОЦЕССОВ ПОРТФЕЛЬНОЙ ИНВЕСТИЦИОННОЙ ДЕЯТЕЛЬНОСТИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЙ ОРГАНИЗАЦИИ. ПОСТАНОВКА ПРОБЛЕМЫ

1.1. Информационный процесс принятия решений в портфельной инвестиционной деятельности социально-экономической организации

ПИД СЭО представляет собой определенную совокупность управленческих решений и действий ее органов управления, которые направлены на оптимизацию вложения финансовых инвестиций в активы РЦБ, в форме ИП, с целью получения прибыли в виде процентов (дивидендов). Принятие портфельных инвестиционных решений (ПИР) основано на общей теории принятия решений, представляющей собой комплексную дисциплину, направленную на разработку методов и средств, помогающих ЛПР, сделать обоснованный выбор наилучшего решения из имеющихся вариантов. В систематизированном виде она сформировалась в XX веке, а появление отдельных ее положений (например, голосования как коллективного выбора) относится к XVII веку [199]. Решающую роль на ее формирование оказал и оказывает системный анализ [197, 238]. По определению академика Н.Н. Моисеева: «Системный анализ - это дисциплина, занимающаяся проблемами принятия решений в условиях, когда выбор альтернативы требует анализа сложной информации различной физической природы» [82].

Сложность принятия ПИР обусловлена наличием большого числа неопределенных факторов, влияющих на формальную постановку соответствующих задач и способы их решения. Множество неопределенных факторов можно разделить на три группы (см. рисунок 1.1).

Факторы из группы «Неопределенность природы» человеку неизвестны и от него не зависят [107, 176].

Факторы из группы «Неопределенность человека» связаны с непоследовательностью его действий, допускаемыми ошибками, зависимостью его действий от других людей (конкурентов, партнеров и др.) [88, 89, 138].

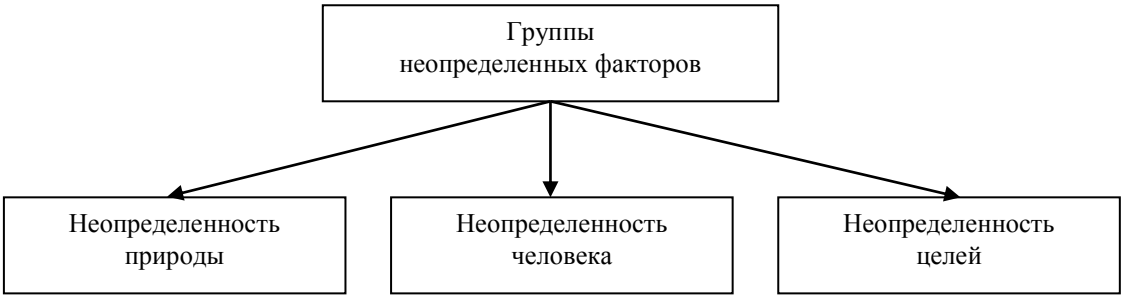


Рисунок 1.1. – Группы неопределенных факторов при принятии решений

Факторы из группы «Неопределенность целей» отражают наличие несовпадающих, а зачастую и противоречивых целей. Для их достижения необходимо наличие определенного компромисса.

Классификация неопределенностей приведена на рисунке 1.2.

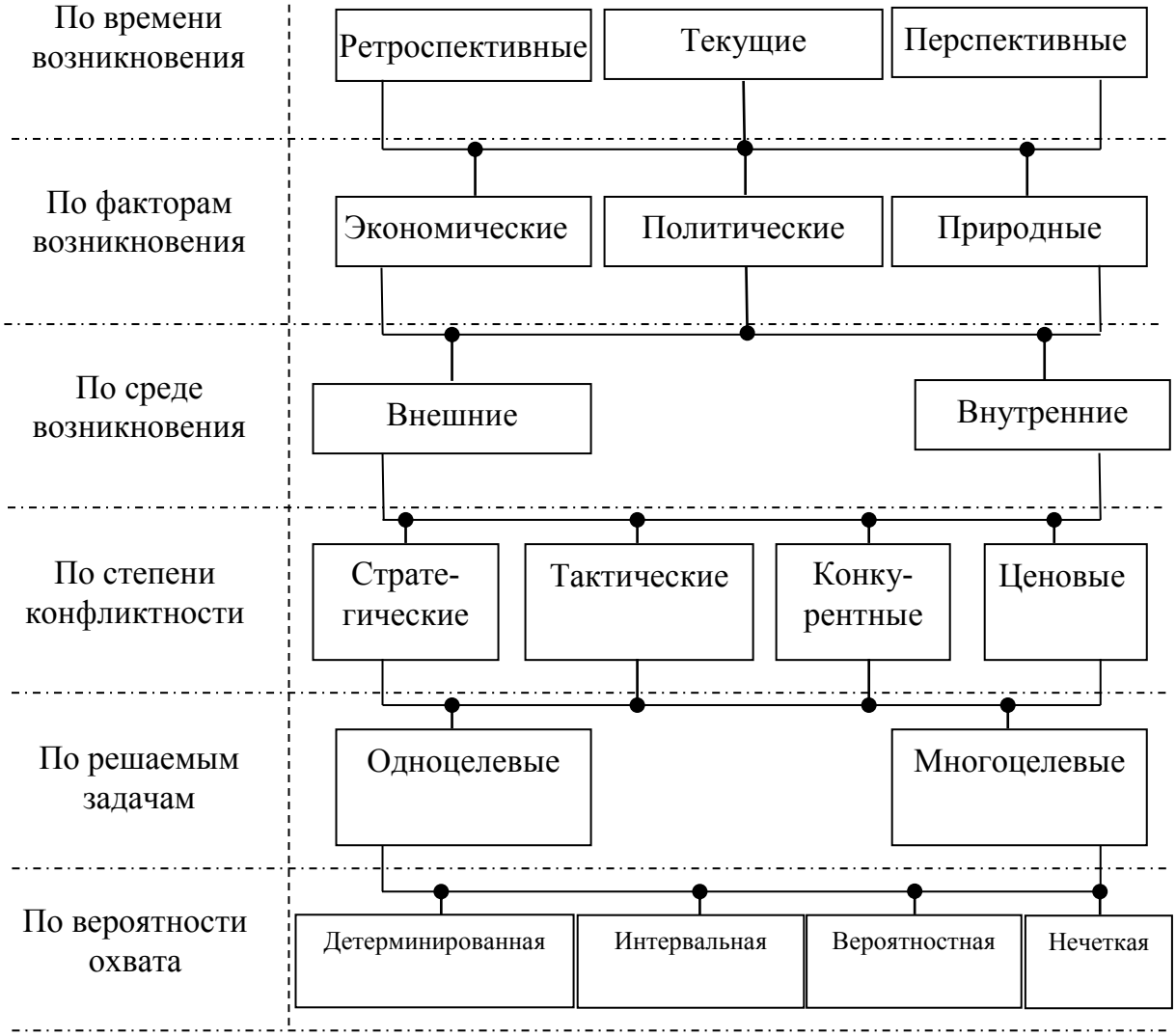


Рисунок 1.2. – Классификация неопределенностей

Из схемы видно, что в качестве классифицирующих признаков использованы: время возникновения; факторы возникновения; среда возникновения;

степень конфликтности; решаемые задачи; вероятность выпадения.

По времени возникновения неопределенности могут быть разделены на ретроспективные, текущие и перспективные. Учитывать данные виды неопределенности необходимо по причине того, что, как эффект, так и затраты могут быть распределены во времени по-разному. Поэтому может быть различный эффект.

Применительно к факторам возникновения, неопределенности подразделяются на политические, экономические и природные. К политическим неопределенностям относятся изменения в политической обстановке, которые оказывают влияние на коммерческую деятельность. Неблагоприятные изменения, происходящие в экономике страны, влияющие на деятельность социально-экономических организаций, относят к экономическим неопределенностям. Это могут быть неопределенности: цен представленных рыночных активов и их производных на фондовом рынке; спроса и предложения; валютных котировок и др. В практической деятельности, между этими двумя видами неопределенности наблюдается тесная взаимосвязь. К природным неопределенностям относят такие факторы, как: погодные условия, сейсмическую активность и др.

По среде возникновения неопределенности могут быть внешними и внутренними. Неопределенности внешней среды (природы) ЛПР неизвестны и от него не зависят. Они могут быть обусловлены социальными, демографическими и др. причинами. Внутренние неопределенности обусловлены личностными характеристиками ЛПР (характер и др.) и внутрисистемными особенностями функционирования (сбои оборудования, ошибки программ и др.).

Некоторые виды неопределенностей могут возникать на почве конфликтности. В зависимости от уровня и вида конфликта, неопределенности могут быть стратегическими, тактическими, конкурентными, ценовыми и др. [120, 123, 128].

В зависимости от решаемых задач (от количества достигаемых подцелей), неопределенности могут быть одноцелевыми и многоцелевыми.

По вероятности охвата неопределенность может быть детерминированной, интервальной, вероятностной и нечеткой [190].

Для корректного принятия решений в условиях воздействия множества

неопределенных факторов, необходимо ограничение или уменьшение их влияния [96, 148, 183]. Теоретически целесообразно полное «снятие» неопределенности (идеальный случай). На практике зачастую удается лишь ее снижение. Учет неопределенности приводит к необходимости содержательного анализа поставленных задач принятия решений и введения различных ограничений.

Информационный процесс (ИФП) принятия решений целесообразно рассматривать в рамках более глобального процесса – управленческой деятельности субъекта, представленной в схематичном виде на рисунке 1.3 [16].

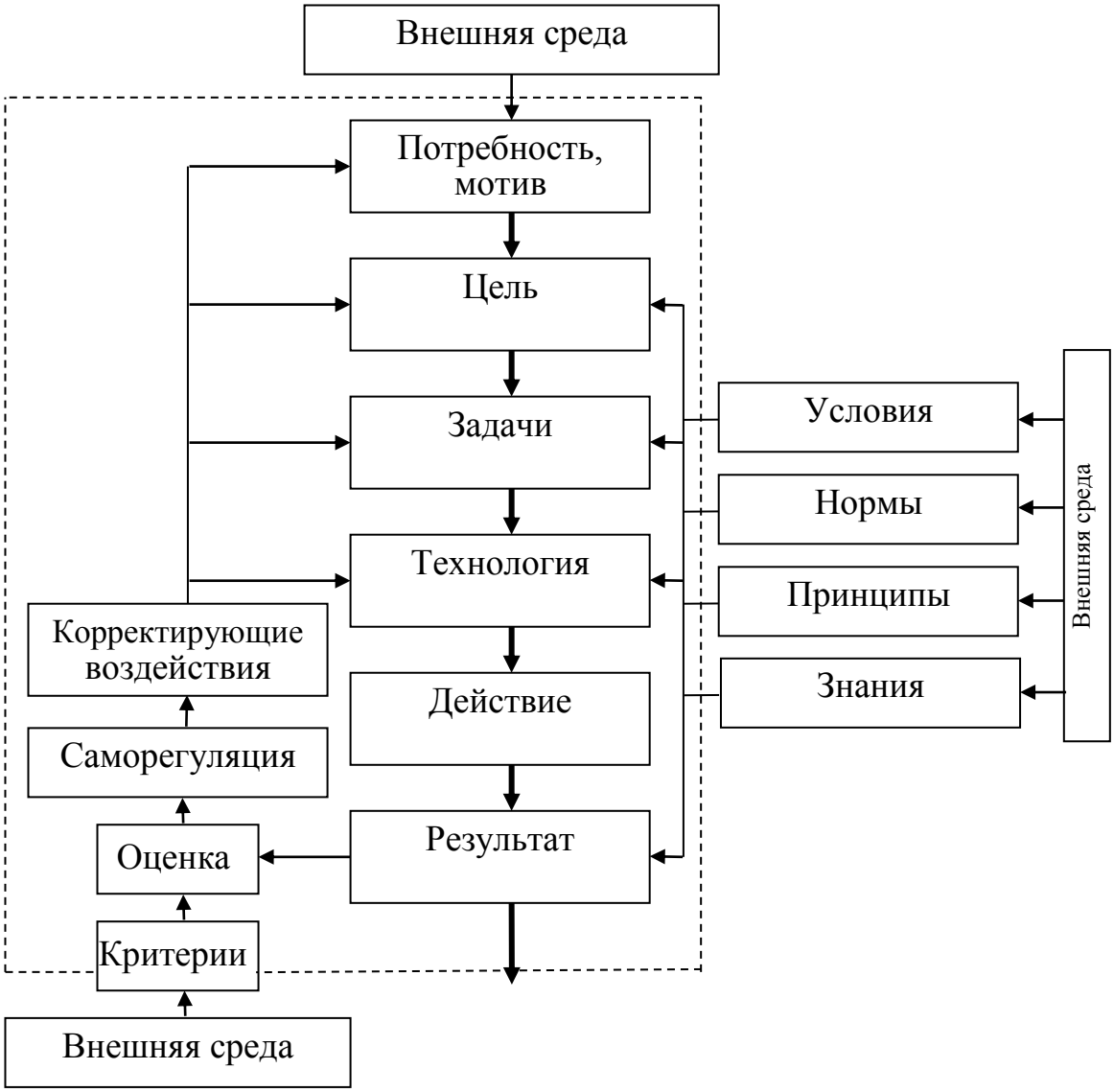


Рисунок 1.3. – Основные структурные компоненты деятельности субъекта

Комплексный набор этапов деятельности включает прямую последовательность действий (мотив – цель – задачи – технология – действие - результат)

и последовательность оценки и реализации уточняющих (корректирующих) действий (оценка – саморегуляция – корректирующие воздействия). Границы деятельности субъекта обозначены пунктирным прямоугольником. Отдельное место в данной схеме занимает набор компонент учета ситуативных особенностей (условия – нормы - принципы) и знаний субъекта. Наличие компонента знания является отличительной особенностью предложенной схемы, от существующей [16]. Его введение является необходимым условием осуществления эффективной деятельности субъекта. С позиций системного подхода, компоненты учета ситуативных особенностей и знания выступают в роли надсистемы над деятельностью субъекта. Действительно, насколько велик «багаж» знаний субъекта (умение формулировать цель и задачи, учет норм и условий, владение принципами и технологией), настолько эффективным будет результат.

Внешняя среда в данной схеме определяется как совокупность всех объектов/субъектов, не входящих в систему, изменение свойств и/или поведение которых влияет на изучаемую систему, а также тех объектов/субъектов, чьи свойства и/или поведение которых меняются в зависимости от поведения системы [24]. Она в определенной степени задает критерии, нормы и принципы деятельности.

Используемые критерии позволяют оценить соответствие полученного результата поставленной цели. Установленные нормы и принципы определяют диапазон деятельности субъекта (насколько далеко он может зайти в правовых, этических, гигиенических и других аспектах).

Условия деятельности представляют собой последовательность действий субъекта в различных складывающихся ситуациях (например, в детерминированных условиях, в условиях неопределенности и др.).

Если управление осуществляет субъект, то управление следует рассматривать как деятельность [16] и управление это вид практической деятельности по организации деятельности [189]. Действительно, если управление – это деятельность, то осуществление этой деятельности является функцией управляющей системы (в данном случае ЛПР), процесс реализации управления соответствует

ИФП деятельности, а управленческое воздействие – ее результату и т.д. Другими словами, в СЭО, где и управляющий орган (ЛПР) и управляемая система (система поддержки принятия решений) являются субъектами, управление является деятельностью по организации деятельности [189].

Классификация общих управленческих решений приведена в [173].

В области портфельных финансовых инвестиций (ПФИ) управленческая деятельность имеет свою специфику [3]. Она базируется на двух фундаментальных понятиях – портфельной инвестиционной среде (ПИС) и портфельном инвестиционном процессе (ПИП) [234].

ПИС характеризуется типами ЦБ, обращающихся на РЦБ, условиями их продажи и приобретения [234].

Понятие ПИП связано с тем, как ЛПР принимает решения по выбору ЦБ, сроков и объемов вложения [234].

Суть основных понятий, применительно к ПИС, достаточно подробно изложена в [141, 172, 173].

Под ПФИ будем понимать денежные вложения в ЦБ, иностранные валюты, банковские депозиты, объекты тезаврации и др. [56]. В результате подобных вложений инвестор может получить дивиденды и другие доходы, тем самым увеличивая свой финансовый капитал. Основные объекты ПФИ (другое название – финансовые инструменты) представлены на рисунке 1.4.

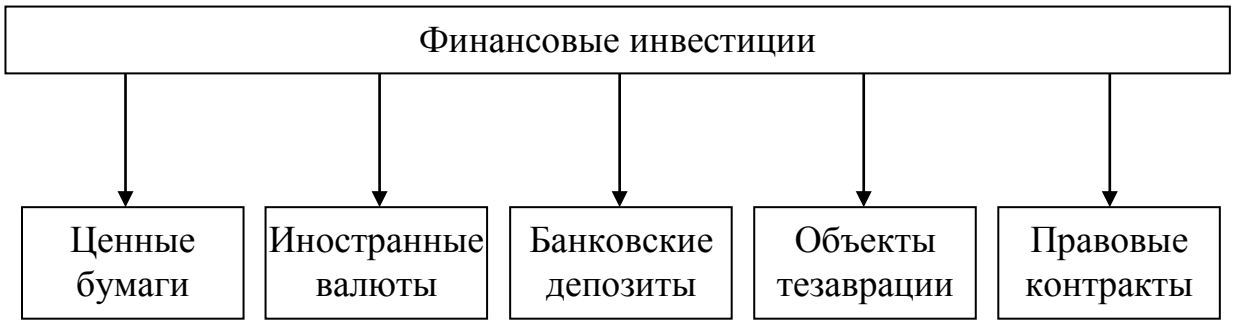


Рисунок 1.4. – Основные объекты финансового инвестирования

ЦБ это документы, подтверждающие право их владения собственником. Они позволяют владельцу в некоторых ситуациях получить доход при их обмене на деньги [194, 241]. По числу участников инвестирования ЦБ могут быть индивидуальными и коллективными [30].

Индивидуальные ЦБ приобретаются либо на первичном (биржевом), либо на вторичном (внебиржевом) рынках [242]. Коллективные ЦБ представляют собой приобретенные паи или акции различных инвестиционных фондов и организаций [254]. В настоящее время биржевой рынок РФ оперирует акциями, облигациями, казначейскими обязательствами, сертификатами, векселями, варрантами, коносаментами, опционами, фьючерсами и др.

Иностранные валюты – денежные единицы стран (рубль, доллар, евро и др.) и их типы (золото, серебро и др.) [216]. Наиболее распространенными способами инвестирования в иностранные валюты являются [172, 173]:

- приобретение иностранной валюты в банках;
- покупка валюты на бирже (сделки спот);
- заключение биржевого фьючерсного контракта;
- открытие валютного банковского счета.

Банковские депозиты – традиционный способ хранения капитала и получения дивидендов в виде процентов. При реализации данного инструмента риски не высоки, однако также не высоки и получаемые дивиденды.

Тезаврационные инвестиции – вложение капитала в ценности (серебро, золото, платину, палладий, драгоценные камни, антиквариат, картины, монеты и др.) [11, 14]. Дивиденды формируется за счет разницы между ценой их покупки и продажи [249]. В условиях нестабильной экономики данный инструмент может быть успешным. Золото является одним из наиболее распространенных объектов вложения. Среди форм вложений в золото встречаются: приобретение золотых слитков; покупка золотых монет; приобретение изделий из золота; вложения в акции золотодобывающих организаций; приобретение акций золотодобывающих организаций. Данный финансовый инструмент характеризуется высоким риском и значительной капиталоемкостью. Ввиду последнего обстоятельства их проводят крупные инвесторы. Меньшее распространение получили инвестиции в драгоценности [21]. Они обладают большими издержками, которые возникают при перепродажах. Подобными инвестициями занимаются в основном крупные организации [72]. Основными недостатками тезаврационного

инвестирования являются их сложность и относительно узкий рынок [173].

Правовые контракты дают право владельцу купить/продать определенный актив по некоторой, заранее оговоренной цене в течение определенного промежутка времени [8]. К ним относятся опционы, фьючерсы и др.

Вопросами второго фундаментального понятия инвестиций в ЦБ в ПИП занимается инвестиционный менеджмент (ИМ) [10, 247].

Целью ПИП является получение максимальных дивидендов от их (инвестиций) вложений и минимизация финансовых рисков [11, 157].

Основными задачами ИМ являются [13]:

- прогнозирование состояний рынка инвестиций;

- исследование влияния внешней среды на динамику инвестиционного рынка;

- оценка текущего состояния инвестиционного рынка;

- формирование рациональных решений в инвестиционной деятельности;

- определение стратегических решений в рамках инвестиционных ресурсов;

- сравнительный анализ и выбор лучших инвестиционных проектов;

- эффективное управление ИП;

- оперативное, тактическое и стратегическое планирование управлением инвестиционными проектами и программами;

- оперативная оценка текущих проектов, своевременная реструктуризация неэффективных инвестиционных проектов.

Портфельные инвестиционные управленческие решения (ПИУР), принимаемые в рамках ИМ, отличаются своим многообразием [147]. Их классификация приведена в [172, 173].

ИФП принятия ПИУР базируется на типовом процессе принятия решений [173]. В схематичном виде, с учетом специфики области ПФИ, он представлен на рисунке 1.5.

На данной схеме представлены основные (отражены сплошными линиями) и вспомогательные (отражены пунктирными линиями) этапы принятия ПИУР.

Начальный этап ИФП принятия ПИУР - сбор информации (блок 1).

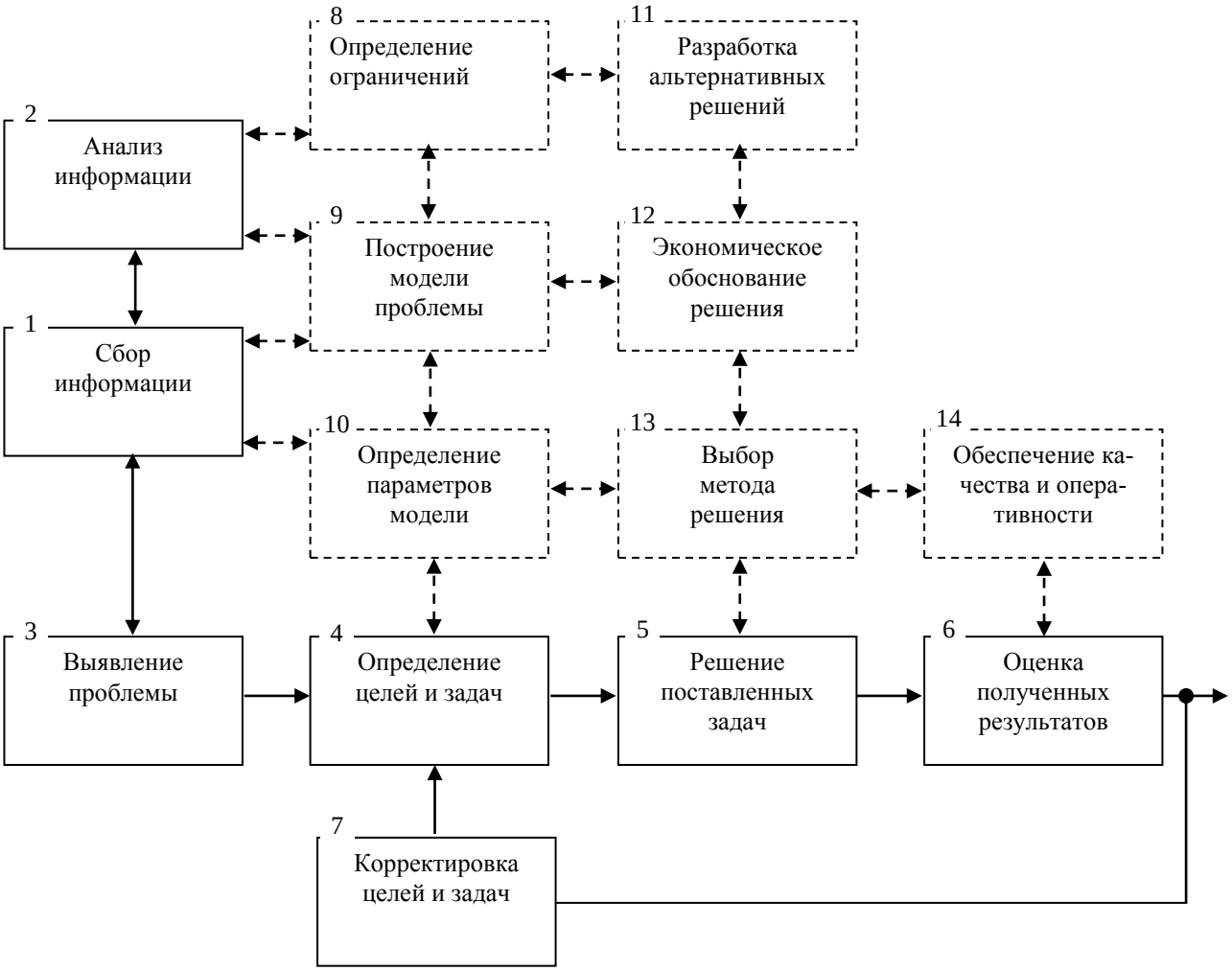


Рисунок 1.5. – Схема информационного процесса принятия ПИУР

На втором этапе (блок 2) проводится анализ информации.

На третьем этапе (блок 3) выявляется проблема. Содержимым четвертого этапа (блок 4) является формулировка целей и решаемых задач, обеспечивающих их достижение. Перед постановкой целей и решаемых задач проводится детальный анализ исходной информации и способов решения возникшей проблемы. Применительно к принятию ПИУР детальный анализ включает следующие этапы [101]: выявление ограничений (блок 8); построение модели (блок 9); определение параметров модели (блок 10). Детальный анализ способен повысить качество принимаемых ИУР, однако требует дополнительных затрат. Решение поставленных задач (блок 5) является очередным этапом общего информационного процесса принятия ИУР. Данный этап может детализироваться информационным подпроцессом, содержащим: разработку альтернатив (блок 11), экономическую оценку альтернатив (блок 12)

и выбор рациональной альтернативы (блок 13). При выборе альтернативных решений целесообразно учитывать такие факторы, как: время; качество; масштаб; степень освоенности объекта; метод получения информации; условия применения объекта; уровень инфляции; степень риска; уровень неопределенности [54, 156, 173]. При этом среди правил анализа альтернатив выделяют следующие [226]:

- число различных вариантов должно быть больше трех;

- последний по времени вариант должен приниматься в качестве базового, остальные, корректирующими коэффициентами, приводятся к нему;

- альтернативные варианты должны формироваться с учетом более высоких эффективности и качества, принимаемых решений;

- для повышения качества и оперативности принимаемых решений, а также для снижения затрат целесообразно применение ИС.

Следующим этапом общего ИФП принятия ПИУР является оценка полученных результатов (блок 6). Вспомогательным информационным подпроцессом при его реализации является оценка качества и оперативности полученных результатов (блок 14). Завершающим этапом в общем ИФП принятия ПИУР является корректировка поставленных целей и задач (блок 7). Если оценка эффективности принятого решения низкая (низкое качество и оперативность), оно может быть пересмотрено. В случае возникновения новых существенных условий текущие цели и решаемые задачи могут быть уточнены. Затем весь ИФП принятия решения или его отдельные этапы могут повторяться. Способность повторения указывает на итеративность ИФП принятия ИУР. Данный ИФП в определенной степени идеализирован. На практике, его некоторые этапы могут быть объединены, некоторые видоизменены, а некоторые могут отсутствовать [172, 173].

В конечном итоге вышеизложенный ИФП реализует многоэтапную процедуру – основу ПИП, включающую: анализ РЦБ в интересах выбора актуальных и перспективных направлений вложения инвестиций (инвестиционной политики), определение конкретных активов и долей вкладываемого в

них капитала (анализ ЦБ); реализация микропрогнозирования (анализ динамики цен отдельных видов ЦБ) и макропрогнозирования (анализ изменения уровня цен акций относительно ЦБ с фиксированным доходом, например корпоративных облигаций), проведение диверсификации (выбор ЦБ с низкой взаимной корреляцией), расчет значений доходности и риска сформированного ИП, корректировка состава и структуры ИП с учетом изменений факторов внешней и внутренней среды [136, 234].

Выбор инвестиционной политики зависит от цели инвестора и объема инвестируемых средств. Цели инвестирования формулируются с учетом, как доходности, так и риска. Данный этап завершается выбором потенциальных видов финансовых активов для включения в ИП [234].

Суть анализа ЦБ заключается в изучении отдельных их (ЦБ) видов с целью выявления наиболее перспективных в плане получения дивидендов, а также в определении неверно оцененных ЦБ. Основными направлениями анализа ЦБ являются технический и фундаментальный анализ [89, 131, 22, 155].

Технический анализ в его простейшей форме (на примере акции фирмы) включает изучение конъюнктуры курсов рынка акций, с тем, чтобы дать прогноз динамики курсов акций конкретной фирмы [124]. Первоначально проводится исследование курсов за прошедший период с целью выявления повторяющихся тенденций или циклов в динамике курсов. Затем анализируются курсы акций за последний период времени, с тем, чтобы выявить текущие тенденции, аналогичные обнаруженным ранее [210]. Это сопоставление существующих тенденций с прошлыми осуществляется, исходя из предположения, что ценовые тренды периодически повторяются. Таким образом, выявляя текущие тенденции, есть возможность дать достаточно точный прогноз будущей динамики курсов рассматриваемых акций [234].

Фундаментальный анализ исходит из того, что «истинная» (или внутренняя) стоимость любого финансового актива (например, акций) равна приведенной стоимости всех наличных денежных потоков, которые владелец актива рассчитывает получить в будущем [234]. В соответствии с этим ЛПР

стремится определить время поступления и величину этих наличных денежных потоков, а затем рассчитывает их приведенную стоимость, используя соответствующую ставку дисконтирования [234]. Он должен не только оценить ставку дисконтирования, но также спрогнозировать величину дивидендов, которая будет выплачена в будущем по данной акции. Последнее эквивалентно вычислению показателей прибыли фирмы в расчете на одну акцию и коэффициента выплаты дивидендов [234]. Более того, необходимо дать оценку ставки дисконтирования [234]. После того, как внутренняя стоимость акции данной фирмы определена, она сравнивается с текущим рыночным курсом акций с целью выяснить, правильно ли оценена акция на рынке. Акции, внутренняя стоимость которых меньше текущего рыночного курса, называются переоцененными, а те акции, рыночный курс которых ниже внутренней стоимости, — недооцененными [234]. Разница между внутренней стоимостью и текущим рыночным курсом также представляет собой важную информацию, поскольку обоснованность заключения ЛПР о неправильности оценки данной акции зависит в значительной степени от этой величины [234]. Случаи неверной оценки активов исправляются впоследствии рынком: курсы недооцененных акций растут быстрее, а переоцененных - медленнее, чем средние рыночные курсы [234].

Этап формирования ИП ЦБ заключается в определении перечня ЦБ для вложения инвестиций и пропорций размещения инвестируемого капитала между ними. При этом возникают задачи селективности, выбор операционного времени и диверсификации. Селективность связана с прогнозом цен отдельных видов ЦБ. Выбор операционного времени включает прогноз изменения величин цен на ЦБ, в сравнении с ценами для ЦБ с фиксированным доходом, такими, как корпоративные облигации. Суть диверсификации заключается в формировании ИП так, чтобы при некоторых ограничениях его был бы минимизирован [234].

Процесс пересмотра ИП заключается в периодическом повторении вышеперечисленных этапов. Ведь, через некоторое время цели инвестора мо-

гут поменяться. В итоге, текущий ИП может потерять свойство оптимальности [155]. Тогда инвестор будет формировать новый ИП. При этом, он может продать часть имеющихся ЦБ и приобрести новые. Другим основанием для пересмотра ИП является изменение курса ЦБ с течением времени. В связи с этим некоторые ЦБ, первоначально бывшие непривлекательными для инвестора, могут стать выгодным объектом вложения, и наоборот [83, 155]. Тогда инвестор захочет приобрести первые, одновременно продав последние из своего ИП. Решение о пересмотре ИП зависит кроме других факторов от размера транзакционных издержек и ожидаемого роста доходности пересмотренного ИП [234].

Оценка эффективности ИП заключается в периодической оценке, как показателей риска, так и полученной доходности. При этом необходимо для сравнения использовать соответствующие стандарты (своеобразные «эталонные» значения), а также приемлемые показатели доходности и риска [204, 234].

Существенную помощь ЛПР в процессе ПИУР способна оказать ИСППИД СЭО – разновидность ИС [108, 109, 223].

1.2. Определение информационной системы поддержки портфельной инвестиционной деятельности социально-экономической организации

Известно, что ИС представляет собой совокупность технических и программных средств, математического обеспечения и персонала, объединенных в единое целое в интересах добычи, хранения, обработки и выдачи информации в соответствии с поставленной целью [108 – 110, 223].

В рамках существующей классификации ИС [47] выделяются системы поддержки принятия решений (СППР). Появление данного класса ИС обусловлено спецификой задач, решаемых в управленческой деятельности. Они могут быть условно разделены на три класса: структурируемые, полуструктурируемые и неструктурируемые [228].

В качестве инструмента для решения неструктурируемых и полуструктурируемых задач, наиболее характерных для ИМ, уже более сорока лет, ис-

пользуются СППР.

Единого определения СППР в настоящее время не существует, по причине их (систем) многогранности и многообразия. Некоторые из определений выделяют различные аспекты применения СППР. Их необходимо учитывать при определении ИСППИД.

СППР — это «совокупность процедур по обработке данных и суждений, помогающих руководителю в принятии решений, основанная на использовании моделей» [7, 66, 251].

СППР — это «система, которая обеспечивает пользователям доступ к данным и/или моделям, так что они могут принимать лучшие решения» [243].

СППР — это «интерактивные автоматизированные системы, помогающие лицу, принимающему решения, использовать данные и модели для решения слабо структурированных проблем» [74, 255].

В [74] под СППР понимаются компьютерные ИС, помогающие управляющему в принятии решений, при решении плохо структурированных задач посредством прямого диалога с машиной с использованием данных, знаний и математических моделей.

Отличительными особенностями СППР являются [74, 173]:

- ориентация на решение плохо структурированных (формализованных) задач, характерных, главным образом для высоких уровней управления;

- возможность сочетания традиционных методов доступа и обработки компьютерных данных с возможностями математических моделей и методами решения задач на их основе;

- направленность на непрофессионального конечного пользователя ЭВМ посредством использования диалогового режима работы;

- высокая адаптивность, обеспечивающая возможность приспосабливаться к особенностям имеющегося технического и программного обеспечения, а также требованиям пользователя.

В [228] СППР определяются как особые интерактивные ИС менеджмента (институциональные или на случай), использующие оборудование,

программное обеспечение, данные, базу моделей и труд менеджера с целью поддержки всех стадий принятия полуструктурируемых и неструктурируемых решений непосредственными пользователями – менеджерами в процессе аналитического моделирования на основе предоставленного набора технологий.

Современное определение СППР, на котором базируется определение ИСППИД представлено в [248].

Система поддержки принятия решений (СППР) (англ. Decision Support System, DSS) — компьютерная автоматизированная система, целью которой является помощь людям, принимающим решение в сложных условиях для полного и объективного анализа предметной деятельности. СППР возникли в результате слияния управленческих ИС и систем управления базами данных. В большинстве случаев СППР — это интерактивная автоматизированная система, которая помогает ЛПР использовать данные и модели для идентификации и решения задач и принятия решений. Система должна обладать возможностью работать с интерактивными запросами с достаточно простым для изучения языком запросов [248].

Исходя их данного определения, ИСППИД представляет собой программно-технический комплекс, использующий оборудование, специальное математическое и программное обеспечение, данные, знания, базу моделей для информационного управления (ИУ) ЛПР с целью выбора оптимального ИП [127, 177].

Под ИУ понимается сообщение определенной информации [188].

Применительно к ИФП принятия ПИУР (см. схему на рис. 1.4), ИСППИД призвана решать задачи в рамках подпроцессов (блоки 8-10): определение ограничений; выбор модели; определение параметров модели. Целевое назначение этих подпроцессов - повысить качество информации, выдаваемой ЛПР для дальнейшего принятия решений. При этом качество выдаваемой информации, в свою очередь, зависит от качества общего используемого информационного обеспечения (ИО).

ИО представляет собой набор формализованной информации, хранящейся и циркулирующей в ИСППИД [92, 104, 105]. Данная информация подразделяется на системную и пользовательскую. Основное внимание в диссертации уделено пользовательской информации. Она представляет собой сведения об явлениях и объектах внешней и внутренней среды, их свойствах и состоянии, параметрах, которые уменьшают степень неполноты и неопределенности, имеющихся знаний о них в области ПИД [103, 173].

Информация о внешней среде включает [172, 173]:

- параметры (котировки и др.) ЦБ;
- значения индексов фондового рынка;
- характеристику политической обстановки;
- параметры экономической обстановки.

Информация о внутренней среде содержит [173]:

- схемы представления данных;
- форматы представления данных в базе;
- параметры обмена данных внутренней локальной сети;
- набор используемых моделей оценки инвестиций.

1.3. Анализ методического обеспечения информационных процессов и систем портфельной инвестиционной деятельности социально-экономической организации

Под методическим обеспечением (МО) ПИД понимается совокупность подходов, стратегий, методов, моделей, методик и алгоритмов, используемых для формализации ИФП портфельной инвестиционной деятельности СЭО.

Сравнительный анализ классических моделей ПИД приведен в приложении Б. Сравнивались модели: Г. Марковица, У. Шарпа, Д. Тобина, САРМ, Блэка-Шоулса. Проведенный анализ показал, что каждая из моделей, в силу своих возможностей, достоинств и недостатков, целесообразна для использования в определенных условиях. Универсальной модели формирования ИП СЭО для различных данных (детерминированных и стохастических) об активах не существует. Однако каждая из них может служить исходной основой для создания новых (модернизированных) моделей, улучшающих их меха-

низм функционирования, парирующих некоторые недостатки (ограничения) и обеспечивающих адаптацию использования в новых условиях (например, для стохастических данных об активах). В данной работе, в силу своей простоты и предпочтительности практического использования, для модернизации выбраны модели: Г. Марковица и У. Шарпа.

Характеристика МО ПИД СЭО приведена в приложении В.

Среди составляющих, разработанного МО ПИД СЭО, носящих фундаментальный характер, выделяется долгосрочная стратегия развития РЦБ РФ, приведенная в [83]. Она базируется на теоретическом исследовании фундаментальных факторов, воздействующих на конкурентоспособность фондовых рынков, объективном анализе результатов формирования отрасли ЦБ РФ в 90-е годы, оценке мирового опыта развития РЦБ и исследовании государственных макроэкономических стратегий в данной области. В рамках разработки данной стратегии, впервые систематизированы результаты исследования и прогнозирования РЦБ РФ. Кроме того, в [2] обоснованы концептуальные подходы, обеспечивающие принятие инвестиционных решений, более адекватных государственным особенностям движения финансовых потоков и динамики цен активов на финансовых рынках, основанные на представлении ПИР как системного процесса с элементами фундаментального и технического анализа, поведенческих финансов и факторов ликвидности.

Ряд работ посвящен разработке и исследованию новых мер волатильности [217, 51] и риска [28, 66, 232, 233] ЦБ и, связанных с их учетом, новых моделей, стратегий, методик и алгоритмов формирования ИП. Последние направлены на повышение адекватности и точности формализации происходящих информационных процессов при формировании ИП, что, в конечном счете, оказывает влияние на точность проводимых расчетов и величины выходных параметров (доходность и риск).

Определенное внимание в работах исследователей уделено совершенствованию отдельных этапов общего процесса информационного управления ИП, в частности:

- предложена методика реструктуризации ИП, основанная на его потенциальной ликвидационной стоимости, используемой в качестве меры эффективности. Разработаны стратегии ликвидации ЦБ для различных временных интервалов, основанные на их длинах [187];

- разработана модель прогнозирования распределений доходностей ЦБ и формирования ИП на основе многомерного *t*-распределения с вектором степеней свободы. Выведены формула смешанного момента общего вида с условиями существования, формулы одномерных маргинальных функций плотности и характеристических функций, получена методика построения произвольной многомерной GARCH модели, предложен алгоритм симулирования, асимметричной модификации и построена копула [6];

- разработаны модели прогнозирования доходности ЦБ на финансовых рынках на основе ИНС. При этом использовались: а) самоорганизующаяся ИНС Кохонена – Хакена, содержащая конкурентные механизмы обработки входных данных, использующие методы независимых и главных компонент в совокупности с применением относительной энтропии Кульбака – Ляйблера; б) роевая модель, реализующая предварительную обработку эмпирических данных на основе кластеризации с использованием расстояния Дженсона-Шеннона [32, 134];

- предложен метод сверхкраткосрочного прогнозирования цен активов на РЦБ, обеспечивающего доходность ИП на основе реализации высокочастотных операций [25].

Значительное внимание уделено совершенствованию методического обеспечения ТА, среди которого выделяются:

- методический подход инвестирования, основанный на алгоритмической торговле, содержащий: а) прогнозные математические модели динамики актива EUR/USD для различных тайм-фреймов; б) стратегии инвестирования для механических торговых систем; в) метод ситуационного риск-контроля, основанный на показателе – «коэффициенте покрытия убытка текущей операции» [64];

- моментум стратегия управления ИП, основанная на портфельном эффекте, проявляемом на моментуме и развороте, представляющим собой такую ценовую аномалию на активах одного класса (например, на акциях), когда ИП, реализованные с использованием прошлых результатов инвестирования на определенном временном интервале (краткосрочном, среднесрочном и долгосрочном) демонстрируют статистически значимую прибыль [81];

- торговая стратегия формирования ИП долгосрочных ЦБ, основанная на индикаторах ТА («баланс оборота», «относительная сила», «полоса Боллинджера», «стохастический осциллятор») [215];

- синтетическая модель анализа финансовых активов, фиксирующая временные периоды эффективной работы функционирующей торговой системы на основе синтетического индикатора Хольта и нелинейной статистической модели Веге [221].

Значительная часть исследований посвящена уточнению определения характеристик различных финансовых инструментов и параметров агрегированных ИП, наиболее значимыми среди которых являются:

- теоретический подход аналитической оценки справедливой стоимости сложных финансовых инструментов, базирующийся на двумерных негауссовских асимметричных совместных распределениях вероятностей цен базовых активов, учитывающий негауссовский характер ценовых процессов и не предполагающий дробление структурированного деривата на составные части [45];

- аналитическое решение задачи определения стоимостей стандартных и специфических Европейских опционов для ИП, обеспечивающей хеджи капиталов с выплатой процентов по активу с высоким риском при квантильном хеджировании [37];

- подход к формированию агрегированных ИП, содержащих наряду с ЦБ, стоимостные значения объектов недвижимости. Разработана стоимостная модель оценки капитала при инвестировании в объекты недвижимости на рынке РФ [200];

- интегрированная модель формирования ИП пенсионных накоплений, основанная на портфельной теории Марковица и безрисковых ЦБ [210].

Представляет интерес динамическая модель управления ИП [29], обеспечивающая синтез оптимальных управляющих стратегий, реализующих максимальную гладкость кривой роста капитала, заданной ЛПР на всех горизонтах инвестирования при наличии ограничений на инвестируемый капитал, с учетом различных моделей динамики цен активов (Блэка-Шоулса, стохастической волатильности, финансового рынка с варьируемыми режимами). Данная модель является единственной, реализующей обработку стохастических данных об активах. Однако она ориентирована на высокорисковые активы (опционы, фьючерсы и др.), поскольку именно для них разработана модель Блэка-Шоулса и является закрытой, что не позволяет использовать ее программное обеспечение для встраивания в другие ИС или модификации.

Характеристика разработанных частных СППР и программных решений в ПИД СЭО представлена в приложении Г.

Разработанные СППР реализуют следующий функционал [28, 66, 206, 221, 232, 233]:

- расчет различных мер риска и оптимизация формирования ИП с учетом полученных значений мер [66, 233];

- формирование эффективной структуры ИП на основе применения различных подходов к измерению риска и доходности, а также с нечетко-множественными входными параметрами [28];

- управление ИП пенсионных накоплений, обеспечивающее оптимальные оперативные воздействия на сформированный ИП [206];

- управление и оценка рисков в финансовой деятельности брокерской компании, формирование оптимального хеджирующего ИП на основе данных о торговых сделках, оценка структуры ИП и определение оптимальной стратегии инвестиционного управления [232];

- управление ИП на основе формализованного механизма синхронной коррекции возникающих торговых и модельных сигналов, обеспечивающего

снижение риска инвестора [221].

В [4] разработан программный комплекс обеспечения проведения торгов на фондовом рынке, реализующий механизм двойного непрерывного аукциона, позволяющего оценивать влияние регулирования на взаимодействие участников в процессе ценообразования и проводить валидацию построенных моделей.

Безусловным достоинством разработанных СППР и программного комплекса является расширение возможностей (функционала) формирования и управления ИП. К недостаткам, ограничивающих их применение и модернизацию, следует отнести реализацию каждым из них только одного (частного) этапа в общем ИФП управления ИП и их закрытость.

Результаты сравнительного анализа зарубежных и отечественных ИС и программ для работы на РЦБ представлены в приложении Д.

Проведенный анализ показал, что наибольшая часть ИС разработана для использования в качестве инструментов ТА для трейдеров и профессиональных инвесторов на РЦБ. Они обладают следующими функциональными возможностями:

- обработка ценовых графиков и индикаторов, работа с опционами, сортировка инструментов в реальном масштабе времени в соответствии с заданными критериями, обработка данных о ЦБ, создание уникальных функций (ИС «OmegaResearchProSuite»);

- формализация торговых стратегий, конструирование индикаторов и новых торговых систем, тестирование торговых систем на основе исторических данных, математическая формализация и включение дополнительных компонент (комиссия, StopLoss, TakeProfit), анализ разномасштабных графиков (ИС «MetaStock»);

- конструирование индикаторов и новых торговых систем, нейросетевые прогнозы, выбор целевой функции для нейросетевой модели, тестирование торговых систем на основе исторических данных, автоматическая загрузка данных, экспорт и импорт данных, генетическая оптимизация набора

и параметров торговых правил, возможность внутридневной работы тестирование стратегий (ИС «Neuroshell DayTrader»);

- предоставление стандартных и разработка уникальных индикаторов, автоматическая загрузка данных из различных источников, настройка интерфейса, отрисовка примитив, поддержка различных внутридневных интервалов, управление позициями (ИС «AmiBroker»);

- предоставление встроенных функций, настройка временных и тиковых фреймов, настройка уникальных стратегий, трехмерное представление результатов, формирование и тестирование торгового портфеля на исторических данных (ИС «Multicharts»);

- прогнозирование состояния рынка и цен, автоматическое определение возможности заключения потенциально прибыльной сделки и моментов закрытия позиций, визуализация полученных результатов, определение «точек разворота» тренда, обучение персонала (ИС «ELWAVE»).

Среди найденных ИС ФА отличается «FXAnalytics». К ее основным возможностям относятся: получение новостей в реальном режиме времени; предоставление котировок активов с различных рынков; наличие календаря экономических событий; формирование торговых рекомендаций; скачивание торговых систем и др.

Среди отечественных систем, непосредственно используемых для ПИД, выделяются ИС «София» и «GAMA». Они обладают следующими функциональными возможностями:

- ведение ИП инвесторов;

- мониторинг состава, структуры и состояния ИП;

- расчет характеристик ИП (число ЦБ, средняя цена покупки, балансовая стоимость ЦБ, сумма выплаченных налогов, средний инвестированный капитал за период и др.);

- проведение операций с ЦБ (покупка, продажа, закрытие, оприходование, списание);

- предоставление графиков изменения рыночных параметров;

- выдача сведений о ЦБ (котировки, объявляемые дивиденды и др.);

- отображение курсов валют;

- отображение экономических и биржевых индексов;

- генерация отчетов и предоставление графиков.

Важной отличительной особенностью ИС «GAMA» является наличие имитационных моделей и модели Блэка-Шоулса. Фактически данная ИС может считаться прототипом полноценной СППР, поскольку содержит базу моделей (БМ). Однако имеется ряд недостатков, как и у ИС «София», ограничивающих их применение. К ним следует отнести: отсутствие возможности работы со стохастическими данными, высокая стоимость и закрытость.

Проведенный анализ методического обеспечения информационных процессов и систем ПИД СЭО показал отсутствие системных проработок в данной предметной области и выявил их следующие недостатки [40, 69]:

- сложность и по этой причине отсутствие динамических постановок задач формирования оптимального ИП СЭО;

- отсутствие методического обеспечения решения информационных задач формирования оптимального ИП для стохастических данных об активах;

- реализация лишь отдельных фрагментов некоторых частных этапов общего ИФП управления ИП СЭО;

- отсутствие полнофункциональных ИСППИД, имеющих базы моделей;

- отсутствие возможности предоставления ЛПР объективной прагматической информации, интегрированный в общий процесс управления ИП СЭО;

- отсутствие методического обеспечения поиска, обработки и хранения больших массивов исходных данных и прагматической информации об активах;

- несоответствие в существующих программных системах фактических значений показателей качества информации (полнота, адекватность, достоверность) требуемым.

Усугубляют сложившееся положение недостатки готовых программных решений. Программные продукты отечественного производства в дан-

ной предметной области ограничены, закрыты (не допускают модификаций и доработок) и дороги. Для зарубежных программ характерны: значительная стоимость; избыточность функциональных возможностей; закрытость; отсутствие учета особенностей отечественного РЦБ [106, 173].

Вышеизложенные недостатки приводят к затруднению своевременной качественной оценки информации и выработки эффективных ИУР. Поэтому необходимость учета стохастических данных об активах в ходе формирования оптимального ИП и информационного управления им в условиях оперативного изменения обстановки на РЦБ, актуализирует проблему оптимизации данного процесса, которую можно разрешить на основе разработки ИСППИД, которая в структурном плане представляет собой совокупность объединенных единством цели методов, моделей, алгоритмов и программ поиска, анализа, обобщения, обработки, хранения и предоставления инвестору СЭО прагматической информации о целесообразных активах для включения в ИП, а также рациональном составе и структуре активов ИП, обеспечивающих максимизацию его доходности с учетом заданного риска, требуемой точности, полноты, достоверности и оперативности проведения расчетов.

Постановка проблемы разработки методов, моделей и алгоритмов синтеза ИСППИД для повышения эффективности информационного обеспечения портфельной инвестиционной деятельности СЭО при принятии решений формирования оптимального ИП в условиях стохастических данных об активах формулируется следующим образом [168].

1.4. Постановка проблемы

Пусть ПИД СЭО основывается на информационной подсистеме управления (ИПУ) портфельным инвестированием (ПИ) в составе центра, субъекта и объекта информационного управления (соответственно ЦИУ, СИУ и ОИУ). ЦИУ представлен топ менеджером (например, финансовым директором), реализующим информационное управление (ИУ) СИУ, в качестве которого выступает лицо, принимающее решения (ЛПР) в области ПИ. ЦИУ - орган ИУ, а ЛПР – управляемый субъект ИУ. В качестве ОИУ в ИПУ выступает ИП,

который формируется СИУ (ЛПР) на основе ИСППИД. Общая схема взаимодействия ЦИУ, СИУ, ОИУ, ИСППИД с РЦБ и внешней средой (в виде воздействия конкуренции, а также политических, экономических, социальных и др. факторов) представлена на рисунке 1.6 [150].

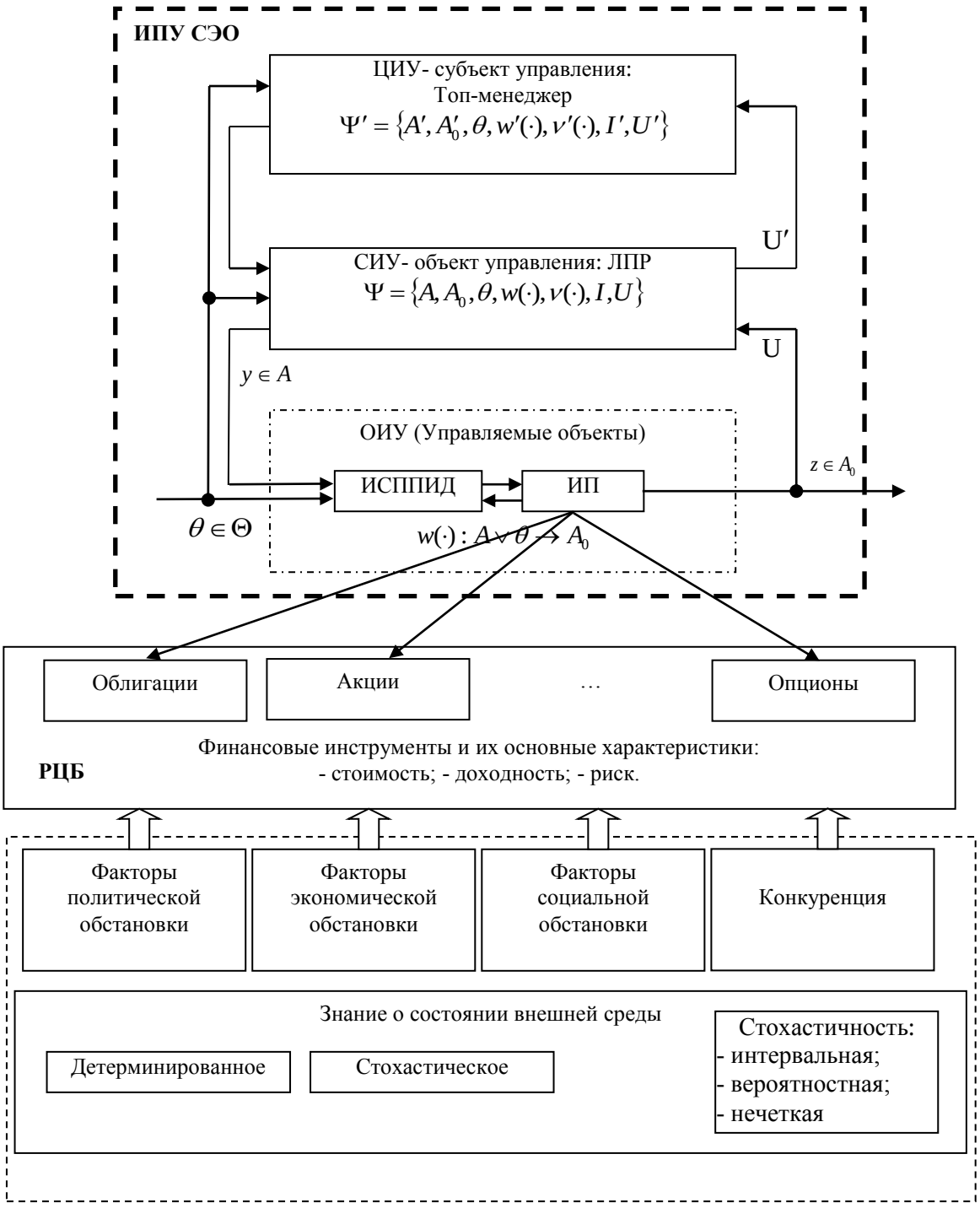


Рисунок 1.6. – Взаимодействие ЦИУ, СИУ, ОИУ, ИСППО с РЦБ и внешней средой

Формируя ИП, ЛПР реализует определенное действие *y* из множества

допустимых $y \in A$ (например, продает акции). Результат данного действия (измененный набор акций) $y\,(y \in A)$ представляется, как $z(y) \in A_0$, где A_0 - множество возможных результатов действий ЛПР. Вследствие стохастичности данных об активах РЦБ, желаемый (ожидаемый) результат действия (например, ожидаемая доходность портфеля) и фактический (реальная доходность портфеля) могут не совпадать. ЛПР при принятии решения сравнивает результаты своих действий, что выражается в его предпочтениях C_{A_0} на множестве $z(y) \in A_0$. Данные предпочтения C_{A_0} параметризируются переменной *c*, принимающей значения из подмножества Ω действительной оси, $\Omega \subseteq C$. Каждому возможному предпочтению ЛПР соответствует фиксированное значение параметра $c \in \Omega$ – финансового ресурса (ФР). Общий ФР *C*, выделяемый ЛПР для реализации ПИД в соответствии с финансовым планом, является ограниченным. Свои действия ЛПР выбирает в соответствии с некоторым законом $W_I(\cdot)$, учитывающим состояние внешней среды $\theta \in \Theta$ в виде $z = w(y, \theta)$, влияющее на результат, где I – переменная, характеризующая объем информации о состоянии внешней среды, которой владеет ЛПР, а также предпочтениями, задаваемыми функцией полезности $v(\cdot)$. Выбор действия ЛПР осуществляется в соответствии с правилом *P* [150]:

$$P^{W_1}(C_{A_0}, A, I) \subseteq A. \qquad (1.1)$$

Правило (1.1) базируется на гипотезе детерминизма, заключающейся в стремлении ЛПР снизить текущую стохастичность (неопределенность) с учетом информации *I,* которой он обладает и которую ему предоставляет система (ИСППИД) и принимать решения в условиях близких к детерминированным. Под текущей стохастичностью (неопределенностью) понимается объективная неопределенность внешней среды («неопределенность природы»), которая может быть интервальной, вероятностной и нечеткой. Установлено, что основной является интервальная стохастичность, парируемая ИСППИД. Закон $W_I(\cdot)$ определяется функцией $w(\cdot)$, зависящей от структуры ИП и информации *I*, которой обладает ЛПР на момент выбора действия [150].

В формализованном виде модель принятия решений СИУ СЭО в лице ЛПР представляется кортежем [150]:

$$\Psi = \{A, A_0, \theta, w(\cdot), \nu(\cdot), I, U\}. \quad (1.2)$$

Аналогичные рассуждения приводятся и для ЦИУ СЭО, объектом управления которого является ЛПР.

Основными управляющими функциями ЦИУ являются планирование, организация, стимулирование и контроль ЛПР. При этом определяющими являются функции организации (предоставление финансовых ресурсов на разработку C_0 и эксплуатацию C_e, а также инвестируемый капитал - C) и контроля (получение дивидендов – U') [150].

Модель принятия решений ЦИУ СЭО в формализованном виде, представляется кортежем [150]:

$$\Psi' = \{A', A_0', \theta, w'(\cdot), \nu'(\cdot), I', U'\}. \quad (1.3)$$

Требуется на множестве допустимых конфигураций (вариантов) $\{K_d\}$, синтезировать такой структурно-функциональный облик ИСППИД K, включающий оптимальный состав компонент N_k (модулей, подсистем), их взаимосвязей (структура) S_k, режимов работы J_k и характеристик функционирования X_k, для обеспечения решения множества задач $\{M_z\}$, максимизирующих значение доходности ИП D_p ПИД СЭО при заданных: уровне риска R_p; количестве прагматической информации, требуемой ЛПР Z_T; времени решения T_h; выделенного финансового ресурса C_1 на разработку всей системы.

Векторная форма представления облика ИСППИД СЭО имеет следующий вид:

$$K = < M_z, J_k, N_k, S_k, X_k >. \quad (1.4)$$

Существенное влияние на облик системы оказывают выделяемые ресурсы $R(K,U_p)$ и стратегии формирования ИП U_p. Для ИСППИД, выделяемые ресурсы, выраженные в стоимостном виде $C(K,U_p)$, ограничены C_1, то есть $C(K,U_p) \leq C_1$.

Важным показателем предпочтительности определенного варианта облика ИСППИД, по сравнению с остальными, является его эффективность W_{ki},

$i = \overline{1, I}$, представляющая собой оптимальное значение отношения степени соответствия его результирующих характеристик, требуемым, к величине затрат. Ее значение определяется обликом ИСППИД и стратегией формирования ИП $W_{ki}= W_{ki}(K,U_p)$.

Целевая функция (ЦФ) $F(K,U_p)$, решаемой проблемы характеризуется набором функциональных задач, обеспечивающих, с учетом накладываемых ограничений (заданного уровня риска $\Re_p$, количества прагматической информации, требуемой ЛПР Z_T, необходимого времени решения T_h), информационную поддержку ИФП, максимизирующую величину доходности D_p, формируемого ИП.

ИСППИД, характеризующаяся обликом $K^*=<M_z^*,\ J_k^*,\ N_k^*,\ S_k^*,\ X_k^*>$, обеспечивающая достижение максимального значения ЦФ, при минимальном уровне затрат, является предпочтительной. С учетом последнего, формализованная постановка проблемы синтеза ИСППИД имеет вид:

$$K^* = Arg \min_{K\in\{K_o\}} C(K, \mathrm{Up}) \tag{1.5}$$

$$\{K_d\}=\{K: W(K,U_p)\geq W_{ki},\ R(K,U_p) \supseteq R\}, \tag{1.6}$$

при:

$$R(M_z)\subseteq R_M; \tag{1.7}$$

$$R(J_k)\subseteq R_J; \tag{1.8}$$

$$R(N_k)\subseteq R_N; \tag{1.9}$$

$$R(S_k)\subseteq R_S; \tag{1.10}$$

$$R(X_k)\subseteq R_X; \tag{1.11}$$

$$R=R_M\cup R_J \cup R_N \cup R_S \cup R_X; \tag{1.12}$$

$$r \leq \Re_p; \tag{1.13}$$

$$z \leq Z_T; \tag{1.14}$$

$$t_r \leq T_h; \tag{1.15}$$

$$\theta \in \Theta = 2; \tag{1.16}$$

$$c\leq C_1, \tag{1.17}$$

где $\{K_d\}$ – множество возможных вариантов комплектаций ИСППИД; R – ограничения, накладываемые на разработку ИСППИД, в части задач $R(M_z)$,

режимов работы $R(J_k)$, состава $R(N_k)$, структуры $R(S_k)$ и характеристик $R(X_k)$.

Данная задача относится к классу оптимизационных многопараметрических задач с нелинейной ЦФ и множествами зависимых аргументов и ограничений. Решение подобных задач основано на их иерархической декомпозиции на составные элементы с применением правил: «часть-целое»; «зависимость целевых решений верхних уровней от качества решаемых задач на нижних»; «право влияния управляющего структурного элемента на действия остальных» и др. В технологическом плане их решение может представлять собой творческий итерационный возвратно-поступательный процесс принятия частных решений и учетом существующих ограничений, либо решение некоторой последовательности задач дискретного программирования, конечной целью которых является сформированный облик оптимальной ИСППИД.

Значительное влияние на облик ИСППИД оказывают различные внешние факторы. Из рисунка 1.6 видно, что на принятие решений ЛПР существенную роль играют внешняя среда и конкуренты. Априорная информация о состояниях внешней среды для ЛПР может быть представлена в виде следующих вариантов [210]:

- внешняя среда $\theta \in \Theta = 1$ является пассивной по отношению к целям ЛПР. Данное состояние может рассматриваться как детерминированный случай. Для $i = 1,...,m,$ известны априорные распределения вероятностей $p_i = (p_{i1},..., p_{iqj}),$ определенные на множествах $A_i = \left\{ p_i = (p_{i1},..., p_{iqj}) : 0 \le p_{ij} \le 1,\ j = 1,..., q_i, \sum_{j=1}^{q_i} p_{ij} = 1 \right\}$ элементов $s_{ij} \in S_i,\ j = 1,..., q_i,$ состояний среды;

- внешняя среда $\theta \in \Theta = 2$ оказывает активное противодействие целям ЛПР. Это характерно действиям со стороны конкурентов. Среда стремится принять такое состояние, при котором значение функции потерь максимально:

$$V_i = \left\| \upsilon_i(x, s_{ij}) \right\|, x \in X, s_{ij} \in S_i \to max.$$

Данное состояние рассматривается как неопределенность. Как известно, неопределенность бывает интервальной, вероятностной и нечеткой [190]. В [173, 188] показано, что ЛПР, стремится устранить (снизить) неопределенность, пытаясь свести задачу к детерминированной и принимать решение в детерминированных условиях. Методами максимального гарантированного результата, гипотезы благожелательности и комбинированными методами снимается интервальная неопределенность. Математическим ожиданием, введением риска, дисперсии, средним квадратическим отклонением и др. устраняется вероятностная неопределенность. Для устранения нечеткости используется выделение множества максимально недоминируемых действий [188].

Важную роль в процессе принятия решений в условиях неопределенности играет информация, которую ЛПР может получать из внешних (книги, публикации, интернет и др.) и внутренних (личные знания, внутренние эксперты и др.) источников информации. Однако решать аналитические вопросы и искать требуемую информацию в условиях жестких временных ограничений, ЛПР не в состоянии. Для этой цели необходима ИСППИД.

Таким образом, необходимо, чтобы ИСППИД являлась эффективным аналитическим средством формирования инвестиционных портфелей и средством добычи качественной прагматической информации в данной предметной области, что в конечном итоге должно снизить неопределенность внешней среды.

Качество функционирования ИСППИД для обеспечения ПИД СЭО характеризуется многоуровневой системой показателей эффективности (интегральных, интегрально-системных, системных, информационно-системных и информационных), приведенной в таблице 1.1 [150]. Их детализация приведена в главе 6 диссертационной работы.

Постановка задачи снижения неопределенности внешней среды на основе ИСППИД формулируется следующим образом [177].

Известно множество возможных значений состояния среды:

Таблица 1.1 – Система показателей эффективности ИСППИД СЭО

Интегральные показатели эффективности																	
Долевой коэффициент чистой прибыли – (W_k)																	
Интегрально-системные показатели эффективности																	
Портфельная доходность – (D_P) и портфельный риск – (R_P)																	
Системные показатели эффективности																	
Инвестиционная точность – (Q_I)				Инвестиционная оперативность – (T_I)					Понятийная точность – (Q_P)		Понятийная полнота – (L_P)		Понятийная оперативность – (T_P)				
Информационно-системные показатели эффективности																	
k_i^h	k_i^{nh}	k_i^m	k_i^g	T_p	T_h	T_{nh}	T_m	T_g	a_q^r	a_q^a	a_w^r	a_w^a	t_p^r	t_p^a	t_p^v	t_d^r	t_d^a	t_d^v
Коэффициент отклонения средней доходности ИП Шарпа	Коэффициент отклонения средней доходности ИП НКМШ	Коэффициент отклонения средней доходности ИП Марковица	Коэффициент отклонения средней доходности ИП МГА	Усредненное время управления ИП ЛПР	Усредненное время управления ИП Шарпа	Усредненное время управления ИП НКМШ	Усредненное время управления ИП Марковица	Усредненное время управления ИП МГА	Коэффициент идентификационной точности ручного способа	Коэффициент идентификационной точности на основе ПТП	Коэффициент идентификационной полноты ручного способа	Коэффициент идентификационной полноты на основе ПТП	Время построения ТПТ на основе ручного способа, с	Время построения ТПТ на основе ПТП, с	Время построения ТПТ на основе ПТП с использованием виртуального пространства, с	Время сбора новой информации на основе ручного способа, с	Время сбора новой информации с использованием ПТП, с	Время сбора новой информации на основе ПТП с использованием виртуального пространства, с
Информационные показатели эффективности																		
Объем детерминированных обрабатываемых данных - V_d и объем стохастических динамических данных - V_s									Объем текстовой информации - V_t				Время загрузки информации - T_r					
Объем финансовых ресурсов, выделенных на разработку - C_1																		

$S=\{s_1,...,s_l\}, l=1,...,L$. Считается, что реализуется наихудшее состояние природы. Для каждого состояния существует набор портфелей ЦБ *{X}*, $(x \in X)$, различающихся своей доходностью – $D=\{d_1,\dots d_e\}, e=1,\dots E$. Она

определяется в соответствии с выражением (1.18):

$$D_p = \sum_{i=1}^{M} \sum_{J=1}^{N} D_{ij} n_{ij} k_i^p , \qquad (1.18)$$

где D_p - доходность портфеля ЦБ; D_{ij} – доходность *j*-й ЦБ *i*-го типа; n_{ij} - количество *j*-х ЦБ *i*-го типа; k_i^p - коэффициент прагматической информативности для *i*-го типа ЦБ.

Коэффициент прагматической информативности (КПИ) указывает на наличие в ИСППИД совокупности данных, используемых для управления портфелем ЦБ. Управление ИП представляет собой совокупность действий ЛПР обеспечивающих его (ИП) формирование, манипулирование и оценку эффективности.

Значение КПИ находится в пределах от 0 до 1. Если $k_i^p = 1$, то это свидетельствует о том, что у ЛПР имеется вся требуемая информация о стратегиях манипулирования и характеристиках ЦБ *i*-го вида. Доходность ИП определяется как произведение доходности *i*-й ЦБ на их количество *M*. Если $k_i^p = 0$, то это свидетельствует об отсутствии у ЛПР необходимой информации о характеристиках ЦБ и стратегиях управления ИП. Это характерно для полной неопределенности внешней среды. В таком случае ЛПР ведет игру на рынке ЦБ «вслепую». Может случиться, что ему повезет и его ИП принесет доход. Он может не выиграть ничего. В этом случае у ИП будет нулевой доход. Он может проиграть, что бывает довольно часто. Тогда у ИП будет отрицательный доход. В практической деятельности, значение КПИ обычно лежит в диапазоне от 0 до 1, $0 \le k_i^p \le 1$.

Цель ИСППИД –достичь максимального значения k_i^p, т. е:

$$\lim_{i \to \infty} k_i^p = 1. \qquad (1.19)$$

ИСППИД должна накапливать полезную информацию для ЛПР об используемых ЦБ, стратегиях управления ИП и моделях его анализа, что снижает неопределенность о природной среде.

В свою очередь, значение КПИ зависит от семантической и синтаксической

мер информации. Формально это представляется в следующем виде:

$$k_i^p = g_{1i} k_i^{sn} + g_{2i} k_i^{se}, \qquad (1.20)$$

где k_i^{sn} – коэффициент синтаксической информативности (КСИИ) для ЦБ i-го типа, k_i^{se} – коэффициент семантической информативности (КСЕИ) для ЦБ i-го типа, g_{1i} и g_{2i} – весовые коэффициенты для ЦБ i-го типа.

КСИИ показывает, что данные, поступившие в ИСППИД в ν – й момент времени (t_ν), представляют собой логически связанную информацию. Однако не ясно, имеет ли она отношение к *i*-му типу ЦБ. Его значение определяется в соответствии с выражением:

$$k_i^{sn}(t_\nu) = \frac{k_i^{\inf}(t_\nu)}{V_i^d(t_\nu)}, \qquad (1.21)$$

где $k_i^{\inf}$ – количество информации в данных по *i*-му типу ЦБ, байт; V_i^d - объем данных по *i*-му типу ЦБ, байт.

Если ИСППИД в течение некоторого времени *T* накапливает информацию, то целесообразно использовать коэффициент накопленной синтаксической информативности (КНСИИ) k_{iT}^{sn}, рассчитываемый в соответствии с выражением:

$$k_{iT}^{sn} = \sum_{\nu=1}^{T} k_{i\nu}^{sn} = \sum_{\nu=1}^{T} \frac{k_{i\nu}^{\inf}}{V_{i\nu}^d}. \qquad (1.22)$$

КСЕИ показывает, что в приходящем в ν–й момент времени (t_ν) сообщении содержится логически связанная информация по данной теме. Его значение определяется в соответствии с выражением:

$$k_i^{se}(t_\nu) = \frac{k_i^{\sin f}(t_\nu)}{V_i^d(t_\nu)}, \qquad (1.23)$$

где $k_i^{\sin f}$ – количество тематической информации в сообщении по i-му типу ЦБ, байт; V_i^d – объем данных в сообщении по *i*-му типу ЦБ, байт.

Если ИСППИД в течение некоторого времени *T* накапливает информацию, то целесообразно использовать коэффициент накопленной семантической информативности (КНСЕИ) k_{iT}^{se}, рассчитываемый в соответствии с выражением:

$$k_{iT}^{se} = \sum_{\nu=1}^{T} k_{i\nu}^{se} = \sum_{\nu=1}^{T} \frac{k_{i\nu}^{\sin f}}{V_{i\nu}^{d}} \,. \qquad (1.24)$$

Для пояснения механизма снижения неопределенности внешней среды на основе ИСППИД, необходимо рассмотреть понятия тезауруса и терминологического портрета.

Тезаурус представляет собой совокупность сведений о понятиях данной предметной области, которыми располагает ЛПР или ИСППИД.

Первым отличием терминологического портрета от тезауруса, является то, что он представляет собой структурированную, иерархически связанную между собой систему понятий.

Пример иерархической структуры терминологического портрета МО управления ИП приведен на рисунке 1.7.

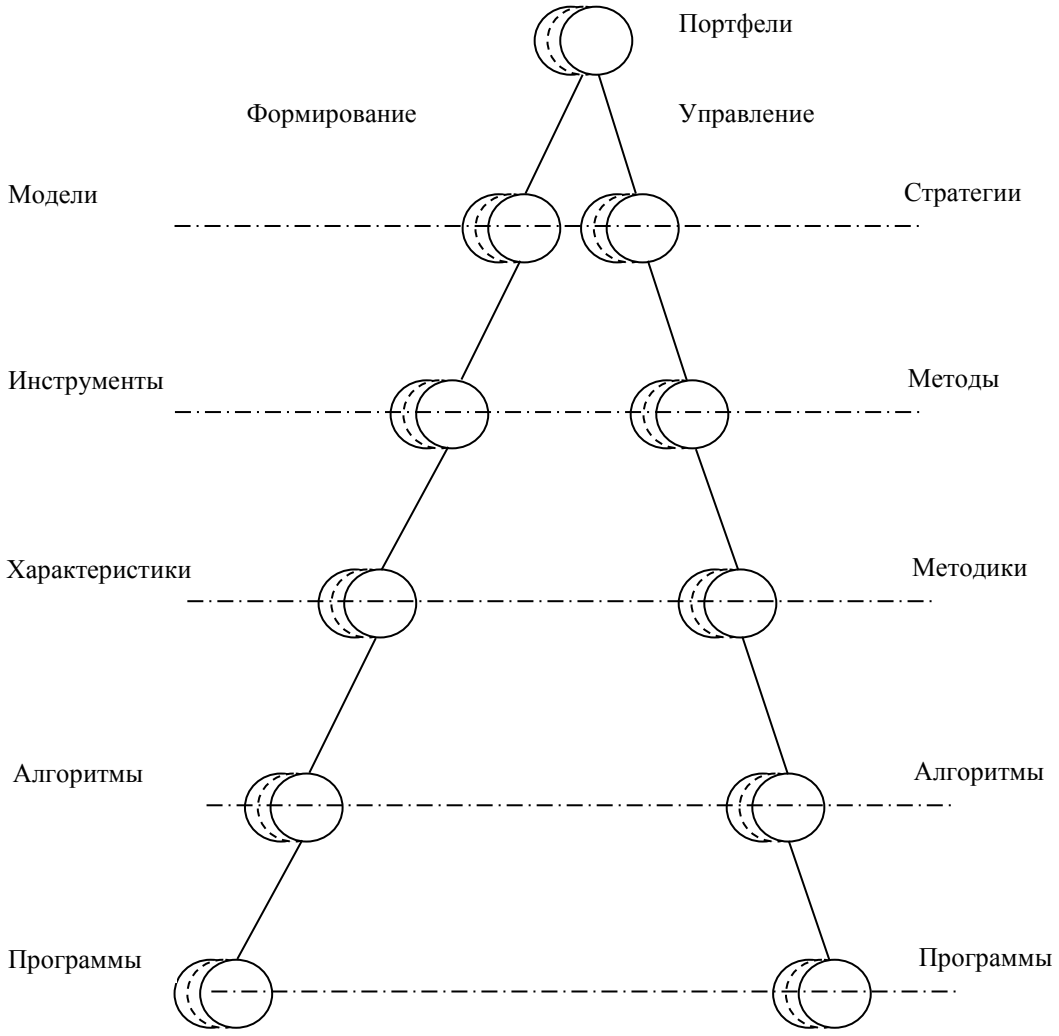


Рисунок 1.7. – Структура терминологического портрета портфеля ЦБ

Данная структура представляет собой сложенное шестиуровневое дерево, состоящее из ветвей «формирование» и «управление». Верхний (первый) уровень представлен портфелями. У ЛПР их может быть несколько. Второй уровень представлен моделями в рамках формирования портфелей и стратегиями в рамках управления ими (портфелями). Узловыми понятиями для моделей формирования портфелей являются: модель Марковица, модель Шарпа, нейрокомитетная модель Шарпа и др. Узловыми понятиями для стратегий управления портфелями являются: активное управление, пассивное управление и др. На третьем уровне для формирования портфелей представлены инструменты – виды ценных бумаг. К узловым понятиям данного уровня относятся облигации, акции и др.

Для управления портфелем на этом уровне представлены методы. Для них характерны свои понятия, например – диверсификация. Четвертый уровень представлен узловыми понятиями, представляющими характеристики используемых инструментов, например: курс, доходность, дюрация Маколея и др. Применительно к управлению, на данном уровне находятся методики. Пятый и шестой уровни представлены узловыми понятиями, обеспечивающими практическую реализацию понятий вышестоящих уровней. На пятом уровне в рамках, как формирования портфелей, так и для их управления, находятся алгоритмы, а на шестом – программы.

Вторым отличием терминологического портрета от тезауруса является то, что понятия, находящиеся на различных уровнях имеют свои весовые коэффициенты. Сумма весов всех понятий терминологического портрета должна быть равна 1.

На рисунке 1.8 приведена зависимость степени усвоения ЛПР новых понятий от их числа [68].

Число поступающих понятий в данной зависимости представлено коэффициентом семантической информативности k_i^{se} , а число усвоенных понятий - числом понятий, содержащихся в терминологическом портрете N_{ptp}.

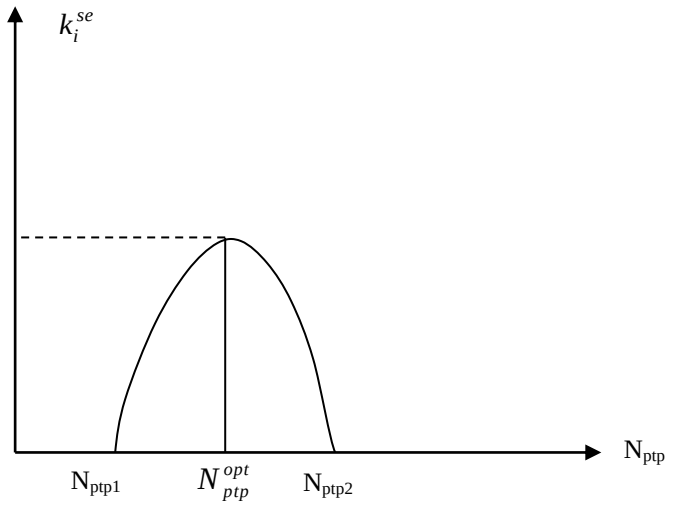


Рисунок 1.8. – Зависимость степени усвоения ЛПР новых понятий от их числа

Если рассмотреть предельные случаи, т.е когда N_{ptp}=0 (ЛПР ничего не знает о предметной области) и когда $N_{ptp} = \infty$ (ЛПР знает все о предметной области), то видно, что для них k_{it}^{se}=0. Это соответствует случаю, когда ЛПР ничего не воспринимает.

Из данного графика видно, что если поступающие сообщения согласованы с терминологическим портретом ЛПР, то степень восприятия их максимальна. В свою очередь, избыточное и несогласованное число поступающих сообщений с терминологическим портретом ЛПР приводит к резкому снижению эффективности восприятия. Следовательно, при функционировании ИСППИД, необходимо согласование между k_{it}^{se} и терминологическим портретом ЛПР. График приведенной зависимости носит квадратичный характер и представляет собой параболу. Обозначим точки пересечения графика данной функции с осью абсцисс соответственно N_{ptp1} и N_{ptp2}. Тогда уравнение представленной зависимости можно записать следующим образом:

$$k_{it}^{se} = (-N_{ptp} - N_{ptp1})(N_{ptp} - N_{ptp2}). \quad (1.25)$$

После некоторых преобразований, получим:

$$k_{it}^{se} = N_{ptp}^2 + N_{ptp}(N_{ptp1} - N_{ptp2}) + N_{ptp1}N_{ptp2}. \quad (1.26)$$

Подставив (1.26) в (1.20) получим выражение детального расчета КИП:

$$k_i^p = g_{1i}k_i^{sn} + g_{2i}(N_{iptp}^2 + N_{iptp}(N_{iptp1} - N_{iptp2}) + N_{iptp1}N_{iptp2}). \quad (1.27)$$

В соответствии с выражением (1.27) значение КИП рассчитывается для *i*–го понятия терминологического портрета ЛПР. Значение КИП для всего терминологического портрета ЛПР будет определяться в соответствии с выражением:

$$k^p = \prod_{i=1}^{M} (g_{1i} k_i^{sn} + g_{2i} (N_{iptp}^2 + N_{iptp} (N_{iptp1} - N_{iptp2}) + N_{iptp1} N_{iptp2})) . \qquad (1.28)$$

Если значение $k^p = 1$, то это свидетельствует о том, что в ИСППИД имеется полная информация о характеристиках используемых финансовых инструментов (ЦБ), моделях портфельного анализа и стратегиях его (портфеля) управления, требуемая для принятия решений. Неопределенность внешней среды устранена и ЛПР принимает решение в детерминированных условиях.

С учетом вышеизложенного, общая математическая постановка задачи снижения неопределенности внешней среды на основе ИСППИД, формулируется следующим образом.

Дано. Набор типов ЦБ – *M* и число ЦБ – *N i*-го которыми может воспользоваться ЛПР для формирования инвестиционного портфеля. Известен перечень характеристик ЦБ.

Существуют альтернативные портфели *{X}*, где $(x \in X)$, сформированные в результате учета ряда условий и ограничений. Для описания качества альтернативных портфелей используются два локальных критерия (характеристики): k^p – коэффициент прагматической информативности; t_r – время накопления информации. Внешняя среда описывается множеством дискретных состояний: $S_l = \{s_{l1}, ..., s_{lq}\}, l = 1, ..., m$. Известно, что состояние внешней среды в данный момент времени стремится минимизировать значение функции доходности портфеля:

$$D_{pl} = \sum_{i=1}^{M} \sum_{j=1}^{N} D_{ij} n_{ij} k_i^p \to \min \qquad (1.29)$$

Для множества альтернатив *{X}* и состояний среды *{S}* существует множество значений коэффициента прагматической информативности k_i^p и число понятий N_{iptp}, содержащихся в терминологическом портрете. Для каждого набора k_i^p и N_{iptp}, т. е альтернативы x_i существует свое значение доходности:

$$D_i = \left\| d_i(k_i^p, N_{iptp}, s_{ij}) \right\|, k_i^p \in K^p, N_{iptp} \in N_{ptp}, s_{ij} \in S_i . \quad (1.30)$$

Найти: управляющее воздействие y^*, при котором:

$$y^* \in Arg \max_{y \in A} \min_{s_{il} \in S} \sum_{i=1}^{M} \sum_{j=1}^{N} \prod_{l=1}^{L} D_{ij} z_{ij} k_l^p \quad (1.31)$$

при этом время решения задачи t_r не должно превышать требуемое T_h:

$$t_r \le T_h . \quad (1.32)$$

Данная задача относится к классу принятия многокритериальных решений в условиях неопределенности. Учитывая факт использования человеко-машинной системы - ИСППИД, для поиска лучших альтернатив целесообразно использовать один из методов выбора наилучшей паретовской точки, например, метод ограничений - STEM.

Реализация метода ограничений (STEM) заключается в выполнении следующей последовательности действий.

1. Исследование области допустимых значений; оптимизация по каждому из критериев; определение вектора, объединяющего оптимальные значения для каждого критерия $Z_{\max}$.

2. Определение весов критериев; оптимизация свертки критериев; получение вектора Z_{opl}.

3. Диалог с ЛПР по $Z_{\max}$. Определение «хороших» и «плохих» компонент Z_{opl}.

4. Выбор Z_i с наименее удовлетворительным значением. Определение ЛПР удовлетворительного значения для Z_i.

5. Максимизация значения доходности портфеля D_{pl} с учетом значения коэффициента информационной прагматичности k_i^p при ограничении времени накопления информации - t_r.

6. Выбор ЛПР ограничений на k_i^p; переход к новой области допустимых значений.

7. Если ЛПР удовлетворено решением, то следует остановка поиска; в

противном случае необходимо перейти к п. 1.

В таблице 1.2 приведены основные способы снижения неопределенности внешней среды [17].

Таблица 1.2 – Способы снижения неопределенности внешней среды

Наименование	Краткое содержание
Гарантированного результата	При проведении расчетов учитывается наихудшее значение неизвестного параметра, и ищутся способы и средства максимизации целевой функции
Оптимистичного результата	Целевая функция оптимизируется для наилучших условий, применительно к неизвестному параметру
Статистический	Его реализация предполагает знание статистических характеристик неизвестного параметра. Снижение неопределенности происходит путем усреднения целевых функций по известному распределению неизвестного параметра
Асимметричной информированности	Один из исполнителей лучше осведомлен о неизвестном параметре, чем проект – менеджер. В этом случае проект – менеджер может устранить неопределенность, попросив исполнителя сообщить информацию о неизвестном параметре и использовать эту информацию при принятии управленческих решений. При этом возникает задача манипулирования
Сравнения	Способ сравнения результатов деятельности одинаковых исполнителей в одинаковых условиях. При этом возможно получение информации о неизвестных условиях деятельности исполнителей.
Системной информированности (предлагаемый)	Все участники проекта не осведомлены о значении неизвестного параметра. В роли информатора в проекте выступает система поддержки принятия решений. Она реализует поиск, обработку и получение информации из различных источников (Интернет, научная литература, СМИ и др.) о возможном значении неизвестного параметра или способах его расчета. В отличие от остальных участников проекта, СППР не обладает свойством манипулирования (не ищет выгоды для себя, и по этой причине не скрывает истинного положения дел).

Решение данной задачи позволит дополнить способ асимметричной информированности, предложенный В.Н. Бурковым в [17], способом системной информированности. В данном случае информирует ИСППИД. Отличительной особенностью данного способа является отсутствие манипулирования (ИСППИД не ищет выгоды для себя, и по этой причине не скрывает истинного положения дел).

Дальнейшим развитием данной модели является вероятностная модель учета неопределенности внешней среды. Она позволяет оценить объем требуемых ресурсов для разработки ИСППИД.

1.5. Вероятностная модель учета влияния неопределенности внешней среды при синтезе информационной системы поддержки портфельной инвестиционной деятельности

Впервые данная модель была разработана применительно к системам интегрированного менеджмента [91, 96, 97, 98, 100, 122, 138, 151-153].

Интегрированный менеджмент (ИМ) является дальнейшим развитием классического [121, 179]. Его основное назначение – обеспечить стабильное функционирование СЭО в сложных экономических условиях (в условиях предкризисных и кризисных состояниях внешней экономической среды) [139]. Он базируется на мониторинге внешней среды, анализе процессных состояний СЭО (предыдущих, текущих и прогнозных), синтезе и реализации эффективных оперативных и упреждающих мероприятий (проектного управления, формы хозяйствования 5С, комплексирования методов антикризисного управления, инновационно - инвестиционного проектирования) [90, 95, 99, 121, 152, 153].

Структурная схема информационной системы интегрированного менеджмента (ИСИМ) СЭО представляет собой надсистему для системы управления традиционного менеджмента. В схематичном виде она представлена на рисунке 1.9 [140, 152].

Данная система более узконаправленна и предназначена для информационного обеспечения решения задач ИМ. ИСИМ может функционировать в режимах прямого и косвенного воздействий. На начальном этапе её формирования, преобладает косвенный режим работы. В данном режиме работы ЛПР субъекта управления традиционного менеджмента сложившегося в СЭО, периодически обращается за консультацией к субъекту управления интегрированного менеджмента [152].

По мере настройки ИСИМ, начнет преобладать режим прямого воздействия. Технология его аналогична вышеизложенной, только объектом воздейст-

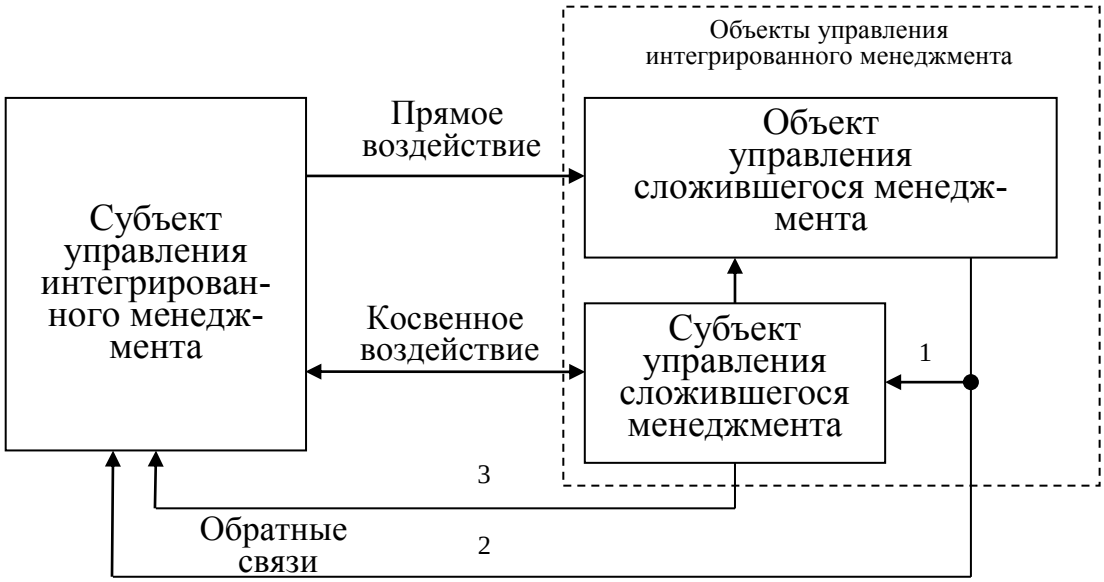


Рисунок 1.9. – Структурная схема ИС интегрированного менеджмента

вия будут непосредственно объекты управления системы традиционного менеджмента [122, 152, 175].

ИСППИД и ИСИМ подобны, в силу следующих трех основных причин.

Во-первых, методологические основы ИМ и предметной области принятия инвестиционных решений, где используются данные системы в качестве инструментов, достаточно близки [97, 98, 122].

Во-вторых, в рамках технологии применения, ИСППИД и ИСИМ имеют общий режим работы (косвенный).

В-третьих, набор их функциональных подсистем практически одинаков [143, 172]. Различие заключается в способах реализации подсистем прогнозирования. В ИСППИД для прогнозирования временных рядов используются ИНС (ПКНПВР) [94], а в ИСИМ - набор моделей и алгоритмов, базирующихся на других принципах [85, 212].

Подобие данных систем позволяет использовать соответствующие наработки, сделанные для ИСИМ [152].

Возвращаясь к содержанию вероятностной модели учета влияния неопределенности внешней среды при разработке и проектировании ИСППИД, следует отметить следующее. В интересах повышения качества выдаваемой информации и оперативности ее получения, необходимо определить объем

оптимальных выделяемых ресурсов (финансовых, временных и др.) на разработку ИСППИД, которые (ресурсы), в свою очередь, зависят от ее (ИСППИД) структурных и функциональных особенностей. Фактически имеет место оптимизационная задача, постановка которой может быть сформулирована следующим образом [153].

Пусть имеется ИСППИД, состоящая из множества подсистем $\{S\}$. Для каждой подсистемы $\mathrm{m}_i \in \{S\}, i = \overline{1;N}$ заданы [153]:

- множество необходимых операций, выполняемых *m* подсистемой для достижения поставленной цели:

$$\mathrm{O}_0^{\mathrm{m}} = \{\mathrm{o}_1^{\mathrm{m}},\ldots,\mathrm{o}_v^{\mathrm{m}},\ldots,\mathrm{o}_{n_1^{\mathrm{m}}}^{\mathrm{m}}\}, \quad \mathrm{m} \in S;$$

- множество типов требуемых ресурсов для реализации множества необходимых операций *m* подсистемой:

$$\mathrm{R}_t^m = \{\mathrm{r}_{1t}^m,\ldots,\mathrm{r}_{lt}^m,\ldots,\mathrm{r}_{n_2^m t}^m\}, \quad m \in S;$$

- множество типов наличных ресурсов для реализации множества необходимых операций *m* подсистемой:

$$\mathrm{R}_n^m = \{\mathrm{r}_{1n}^m,\ldots,\mathrm{r}_{\mathrm{ln}}^m,\ldots,\mathrm{r}_{n_2^m n}^m\}, \quad m \in S;$$

- множество временных интервалов функционирования подсистем:

$$T^m = \{\Delta t_1,\ldots,\Delta t_{n_4}\},$$

где $\Delta t_i = |t_{in} - t_{ik}|$, t_{in} - момент начала функционирования подсистемы на *i*-м интервале, t_{ik} - момент окончания целевого функционирования подсистемы на *i*-м интервале.

Введем в рассмотрение следующие вероятности [153]:

- вероятность выполнения требуемых операций *m* подсистемой для решения поставленных задач - P_o^m;

- вероятность наличия требуемого ресурса для выполнения требуемых операций *m* подсистемой - P_r^m;

- вероятность появления дополнительных операций требуемых для решения поставленных задач *m* подсистемой, обусловленных влиянием неопределенности внешней среды - P_{do}^m;

- вероятность выполнения дополнительных операций *m* подсистемой - P_{vdo}^{m};

- вероятность наличия требуемого ресурса для выполнения дополнительных операций m подсистемой - P_{dr}^{m}.

Вероятность выполнения *i*-й задачи *m* подсистемой в условиях неопределенности будет выражаться следующим образом [153]:

$$P_i^m = P_o^m \, P_r^m \, P_{do}^m \, P_{vdo}^m \, P_{dr}^m .$$

Данная вероятность выступает в роли показателя качества функционирования *m* подсистемы. Показатель качества функционирования всей системы, представляющий собой вероятность достижения цели ОС в условиях неопределенности, определяется в соответствии с формулой (1.33) [153]:

$$P_c = \prod_{m=1}^{S} \prod_{i=1}^{K} P_{io}^m P_{ir}^m P_{ido}^m P_{ivdo}^m P_{idr}^m . \qquad (1.33)$$

Функция ресурсного обеспечения системы R^S может быть представлена в следующем виде [153]:

$$R^S = f(O_o, O_d) = \sum_{i=1}^{K} (r_{io} + r_{id}), \qquad (1.34)$$

где O_o – объем ресурсов, требуемых системой для достижения цели, в детерминированных условиях, O_d – объем дополнительных ресурсов, требуемых системой для достижения цели, в условиях неопределенности; r_{io} – объем ресурсов, требуемых *i* –й подсистемой для достижения целевого назначения, в детерминированных условиях; r_{id} - объем дополнительных ресурсов, требуемых *i* –й подсистемой для достижения целевого назначения в условиях неопределенности.

В условиях неопределенности внешней среды требуется уточнить число дополнительных операций, необходимых для достижения поставленной цели, а, следовательно, объем дополнительных ресурсов. Возможности ИСППИД по уточнению числа дополнительных операций зададим соответствующей вероятностью – P_{ydo}^m. При этом выражение для показателя качества

функционирования системы в условиях неопределенности для случая использования ИСППИД примет вид [153]:

$$\hat{P}_c = \prod_{m=1}^{S} \prod_{i=1}^{K} P_{io}^m P_{ir}^m P_{iydo}^m P_{ido}^m P_{ivdo}^m P_{idr}^m . \quad (1.35)$$

В представленной модели показатель качества функционирования системы (1.35) выступает в роли ограничения, а функция ресурсного обеспечения (1.34) является целевой функцией. В символьном виде данная оптимизационная задача может быть представлена следующим образом [153]:

$$R^S = f(O_o, O_d) = \sum_{i=1}^{K} (r_{io} + r_{id}) \quad \to \min_{r_d}, \quad (1.36)$$

$$\hat{P}_c \geq P^+, \quad (1.37)$$

где P^+ – пороговое значение функции ограничений (требуемое значение вероятности выполнения системой поставленной цели).

Данная оптимизационная задача относится к классу дискретных с нелинейной целевой функцией и невыпуклой функцией ограничений [148].

Для решения данной задачи могут использоваться следующие методы: ветвей и границ, динамического программирования, дихотомического программирования и др. Исходя из простоты своей реализации, для решения данной задачи целесообразно использовать метод дихотомического программирования, описанного Бурковым В.Н. [16]. Детализация данного метода и технология применения приведены в разделе 1.8 диссертационной работы.

Результатами решения данной задачи могут быть синтезированные алгоритмы построения оптимальной структуры ИСППИД, оптимизации ее функционала и др.

Предложенная модель в теоретическом плане найдет применение при обосновании требований к ИСППИД на этапе ее синтеза.

На начальном этапе синтеза, к ИСППИД предъявляется ряд требований [46, 219, 220].

1.6. Требования, предъявляемые к информационной системе поддержки портфельной инвестиционной деятельности социально-экономической организации

Одной из прагматических целей разработки ИСППИД является формирование качественной информации. Обеспечение качественной информации для ЛПР является первым требованием, предъявляемым к ИСППИД. Качество информации представляет собой набор свойств (показателей), удовлетворяющих потребности ЛПР при принятии решений [130, 219, 220]. В схематичном виде они представлены на рисунке 1.10.

Рисунок 1.10. – Показатели качества информации

Репрезентативность информации представляет собой показатель, характеризующий адекватность представления свойств объекта (понятия). Встречается также ее определение, как правильность отбора и компоновки информации с целью наиболее адекватного описания объекта (отражения его основных свойств) [46]. Репрезентативность информации зависит от корректности формулировки исходного используемого понятия и отбираемых существенных связей и признаков. Область финансовых инвестиций не нова. Сформулированные понятия данной предметной области, а также их существенные признаки и связи являются установившимися. Они проверены практикой и временем. Учет данного показателя целесообразен в виде соответствующих справок и глоссария, включаемых в состав ИСППИД [171]. Особенно это важно для молодых специалистов, начинающих работать в данной предметной области.

Содержательность информации интерпретируется как отношение количества семантической информации к общему объему сообщения. Данный показатель уже учтен в коэффициенте семантичности.

Полнота информации представляет собой минимальный объем сведений о минимальном наборе необходимых характеристик, достаточном для принятия решений. Данный показатель учитывается в ИСППИД на этапе формирования терминологического портрета предметной области. Следует отметить, что данный показатель относительный и изменяется после получения дополнительных сведений и принятия решения о возникновении нового понятия и включении его в терминологический портрет.

Под доступностью информации понимается ее представление в понятном для ЛПР виде. Данный показатель заложен на этапе синтеза ИСППИД.

Актуальность информации характеризует степень сохранения ее ценности на момент использования. Данный показатель зависит от степени «устаревания» информации и тесно связан с оперативностью ее добывания. Динамика старения информации зависит от используемого финансового инструмента (вида ЦБ). Показатель оперативности добывания информации является важной характеристикой ИСППИД и в виде соответствующего ограничения $t_r \le T_z$, указывающего на то, что время решения задач и реализации принятых решений t_r не должно превышать требуемое T_z, используется в общей постановке задачи.

Точность информации характеризует степень соответствия полученной информации реальному состоянию понятия. В рамках ИСППИД используется номинальная точность, зависящая от функционального назначения применяемого показателя.

Достоверность информации используется в интересах оценки возможности ее получения с заданной точностью. Для измерения достоверности используется доверительная вероятность, показывающая, что представленное значение параметра (характеристики) и его истинное значение находятся в пределах заданной точности.

Устойчивость информации показывает, что определенные изменения данных не приводят к потере ее точности.

Требуемая достоверность и устойчивость в ИСППИД достигается использованием множества источников информации, проведением расчетных экспериментов и опытом ЛПР.

Обеспечение оперативности формирования информации является вторым требованием, предъявляемым к ИСППИД. Важно обеспечить оперативность проведения финансового анализа (выбор модели оценки, проверка полученных результатов, корректировка параметров и др.) и оперативность получения информации, связанной с терминологическим портретом (уточнение понятий, получение новых знаний и др.). Достижение необходимой оперативности зависит от алгоритмов и процедур обработки информации [161]. Кроме того, оказывают влияние используемые технические средства, операционная система и среда разработки программ.

Возможность наращивания функциональных возможностей – третье требование, предъявляемое к ИСППИД. Оно достигается использованием открытой модульной структуры ИСППИД, позволяющей реализовать на начальном этапе ее прототип (ИСППИД с ограниченным функционалом), а затем полнофункциональную ИСППИД, путем встраивания соответствующих программных модулей. Подобную реализуемость обеспечивает открытый интерфейс, основанный на стандартизации и унификации [69, 208].

Четвертое требование, предъявляемое к ИСППИД, связано с выбором специального математического и программного обеспечения (СМПО). К специальному математическому обеспечению относятся модели и алгоритмы, реализующие полный функционал ИСППИД. Полный функционал включает основной, связанный с решением основных задач, и вспомогательный, связанный с обеспечением решения основных задач (хранение и резервирование данных, оптимизация БД и запросов и др.). Суть данного требования состоит в выборе рациональных моделей и алгоритмов, обеспечивающих: формирование качественной информации (первое требование), опера-

тивный финансовый анализ и управление инвестиционным портфелем (второе требование), рациональное решение вспомогательных задач [145].

Удовлетворение четвертого требования предполагает определение критериев оптимизации синтеза ИСППИД СЭО.

1.7. Критерии оптимизации, постановка и решение задачи структурно-функционального синтеза систем поддержки портфельной инвестиционной деятельности социально-экономической организации

К основным видам обеспечения ИСППИД СЭО относятся: специальное математическое (*M)*; информационное (*I)*; программное (*P)*; техническое (*T)* [46, 48, 213]. Каждый вид обеспечения вносит свой вклад в общую эффективность функционирования ИСППИД СЭО. Учитывая вышеизложенные требования к ИСППИД СЭО, целесообразно использовать следующие критерии ее функционирования:

- точность финансового анализа – δ_i;

- время финансового анализа – t_a;

- стоимостные затраты на разработку СМПО, информационного и технического обеспечения - c_Σ.

Точность анализа финансовых инвестиций представляет собой величину обратную ошибке ε_r между прогнозными и фактическими значениями стоимостных характеристик выбранных финансовых инструментов (курсы валют, котировки акций и др.). Данный параметр в рассматриваемой задаче будет выступать в качестве критерия целевой эффективности. Он зависит от значения коэффициента прагматической информативности k^p и характеристик, используемых видов специального математического, программного, информационного и технического обеспечения.

Специальное математическое обеспечение (СМО) представляет собой множество математических моделей, методов и алгоритмов, необходимых для управления ИП. В символьной нотации СМО M_O может быть представлено следующим образом [173]:

$$M_O = \left\{ M_d^a, M_e^a, A_l^a \right\}, \quad (1.38)$$

где M_d^a – модели, M_e^a – методы и A_l^a – алгоритмы управления ИП.

Работоспособность используемого на практике СМО управления ИП обеспечивают программы, для реализации элементов которых используется соответствующий программный код. Совокупность программ представляет собой программное обеспечение P_o. Оно состоит из специального программного обеспечения P_{os}, непосредственно реализующего используемые алгоритмы, и общего программного обеспечения P_{oo} (операционная система, среда программирования и т.д.), используемого при разработке специального программного обеспечения $P_o = \{P_{os}, P_{oo}\}$ [173].

ИО включает в себя выбранную технологию обработки и взаимообмена данными [220]. В символьном виде ИО представляется как [173]:

$$I_o = \{I_m\}, \quad (1.39)$$

где I_m – модель представления данных при проведении ПИА.

Техническое обеспечение - комплекс технических средств (ПК, средства телекоммуникаций и др.), используемых для решения задач управления ИП.

Значения характеристик специального математического и информационного обеспечения, а также значение коэффициента прагматической информативности в решаемой задаче будут выступать в качестве варьируемых параметров, а значения характеристик программного и технического обеспечения – в качестве ограничений.

Как отмечалось выше, информация, необходимая для принятия решений, должна предоставляться пользователю своевременно или даже с некоторым временным запасом (заблаговременно). Поэтому время анализа t_a, характеризующее оперативность проведения расчетов, не должно превышать требуемое. При задании требуемого времени должны учитываться временные параметры всего процесса ПИУР. Данный параметр в решаемой задаче будет выступать также в качестве ограничения [173].

Учитывая вышесказанное, общая научная задача структурно-

функционального синтеза ИСППИД формулируется следующим образом.

Максимизировать точность принимаемых ПИУР ЛПР δ, путем оптимизации параметров специального математического $M_O = \{M_d^a, M_e^a, A_l^a\}$ и информационного $I_o = \{I_m\}$ обеспечения, а также значения коэффициента прагматической информативности - k^p, синтезируемой ИСППИД, с учетом заданных суммарных затрат C на ее разработку. При этом время проведения анализа t_a (оперативность) не должно превышать требуемое время T_T [173]:

$$\delta\left(k^p, M_d^a, M_e^a, A_l^a, I_m\right) \to \max, \tag{1.40}$$

$$C \le C_\Sigma \tag{1.41}$$

$$t_a \le T_T. \tag{1.42}$$

Фактически задача (1.40)-(1.42) представляет собой поиск оптимального варианта комплектации ИСППИД СЭО, удовлетворяющего вышеизложенным требованиям (точность-цена-время). Для решения данной задачи введем двоичную переменную x_{ij}. Она отражает следующие моменты.

Если $x_{ij} = 1$, то для i-ой ИСППИД принят j-ый вариант комплектации, в противном случае $x_{ij} = 0$. При этом, сумма затрат на j-ый вариант комплектации будет составлять c_{ij}. Под комплектацией j-го варианта будем понимать набор $< k_j^p, M_{dj}^a, M_{ej}^a, A_{lj}^a, I_{mj} >$.

С учетом вышеизложенного, задача (1.40)-(1.42) может быть представлена в упрощенном виде:

$$\begin{aligned} &\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}x_{ij} \to \min, \\ &\sum_{i=1}^{n}\sum_{j=1}^{m} T_{ij}x_{ij} \ge T, \\ &\sum_{j=1}^{m} x_{ij} = 1,\ i = \overline{1,n}, \end{aligned} \tag{1.43}$$

где T – нормативное значение оценки точности; n – количество ИСППИД; m –число вариантов комплектации.

Вербальное описание данной задачи представляется следующим образом: определить такие варианты комплектации ИСППИД СЭО, при которых нормативное значение оценки точности превышало бы заданную величину и при этом значения затрат были бы минимальными.

Данная задача относится к классу математических задач комбинаторного программирования. Решение данной задачи возможно несколькими методами, в частности: методом ветвей и границ, методом динамического программирования, методом дихотомического программирования и др. В силу наибольшей эффективности, для решения данной задачи целесообразно использовать метод дихотомического программирования, разработанный Бурковым В.Н [16].

В интересах упрощения расчетов, чтобы не рассматривать 3-х мерные результирующие таблицы, зафиксируем значение требуемого времени t_a решения задачи. Для других его значений будут аналогичные таблицы с другими числовыми значениями.

Рассмотрим случай для четырех версий ИСППИД с тремя вариантами их комплектации. Данные числовые значения, определяющие размерность задачи, приняты с целью облегчения проведения ручных вычислений.

Исходные данные для решения примера оптимизации ИСППИД методом дихотомического программирования представлены в таблице 1.3 [16].

Таблица 1.3 – Исходные данные для примера

i / Вариант	1	2	3	4
I вариант	**2/200**	**1/100**	**4/400**	**10/1000**
II вариант	**3/400**	**3/150**	**8/600**	**15/1500**
III вариант	**4/400**	**5/200**	**13/800**	**17/2000**

В числителе ячеек таблицы представлены данные о затратах, а в знаменателе – значения бальных экспертных оценок точности.

Графическая интерпретация поиска решения для рассматриваемого примера приведена на рисунке 1.11 [16].

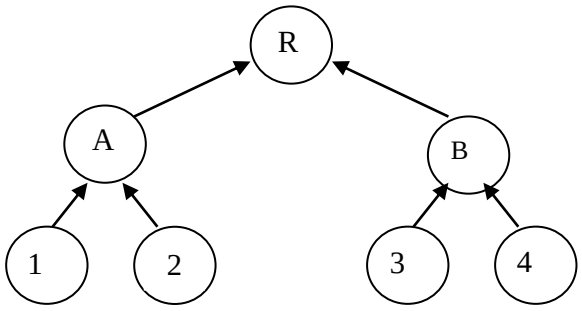


Рисунок 1.11 – Графическая интерпретация примера поиска решения

Из приведенного графа видно, что если рассматривается одна ИСППИД, то решение тривиальное. Вариант комплектации однозначно определяется с учетом наличных ресурсов. Если рассматриваются две ИСППИД, то возможно несколько вариантов их комплектации. На приведенном рисунке это вершина А. Для рассматриваемого примера используются вершины A-R-B. Возможные варианты комплектации составленные для наборов А и В представлены в таблицах 1.4 и 1.5.

Таблица 1.4 – Варианты комплектаций для набора А

5 / 200	**7 / 400**	**8 / 500**	**9 / 600**
⇒ **3 / 150**	**5 / 350**	**6 / 450**	**7 / 550**
1 / 100	**3 / 300**	**4 / 400**	**5 / 500**
2 / 1	**2 / 200**	**3 / 300**	**4 / 400**

Таблица 1.5 – Варианты комплектаций для набора В

17 / 2000	**21 / 2400**	**25 / 2600**	**30 / 2800**
⇒ **15 / 1500**	**19 / 1900**	**23 / 2100**	**28 / 2300**
10 / 1000	**14 / 1400**	**18 / 1600**	**23 / 1800**
4 / 3	**4 / 400**	**8 / 600**	**13 / 800**
			⇑

Проводя отбор конкурентоспособных вариантов, используя результаты приведенные в таблицах 1.4 и 1.5, можно определить значения комплектаций

для наборов *A* и *B*. Они представлены в таблицах 1.6 и 1.7. При определении комплектаций учитывалось, что для каждой ИСППИД возможен единственный вариант комплектации. Случай, когда одна из ИСППИД не функционирует и поэтому затраты на нее нулевые, а другая ИСППИД укомплектована по улучшенному варианту, не рассматривался.

Таблица 1.6 – Значения комплектаций для набора *A*

Затраты	5	7	8	9
Оценка точности	350	400	500	600

Таблица 1.7 – Значения комплектаций для набора *B*

Затраты	14	18	23	28	30
Оценка точности	1400	1600	2100	2300	2800

На основе данных, представленных в таблицах 1.6 и 1.7, определим значения комплектаций для набора *R*, содержащего *A* и *B*. Полученные результаты приведены в таблице 1.8.

Таблица 1.8 – Варианты комплектаций для набора *R*

30 / 2800	**35 / 3150**	**37 / 3200**	**38 / 3300**	**39 / 3400**
⇒ **28 / 2300**	**33 / 2650**	**35 / 2700**	**36 / 2800**	**37 / 2900**
23 / 1800	**28 / 2150**	**30 / 2200**	**31 / 2300**	**32 / 2400**
18 / 1600	**23 / 1950**	**25 / 2000**	**26 / 2100**	**27 / 2200**
14 / 1400	**19 / 1750**	**21 / 1800**	**22 / 1900**	**23 / 2000**
B / A	**5 / 350** ⇑	**7 / 400**	**8 / 500**	**9 / 600**

На основании результатов представленных в таблице 1.8, можно реализовать оптимальную стратегию комплектации четырех рассматриваемых ИСППИД, с учетом выделенного объема финансовых ресурсов. С этой целью в таблице 1.8 находим ячейку с записанной дробью, значение числителя ко-

торой максимально приближено к значению выделенного объема финансовых ресурсов. Значение знаменателя найденной дроби будет соответствовать величине максимально возможной оценке точности. Из этой же таблицы находим значения наборов A и B для соответствующего значения оценки точности, как абсциссу и ординату точки пересечения найденной ячейки. Далее осуществляются переходы в таблицы 1.4 и 1.5 для наборов A и B соответственно. По значениям найденных координат точек пересечения определяются соответствующие варианты комплектации.

Рассмотрим применение вышеизложенной стратегии на конкретном числовом примере. Допустим, что имеется финансовый ресурс в размере 33 единиц. В таблице 1.8 данному ресурсу соответствует значение максимально возможной оценки точности равное 2650. Ячейка, содержащая данную дробь в таблице 1.8, выделена серым цветом. Координатами ячейки с данной дробью являются значения $\frac{\mathbf{5}}{\mathbf{350}}$ и $\frac{\mathbf{28}}{\mathbf{2300}}$, которые отмечены стрелками. Далее в таблице 1.4 находим ячейку с дробью $\frac{\mathbf{5}}{\mathbf{350}}$ (выделена серым цветом). Это значение промежуточной свертки A. Координаты этой ячейки (обозначены стрелками) соответствуют стратегии комплектования. Ее суть заключается в том, что первой ИСППИД соответствует 1 вариант комплектации, а второй ИСППИД вариант 2. Аналогичным образом определяются стратегии комплектования для третьей и четвертой ИСППИД. При этом используется ячейка с дробью $\frac{\mathbf{28}}{\mathbf{2300}}$ в таблице 1.5. Из нее видно, что третьей ИСППИД соответствует 3 вариант комплектации, а четвертой ИСППИД вариант 4.

На основе метода дихотомического программирования может быть решена многокритериальная задача, представляющая большой практический интерес. В данной задаче требуется оптимизировать два противоречивых критерия: оценки точности и величины затрат. Требуется выбрать такой вариант комплектации ИСППИД, при котором оценка точности принимает максимальное значение, а размер финансовых ресурсов должен быть мини-

мальным. В символьном виде это представляется следующим образом [16]:

$$\begin{cases} \sum_{i=1}^{n}\sum_{j=1}^{m} T_{ij}x_{ij} \to \max, \\ \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}x_{ij} \to \min \\ \sum_{j=1}^{m} x_{ij} = 1,\ i = 1,\overline{n}\ , \end{cases} \quad . \qquad (1.44)$$

Для решения подобных многокритериальных задач применяют методы интегральной оценки. При этом стремятся привести исходную задачу к разновидности задач математического программирования. Однако получение интегральной оценки при подобном сведении является субъективным и сложным. Более интересен следующий подход. Формируется ограниченный набор конкурентоспособных решений оптимальных по Парето. Из данного набора ЛПР выбирается единственное, которое в максимальной степени удовлетворяет сложившейся ситуации.

В решаемой задаче (1.33) формирование оптимальных решений по Парето осуществляется следующим образом.

Построим на плоскости оптимальных решений по Парето, представленной на рисунке 1.12, «антиидеальную» точку с координатами (C_{max}; T_{min}). Ей соответствует ИСППИД с максимальными затратами (C_{max}) и минимальной точностью (T_{min}) [16].

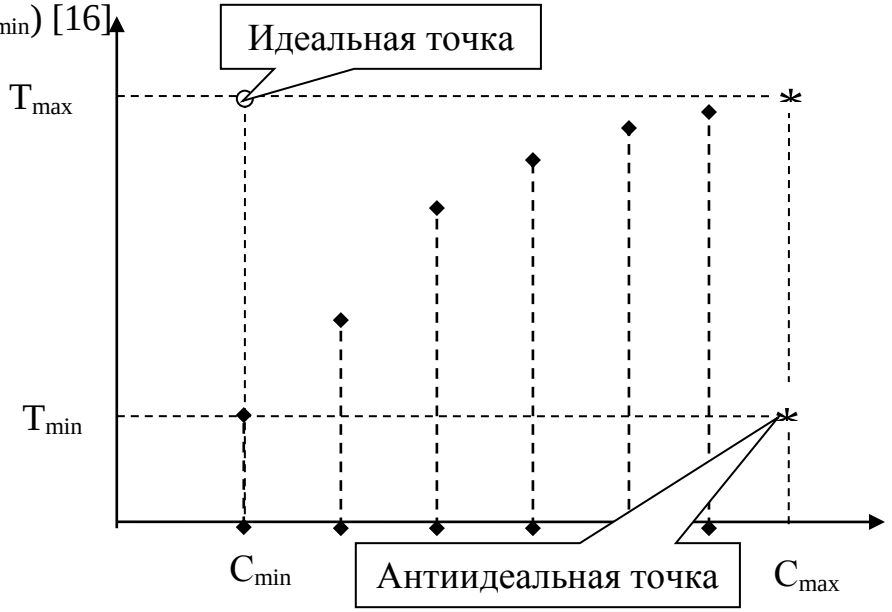


Рисунок 1.12. – Плоскость оптимальных решений по Парето

Противоположная точка с координатами $(C_{min};\ T_{max})$ является «идеальной». Ей соответствует ИСППИД с минимальной ценой (C_{min}) и максимальной точностью (T_{max}) [16].

Для определения промежуточных точек – оптимальных решений по Парето, требуется решить множество соответствующих задач целочисленного программирования, которые имеют вид [16]:

$$\begin{aligned} &\sum_{i=1}^{n}\sum_{j=1}^{m} T_{ij}x_{ij} \to \max, \\ &\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}x_{ij} \le R_k, \quad k \in [c_{\min}; c_{\max}] \\ &\sum_{j=1}^{m} x_{ij} = 1,\ i = \overline{1,n}, \end{aligned} \tag{1.45}$$

Каждой точке соответствует одно решение. Для достижения требуемой точности получаемого решения необходимо задать соответствующий шаг дискретизации на интервале $[c_{min};\ c_{max}]$.

В силу простоты, для решения данной задачи целесообразно использовать метод дихотомического программирования. Результаты его применения приведены на рисунке 1.13 [16].

Рисунок 1.13 – Результаты применения метода дихотомического программирования

В конкретной ситуации ЛПР, на основании ограниченного множества решений, способен принять обоснованное решение, путем выбора из множества решений по Парето наиболее соответствующего рассматриваемым условиям внешней среды.

Лемма 1. В результирующей таблице, полученной на основе метода дихотомического программирования, содержится множество оптимальных решений исходной задачи по Парето [16].

Доказательство базируется на методике формирования результирующей таблицы. Она построена на отсеве очевидных доминируемых стратегий поведения на предварительных этапах исследований.

На практике важна возможность получения единственного оптимального решения. Она достигается путем задания значимости используемых критериев. Это осуществляется путем задания соответствующих весов экспертами и применения принципов оптимальности. Исследуем их использование в условиях применения равнозначных критериев, то есть одинаковой значимости параметров затрат и оценки точности. Для этого проведем их (критериев) нормализацию в соответствии с выражением [16]:

$$\bar{y} = \frac{y - y^{\min}}{y^{\max} - y^{\min}} \qquad (1.46)$$

Экстремальные значения параметров находятся их графика представленного на рисунке 1. 13. Им соответствуют следующие числовые значения:

$T^{min}=17;\ T^{max}=39;\ C^{min}=1700;\ C^{max}=3400.$

Наибольшей наглядностью обладают принципы демонстрируемые «идеальной» и «антиидеальной» точками. Суть реализации данных принципов состоит в том, что определяются их экстремальные значения, после чего рассчитываются расстояния от каждой Парето-оптимальной точки до экстремальных. Для случая «идеальной» точки оптимальное решение соответствует минимальному расстоянию до нее, а для «антиидеальной» точки наоборот. В нормализованном пространстве, «идеальной» точке соответствуют координаты (0; 1), а «антиидеальная» точка будет иметь координаты

(1; 0) [16].

Значения исходных данных, величины нормализованных данных и значения расстояний до «идеальной» и «антиидеальной» точек представлены в таблице 1.9. Оптимальные решения выделены в ней серым цветом [16].

Решение, соответствующее принципу «идеальной точки», ранее рассматривалось, а решение, соответствующее принципу «антиидеальной точки» находится вблизи точки с максимальным значением затрат. При этом максимальному значению затрат соответствует максимальная оценка (индекс) точности.

Таблица 1.9 – Результаты реализации различных принципов оптимальности

Финансовые ресурсы	19	21	22	23	25	26	28	30
Точность	1750	1800	1900	1950	2000	2100	2150	2200
Нормализованные финансовые ресурсы	0,091	0,182	0,227	0,273	0,364	0,409	0,5	0,591
Нормализованная точность	0,029	0,059	0,118	0,147	0,176	0,235	0,265	0,294
Идеальная точка	0,975	0,959	0,911	0,895	0,9	0,867	0,889	0,921
Антиидеальная точка	0,91	0,82	0,782	0,742	0,66	0,636	0,566	0,504
Финансовые ресурсы	31	32	33	35	36	37	38	39
Точность	2300	2400	2650	2700	2800	2900	3300	3400
Нормализованные финансовые ресурсы	0,636	0,682	0,727	0,818	0,864	0,909	0,955	1
Нормализованная точность	0,353	0,412	0,559	0,588	0,647	0,706	0,941	1
Идеальная точка	0,908	0,9	0,851	0,916	0,933	0,955	0,956	1
Антиидеальная точка	0,507	0,52	0,622	0,616	0,661	0,712	0,942	1

Представляет интерес точка соответствующая минимуму объема выделяемого финансового ресурса (первая координата) и минимуму оценки точности (вторая координата). Данная точка представляет первую стратегию комплектации ИСППИД, в соответствии с которой они требуют минимальных финансовых вложений и обеспечивают минимальную точность оценок [16].

Если уровни значимости исследуемых критериев будут изменены, то полученные результаты (финансовые затраты и оценки точности) изменятся.

Таким образом предложенный метод дихотомического программирова-

ния на конкретных числовых примерах показал свою работоспособность и эффективность для решения поставленной проблемы в части обоснования требуемой комплектации ИСППИД по критериям минимальных финансовых затрат и максимальной точности принимаемых инвестиционных решений.

Варианты комплектации ИСППИД зависят от ее облика, который является результатом реализации соответствующей технологии проведения исследований.

1.8. Технологии проведения исследований и внешнесистемного синтеза системы

Сложность решаемой проблемы накладывает отпечаток на технологию проведения исследований. В частности возникает необходимость проведения внешнесистемного синтеза ИСППИД, результатом которого должен стать ее облик, и внутрисистемного синтеза, в рамках которого она рассматривается как некоторый проект, выполняемый группой проектировщиков и разработчиков, и его результатом должен стать синтезированный комплект проектной документации на разработку таких систем.

Проведение внешнесистемного и внутрисистемного синтеза основано на применении метода системного анализа (МСА) [60, 82] и вербального моделирования [41].

В соответствии с [60], основными этапами реализации МСА являются: декомпозиция (формирование общего представления о ИСППИД), анализ (детализация представления ИСППИД), синтез (обоснование спецификации ИСППИД), реализация (разработка прототипа ИСППИД, решающего исходную проблему).

Функциональное дерево МСА применительно к ИСППИД, состоящее из четырех ветвей, приведено на рисунке 1.14.

Ветвь «декомпозиция» включает такие составляющие, как: определение целей и задач; выделение ИСППИД из внешней среды; описание факторов воздействия; учет неопределенности и тенденций ее развития; системное представление ИСППИД в виде «черного ящика»; функциональное, компонентное и структурное представление ИСППИД.

Аналитическая ветвь МСА представлена такими видами анализа, как: структурно-функциональный; информационный; генетический; сравнительный (анализ показателей качества аналогов); интегральным (анализ эффективности); оценочным (обоснование требований к разрабатываемой системе).

Ветвь синтеза МСА включает такие компоненты, как: структурно-функциональный синтез (разработка модели); синтез вариантов структуры системы; параметрический синтез; оценка качества вариантов структур системы.

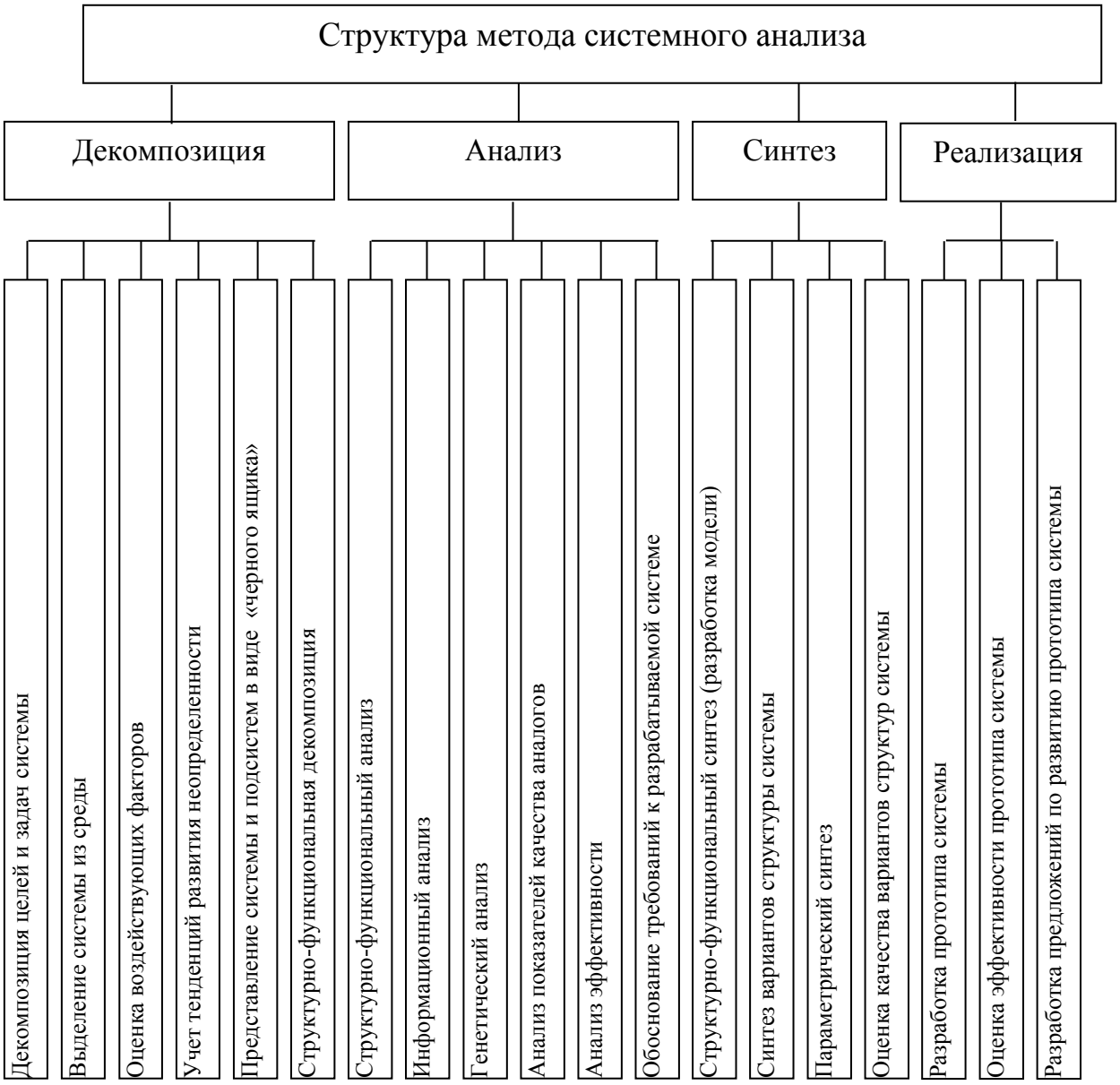


Рисунок 1.14 – Функциональное дерево МСА

В состав ветви «реализация» входят компоненты, связанные с разработкой в соответствии с синтезированным обликом прототипа ИСППИД, решающего исходную проблему, оценкой эффективности его функциони-

рования и разработкой предложений по его дальнейшему развитию.

Технологическая схема проведения диссертационных исследований, основанная на МСА, представлена на рисунке 1.15. Она предполагает реализацию внешнесистемного синтеза, результатом которого является разработанный прототип системы в соответствии со сформированным обликом в совокупности с проведенной оценкой его эффективности и обоснованием предложений по его развитию, а также внутрисистемного синтеза, завершающегося разработкой комплекта проектной документации на разработку ИСППИД.

Первый этап внешнесистемного синтеза заключается в декомпозиции ИСППИД, представляющей собой процесс, состоящий из шести основных стадий развития.

На первой стадии, исходя из формулировки общей проблемы, определяется область допустимых состояний ИСППИД. В дальнейшем, с целью сокращения, системы. Известно, что состояние характеризуется множеством существенных свойств системы [60]. К числу существенных свойств рассматриваемой системы относятся точность, оперативность, устойчивость, адекватность и др. Установлено, что система должна обеспечить поддержку формирования оптимальных ИП и предоставление ЛПР прагматической информации в области ПИД СЭО. В рамках поддержки формирования ИП она должна обеспечить принятие эффективных решений для детерминированных и стохастических данных об активах в случаях их статики и динамики. Реализуется данная стадия путем построения дерева целей (ДЦ), которое на данной стадии носит обобщенный характер (фактически определяются его контуры).

Суть второй стадии декомпозиции заключается в представлении ИСППИД в качестве подсистемы, у которой есть надсистема, в данном случае ИПУ ПИ СЭО, содержащая ЦИУ, СИУ и ОИУ. Необходимость данной стадии обусловлена соблюдением нескольких основных принципов системного анализа (СА) – системности и связности, обеспечивающих рассмотрение ИСППИД совместно с ее окружением в виде совместных пар «ИСППИД - ИПУ ПИ СЭО», «ИСППИД – внешняя среда» и др. (принцип системности) и

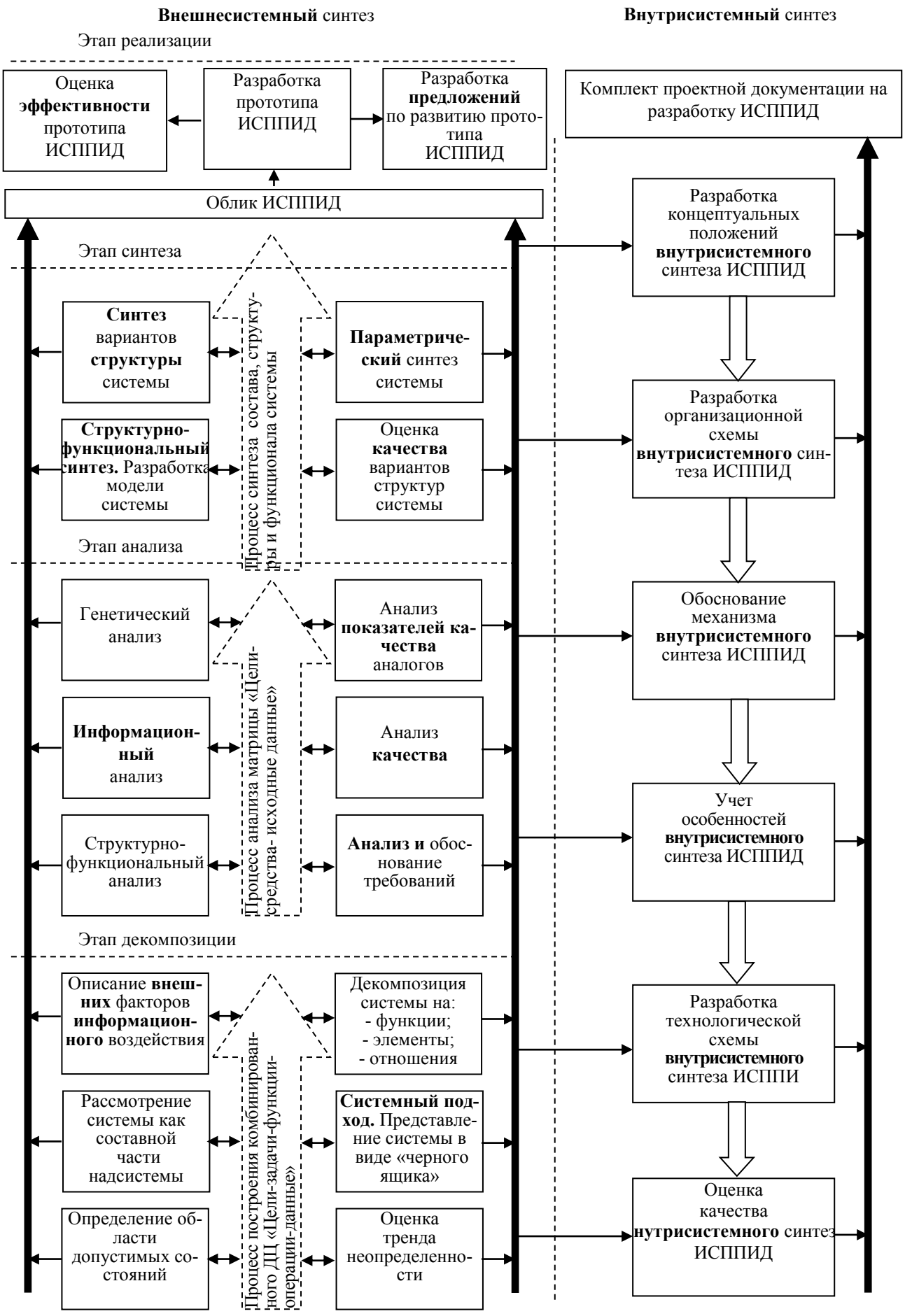


Рисунок 1.15 – Технологическая схема проведения исследований

выявление ее межэлементных связей и связей с внешней средой (принцип связности). Данная стадия влияет на конфигурацию отношений между элементами внутри ДЦ.

Третья стадия представлена описанием влияния на систему внешних факторов информационного воздействия (экономических, политических, социальных и др.), которые могут выступать в качестве ограничений, накладываемых на ее элементы (подсистемы) и систему в целом (например, дорогая и эффективная ИСППИД не принесет больших дивидендов в бедном государстве).

На четвертой стадии учитывается неопределенность внешней среды и тренд ее развития. В разделе 1.4 представлена характеристика видов неопределенности, а в разделе 1.5 приведена вероятностная модель учета влияния неопределенности внешней среды при синтезе ИСППИД. Установлено, что учет неопределенности влияет на структурные и функциональные особенности системы, а, следовательно. на данном этапе на состав и структуру ДЦ.

Пятая стадия характеризуется построением модели системы в виде «черного ящика», в которой известны величины входных и выходных потоков информации, а механизм устройства системы, нет. Данная модель обеспечивает изучение поведения системы без знания ее внутреннего функционирования. Это необходимо для детализации режимов работы системы, разработки ее интерфейса и др. В результате исследования данной модели установлено, что ИСППИД должна иметь несколько режимов работы: «Обработка статических детерминированных данных»; «Обработка динамически изменяющихся детерминированных данных»; «Обработка статических стохастических данных»; «Обработка динамически изменяющихся стохастических данных»; «Анализ РЦБ»; «Прогнозирование»; «Диверсификация»; «Расчет портфеля»; «Корректировка портфеля»; «Исследование ИНС»; «Настройка НКМШ»; «Выбор модели формирования ИП»; «Автоматический режим»; «Загрузка данных»; «Поиск прагматической информации». Каждый из них выбирается ЛПР в зависимости от условий, складывающихся на РЦБ.

Завершающей стадией этапа декомпозиции является разделение системы

на совокупности (множества) функций, элементов, отношений и др. Они отображаются в виде ветвей, листьев и связей между ними, при построении ДЦ «Цели-задачи-функции-операции-данные». Анализ общности функций, выполняемых элементами, а также информационных, логических и иерархических связей между ними, позволил выделить в ИСППИД подсистемы: формирования ИП; прогнозирования значений активов; терминологического поиска; сбора, обработки и хранения информации также и четыре обособленных модуля: ввода исходных данных; автоматизированной загрузки данных; управления; формирования результатов.

Стадии структурно-функционального, информационного, генетического анализа, а также анализа показателей качества аналогов, системы в целом и обоснования, предъявляемых к ней требованиям, составляют суть второго этапа внешнесистемного синтеза, обеспечивающего детальное представление системы.

Стадия структурно-функционального анализа заключается в уточнении: состава и особенностей (закономерностей) функционирования элементов; взаимного влияния подсистем; алгоритмов функционирования подсистем; постоянных и переменных данных. Кроме того, на данной стадии задаются (обосновываются) пространства состояний и параметров системы, описывающие ее поведение, проводится анализ ее целостности и формулируется первичный перечень, предъявляемых к ней требований.

Стадия информационного анализа заключается в исследовании информационных процессов системы, выделении в них разнородных информационных потоков данных (цифровых, текстовых, графических), предъявлении требований к пропускной способности и быстродействию соответствующих каналов и подсистем обработки, выявлении дублирующих связей и избыточных элементов.

На стадии генетического анализа проводится изучение исторических аспектов возникновения данной проблемы, состояние текущих исследований и перспектив развития в данной предметной области.

Содержанием стадии анализа показателей качества аналогов является оценка возможностей (структурных, функциональных, информационных и др.)

существующих систем, функционирующих в данной предметной области с целью использования готовых методических и инструментальных средств (моделей, методов, методик, алгоритмов, подсистем, модулей и др.) в интересах решения данной проблемы.

На стадии анализа качества системы осуществляется выбор шкалы и критериев оценки качества (точность, оперативность и др.), непосредственная их оценка и анализ полученных результатов.

Завершающей стадией этапа анализа является обоснование требований, предъявляемых к системе в совокупности с выбором критериев оценки и учетом ограничений. Результаты обоснования приведены в разделе 1.7.

Третьим этапом процесса создания ИСППИД является синтез, включающий четыре стадии: структурно-функциональный синтез; синтез вариантов структуры системы; параметрический синтез; оценка качества вариантов структур системы.

На стадии структурно-функционального синтеза осуществляется разработка модели ИСППИД, включающей: выбор ее математического представления (математического аппарата); проведение моделирования; оценку разработанной модели в соответствии с критериями точности, простоты, адекватности, инвариантности реализаций, модульности построения.

Стадия синтеза вариантов структуры системы завершается множеством сгенерированных экземпляров ИСППИД, различающихся составом и структурой, то есть набором, используемых компонентов (моделей, методов, алгоритмов и др.). Каждый вариант обладает возможностью решения общей проблемы и имеет свою затратную стоимость.

На стадии параметрического синтеза обосновывается номенклатура параметров ИСППИД и их количественные значения, требуемые для разрешения проблемы.

На завершающей стадии синтеза проводится оценка качества сгенерированных вариантов структур системы по критерию «эффективность-стоимость», их сравнение, обработка результатов сравнения, анализ полученных результатов

и выбор наиболее оптимальной структуры в соответствии с заданным критерием. Выбранный вариант представляет собой синтезированный облик ИСППИД.

Завершающий этап общего синтеза системы - этап реализации, включает стадии разработки прототипа системы в соответствии с синтезированным обликом, оценки его эффективности и обоснования предложений по его дальнейшему развитию. Детализация данных стадий приведена в главах 2 – 6 диссертационной работы.

Рассмотрим наиболее важные процессы синтеза ИСППИД СЭО - построение ДЦ и его анализ.

Известно, что ИСППИД должна обеспечить достижение двух целей. Одна из них заключается в формировании оптимального ИП с выдачей соответствующих рекомендаций ЛПР, другая, в поиске и предоставлении ему прагматической информации, требуемой для повышения его уровня знаний (описание нового методического обеспечения и программных решений) в области ПИД СЭО. Уровень подцелей представлен набором задач, требующих решения для достижения целей. На следующем уровне ДЦ находятся функции, реализация которых, позволит решить задачи вышестоящего уровня. В свою очередь, реализация каждой функции предполагает выполнение определенного множества операций, размещенных на соответствующем уровне иерархии. Самый нижний уровень представлен наборами исходных данных, требуемых для решения задачи и реализации некоторых функций и операций. В результате строится шестиуровневое ДЦ, содержащее цели, задачи, функции, операции и исходные данные. Определенные последовательности операций, функций и задач представляют собой средства, обеспечивающие достижение целей. Следовательно, механизм, предназначенный для формирования ДЦ, определяется следующей триадой: «цели-средства-данные». Его отличие от ранее предложенного механизма «цели-средства» заключается в применении элемента «данные». Целесообразность выделения данного элемента обусловлена особенностями функционирования СЭО на РЦБ. Применительно к детерминированным условиям принятия ПИР, что свойственно стабильному РЦБ, решаются в основном структурированные задачи с применением методического обеспечения

и программных средств ТА. При этом исходные данные имеют достаточно высокое качество (полноту, репрезентативность, содержательность, актуальность, точность, доступность, устойчивость, достоверность) [130, 219]. Для условий НВС, присущих нестабильному РЦБ из-за наличия кризиподобных состояний экономического развития, исходные данные об активах теряют свою детерминированность и носят стохастический характер, что делает нецелесообразным применение инструментов ТА. Возникает необходимость применения методического обеспечения и программных средств других видов анализа (например, ПИА, ФА и др.) [236], в рамках которых решаются другие задачи и используются иные функции и операции.

Обобщенное ДЦ для ИСППИД приведено на рисунке 1.16 [154].

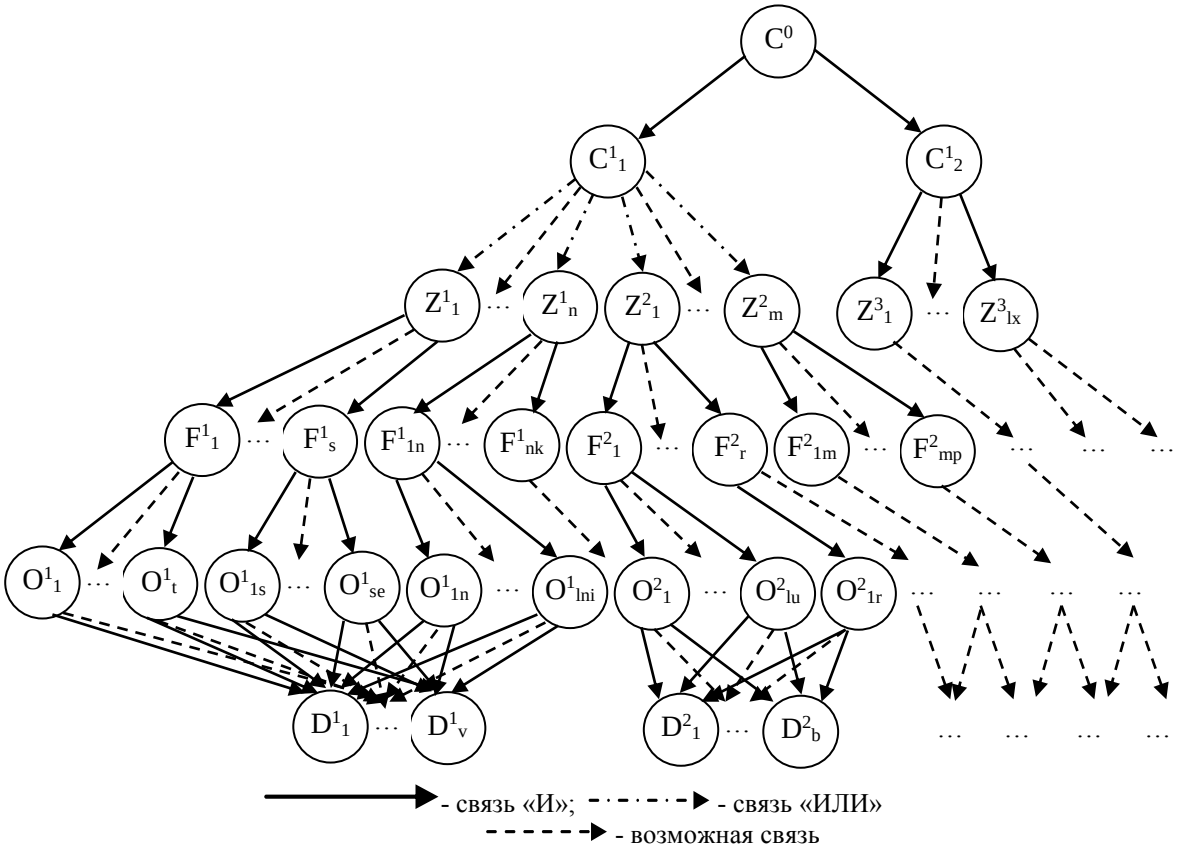


Рисунок 1.16 – Обобщенное дерево целей ИСППИД

Для наглядности ряд ветвей (функции, операции и наборы данных для задач терминологического поиска) на ДЦ не показан. Из рисунка следует, что иерархическое ДЦ включает шесть уровней. Самый высокий (нулевой) отведен глобальной цели C^0. Первый уровень представляют подцели C_1^1 и C_2^1. На втором уровне находятся задачи, решаемые в интересах достижения по-

ставленных подцелей и глобальной цели. Множество задач $\{Z^1\}=\{Z_1^1,\ldots,Z_n^1\}$ и $\{Z^2\}=\{Z_1^2,\ldots,Z_m^2\}$ решаются в интересах достижения цели C^1_1, а задачи $\{Z^3\}=\{Z_1^3,\ldots,Z_{1x}^3\}$ необходимо решить для реализации цели C_2^1.

Задачи множества $\{Z^1\}$ структурированные, характерные для ТА (например, определение тренда цены ЦБ на основе метода «голова-плечи», определение корректируемых уровней Фибоначчи, и др.). Их решение основано на множестве функций $\{F^1\}=\{F_1^1,\ldots,F_s^1\}$ и $\{F_n^1\}=\{F_{1n}^1,\ldots,F_{nk}^1\}$, которые представлены на более низком уровне иерархии (третьем). Реализация данных функций базируется на множестве операций $\{O^1\}=\{O_1^1,\ldots,O_t^1\},\{O_s^1\}=\{O_{1s}^1,\ldots,O_{1se}^1\}$ и $\{O_n^1\}=\{O_{1n}^1,\ldots,O_{1nt}^1\}$. Они размещены на еще более низком уровне иерархии (четвертом). Для решения задач $\{Z^1\}$ требуются исходные данные $\{D^1\}=\{D_1^1,\ldots,D_v^1\}$, представленные на самом низком уровне иерархии (пятом).

Множество задач $\{Z^2\}=\{Z_1^2,\ldots,Z_m^2\}$ являются не структурированными. Они характерны для ПИА (например, прогнозирование параметра ЦБ на основе ИНС, формирование ИП на основе модели Блека-Шоулса и др.). Их решение основано на множестве функций $\{F^2\}=\{F_1^2,\ldots,F_r^2\}$ и $\{F_m^2\}=\{F_{1m}^2,\ldots,F_{mp}^2\}$, находящихся на третьем уровне. Реализация данных функций предполагает использование набора операций $\{O^2\}=\{O_1^2,\ldots,O_{1r}^2\}$, размещенных на четвертом уровне. Данные, которые необходимы для решения $\{D^2\}=\{D_1^2,\ldots,D_b^2\}$ находятся на пятом уровне.

Аналогичная структуризация может быть проведена для множества задач $\{Z^3\}=\{Z_1^3,\ldots,Z_{1x}^3\}$, относящихся к области терминологического поиска.

Построенное ДЦ представляет собой графическую модель взаимосвязи глобальной цели, подцелей, решаемых задач, используемых функций и операций, а также исходных данных, требуемых для формирования облика ИСППИД [198]. Его структура по своей сути избыточна, поскольку содержит множество

альтернативных вариантов, способных реализовать требуемый функционал (решить выявленную проблему), отличающихся временными, материальными и другими видами затрат, а также множеством ограничений. Поэтому для определения наиболее рационального варианта состава и структуры ИСППИД необходимо оптимизировать ДЦ, результаты которой представлены в [149].

Для оптимизации ДЦ необходимо построить структурный граф $\tilde{A} = (C, U)$, реализующий логику «И», приведенный на рисунке 1.17. Данная логика позволяет провести предварительный отсев несовместных наборов (вариантов) компонентов системы [149].

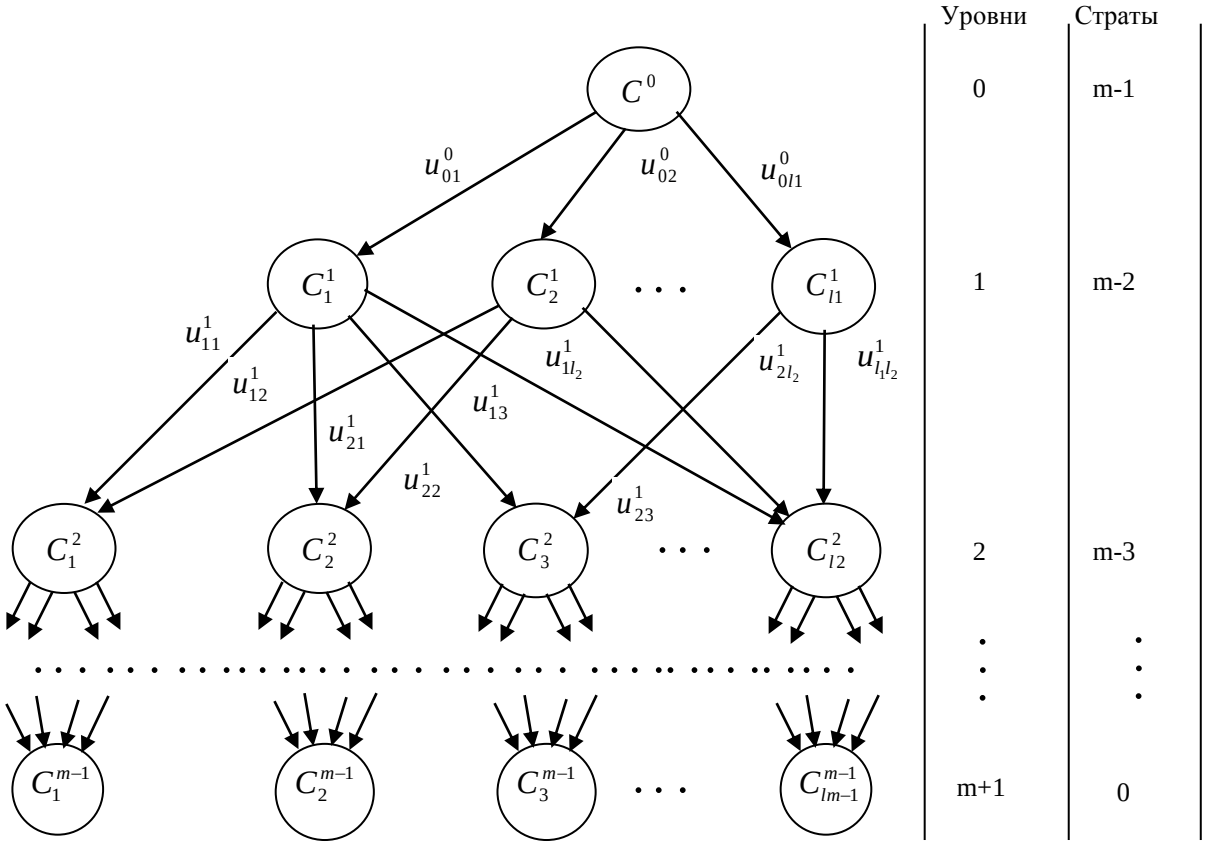


Рисунок 1.17 – Структурный граф ИСППИД

Анализ графа $\tilde{A} = (C, U)$ начинается с рассмотрения кортежа $C = \{C^0, C^1, \ldots, C^{m-1}\}$ на множестве целей различных уровней: C^0 – глобальная цель; $C^i = \{C_1^i, C_2^i, \ldots, C_{l_i}^i\}$, $i=0\ldots m-1$ – множество целей i-го уровня. Дуги на графе $U = \{u_{o\mu}^0, u_{j\nu}^j\}$, $\mu = 1, \ldots, l_1, i = 1, \ldots, m-2, j = 1, \ldots, l_j, \nu = 1, \ldots, l_{i+1}$, представляют связи (в том числе и перекрестные), определяющие направления целей вы-

шестоящих уровней. Перекрестные связи представляют характер зависимости целей *i*-го уровня от решения задач *(i+1)*–го уровня. С целью весовой обработки графа для связей устанавливаются отношения значимости, показывающие важность решения ν-й задачи (*i+1*)-го уровня для достижения *j*-ой цели *i*-го уровня. Для этого дугам $u_{j\nu}^{j}$ ставятся в соответствие весовые коэффициенты $q_{j\nu}^{i}$, диапазон изменения значений которых лежит в пределах от 0 до 1, при этом $\sum_{\nu=1}^{l_{i+1}} q_{j\nu}^{i} = 1$ [149].

Множество значений весов дуг $q_{j\nu}^{i}$ представляется в матричной форме $\mathbf{Q}^{i} = (q_{j\nu}^{i})$, которая в отличие от традиционной матрицы «цели-средства» [178], имеет структуру «цели-средства-данные», позволяющую учитывать вид обрабатываемых данных, влияющих на конфигурацию ИСППИД [149].

В предположении известных значений коэффициентов важности целей *i*-го уровня в виде вектор-столбца $\boldsymbol{K}^{i} = \left(K_1^{i}, K_2^{i}, \ldots, K_j^{i}, \ldots, K_{l_i}^{i}\right)$, удовлетворяющих условиям $0 \le k_j^{i} \le 1$, $\sum_{j=1}^{l_i} k_j^{i} = 1$. С помощью матрицы «цели-средства-данные» $\boldsymbol{Q}^{i}$ определяется вектор коэффициентов важности целей (*i+1*) –го уровня по формуле [149]:

$$\boldsymbol{K}^{i+1} = \boldsymbol{Q}^{iT}\,\boldsymbol{K}^{i}, \qquad (1.47)$$

где $\boldsymbol{Q}^{iT}$ – матрица, транспонированная по отношению к $\boldsymbol{Q}^{i}$.

С использованием (1.47) определяются вектор-столбцы коэффициентов важности для различных целей от *i=2* и до *i=m-1* уровней. Для первого уровня значения вектор-столбца заданы, поскольку степень важности цели первого уровня C_{ν}^{i} определяется весовым коэффициентом дуги $q_{0\nu}^{0}$, которая исходит из вершины C^{0} и входит в вершину C_{ν}^{i} [149].

На основании (1.47), значение ν-й компоненты вектора $\boldsymbol{K}^{i+1}$ определяется в виде:

$$K_{\nu}^{i+1} = \sum_{j=1}^{l_i} k_j^{i} q_{j\nu}^{i},\ i = 1 \ldots m-2, \nu = 1 \ldots l_{i=1}.$$

Таким образом, коэффициент важности цели C_{ν}^{i+1} определяется суммой

коэффициентов значимости всех целей уровня i, которые взвешены весами дуг, которые входят в вершину C_v^{i+1} из вершин нижестоящего уровня i. При этом

$$\sum_{v=1}^{l_{i+1}} k_v^{i+1} = \sum_{v=1}^{l_{i+1}} \sum_{j=1}^{l_i} k_j^i q_{jv}^i = \sum_{j=1}^{l_1} \sum_{v=1}^{l_{i+1}} k_j^i q_{jv}^i = \sum_{j=1}^{l_1} k_j^i \sum_{v=1}^{l_{i+1}} q_{jv}^i = 1.$$

Итерационный процесс анализа операционного графа позволил сформировать облик ИСППИД СЭО в составе подсистем формирования ИП, прогнозирования значений активов, терминологического поиска, а также сбора, обработки и хранения информации.

ВЫВОДЫ

Неопределенность внешней среды приводит к возникновению на РЦБ динамически изменяющихся стохастических данных об активах, что делает не эффективным применение существующего методического обеспечения и ИС, ориентированных на обработку детерминированных данных. ЛПР стремится снизить (устранить) неопределенность и принимать решения в детерминированных условиях.

В качестве инструмента снижения неопределенности предлагается использовать ИСППИД - набор программно-технических комплексов, использующих оборудование, специальное математическое и программное обеспечение, данные, знания, базу моделей для ИУ ЛПР и поиска прагматической информации в области ПИД СЭО с целью формирования оптимального ИП.

Проведенный анализ методического обеспечения информационных процессов и систем ПИД СЭО показал отсутствие системных проработок в данной предметной области и выявил следующие основные проблемы:

- отсутствие методического обеспечения решения информационных задач формирования оптимального ИП для стохастических данных об активах;

- сложность и по этой причине отсутствие динамических постановок задач формирования оптимального ИП СЭО и их решения;

- отсутствие полнофункциональных ИСППИД, имеющих базы моделей и реализующих информационную поддержку ЛПР на всех этапах процесса ИУ ИП;

- отсутствие методического обеспечения поиска, обработки и хранения прагматической информации в предметной области ПИД и возможности ее предоставления ЛПР в процессе информационного управления ИП СЭО;

- несоответствие фактических значений показателей качества информации (полнота, адекватность, достоверность) требуемым, в существующих программных системах.

В силу выявленных недостатков, постановка задачи синтеза структурно-функционального облика ИСППИД, потребовала, наряду с разработкой методического обеспечения ориентированного на обработку динамически изменяющихся стохастических данных об активах, включения в состав целевой функции, числа понятий, содержащихся в терминологическом портрете – новой категории отличной от тезауруса, и значения коэффициента прагматической информированности – величины характеризующей наличие в системе информации, используемой в интересах ИУ ЛПР в процессе формирования ИП СЭО. В качестве ограничений выступают: выделенный финансовый ресурс, допустимый риск ИП и требуемая оперативность решения данной задачи. Данная задача принадлежит к классу мультипараметрических оптимизационных задач с взаимозависимыми ограничениями и нелинейной целевой функцией, включающей связанные переменные. Для ее решения использован инструментарий внешнесистемного синтеза, основными технологическими аспектами которого являются построение и анализ ДЦ.

Основополагающее положение вероятностной модели учета влияния неопределенности внешней среды при разработке ИСППИД заключается в том, что неопределенность приводит к необходимости увеличения числа дополнительных операций для достижения поставленной цели. Поэтому требуется уточнение их числа и объема дополнительных ресурсов, выделяемых для их реализации. В качестве ЦФ выступает функция ресурсного обеспечения си-

стемы, а ограничением является показатель качества функционирования всей ИСППИД в виде вероятности выполнения ею поставленной цели. Необходимо минимизировать объем выделяемых дополнительных ресурсов, при этом ИСППИД должна выполнять поставленную задачу с заданной вероятностью. Данная задача относится к классу математических задач комбинаторного программирования. Для ее решения использован метод дихотомического программирования.

Модель обоснования состава ИСППИД призвана решить задачу максимизации точности принимаемых инвестиционных управленческих решений ЛПР путем оптимизации параметров специального математического и информационного обеспечения, а также значения коэффициента прагматической информативности, разрабатываемой ИСППИД, с учетом заданных суммарных затрат на ее разработку. При этом время проведения анализа не должно превышать требуемое. Решаемая задача является представителем класса математических задач комбинаторного программирования. Предложенный для ее решения метод дихотомического программирования на конкретных числовых примерах показал свою работоспособность и эффективность в части обоснования требуемой комплектации ИСППИД по критериям минимальных финансовых затрат и максимальной точности принимаемых инвестиционных решений.

2. МОДЕЛИ, МЕТОДИКИ И АЛГОРИТМЫ ИНФОРМАЦИОННОГО ПРОЦЕССА ФОРМИРОВАНИЯ ИНВЕСТИЦИОННОГО ПОРТФЕЛЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЙ ОРГАНИЗАЦИИ

2.1. Основные модели портфельного инвестиционного анализа

В различные исторические периоды разрабатывались различные модели портфельного инвестиционного анализа. Результаты сравнительной оценки основных из них, представлены в приложении Б [49].

Хронология разработки данных моделей следующая.

Модель Г. Марковица берет свое начало из авторской статьи «Выбор портфеля», опубликованной в 1952 году. В ней содержится описание математической модели формирования оптимального портфеля ЦБ, а также методов его построения. Тем самым задача выбора оптимальной инвестиционной стратегии получила строгое математическое обоснование. На ее основе показано, каким образом инвесторы могут обеспечить стабильную доходность портфеля, варьируя акциями.

У. Шарп - ученик Г. Марковица в первой половине 60-х годов предложил однофакторную модель рынка капиталов. В ней использовался усеченный метод выбора оптимального портфеля, сводивший задачу квадратичной оптимизации, используемой в модели Г. Марковица, к линейной. В силу своей простоты эта модель нашла широкое практическое использование. На ее основе появились первые программные продукты для решения задач управления портфелем ЦБ.

Работы Д. Тобина были опубликованы в конце 50-х и начале 60-х гг. Различие в моделях Г. Марковица и Д. Тобина обусловлены используемыми подходами. Г. Марковиц использовал микроэкономический анализ, а Д. Тобин – макроэкономический. Суть макроэкономического подхода заключалась в анализе и включении в инвестиционный портфель безрисковых активов (например, государственных облигаций). Д. Тобин провел анализ адекватности количественных характеристик активов и портфелей – исходных данных для модели Марковица.

Период с 1964 г. по 1966 г ознаменован появлением модели оценки ка-

питальных активов CAPM (Capital Asset Price Model). Она базировалась на моделях Линтнера, Моссина и Шарпа. Ее основные положения состоят в следующем. 1. Обладая одной и той же информацией, все инвесторы способны одинаково оценить риск и доходность отдельной акции, а также инвестиционный портфель. 2. Выбор оптимального портфеля базируется на учете систематического (недиверсифицируемого) риска. 3. Взаимосвязь между риском и доходностью описывается линейной зависимостью. В настоящее время CAPM является одной из распространенных моделей, несмотря на ряд критических замечаний.

В 1973 г. Была разработана авторская модель Блека-Шоулса. Ее предметной областью являлись опционы. Она явилась началом развития нового класса - динамических теоретико-вероятностных моделей, базирующихся на теории случайных процессов. Ее основу составляли следующие фундаментальные положения. 1. При использовании акции и оформленным на нее опционом возможна безрисковая сделка. 2. Цена безрисковой сделки соответствует стоимости включенных рыночных безрисковых активов. 3. В связи с изменением цены актива во времени, изменяется стоимость оформленного опциона, обеспечивающего безрисковую сделку. Данная модель нашла широкое распространение в 70–х годах и является фундаментальной для анализа опционов.

Проведенный анализ разработанных моделей показал, что какого-то универсального инструмента позволяющего проводить анализ финансовых активов (портфелей) в различных условиях, не существует. Каждая из моделей, разработанных в различные периоды времени, эффективна в определенных условиях. Они описаны в различных источниках и сведения о них носят разрозненный характер. Представляет интерес объединение ряда из них в рамках ИСППИД с целью использования в качестве инструментария портфельного анализа [173, 174].

В рамках реализации моделей и алгоритмов инвестиционного анализа целесообразно использовать технологию прототипирования.

Под прототипом будем понимать первоначальный вариант ИСППИД с

ограниченным функционалом, призванный проверить работоспособность, заложенных в ней идей, на практике. В перспективе, прототип доводится до полнофункционального лабораторного образца, готового для дальнейшего промышленного изготовления и внедрения. Разрабатываемый прототип строится как открытая модульная система, реализующая ее основной функционал. В дальнейшем она доводится до полномасштабного образца, путем разработки и встраивания модулей, реализующих дополнительный функционал. В состав действующего прототипа включены следующие модели и алгоритмы инвестиционного анализа: модель Марковица; обобщенный алгоритм реализации модели Марковица; одноиндексная модель Шарпа; нейрокомитетная модель Шарпа; алгоритмы расчета характеристик ЦБ [113, 114]; модифицированный генетический алгоритм распределения инвестиций; островная модель параллельных вычислений. Рассмотрим детализацию характерных для них информационных процессов.

2.2. Анализ информационных процессов модели Марковица

Модель Марковица реализует ИФП, определяющий набор эффективных ИП, обеспечивающих наибольшие ожидаемые доходности для определенных уровней риска [9]. Данная модель имеет ряд условий.

1. ИФП инвестирования, в рамках данной модели, рассматривается как одноперiодный процесс, т.е. полученный в результате инвестирования доход не реинвестируется [59].

2. Считается, что рынок ценных бумаг является эффективным. Это означает, что изменение информации о состоянии внешней среды (информации о политической ситуации, данных об экономической обстановке и др.) практически мгновенно отражается на котировках ЦБ [55].

3. Предполагается, что значения доходности ЦБ являются случайными величинами, распределенными по нормальному закону. Поэтому при формировании портфеля достаточно учитывать два показателя: ожидаемую доходность r и стандартное отклонение σ, как меру риска (только эти два показателя определяют плотность вероятности случайных чисел при нормальном распределении) [59].

4. Модель позволяет определить набор эффективных ИП [18]. Если реализовать геометрическую интерпретацию данной модели, то в определенных случаях можно показать существование единственного оптимального ИП, что имеет практическое значение для инвестора и подтверждает целесообразность применения данной модели в ИСППИД. Рассмотрим доказательство данного утверждения.

Для каждого ИП ЛПР задает функцию полезности *U(σ,r)*, аргументами которой являются ожидаемая доходность *r* и риск *σ*. В качестве риска, как правило, выступает стандартное отклонение *r* [59]. В процессе своей ПИД ЛПР стремится максимизировать функцию полезности $U(\sigma,r)\to\max$. С математической точки зрения это эквивалентно решению уравнения вида $U(\sigma,r)=C$. Решения *{r (σ)}* данного уравнения представляют собой линии уровня функции *U*. В литературе [15] линии уровня называют кривыми безразличия (indifference curves). Функция полезности обладает рядом свойств:

1) нерасположенность к риску $U_{\sigma}^{`}(\sigma,r)<0$;
2) ненасыщаемость $U_{\sigma}^{`}(\sigma,r)>0$;
3) выпуклость.

На рисунке 2.1 представлен вид линий уровня функции полезности вида $a^2r-b^2\left(r^2+\sigma^2\right)$. Стрелкой отмечено направление возрастания функции полезности.

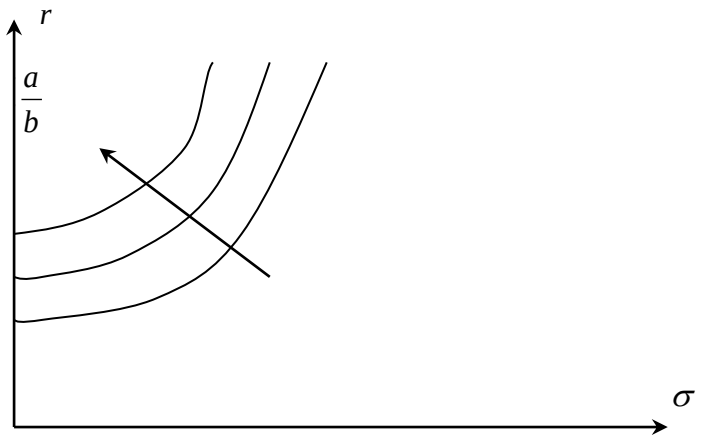


Рисунок 2.1. – Линии уровня функции полезности

ИП $\widehat{\Pi}$ с ожидаемой доходностью $\tilde{r}$ и риском $\tilde{\sigma}$ является оптимальным, если на множестве допустимых портфелей $\{\Pi\}$ для него функция полезности

достигает максимума $\tilde{U}(\tilde{\sigma},\tilde{r}) = \max$. Данное утверждение может быть сформулировано в виде следующей теоремы [173, 201].

Теорема 1. Оптимальный ИП находится в точке $\hat{\Pi}$ касания области E эффективного множества ИП с определенной кривой безразличия.

Доказательство. На рисунке 2.2 приведены кривые безразличия и область эффективных ИП.

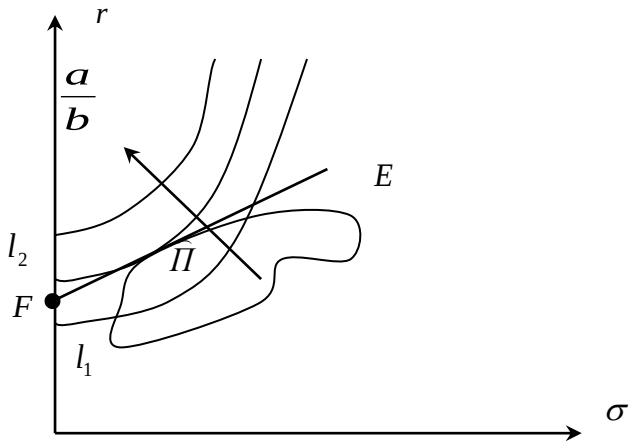


Рисунок 2.1. – Кривые безразличия и область эффективных ИП

Если область эффективных ИП E пересекается с кривой безразличия l_1, то существуют допустимые ИП, лежащие выше и левее кривой l_1, для которых значение функции полезности будет больше, чем на самой кривой или ниже её. Следовательно, оптимальный ИП лежит левее и выше линии уровня l_1, пересекающей область E [173].

Рассмотрим произвольную линию уровня l_2, лежащую левее и выше области E. Значения функции полезности на линии уровня l_2 больше чем на линии уровня l_1. Однако, как видно из рисунка, на ней не лежит ни одного ИП (она не имеет точек пересечения с областью эффективных ИП E). Следует предположить, что оптимальный ИП находится правее и ниже l_2. А это есть точка касания $\hat{\Pi}$, что и требовалось доказать [173, 174].

Важным инструментом в рамках практической реализации данной модели являются безрисковые ЦБ. Они существенным образом влияют на множество эффективных ИП. Представим на координатной плоскости (риск, доходность - *(σ,r))* множество эффективных ИП E_r (рисунок 2.3) [173].

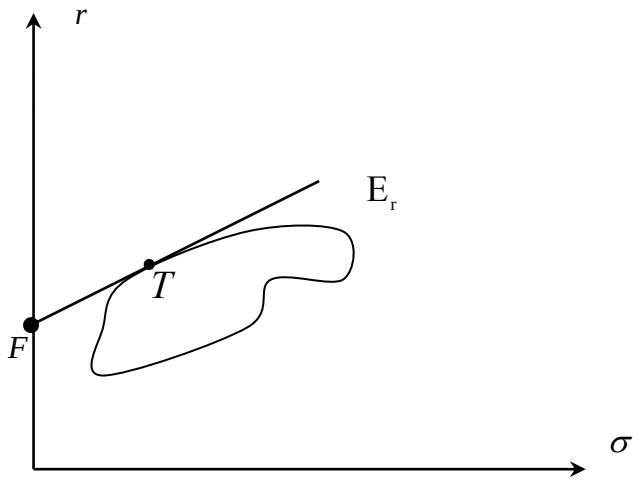


Рисунок 2.3. – Эффективные ИП и безрисковая ЦБ

Они состоят только из рисковых ЦБ $A_1,\ldots,A_n$. Пусть существует безрисковая ЦБ $F=(0, r_F)$. Поскольку множество E_r вогнуто, то существует не более одной касательной к E_r, проходящей через точку F. Если касательная существует, то обозначим ее l, а точку касания – T. ИП, соответствующий точке T, будем называть касательным ИП [173].

Теорема 2. Луч $[FT)$ является множеством эффективных ИП, включающих ЦБ $F, A_1,\ldots,A_n$.

Доказательство. Рассмотрим множество допустимых ИП вида $П=v_0F+v_1A_1+\ldots+v_nA_n$, $v_i\geq 0$ при $i\geq 1$ и $v_0+v_1+\ldots+v_n=1$. Среди них есть ИП $П_r$, который включает лишь рисковые ЦБ $A_1,\ldots,A_n$. Доли рисковых ЦБ в данном ИП повторяют доли ИП $П$. Если предположить, что $v_0<1$, то

$$П_r=\frac{v_1}{1-v_0}A_1+\ldots+\frac{v_n}{1-v_0}A_n. \qquad (2.1)$$

Тогда $П= v_0F+(1- v_0)\,П_r$. При этом ожидаемые доходность и риск будут определяться в соответствии с выражениями [173]:

$$r_П= v_0r_F+(1- v_0)\,r_{П_r},\ \sigma_П=(1-v_0)\sigma_{П_r}. \qquad (2.2)$$

Уравнения (2.2) являются параметрическими. Они задают на плоскости (σ,r) луч. Значит, ИП $П_r$ принадлежит лучу $[FП_r)$. Отсюда следует, что ИП $П$, который включает ЦБ $F, A_1,\ldots,A_n$, является допустимым лишь в том случае, если существует ИП $П_r$ (по определению он состоит только из рисковых ЦБ) такой, что $П\in[FП_r)$. Таким образом, ИП, принадлежащие $[FT)$, являются

допустимыми, поскольку *T* является допустимым ИП, состоящим только из рисковых ЦБ [173].

Допустим, что ИП $П_0=(\sigma_{П_0}, r_{П_0}) \in [FT)$ эффективным не является. Тогда следует предположить, что существует такой ИП $П_1 \neq П_0$, для которого $\sigma_{П_1} \leq \sigma_{П_0}$ и $r_{П_1} \geq r_{П_0}$. Это значит, что $П_1$ находится левее и выше луча $[FT)$. С учетом вышеприведенного доказательства существует такой ИП $П_2$, содержащий рисковые ЦБ, что $П_1 \in [FП_2)$. Но луч $[FП_2)$ находится левее и выше луча $[FT)$, а значит, не пересекает множество допустимых ИП, что неверно. Значит, верно первое предположение, что и требовалось доказать.

В рамках реализации ИФП данной модели на практике, важно обоснование того факта, что, несмотря на общие подходы к формированию ИП, результаты у различных ЛПР (формируемые ИП) оказываются независимыми. В связи с этим рассмотрим теорему о независимости комбинаций рисковых ЦБ в ИП при одинаковой оценке инвесторами рисков и ожидаемых доходностей [173].

Теорема 3. Пусть разные ЛПР одинаково оценивают риски и ожидаемые доходности. Тогда оптимальная для ЛПР комбинация рисковых ЦБ не зависит от его предпочтений относительно риска и дохода [173].

Доказательство. В соответствии с теоремой 2 все ЛПР сформируют ИП [173]:

$$П = v_{0j}F + (1 - v_{0j})T,$$

где *F* – безрисковая ЦБ, *T* – касательный ИП, v_{0j} – доля капитала *j* – го ЛПР, вложенная в безрисковую ЦБ. Поскольку рисковая часть ИП содержит касательный ИП *T*, то это значит, что доля вложений в произвольную рисковую ЦБ по отношению к рисковой части ИП не зависит от предпочтений ЛПР, несмотря на различные доли v_{0j}. Таким образом, теорема доказана.

Из теоремы 3 не следует, что разные ЛПР сформируют один и тот же ИП. Чем больше ЛПР не хочет рисковать, тем ниже на луче $[FT)$ расположен его оптимальный ИП. Поэтому ИП будут различными.

Приведенные выше доказательства подтверждают возможность практического использования модели Марковица в ИСППИД в соответствии со следующим алгоритмом [173]:

Шаг 1. Построение множества допустимых ИП.

Шаг 2. Выделение эффективных ИП на множестве допустимых.

Шаг 3. Построение кривых безразличия ЛПР.

Шаг 4. Выбор кривой безразличия, соприкасающейся с множеством эффективных ИП.

Шаг 5. Определение точки касания. Оптимальный ИП находится в точке касания.

Ожидаемая доходность ИП будет определяться в соответствии с выражением

$$r_{П}=\sum_{i=1}^{n} v_i r_i \,, \qquad (2.3)$$

где r_i – доход, приносимый i- й ЦБ.

Риск ИП будет определяться в соответствии с выражением

$$\sigma_{П}=\sqrt{\sum_{i,j=1}^{n} v_i v_j \rho_{ij}\sigma_i\sigma_j} \,, \qquad (2.4)$$

где ρ_{ij} - коэффициент корреляции между доходностями r_i и r_j, $\sigma_i\sigma_j$ - риски i-й и j – й ЦБ.

Рассмотрим пример, демонстрирующий применение данного алгоритма.

Пусть РЦБ представлен двумя ЦБ *A(0,1;0,1)* и *B(0,2;0,3)* с коэффициентом корреляции взаимных доходностей $\rho_{AB}=0{,}5$. Функция полезности ИП имеет вид: $U(\sigma,r)=0{,}6r - r^2 - \sigma^2$ [173].

Необходимо найти параметры оптимального ИП.

Параметрические уравнения расчета доходности и риска, допустимых ИП $П= tA+(1-t)B$, с учетом (2.3) и (2.4), будут иметь вид [173]:

$$r_{П}= tr_A+(1- t)\, r_B=0{,}3-0{,}2t, \qquad (2.5)$$

$$\sigma_{П}=\sqrt{t^2\sigma_A^2+(1-t)^2\sigma_B^2+2t(1-t)0{,}5\sigma_A\sigma_B}=\sqrt{0{,}03t^2-0{,}06t+0{,}04}\,. \qquad (2.6)$$

Точка касания линии уровня функции $U(\sigma,r)$ и множества допустимых

ИП является точкой размещения оптимального ИП $\hat{П}$.

Касательная к множеству допустимых ИП в точке *П= tA+(1-t)B* будет рассчитываться как [173]

$$k = \frac{dr_П}{d\sigma_П} = \frac{dr_П}{dt} \Big/ \frac{d\sigma_П}{dt}.$$

При этом, $\frac{dr_П}{dt} = -0{,}2$, $\frac{d\sigma_П}{dt} = \frac{0{,}06t - 0{,}06}{2\sigma}$.

Следовательно, $k = \frac{2\sigma}{0{,}3 - 0{,}3t}$.

Линии уровня функции $U(\sigma,r)=0{,}6r - r^2 - \sigma^2$ описываются уравнением вида $-(r-0{,}3)^2-\sigma^2 = C-0{,}09$. Для определения наклона $\frac{dr}{d\sigma}$ касательной к линии уровня продифференцируем данное уравнение почленно по σ. Имеем:

$$2(r - 0{,}3)\frac{dr}{d\sigma} + 2\sigma = 0 .$$

Следовательно, $\frac{dr}{d\sigma} = \frac{\sigma}{(0{,}3 - r)}$.

Если приравнять полученное выражение к выражению для расчета наклона к множеству допустимых ИП, получим [173]

$$\frac{\sigma}{(0{,}3 - r)} = \frac{2\sigma}{0{,}3 - 0{,}3t}.$$

Так как искомая линия уровня не только имеет наклон *k*, но и проходит через точку $\hat{П}$, то согласно (П1.5) *r=0,3-0,2t*. Если подставить это выражение для *r* в предыдущую формулу, получим *t=3/7*. Следовательно, *П=3A/7+4B/7*, $r_П \approx 0{,}141$, $\sigma_П \approx 0{,}214$ [173].

Приведенный пример показывает работоспособность разработанного алгоритма. Однако последний не позволяет определить структуру ИП, то есть доли ЦБ, входящих в ИП. Для этих целей целесообразно использовать аналитический метод [173].

2.2.1. Постановка задачи определения долей ценных бумаг в оптимальном портфеле Марковица

Пусть v_i, $i=1,...,n$ есть доля ЦБ A_i в произвольном ИП $П$: $П=v_1A_1+...+v_nA_n$. Основные параметры ИП определяются в соответствии с выражениями (2.3) и (2.4) [173].

Необходимо выбрать доли ЦБ $v_1+...+v_n$ таким образом, чтобы значение функции U для результирующего ИП $П=(\sigma_П, r_П)$ было максимальным [173]:

$$U(\sqrt{\sum_{i,j=1}^{n} v_i v_j \rho_{ij} \sigma_i \sigma_j}, \sum_{i=1}^{n} v_i r_i) \to \max . \qquad (2.7)$$

Ограничениями в задаче выступают условия (2.8) и (2.9) [173]:

$$v_1+...+v_n=1, \qquad (2.8)$$

$$0 \le v_i \le 1. \qquad (2.9)$$

Условие (2.8) показывает, что доли в сумме должны составлять единицу, а условие (2.9) говорит о том, что ЦБ нельзя брать взаймы и давать в долг. Применительно к безрисковой ЦБ условие $v_i \ge 0$ не учитывается [173].

Данная задача (2.7) – (2.9) определения условного экстремума решается известными методами математического анализа (например, методом множителей Лагранжа, сводящим задачу на условный экстремум к задаче на безусловный экстремум). Для получения конкретных результатов важно знать вид функции полезности. Наименее сложна задача определения ИП с заданной доходностью и минимизацией риска. Доли риска в данном случае будут определяться в соответствии с выражением [173]

$$\vec{v} = \frac{r^* - r_0}{(\vec{r} - r_0\vec{I})^T V^{-1} (\vec{r} - r_0\vec{I})} V^{-1}(\vec{r} - r_0\vec{I}), \qquad (2.10)$$

где $\vec{v} = (v_1,...,v_n)^T$ - вектор долей рисковых ЦБ в ИП; $V = (Cov(R_i, R_j))_{i,j=1}^{n}$ - матрица ковариаций случайных величин R_i, $\vec{r} = (r_1,...,r_n)^T$ - ожидаемые доходности ЦБ, r_0 – доходность безрисковой ЦБ, $\vec{I} = (1,...,1)^T$ - вектор-столбец с единичными координатами [173].

Рассмотрим пример определения долей облигаций в ИП.

Дано. Сформирован ИП, состоящий из облигаций трех эмитентов: Авангард, ВТБ, Газпром. Число временных интервалов наблюдения равно 3. Доходности облигаций в заданных временных интервалах наблюдения распределены следующим образом: Авангард: 0,175; 0,169; 0,178; ВТБ - 0,189; 0,195; 0,197; Газпром - 0,167; 0,171; 0,163.

Найти. Доли облигаций в ИП.

Решение. Средние значения доходностей облигаций найдем по формуле (2.3):

$$\bar{r}_1 = \frac{0{,}175 + 0{,}169 + 0{,}178}{3} = 0{,}174,$$

$$\bar{r}_2 = \frac{0{,}189 + 0{,}195 + 0{,}197}{3} = 0{,}193,$$

$$\bar{r}_3 = \frac{0{,}167 + 0{,}171 + 0{,}163}{3} = 0{,}167.$$

Получим вектор средних доходностей:

$$\bar{r} = \begin{bmatrix} 0{,}174 \\ 0{,}193(6) \\ 0{,}167 \end{bmatrix}$$

Матрица ковариаций определяется следующим образом:

$$\sigma_{11} = ((0{,}175 - 0{,}174)^2 + (0{,}169 - 0{,}174)^2 + (0{,}178 - 0{,}174)^2)/3 = 1{,}4e-05,$$

$$\sigma_{12} = ((0{,}175 - 0{,}174)(0{,}189 - 0{,}193(6)) + (0{,}169 - 0{,}174)(0{,}195 - 0{,}193(6)) + (0{,}178 - 0{,}174)(0{,}197 - 193{,}3(6))/3 = -6{,}6(6)7e-07$$

Аналогично вычисляем остальные значения. Получаем:

$$\sigma = \begin{bmatrix} 1{,}4e-05 & -6{,}6(6)e-07 & -1{,}2e-05 \\ -6{,}6(6)7e-07 & -1{,}15(5)6e-05 & -2{,}6(6)7e-06 \\ -1{,}2e-05 & -2{,}6(6)7e-06 & 1{,}06(6)7e-05 \end{bmatrix}$$

Обратная матрица к матрице ковариаций равна:

$$\sigma^{-1} = \begin{bmatrix} -1{,}225019088e+21 & -2{,}6250409028e+20 & -1{,}4437724965e+21 \\ -2{,}6250409028e+20 & -5{,}6250876488e+19 & -3{,}0937982068e+20 \\ -1{,}4437724965e+21 & -3{,}0937982068e+20 & -1{,}7015890138e+21 \end{bmatrix}$$

Доли облигаций в ИП найдем по формуле (2.10). Получим:

$$X[i]=\frac{\begin{matrix}-1{,}225019088e+21 & -2{,}6250409028e+20 & -1{,}4437724965e+21\\ -2{,}6250409028e+20 & -5{,}6250876488e+19 & -3{,}0937982068e+20\\ -1{,}4437724965e+21 & -3{,}0937982068e+20 & -1{,}7015890138e+21\end{matrix}\begin{bmatrix}1\\1\\1\end{bmatrix}}{\begin{bmatrix}1\\1\\1\end{bmatrix}\begin{matrix}-1{,}225019088e+21 & -2{,}6250409028e+20 & -1{,}4437724965e+21\\ -2{,}6250409028e+20 & -5{,}6250876488e+19 & -3{,}0937982068e+20\\ -1{,}4437724965e+21 & -3{,}0937982068e+20 & -1{,}7015890138e+21\end{matrix}\begin{bmatrix}1\\1\\1\end{bmatrix}}\times 100\%=\begin{matrix}41{,}79\\8{,}96\\49{,}25\end{matrix}$$

Общая доходность ИП, рассчитываемая по формуле (2.3), составит:

$$mp=[0{,}174 \quad 0{,}193(6) \quad 0{,}167]\times\begin{bmatrix}41{,}79\\8{,}96\\49{,}25\end{bmatrix}=17{,}23\%\cdot$$

Вышеизложенный механизм решения задачи определения долей ЦБ в оптимальном ИП Марковица может быть представлен в виде следующего алгоритма.

2.2.2. Обобщенный алгоритм определения долей ценных бумаг в оптимальном портфеле Марковица

Блок-схема обобщенного алгоритма определения долей ЦБ представлена на рисунке 2.4 [172-174].

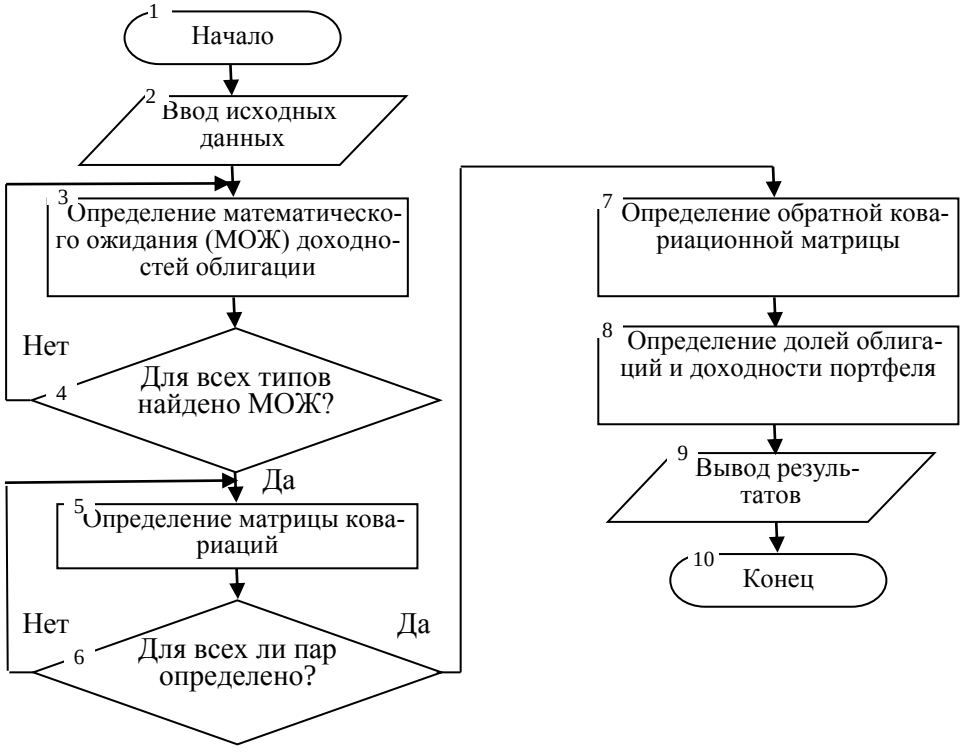


Рисунок 2.4. – Блок-схема обобщенного алгоритма реализации модели Марковица

Блоки 1, 10 реализуют пуск и остановку алгоритма.

Блок 2 обеспечивает ввод следующих исходных данных: вид ЦБ; время

наблюдения; доходности ЦБ.

В блоке 3 рассчитывается математическое ожидание доходностей ЦБ. Для этого используется выражение [173]:

$$\bar{r}_i = \frac{1}{T}\sum_{t=1}^{T} r_{it}\,, \tag{2.11}$$

где r_{it} –доходность i-й ЦБ за время t; t – номер периода накопления информации; T – длительность периода накопления.

В блоке 4 осуществляется проверка условия расчета МОЖ доходностей ЦБ всех типов. В случае выполнения условия, управление предоставляется в блок 5. Иначе управление предоставляется блоку 3.

В блоке 5 определяется матрица ковариаций доходностей ЦБ. Для расчета ковариации эффективных доходностей между i-й и j-й ЦБ (σ_{ij}) используется выражение [173]:

$$\sigma_{ij} = \frac{1}{T}\sum_{t=1}^{T}(r_{it} - \bar{r}_i)(r_{jt} - \bar{r}_j)\,, \tag{2.12}$$

где r_{it} и r_{jt} – соответствующие эффективные доходности i-й и j-й ЦБ за временной период t, %; r_i и r_j – МОЖ эффективных доходностей i-й и j-й ЦБ.

Блок 6 реализует проверку перебора всех пар ЦБ. В случае выполнения условия, управление направляется в блок 7. Иначе, управление передается в блок 5.

Блок 7 предназначен для определения обратной ковариационной матрицы.

Блок 8 обеспечивает расчет долей ЦБ в ИП в соответствии с выражением:

$$X[i] = \sigma^{-1}e/(\sigma^{-1}e,\bar{e}), \tag{2.13}$$

где $x[i]$- вектор долей ЦБ в ИП; σ_{ij}^{-1}- обратная ковариационная матрица; е - единичный вектор; $\bar{e}$- транспонированный единичный вектор. В данном блоке также определяется общая доходность ИП в соответствии с выражением [173]:

$$\sum_{i=1}^{N}\overline{r_{it}} \times x_i = mp\,, \tag{2.14}$$

где $\overline{r_{it}}$ - вектор МОЖ облигаций; $x[i]$- вектор долей ЦБ в ИП; N – число ЦБ; mp - доходность ИП.

В Блоке 9 реализован вывод результатов расчета для дальнейшего анализа.

2.2.3. Программная реализация определения долей ценных бумаг в оптимальном портфеле Марковица

Реализация задачи определения долей ЦБ, входящих в оптимальный ИП Марковица, представлена в виде программного модуля, включенного в состав ИСППИД [165, 172, 173]. В качестве среды разработки модуля использована объектно–ориентированная среда разработки Delphi 7.0. Обоснование выбора данной среды представлено в [7].

Экранная форма расчета долей ЦБ, входящих в ИП Марковица, приведена на рисунке 2.5.

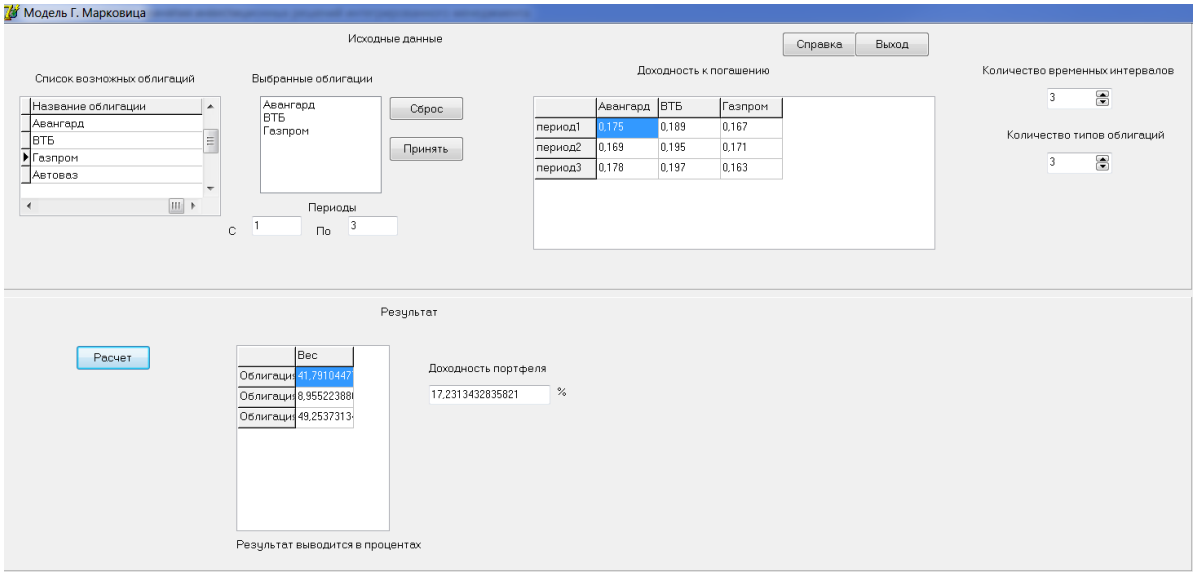


Рисунок 2.5. – Вид экранной формы расчета долей ЦБ по модели Марковица

В верхней части экранной формы вводятся исходные данные. В ее левой части находится окно со списком ЦБ (облигаций), которые могут быть включены в состав ИП. Рядом есть окно, в котором отображаются выбранные облигации. В средней части экранной формы находится окно «Доходность к погашению». В правой части формы размещены два окна, в которых задаются количество типов ЦБ (облигаций) и количество рассматриваемых временных интервалов, соответственно. Они заполняются на начальном этапе работы с программным модулем. Затем осуществляется выбор конкретных ЦБ (облигаций). После выбора необходимой облигации, ЛПР нажимает правую кнопку мыши. Появляется кнопка «Добавить». Нажатием данной кнопки, со-

ответствующая ЦБ, перемещается в окно «Выбранные облигации». Далее необходимо ввести требуемые периоды наблюдения за доходностями облигаций. Соответствующие окна ввода, размещены снизу окна «Выбранные облигации». Кнопки «Сброс» и «Принять», размещенные справа от этого окна, обеспечивают соответственно удаление или принятие списка выбранных ЦБ. Перечень выбранных (принятых) ЦБ и их доходность за выбранные периоды, отобразится в окне «Доходность к погашению», размещенного в правой верхней части экранной формы. Значения доходности облигаций могут корректироваться и вводиться с клавиатуры.

В нижней части экранной формы отображаются результаты расчета. Они появляются после ввода исходных данных и нажатия кнопки «Расчет». В состав результатов расчета включены доходность ИП и процентные соотношения выбранных ЦБ, в виде соответствующих весов.

Кроме того на данной экранной форме имеются кнопки «Справка» и «Выход». Нажатие кнопки «Выход» обеспечивает возврат на главную форму ИСППИД. При нажатии на кнопку «Справка», появляется соответствующая экранная форма, в которой отражаются основные сведения об используемой модели. Вид экранной формы «Справка» приведен на рисунке 2.6 [172, 173].

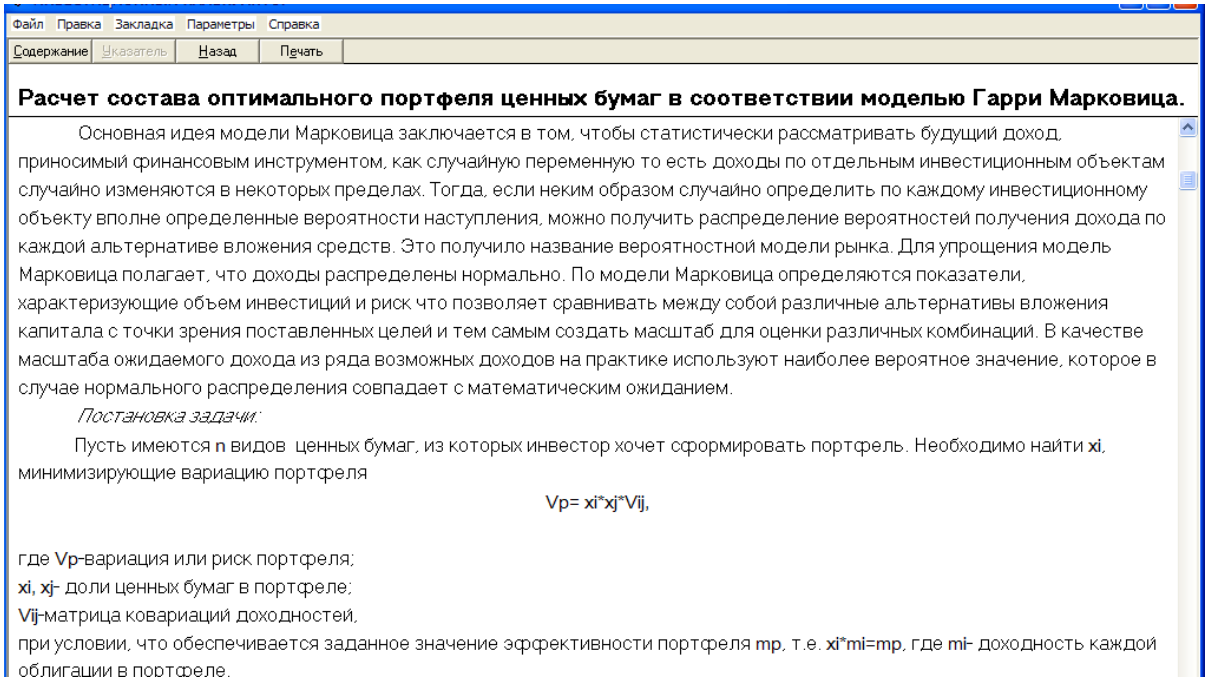


Рисунок 2.6. – Вид экранной формы «Справка»

2.3. Анализ информационных процессов модели Шарпа

Отличительной особенностью данной модели является отсутствие необходимости расчета взаимной ковариации (или корреляции) оцениваемой ЦБ со всеми остальными. Оценивается ее доходность с фондовым индексом (например, RTSI, RBCC и др.). Поэтому безусловным достоинством данной модели является простота проведения расчетов [234]. Модель Марковица является мультииндексной, а модель Шарпа является одноиндексной (диагональной) моделью.

В одноиндексной модели Шарпа (ОМШ) доходность ИП рассчитывается как средне взвешенное значение доходностей, входящих в него ЦБ. При этом учитывается β - риск. Для расчета доходности используется выражение [71, 173, 234]:

$$R_p = R_f + \sum_{i=1}^{N}\left(\alpha_i \cdot W_i\right) + \left(R_m - R_f\right) \cdot \sum_{i=1}^{N}\left(\beta_i \cdot W_i\right), \qquad (2.15)$$

где R_f - доходность без риска; R_m - ожидаемая полная доходность РЦБ.

В основе определения риска ИП лежит оценка среднего квадратичного отклонения функции R_f, определяемая в соответствии с выражением [234]:

$$\sigma_p = \sqrt{\left(\sum_{i=1}^{N}\left(\beta_i \cdot W_i\right)\right)^2 \cdot \sigma_m{}^2 + \sum_{i=1}^{N}\left(\sigma_{\varepsilon i}{}^2 \cdot W_i^2\right)}, \qquad (2.16)$$

где σ_m - СКО рыночной доходности (риск РЦБ); $\beta_i, \sigma_{\varepsilon i}$ - β - риск и остаточный риск i - й ЦБ; W_i – вес i - й ЦБ.

Практическая реализация данной модели в ИСППИД предполагает определение долей ЦБ в ИП.

2.3.1. Постановка задачи определения долей ценных бумаг в оптимальном портфеле Шарпа

Пусть v_i, $i=1,\ldots,n$ есть доля ЦБ A_i в произвольном ИП $П$: $П=v_1A_1+\ldots+v_nA_n$. Основные параметры ИП определяются в соответствии с выражениями (2.15) и (2.16).

Необходимо выбрать доли $v_1+...+v_n$ таким образом, чтобы значение функции доходности R для результирующего ИП $П=(\sigma_{\text{П}}, \text{r}_{\text{П}})$ было максимальным с учетом заданного значения риска, то есть [173]:

$$\begin{cases} R_f + \sum_{i=1}^{N}\left(\alpha_i \cdot W_i\right) + \left(R_m - R_f\right) \cdot \sum_{i=1}^{N}\left(\beta_i \cdot W_i\right) \to \max; \\ \sqrt{\left(\sum_{i=1}^{N}\left(\beta_i \cdot W\right)_i\right)^2 \cdot \sigma_m{}^2 + \sum_{i=1}^{N}\left(\sigma_{\varepsilon i}{}^2 \cdot W_i^2\right)} \le \sigma_{req}; \\ W_i \ge 0; \\ \sum W_i = 1. \end{cases} \quad (2.17)$$

Основные положения данной модели заключаются в следующем.

1. Безрисковой ставкой доходности R_f является доходность государственных ЦБ (например, облигаций внутригосударственного займа).

2. Доходность РЦБ за период t представлена фондовыми индексами. При этом усредненное значение доходности ЦБ, представляющих РЦБ, в период времени t рассчитывается по формуле [71, 173]:

$$r_{mt} = \frac{\sum_{i=1}^{N} r_{it}}{N}, \quad (2.18)$$

где r_{mt} - доходность РЦБ за период времени t; r_{it} - доходность i - й ЦБ за период времени t.

3. Коэффициент β - риск ЦБ, рассчитывается в соответствии с выражением [71, 173]

$$\beta_i = \frac{\sum_{t=1}^{T}\left[\left(r_{mt} - R_{ft} - \frac{\sum_{t=1}^{T}\left(r_{mt} - R_{ft}\right)}{T}\right) \cdot \left(r_{it} - R_{ft} - \frac{\sum_{t=1}^{T}\left(r_{it} - R_{ft}\right)}{T}\right)\right]}{\sum_{t=1}^{T}\left(r_{mt} - R_{ft} - \frac{\sum_{t=1}^{T}\left(r_{mt} - R_{ft}\right)}{T}\right)}, \quad (2.19)$$

где β_i - β-риск i - й ЦБ; R_{ft} - безрисковая доходность за период времени t; T – исследуемое число временных периодов.

4. Переоцененная доходность ЦБ рассчитывается в соответствии с вы-

ражением [232]:

$$\alpha_i = \frac{\sum_{t=1}^{T}\left(r_{it} - R_{ft}\right)}{T} - \beta_i \cdot \frac{\sum_{t=1}^{T}\left(r_{mt} - R_{ft}\right)}{T}. \qquad (2.20)$$

5. Общий риск РЦБ рассчитывается в соответствии с выражением [57]:

$$\sigma_m = \sqrt{\frac{\sum_{t=1}^{T}\left(r_{mt} - R_{ft} - \frac{\sum_{t=1}^{T}\left(r_{mt} - R_{ft}\right)}{T}\right)^2}{T}}. \qquad (2.21)$$

Рассмотрим пример расчета параметров ИП в соответствии с данной моделью. Исходные данные для примера приведены в таблице 2.1. Они представляют собой еженедельные котировки ЦБ (акций) шести организаций за определенные периоды времени. Всего рассматривается восемь периодов времени [173].

Таблица 2.1 –Исходные данные о доходностях ЦБ

	НОМЕР ПЕРИОДА							
ЦБ:	1	2	3	4	5	6	7	8
ЦБ 1	-1,23%	0,01%	4,41%	-12,11%	32,43%	-5,22%	-17,54%	14,01%
ЦБ 2	-15,54%	0,02%	72,12%	10,85%	22,72%	1,13%	6,63%	-8,32%
ЦБ 3	11,22%	0,03%	19,73%	17,33%	-4,04%	4,24%	-25,92%	-9,01%
ЦБ 4	0,01%	0,02%	2,31%	1,11%	0,01%	0,01%	16,54%	0,01%
ЦБ 5	-0,83%	14,86%	33,58%	-27,05%	0,11%	23,25%	-4,43%	2,61%
ЦБ 6	47,34%	-11,93%	-9,44%	4,45%	22,05%	-4,91%	-3,31%	-8,65%
	НОМЕР ПЕРИОДА							
ЦБ:	9	10	11	12	13	14	15	
ЦБ 1	0,01%	-3,44%	-10,1%	30,02%	1,21%	-8,95%	-1,44%	
ЦБ 2	1,13%	3,41%	5,22%	-7,63%	-13,23%	2,77%	8,53%	
ЦБ 3	23,06%	10,63%	0,93%	0,01%	12,33%	0,01%	-34,01%	
ЦБ 4	0,01%	5,74%	1,53%	-0,73%	0,01%	-1,43%	-12,52%	
ЦБ 5	0,30%	13,42%	4,72%	0,02%	-23,53%	9,42%	3,03%	
ЦБ 6	-3,73%	-1,22%	0,01%	32,13%	17,32%	2,00%	-1,95%	

Общая рыночная доходность рассчитывалась на основании оценок экспертов, поскольку данные из внешних источников отсутствовали. Безрисковая доходность R_f соответствовала доходности трехмесячных государственных облигаций, приведенных к недельному сроку. Примеры данных об общей ры-

ночной доходности и безрисковой доходности приведены в таблице 2.2.

Таблица 2.2 – Данные о рыночной и безрисковой доходности

Период	Полная рыночная доходность	Безрисковая доходность
1	5,1 %	0,80 %
2	2,4 %	0,80 %
3	11 %	0,75 %
4	3 %	0,75 %
5	6 %	0,75 %
6	3 %	0,85 %
7	1,6 %	0,85 %
8	3 %	0,85 %
9	4 %	0,85 %
10	3,4 %	0,80 %
11	2,2 %	0,75 %
12	4 %	0,75 %
13	1,4 %	0,75 %
14	3 %	0,75 %
15	2 %	0,75 %

Примеры рассчитанных характеристик ЦБ (β - риск, избыточная доходность α_i и остаточный риск σ_i) приведены в таблице 2.3.

Таблица 2.3 – Рассчитанные значения основных характеристик ЦБ

Тип ЦБ	β - риск	Избыточная доходность	Остаточный риск
ЦБ 1 типа	2,884	-7,03 %	11,88 %
ЦБ 2 типа	5,912	-10,53 %	14,33 %
ЦБ 3 типа	2,671	-6,14 %	11,34 %
ЦБ 4 типа	0,133	-0,32 %	5,52 %
ЦБ 5 типа	3,354	-6,44 %	12,63 %
ЦБ 6 типа	1,564	0,32 %	15,94 %

При проведении расчетов учитывался ряд параметров ИП, таких как:

- требуемая доходность портфеля составляет R_{pt}=3 %,;

- допустимый портфельный риск - σ_{pt}=7 %;

- предполагаемая безрисковая доходность - R_f=1,5 %;

- прогнозируемая рыночная доходность составляет R_p=3,8 %.

Результаты оптимизации инвестиционного портфеля для приведенных

6 типов акций, представлены в таблице 2.4.

Таблица 2.4 – Результаты оптимизации ИП на основе модели Шарпа

Прогноз:	Доходность рынка 3,8 % ; безрисковая доходность 1,5 %	
	Структура инвестиционного портфеля	
	Прямая задача	Обратная задача
Требования:	Риск меньше 7 %	Доходность выше 3 %
ЦБ 1 типа	0 %	0 %
ЦБ 2 типа	17 %	22 %
ЦБ 3 типа	0 %	0 %
ЦБ 4 типа	37 %	22 %
ЦБ 5 типа	10 %	10 %
ЦБ 6 типа	33 %	42 %
Характеристики оптимального портфеля	Доходность 3,39 % Риск 7 %	Доходность 5 % Риск 9,74 %

В рамках проведенной оптимизации решалась прямая задача (максимизация доходности при заданном риске) и обратная (определение минимального риска при заданном уровне доходности).

Вышеизложенный механизм решения задачи определения долей ЦБ в оптимальном ИП Шарпа может быть представлен в виде следующего алгоритма.

2.3.2. Алгоритм определения долей ценных бумаг в оптимальном портфеле Шарпа

Обобщенный алгоритм реализации ОМШ представляет собой выполнение следующей последовательности действий [59, 173]:

Шаг 1. Выбор числа ЦБ - *M*, из которых формируется ИП, и определение исторического промежутка в *N* шагов расчета, в течение которого рассматриваются значения доходности r_{it} для каждой ЦБ.

Шаг 2. Определение рыночных доходностей r_{mt} за рассматриваемый период времени по выбранному рыночному индексу.

Шаг 3. Определение величины дисперсии рыночного показателя σ_m^2, значений ковариаций $\sigma_{i,m}$ доходностей каждой ЦБ с рыночной доходностью

ИП (выбранным фондовым индексом) и расчет величины $\beta_i = \frac{\sigma_{i,m}}{\sigma_m^2}$.

Шаг 4. Вычисление ожидаемых доходностей каждой ЦБ $R_i(r_i)$, доходности рыночного ИП $R_p(r_m)$ и расчет параметра $\alpha_i = R_i(r_i) - \beta R_p(r_m)$.

Шаг 5. Вычисление дисперсии случайных ошибок ε_i.

Шаг 6. Подстановка полученных значений в соответствующие уравнения. Определение весов ЦБ для заданного значения прогнозируемой доходности ИП.

Шаг 7. Отображение границы эффективных ИП и выбор среди них оптимального.

2.3.3. Программная реализация определения долей ценных бумаг в оптимальном портфеле Шарпа

Реализация задачи определения долей ЦБ, входящих в оптимальный ИП Шарпа, представлена в виде программного модуля, включенного в состав ИСППИД [172]. В качестве среды разработки модуля использована объектно–ориентированная среда разработки Delphi 7.0. Обоснование выбора данной среды представлено в [173].

Экранная форма расчета долей ЦБ, входящих в оптимальный ИП Шарпа, приведена на рисунке 2.7.

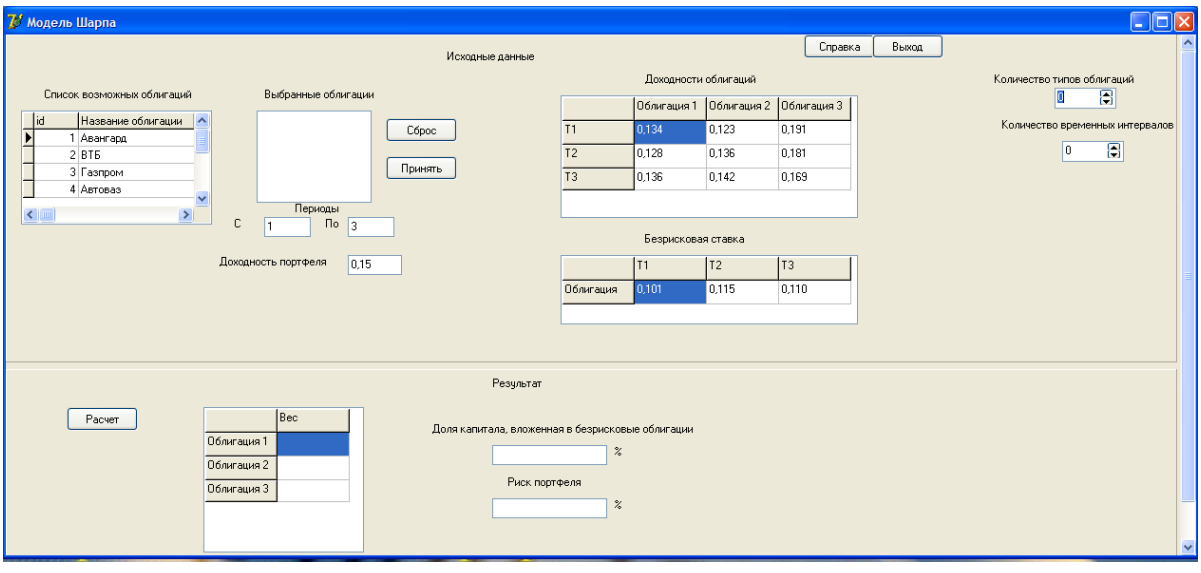


Рисунок 2.7. – Вид экранной формы расчета долей ЦБ по модели Шарпа

Верхняя часть экранной формы предназначена для ввода исходных данных для расчета. В левой верхней части экранной формы находится окно со списком ЦБ (облигаций), которые могут быть включены в состав ИП. Рядом находится окно, в котором отображаются выбранные ЦБ. В средней части экранной формы находится окно «Доходность к погашению». В правой части формы размещены два окна, в которых задаются количество типов ЦБ и количество рассматриваемых временных интервалов, соответственно. Они заполняются на начальном этапе работы с программным модулем. Затем осуществляется выбор конкретных ЦБ. После выбора необходимой ЦБ, ЛПР нажимает правую кнопку мыши. Появляется кнопка «Добавить». Нажатием данной кнопки, соответствующая ЦБ, перемещается в окно «Выбранные облигации». Далее необходимо ввести требуемые периоды наблюдения за доходностями ЦБ. Соответствующие окна ввода, размещены снизу окна «Выбранные облигации». Кнопки «Сброс» и «Принять», размещенные справа от этого окна, обеспечивают соответственно удаление или принятие списка выбранных ЦБ. Перечень выбранных (принятых) ЦБ и их доходность за выбранные периоды, отображается в окне «Доходность к погашению», размещенного в правой верхней части экранной формы. Значения доходности ЦБ могут корректироваться и вводиться с клавиатуры.

В нижней части экранной формы отображаются результаты расчета. Они появляются после ввода исходных данных (заполнения верхней части экранной формы) и нажатия кнопки «Расчет». В состав результатов расчета включены доля капитала, вложенная в безрисковые ЦБ, риск ИП и процентные соотношения выбранных ЦБ, в виде соответствующих весов.

Кроме того на данной экранной форме имеются кнопки «Справка» и «Выход». Нажатие кнопки «Выход» обеспечивает возврат на главную форму ИСППИД. При нажатии на кнопку «Справка», появляется соответствующая экранная форма, в которой отражаются основные сведения об используемой модели.

Результаты проведенных исследований показывают, что использование

в данной модели стохастических данных о ЦБ приводит к значительным ошибкам. Это обусловливает необходимость ее модернизации.

2.4. Информационные процессы нейромодифицированной модели Шарпа

Недостаток ОМШ состоит в том, что ИП, формируемый на ее основе, в упреждающие моменты времени теряет свойства оптимальности. Математическое доказательство данного факта приведено в [36]. При поступлении стохастических данных о ЦБ подобные потери оказываются достаточно продолжительными, что приводит к существенным ошибкам и снижению доходности ИП. Причиной потери оптимальности является отсутствие в ОМШ механизма прогнозирования. Реализация подобного механизма путем использования опыта экспертов предложена в [36]. Однако предложенный механизм не лишен недостатков. Неизвестно сколько нужно экспертов. Кроме того, с передачей опыта экспертами есть определенные трудности. В разработанной нейромодифицированной модели Шарпа (НМШ) [132, 166] в роли экспертов прогнозных оценок используются ИНС. Суть предлагаемого подхода состоит в следующем [173].

ОМШ базируется на уравнении регрессии (2.22), устанавливающем связь между доходностью ЦБ, которая включена в ИП, и доходностью РЦБ в виде определенного значения соответствующего индекса [5]:

$$r_i(t) = \alpha_i + \beta_i r_I(t) + \varepsilon_i(t), \qquad (2.22)$$

где $r_i(t)$ - уровень доходности *i*-й ЦБ в момент времени *t*; $r_I(t)$ - значение индекса доходности РЦБ в момент времени *t*; α_i, β_i - параметры уравнения регрессии; $\varepsilon_i(t)$ - случайная инструментальная погрешность.

Составляющая доходности ЦБ, которая не зависит от динамики РЦБ, представляется параметром α_i. Он носит название сдвига или смещения. Данный коэффициент представляет собой меру недооценки или переоценки РЦБ исследуемой ЦБ. Если его значение положительно, то исследуемая ЦБ переоценена РЦБ, в противном случае имеет место недооценка. Значение данного параметра рассчитывается в соответствии с выражением [173]:

$$\alpha = \frac{\sum_{i=1}^{n} y_i}{n} - \frac{\beta \cdot \sum_{i=1}^{n} x_i}{n}, \qquad (2.23)$$

где y_i - доходность РЦБ в течение *i*-го периода времени; x_i – уровень доходности ЦБ в течение *i*-го периода времени; *n* – число периодов.

Не менее важным параметром является β_i. Это чувствительность исследуемой ЦБ к изменениям РЦБ. Если стоимость ЦБ изменятся быстрее выбранного индекса РЦБ, то β >1. Это указывает на то, что данная ЦБ более рискованная, чем РЦБ в среднем. В противном случае β_i<0 и ЦБ менее рискованная, чем РЦБ в среднем. Оценка данного параметра проводится на основе использования метода наименьших квадратов, путем сравнения доходности исследуемой ЦБ и индекса РЦБ в течение заданного временного интервала [173].

Для учета скачкообразных изменений параметров ЦБ в упреждающие моменты времени, введем в выражение (2.22) дополнительное слагаемое $p_i k_{ti}$. При этом оно примет следующий вид [173]:

$$r_i(t) = \alpha_i + p_i k_{ti} + \beta_i r_I(t) + \varepsilon_i(t), \qquad (2.24)$$

где *p* – значение оценки средневзвешенной величины изменений доходности ЦБ, k_{ti} - дихотомический параметр. Изменения параметра *p* могут носить скачкообразный характер.

Значение дихотомического параметра k_{ti} равно +1, если текущая доходность ЦБ превышает уровень тренда, иначе его значение равно –1 [132]. Формальное представление данного факта имеет следующий вид [173]:

$$k_{ti} = \begin{Bmatrix} +1, & \varepsilon_{ti} \geq 0 \\ -1, & \varepsilon_{ti} \leq 0 \end{Bmatrix}, \; t = \overline{1,T} \;, i = \overline{1,n}. \qquad (2.25)$$

Данное представление указывает на то, что доходность ЦБ определяется индексной доходностью РЦБ и некоторыми скачкообразными изменениями, имеющими место в динамике варьирования доходности ЦБ. Данные скачкообразные изменения интерпретируются как «риск – эффекты». Они не имеют внутрирыночного объяснения (например, волевые политические реше-

ния, результаты санкционных воздействий и др.). Тем не менее, они в каждый момент времени воздействуют на доходность ЦБ. Их изменение может, либо увеличивать доходность ЦБ, либо ее уменьшать. Величина оцененного параметра p [36] представляет собой среднюю величину подобных изменений на временном интервале.

Для прогнозирования значений «риск – эффектов» целесообразно применять ИНС. Они позволяют запоминать значения p для подобных условий, в различные исторические временные периоды. Другими словами ИНС могут выступать в роли носителей знаний экспертов. Необходимо лишь правильно их (знания) выбрать, то есть верно определить подобие параметров текущей ситуации и аналогичной, имевшей место в прошлом [146, 160, 173].

Видом ИНС, наиболее подходящим для решения данной задачи, является многослойный персептрон. Он хорошо изучен и способен обеспечить приемлемую точность прогнозных оценок при достаточно быстром обучении.

Процесс использования НМШ состоит в следующем.

В различные моменты времени в ходе биржевых торгов инициируется ряд ИНС, которые обучаются, а затем тестируется и верифицируются. Параметры обученных ИНС и прогнозные значения «риск – эффектов» доходностей ЦБ, полученных на их основе, записываются в БД. Особое значение данная последовательность действий имеет в период протекания аномальных ситуаций (предкризисных, кризисных и посткризисных состояний экономики). В дальнейшем, параметры текущей рыночной ситуации сравниваются с априорными, и в случае их подобия, из БД извлекается и инициализируется ранее сформированные ИНС. Прогнозные значения «риск – эффектов» доходностей ЦБ, полученные на их основе, используются для текущей оценки доходности соответствующей ЦБ [173].

ИНС может не использоваться в том случае, если несколько раз подряд (более трех) при тестировании отмечаются ошибочные результаты. ИНС переходит в режим дообучения, а оценка доходности осуществляется на основе одноиндексной модели Шарпа. Такой подход актуален, так как обеспечивает

возможность накопления ошибок.

Применение НМШ позволит более точно проводить определение и оценку отклонений доходности ЦБ на упреждающем временном интервале. В случае некорректной настройки ИНС, последовательность применения НМШ позволяет исключить последнюю из процесса проведения определения доходностей ЦБ и передать управление одноиндексной модели Шарпа.

Дальнейшим развитием НМШ явилась нейрокомитетная модель Шарпа (НКМШ).

2.5. Разработка нейрокомитетной модели Шарпа

В НКМШ для установления в информационных потоках временных зависимостей изменений «риск – эффектов» ЦБ и их запоминания применяются КИНС.

КИНС представляет собой определенную структуру, включающую множество объединенных между собой, в соответствии с заданной целью ИНС, каждая из которых реализует некоторый набор функций в соответствии с ее назначением [133].

Технологическая схема обработки стохастических данных ЦБ с использованием КИНС приведена на рисунке 2.8.

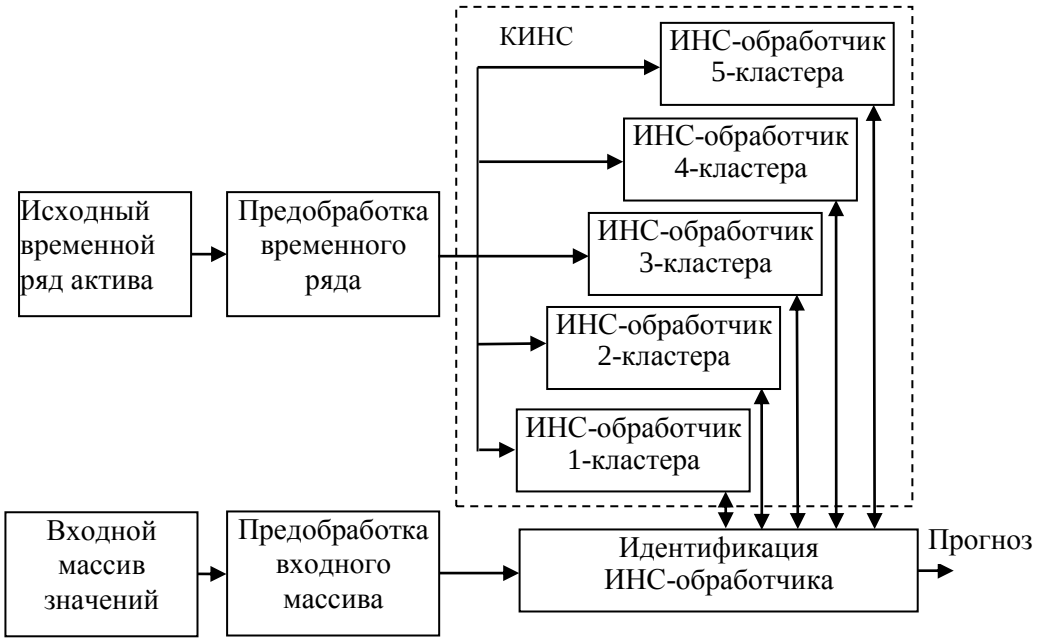


Рисунок 2.8. – Технологическая схема обработки стохастических данных ЦБ с использованием КИНС

Информационный процесс обработки стохастических данных ЦБ с использованием КИНС заключается в следующем. Исследуемый временной ряд ЦБ до поступления в КИНС проходит предварительную обработку. Она требуется для формирования множества подвыборок данных необходимых для обучения ИНС-обработчиков и предполагает проведение сегментации и кластеризации исследуемого временного ряда ЦБ.

Суть сегментации заключается в разделении входного множества данных на *n* сегментов (подмножеств), каждый из которых соответствует некоторому распределению. В состав сегмента входит 10 дискрет. Дистанция между сегментами r_{jh} определяется величиной Дженсона-Шеннона:

$$r_{\mathrm{jh}}(\mathrm{p}_1, \mathrm{p}_2) = \frac{r_{kl}(\mathrm{p}_1(\mathrm{x}), \mathrm{p}_2(\mathrm{x})) + r_{kl}(\mathrm{p}_2(\mathrm{x}), \mathrm{p}_1(\mathrm{x}))}{2}, \qquad (2.26)$$

где p_1, p_2 нормальные распределения.

Данная величина представляет собой среднее расстояние Кульбака-Ляйблера r_{kl}, широко используемого в качестве критерия различия нескольких распределений [239]:

$$r_{kl}(\mathrm{p}_1(\mathrm{x}), \mathrm{p}_2(\mathrm{x})) = \sum_{\mathrm{i}} \ln\left(\frac{\mathrm{p}_1(\mathrm{i})}{\mathrm{p}_2(\mathrm{i})}\right) \mathrm{p}_1(\mathrm{i}). \qquad (2.27)$$

В технологическом плане сегментация реализуется следующим образом.

1. Осуществляется параметрическое разделение массива исходных данных на *n* сегментов. В качестве параметров выступают минимальное число точек, задающих распределение – *m* (в данном случае *m=10*) и значение порогового расстояния межсегментного различия (Дженсона-Шеннона) – *s* (в данном случае *s=2*).

2. Реализуется итеративное разделение скользящего массива-окна, состоящего из *N* дискрет (в данном случае *N=100*) на сегменты. Первая итерация включает разделение массива-окна на две части. Критерием определения точки разделения является расстояние Дженсона-Шеннона *s*. Оно измеряется между входным распределением и данным распределением, представленным в виде двух совмещенных распределений, соответствующих нормальному зако-

ну, размещенных справа и слева от точки деления:

$$s(q) = \log \frac{p_2(q)}{p_1},$$

q - точка деления массива исходных данных; p_1 – функция правдоподобия нормального закона распределения.

Функция правдоподобия нормального закона распределения определяется в соответствии с выражением:

$$p_1 = \prod_{t=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad (2.28)$$

где σ- среднее значение нормального закона распределения; μ - стандартное отклонение нормального закона распределения.

Функция правдоподобия совмещенного распределения определяется в соответствии с выражением:

$$p_2(q) = \prod_{t=1}^{q} \frac{1}{\sigma_L\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu_L)^2}{2\sigma_L^2}} \cdot \prod_{t=q+1}^{N} \frac{1}{\sigma_R\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu_R)^2}{2\sigma_R^2}}, \qquad (2.29)$$

σ_L, μ_L - параметры левого совмещенного распределения; σ_R, μ_R - параметры правого совмещенного распределения.

3. Продолжается итеративное деление скользящего массива-окна путем половинного деления полученных частей до выполнения одного из условий, либо $s<2$ (невозможно различить распределения), либо $m<10$ (протяженность сегмента меньше установленной).

Результатом сегментации является набор сегментов, для обработки которого в дальнейшем используется кластеризация.

Кластеризация заключается в группировании близких по параметрам (подобных по распределению) сегментов. Необходимость ее проведения обусловлена учетом основных режимов работы РЦБ (утреннего, дневного, летнего и др.) и целесообразности (обязательности) обучения ИНС на наборах данных, соответствующих этим режимам. Проведение кластеризации основано на методе *k*-средних. Суть данного метода заключается в группировании множества точек относительно некоторой - центральной по критерию мини-

мума суммарного квадратичного отклонения их взаимного расстояния [261]. Каждый сегмент имеет два характерных параметра: стандартное отклонение и среднее, что соответствует координатам некоторой точки на плоскости. В итоге вся совокупность выявленных сегментов представляется в виде некоторого набора «сгустков» точек. Центры кластеров определяются как минимальное значение суммарного квадратичного отклонения множества точек от предполагаемого центра. Выбор числа кластеров осуществляется на основе оценки кластерного насыщения. Кластерное насыщение представляет собой сумму множества расстояний от центров кластеров до точек, входящих в них сегментов R_s. Оценка кластерного насыщения проводится путем увеличения количества кластеров и определения степени его убывания функции расстояния [261], определяемой в соответствии с выражением:

$$R_s = \sum_{i=1}^{Q} \sum_{j=1}^{L} f(r_i, k_{i,j}) \qquad (2.30)$$

где Q – количество кластеров, L – количество точек в кластере, r_i – положение центра i-го кластера, $f(r_i, k_{i,j})$ - функция изменения расстояния до кластерных центров. Проведенные эксперименты показали, что число выявленных кластеров для 100 дискретного скользящего массива не превышает пяти. Причем их размещение достаточно равномерно. Это обусловило число ИНС – обработчиков кластеров, включенных в состав КИНС. В качестве ИНС – обработчиков использованы однослойные персептроны, включающие двухэлементный вход, двухнейронный скрытый слой и двухнейронный двухвыходной слой. Для активации нейронов используется сигмоидная функция. На один из входов ИНС поступают значения временного ряда доходностей ЦБ, а на другой значения логарифмов разности максимумов и минимумов доходностей ЦБ лага торгового периода, что является эквивалентом ее «риск – эффекта». Два выхода ИНС требуются для прогнозирования трендов возрастания или убывания «риск – эффектов» ЦБ. В интересах преодоления «застревания» ИНС в локальных экстремумах, так называемого эффекта «овражности», для их обучения использован МГА.

Порядок функционирования КИНС заключается в следующем. Как и обычные ИНС, КИНС проходит этапы обучения и верификации на множестве исторических данных ЦБ. После их завершения, КИНС переходит в рабочий режим, в рамках которого на вход системы поступает массив исследуемых (текущих) значений параметров (доходностей) ЦБ, состоящий из десяти дискрет (текущей и девяти предшествующих). Для данного массива проводится предобработка, заключающаяся в определении стандартного отклонения и среднего, что эквивалентно определению точки на кластерной плоскости. Далее, путем сравнения расстояний от данной точки до центров кластеров, определяется ближайший, что эквивалентно идентификации ИНС-обработчика. После чего, на ее основе осуществляется прогноз направления тренда «риск – эффекта» ЦБ за один исследуемый торговый период. Аналогичным образом определяются направления трендов «риск – эффектов» других ЦБ.

В дальнейшем, процесс применения КИНС аналогичен использованию НМШ, подробно изложенному в п. 2.4.

Для исследования ИНС и оценки возможностей их использования в интересах НМШ и НКМШ разработан программный комплекс нейросетевого прогнозирования временных рядов, описание которого приведено в 3-й главе диссертационной работы [167].

Важным этапом применения ИНС является их обучение, в рамках которого с целью преодоления «овражности» ИНС использован МГА. Кроме того МГА использован непосредственно в интересах формирования ИП СЭО.

2.6. Модифицированный генетический алгоритм распределения инвестиций

Основные трудности применения большинства методов оптимизации нелинейных унимодальных функций связаны с проблемами «проклятия размерности» и «застревания» в локальных экстремумах [34, 65].

Одним из способов их преодоления, является использование генетических алгоритмов (ГА), представляющих собой одну из разновидностей случайного поиска, основанную на естественном отборе и размножении [34].

Механизм функционирования ГА достаточно хорошо изучен и представляет собой поэтапную последовательность действий [31].

На первом этапе случайным образом формируется начальная популяция особей. В хромосоме (генотипе) каждая особь представляет собой закодированное возможное решение задачи (фенотип). Качество каждого решения оценивается функцией приспособленности (фитнес-функцией).

На втором этапе осуществляется выбор особей и их скрещивание на основе оператора кроссовера. В результате формируется потомство, генетическая информация которого является результатом обмена хромосом родительских особей.

На следующем этапе из особей вторичного потомства, аналогичным образом формируется новое поколение. В данном поколении могут встречаться мутанты – особи со случайно измененными генотипами. Последнее реализуется на основе генетического оператора мутации. Эволюция популяции включает некоторую последовательность таких поколений [31].

Постановка задачи ПИР в терминах ГА и последовательность ее решения представляется следующим образом.

Пусть задана сложная ЦФ, представляющая собой суммарный доход инвестора. Она зависит от нескольких аргументов (переменных) - объемов инвестиций в каждый проект. Заданы значения минимального и максимального объемов вложений в каждый из проектов. Они определяют область изменения каждой из переменных. Требуется найти такие значения аргументов, при которых ЦФ достигает максимального значения.

Представим каждый вариант инвестирования (набор значений переменных) в виде некоторой особи. Доходность каждого варианта соответствует приспособленности особи. В процессе эволюции приспособленность особей будет возрастать, что означает появление более доходных вариантов для инвестирования. Если остановить эволюционный процесс в некоторый момент времени и выбрать наилучшую особь, то возможно получение хорошего решения задачи.

Функционирование ГА, применительно к решаемой задаче, представляет собой следующую последовательность действий.

На начальной стадии работы ГА формируется исходная популяция [57, 244]:

$$P = \{P_1, P_2, ..., P_j, ..., P_N\},$$

где *j = 0, 1, 2, …* – номер генерации ГА, а *N* – размер популяции (количество особей в ней), не изменяющийся в течение работы ГА. Генерируемая особь представляет собой случайную *L*-битную строку. В данном случае *L* это длина кодировки особи. Она фиксирована и для всех особей неизменна. Каждая строка состоит из генов:

$$P_i = \{g_1, g_2, ..., g_v\}.$$

При выборе способа кодирования необходимо в хромосоме закодировать (с определенной погрешностью) любую точку из пространства поиска. Если этого не сделать, то возможно значительное увеличение времени эволюционного поиска.

Все численные параметры решения задачи кодируются в хромосоме на основе комбинированного кодирования (симбиоза целочисленного и вещественного кодирований).

Для декодирования значений используются следующие формулы:

$$r = \frac{g(x_{\max} - x_{\min})}{2^m - 1} + x_{\min},$$

$$g = \frac{(r - x_{\min})(2^m - 1)}{(x_{\max} - x_{\min})},$$

где r – декодированное (вещественное) значение параметра; g - закодированное целочисленное значение параметра; $x_{\max}$ и $x_{\min}$ - максимум и минимум допустимых значений декодированного параметра [245, 246].

После формирования и кодировки популяции, ГА начинает циклическую итерационную обработку информации. Каждый шаг цикла включает три стадии: «генерация промежуточной популяции - скрещивание особей – мутация».

Набор особей, получивший право размножаться, составляет промежуточную популяцию. Особи, которые приспособились, записываются несколько раз. Ущербные (плохие) особи из дальнейшего рассмотрения исключаются.

Процесс выбора *K*N* хромосом популяции *G(t)* для дальнейших генетических операций составляет существо генерация промежуточной популяции. Хромосомы выбираются случайно с вероятностью пропорциональной её ценности.

Число экземпляров хромосом S_i^t в популяции *G(t+1)* определяется в соответствии с выражением:

$$n_{выб}\left(S_i^t\right) = p_{выб}\left(S_i^t\right) * K * N .$$

Вероятность попадания особи в промежуточную популяцию пропорциональна ее приспособленности. В этом заключается пропорциональный отбор. Его реализация осуществляется следующими способами.

В первом, так называемом стохастическом способе, особи располагаются на рулеточном колесе таким образом, что размер сектора особи прямопропорционален ее приспособленности. На начальном этапе промежуточная популяция пуста. Необходимое число особей для промежуточной популяции реализуется путем запуска рулетки *N* раз. Каждая выбранная особь сохраняется на рулетке.

В рамках пропорционального отбора, особь выбирают с учетом отношения ее приспособленности к средней приспособленности популяции [263].

Целая часть рассчитанного отношения показывает число необходимых записей особи в промежуточную популяцию. Дробная часть соответствует вероятности ее многократного попадания. Реализация этого способа осуществляется в следующем порядке. Разместим особи на колесе рулетки, как и в стохастическом способе. Рулетка имеет *N* стрелок, которые отсекают одинаковые сектора. Однократный запуск рулетки позволяет выбрать *N* особей. Они запишутся в промежуточную популяцию. Для увеличения общей ценности последующей популяции используется операция воспроизводства. Она позволяет увеличить количество строк, представляющих наиболее ценные особи.

В результате воспроизводства, схемы у которых соответствующие под-

популяции имеют более высокую среднюю ценность в популяции, увеличивают число своих особей.

В рамках воспроизводства оперирует с хромосомами, существующими в рассматриваемой популяции, и нет возможности открывать новые области поиска. Чтобы это реализовать необходимо скрещивание.

При скрещивании происходит случайный обмен элементами между произвольно сформированными парами хромосом. С этой целью, выбранные случайным образом на этапе воспроизводства хромосомы, группируются в пары. После этого каждая пара подвергается скрещиванию с заданной вероятностью. На этапе скрещивания, позиции разделителя *d (d=1, 2, ..., l-1*, где *l* - длина строки) выбираются случайным образом. После чего, значения *d* - первых элементов одной хромосомы записываются в соответствующие элементы другой. Аналогично, значения первых *d* элементов второй хромосомы записываются в соответствующие элементы первой. В итоге формируются две новые хромосомы. Каждая из них представляет собой комбинацию частей двух родительских хромосом [250].

В интересах скрещивания использовался одноточечный оператор кроссовера (*1*-point crossover). В результате скрещивания формируются новые хромосомы путём комбинации значений элементов, представляющих наибольшую ценность в популяции G(t) хромосом. Сформированные хромосомы могут быть более ценными, чем их родители, что соответствует эволюционному характеру развития.

Формирование новых областей поиска за счет операции скрещивания реализуется путём перегруппировки комбинаций символов, имеющихся в популяции. Поэтому если использовать только эту операцию, часть потенциально оптимальных областей могут остаться не охваченными. Чтобы этого не происходило, используется операция мутации.

Суть мутации заключается в случайном изменении значений некоторых элементов хромосом. С этой целью, хромосомы сформированные путем скрещивания, поэлементно просматриваются. В процессе просмотра реализуется

процедура мутации, заключающаяся в том, что каждый элемент с заданной вероятностью (вероятностью мутации $p_{мут}$) может претерпеть изменение (измениться на случайно выбранный символ, допустимый для данной позиции). Вероятность мутации составляет менее 1%. Мутация позволяет найти новые комбинации признаков, которые увеличивают ценность хромосом популяции (новые квазиоптимальные решения).

В результате применения описанных операций формируется группа новых хромосом численностью *K*N*. Они, либо полностью формируют новую популяцию *G(t+1)* (при *K=1*), путем замены всех хромосом популяции *G(t)*, либо формируют часть популяции *G(t+1)*, заменяя собой *K*N* наименее ценных хромосом предыдущей популяции. Это значит, что закон $F_0(w_1, w_2, ..., w_n)$ вероятности распределения значений ЦФ определяется и корректируется путём использования популяции (набора) хромосом. Они включают комбинации элементов с наилучшими значениями ЦФ.

Процесс генерации промежуточной популяции, скрещивания и мутации формирует новое поколение. Каждая итерация генетического алгоритма завершается объявлением нового поколения текущим. После чего, над этим поколением все действия повторяются. Данный итеративный процесс эволюции может продолжаться бесконечно долго.

Критерием его завершения может служить, либо заданное число поколений, либо сходимость популяции. Под сходимостью понимается такое состояние популяции, при котором все строки популяции практически одинаковы, а значения целевой функции концентрируются в области некоторого экстремума. При этом кроссовер практически не меняет популяции. Кроме того, вышедшие за счет мутации из этой области особи склонны вымирать. Они, как правило, имеют меньшую приспособленность, особенно в случае наличия глобального максимума. Следовательно, сходимость популяции означает, что найдено лучшее решение.

Блок – схема ГА распределения финансовых инвестиций приведена на рисунке 2.9.

Рисунок 2.9. – Блок-схема ГА распределения финансовых инвестиций

Блоки 1,12 используются для пуска и остановки процесса распределения финансовых инвестиций.

В блоке 2 реализован ввод исходных данных, таких как: количество инвестируемых проектов, общий объем инвестиций; диапазон изменений инвестиций для различных проектов.

В блоке 3 формируются множества начальных популяций, каждая из которых состоит из N индивидуумов (по количеству проектов). Данные популяции формируются случайным образом.

С помощью блока 4 производится расчет приспособленности каждого индивидуума (доходность каждого варианта) и популяции в целом (доходность всех вариантов).

В блоке 5 реализован выбор индивидуумов для скрещивания. Важную роль в организации данного процесса играет модель отбора индивидуумов. Она определяет, каким образом следует строить популяцию следующего поколения. От качества такой модели зависит эффективность (точность и оперативность) ГА. В данном алгоритме использована элитарная модель отбора [65]. Использование данной модели позволило ускорить сходимость ГА.

В блоках 6,7 реализованы процедуры скрещивания, мутации и инверсии.

Блок 8 реализует формирование новой популяции путем циклического добавления новых индивидуумов (вариантов инвестирования) в имеющуюся популяцию.

Блок 9 обеспечивает определение максимального значения ЦФ. В случае достижения глобального максимума (максимальной прибыли инвестора) управление поступает в блок 10. Иначе управление поступает в блок 4 и происходит дальнейшее последовательное выполнение процедур, реализуемых блоками 5,6,7,8.

В блоке 10 раскрывается описание реквизитов лучшего индивидуума во всей популяции.

Блок 11 обеспечивает вывод полученных результатов (рациональный

вариант распределения инвестиций по проектам и соответствующее значение прибыли инвестора).

Для повышения оперативности, ГА подвергся 2-х этапной процедуре модификации.

Первый этап модификации был предложен в [135]. Его обоснование проводилось в терминах постановки и решения классической задачи о рюкзаке [209].

Имеется рюкзак, который способен выдержать вес, ограниченный некоторой константой P. Существует N предметов, каждый из которых имеет определенный вес p_i и стоимость s_i. Необходимо в рюкзак положить такую совокупность предметов m, $m \in N$, чтобы их суммарный вес P_m не превысил константу P ($P_m \le P$), а суммарная стоимость этих предметов S_m была максимальной $S_m \to \max$.

Обычно данную задачу решают методами прямого перебора и ветвей и границ. Эти методы просты и практически всегда дают точное решение. Их основными недостатками является экспоненциальная сложность заложенных алгоритмов и соответственно высокая ресурсоемкость, предъявляемая к вычислительным средствам.

Специфика МГА, реализующего данную задачу состояла в следующем.

Решения кодировались следующим образом. Значение j-го гена хромосомы, где $j = 1, 2 .. k$, равно единице, если i-ый предмет положили в рюкзак, и равно нулю, если предмет не положили.

Учитывалось, что в виду существующего ограничения P (вместимости рюкзака), не всякая комбинация нулей и единиц в векторе длины n может давать допустимое решение. Комбинация предметов, соответствующая произвольно заданной хромосоме, может превысить допустимый вес. Решалась данная проблема следующим образом. Сначала использовались стандартные операторы одноточечного или равномерного кроссовера, а затем полученная хромосома преобразовывалась в допустимую путем замены в ней некоторых

случайно выбранных единиц нулями.

Функция пригодности определяла степень выживаемости индивидуума в популяции. Чем больше значение функции пригодности индивидуума, тем ближе соответствующее ей решение к точному решению задачи о рюкзаке. В качестве функции пригодности конкретного индивидуума, использовалась суммарная стоимость предметов положенных в рюкзак.

Классический метод ветвей и границ, лежащий в основе построения деревьев и оценки перспективности их вершин, нашел своё развитие. Развитие заключалось в том, что строилось не всё дерево решений, а лишь его перспективные поддеревья (первое отличие). Отправной точкой построения поддерева являлась вершина, соответствующая решению, полученному в результате работы ГА. Далее достраивались вышестоящие вершины, которые после проведения соответствующих оценок, в свою очередь становились корневыми вершинами для новых поддеревьев. Другими словами, использовалась стратегия «снизу-вверх», в отличие от стратегии классического метода ветвей и границ - «сверху - вниз» (второе отличие). Полученная высота построенного дерева (совокупности поддеревьев) получила название «уровень приближения» и была принята в качестве числового параметра разработанного ГА. Данный параметр определяет протяженность области определения целевой функции, координаты центра которой (области) определены с использованием ГА. В результате проведенных экспериментов было установлено, что точность ГА прямо пропорционально зависит от заданного значения уровня приближения, а зависимость быстродействия носит обратный характер.

Для апробации разработанного ГА и сравнения его характеристик с параметрами существующих ГА, была реализована клиент - серверная программная система, предоставляющая пользователю следующие возможности:

- проводить загрузку пяти разновидностей ГА (canonical, simple, genitor, island, МГА) [185];

- устанавливать параметры ГА (объем популяции, генетические опера-

торы, количество итераций и т.д.), а также изменять эти параметры в ходе работы программы;

- хронометрировать временные режимы работы ГА;

- выводить результаты сравнения ГА в табличной и графической формах.

ГА обрабатывались популяции размерами 50, 100, 500, 1000, 5000 и 10000 индивидуумов. Для сравнения точности (σ,%) работы ГА, на начальном этапе данная задача решалась классическим методом ветвей и границ. Результат её решения использовался в качестве эталонного (100%). Результаты сравнительного анализа точности и оперативности (T, c) работы пяти ГА и модифицированного ГА приведены в таблице 2.5.

Таблица 2.5 – Результаты сравнительного анализа функционирования ГА

Вид ГА	Число индивидуумов в популяции											
	50		100		500		1000		5000		10000	
	σ,%	T, c	σ,%	T, c	σ,%	T, c	σ,%	T, c	σ,%	T, c	σ,%	T, c
Canonical	88	0,01	90	0,09	92	0,11	94	0,15	97	0,2	99	0,25
Simple	70	0,31	78	0,33	82	0,38	86	0,41	93	0,5	95	0,58
Genitor	72	0,0015	75	0,0019	77	0,0025	81	0,03	88	0,12	90	0,28
Island	80	0,29	84	0,33	88	0,47	91	0,77	94	0,89	96	1,65
МГА	95	0,23	99	0,29	100	0,39	100	0,69	100	0,82	100	1,04

Результаты вычислительных экспериментов показали, что наиболее оперативными оказались алгоритмы Genitor и Canonical, им уступают Island и Simple. МГА по быстродействию занимает промежуточное положение. По точности исследованные алгоритмы распределились следующим образом: МГА – 99%; Canonical – 93%; Island – 89%; Simple – 84%; Genitor – 81%. Наиболее точным оказался МГА.

Установлено, что с увеличением количества индивидуумов в популяциях точность решения задачи увеличивается, а оперативность проведения расчетов падает. Проведенные эксперименты позволили выделить МГА как наиболее рациональный по скорости и точности.

Второй этап модификации МГА состоял в его совмещении с островной моделью параллельных вычислений.

2.6.1. Разработка островной модели параллельных вычислений

Механизм функционирования островной модели параллельных вычислений (ОМПВ) заключается в выполнении следующей последовательности действий. На начальном этапе популяция делится на несколько одинаковых по размеру подпопуляций. Подпопуляции обрабатываются соответствующими процессорами, реализующими набор операций ГА. Через несколько поколений между ними производится обмен особями. Такие обмены называются миграциями. Они делают возможным совместное использование генетического материала подпопуляциями. Схематичное представление данного механизма приведено на рисунке 2.10 [157].

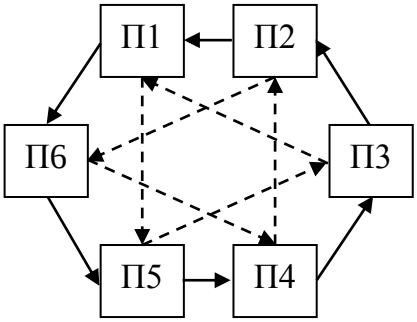


Рисунок 2.10. – Механизм функционирования ОМПВ

Рассмотрим особенности механизма функционирования ОМПВ на следующем примере. Пусть на 16 процессорах выполняются соответствующие ГА. В каждом ГА задействованы подпопуляции, включающие 1000 особей. В случае отсутствия миграций, реализуется 16 независимых поисков решения. Поиски проводятся в различных начальных популяциях. В конечном итоге все поиски сходятся к определенным доминирующим особям. Это показывают проведенные многочисленные исследования. Причина этого связана в следующем. Во-первых, число островов, осуществляющих прием мигрирующих доминирующих особей, ограничено (2-5 островов). Во-вторых, наблюдается односторонний обмен особями. Следовательно, в большой общей популяции возникнут островные группы с различными доминирующими особями. Когда популяции невелики, ложные доминирующие особи могут быстро мигрировать. В итоге, точное решение может находиться только на одном

острове, а на других островах будут находиться несколько ложных доминант. По причине миграции число ложных особей на островах возрастет. На каждый остров мигрируют особи, как минимум с 2-х соседних островов. Следовательно в малой популяции при функционировании ГА возможно появление ошибочных доминирующих особей. Последнее означает схождение ГА к ложному оптимуму. Для предотвращения этого нежелательного явления целесообразно увеличить частоту миграции. Частое мигрирование большого числа особей приводит к смешиванию особей (генетического материала). Это приводит к стиранию локальных различий между островами. В используемой модели миграции с каждого острова могут происходить лишь на определенное расстояние: 2-5 острова в зависимости от числа подпопуляции. В результате, каждый остров оказывается почти изолированным. Число островов, на которые могут мигрировать особи одной подпопуляции, называют расстоянием изоляции.

Блок – схема ГА распределения инвестиций, совмещенного с островной моделью параллельных вычислений, приведена на рисунке 2.11.

Функциональное назначение его блоков аналогично алгоритму, представленному на рисунке 2.9. Особенность данного алгоритма состояла в том, что каждая подпопуляция обрабатывалась (блоки 3,4,5,6,7,8,9) на отдельном процессоре (данные блоки в приведенном алгоритме выделены двойными линиями). Между подпопуляциями осуществлялся обмен (1 раз в итерацию общего цикла) индивидуумами посредством миграции. В результате, отыскивается лучший индивидуум во всей популяции [34].

В процессе обработки информации выделялись две группы операций:

- локальные операции, независимо выполняемые на каждом процессоре, в частности: расчет значений функций приспособленности для каждого индивидуума; выбор индивидуумов; скрещивание; мутации; формирование родительских пар;

- глобальные операции – обмен индивидуумами между процессорами, в частности операции коммуникации и синхронизации.

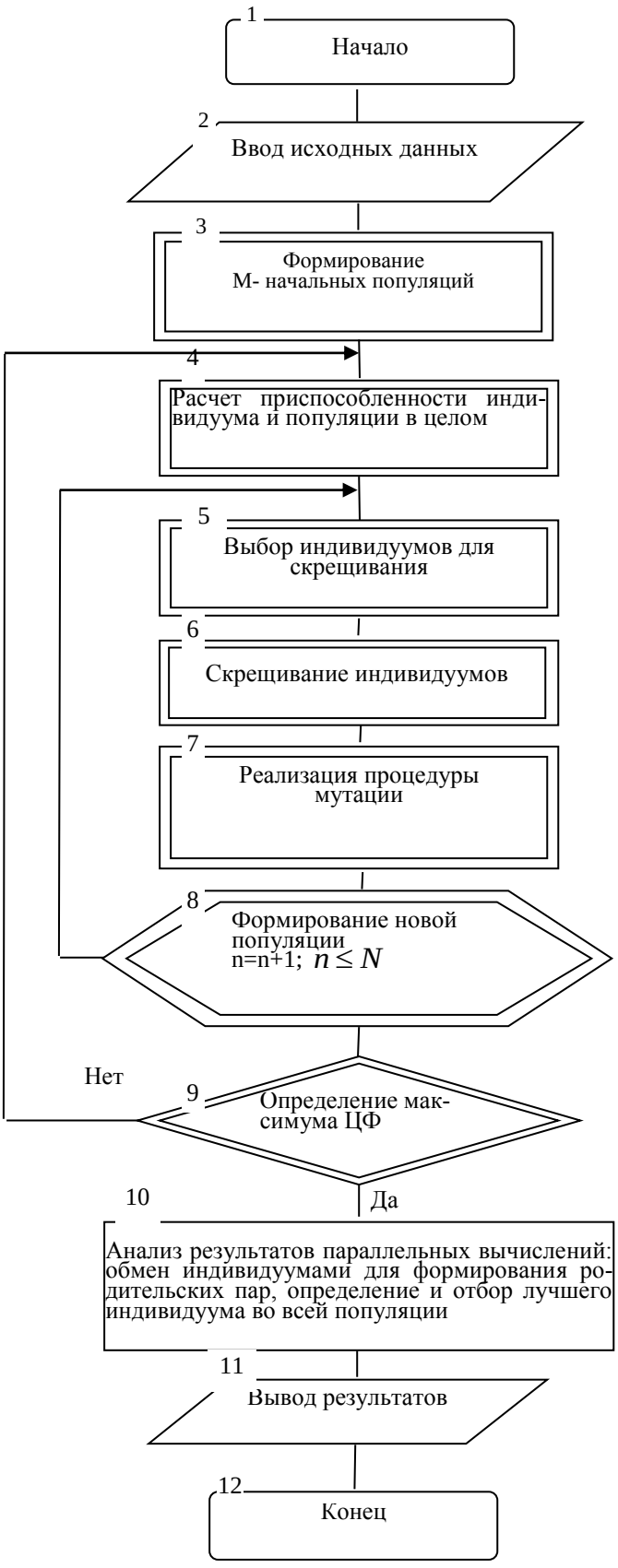


Рисунок 2.11. – Блок-схема ГА совмещенного с ОМПВ

С целью апробации данного алгоритма, в среде Delphi была разработана специальная программа, реализующая интерфейс пользователя, комплекс процедур описанных выше и набор инструментальных средств, используемых в интересах получения и анализа его (алгоритма) динамических свойств. В качестве аппаратной платформы использовались два персональных компьютера (сервера) Intel Pentium IV Dual Xeon 2.8 (двухпроцессорные) с объемом оперативной памяти – 1 Gb, объединенных в единую сеть. Это позволило реализовать проведение параллельных вычислений на 1, 2, 3, и 4 процессорах.

На рисунке 2.12 приведены зависимости изменения ускорения проведения вычислений от числа используемых процессоров для теоретических и реальных условий. Результаты получены относительно расчетов проведенных на одном процессоре. Данные зависимости имеют различия.

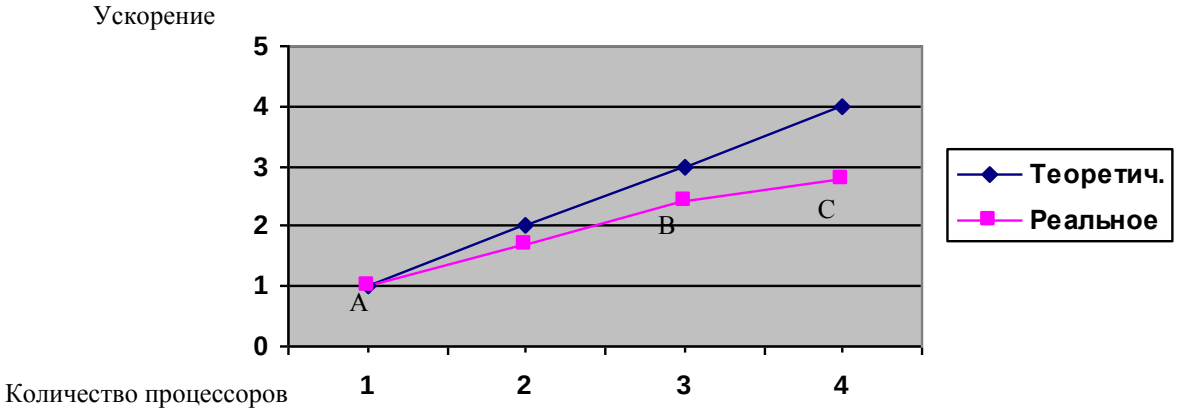


Рисунок 2.12. – Зависимости изменения ускорения проведения вычислений от количества используемых процессоров

На участке |AB| различие незначительное, оно обусловлено несовершенством используемого программного обеспечения (используется относительно «медленная» среда разработки Delphi). На участке |BC| различие более существенно, поскольку для реализации глобальных операций параллельных вычислений используется сетевой обмен данными, что существенно замедляет оперативность. Тем не менее, прирост в быстродействии при использовании 4 процессоров значителен (примерно на 3 порядка по скорости).

Таким образом, применение МГА распределения финансовых инвестиций, базирующегося на ОМПВ, позволило при приемлемой точности искомого

решения, почти на 3 порядка повысить оперативность проведения расчетов.

2.6.2. Программная реализация модифицированного генетического алгоритма распределения инвестиций

Реализация генетического алгоритма распределения финансовых инвестиций, представлена в виде ПК [102]. В качестве среды разработки ПК использована объектно-ориентированная среда разработки Delphi 7.0. Обоснование выбора данной среды приведено в [93, 146, 173].

Для запуска ПК необходимо выполнить файл ProjectGene.exe, расположенный в корневом каталоге клиентской составляющей. В случае его успешного выполнения на экране монитора появится окно главной экранной формы ПК, внешний вид которого представлен на рисунке 2.13.

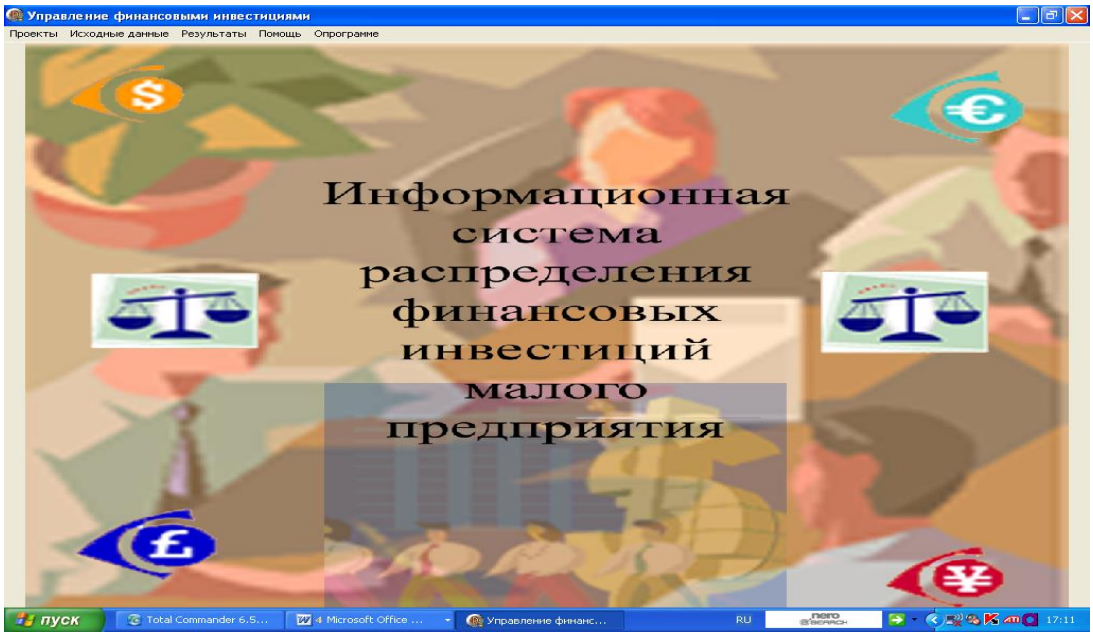


Рисунок 2.13. – Вид главной экранной формы программы

Главная экранная форма ПК содержит 5 закладок: «проекты»; «Исходные данные»; «Результаты»; «Помощь»; «О программе».

После выбора закладки «Проекты» пользователю необходимо пройти процедуры аутентификации и идентификации. В результате появится меню «Проекты», содержащее опции: «создать»; «открыть»; «закрыть»; «сохранить»; «печать»; «свойства»; «выход». Пользователь имеет возможность загружать, редактировать и сохранять готовые проекты или формировать но-

вые. Доступ к проектам обеспечивается в интерактивном режиме, путем выбора соответствующих дисков, каталогов и папок в которых они хранятся. При открытии существующего проекта пользовать должен ввести свою учетную запись и пароль, после чего ему будет обеспечен доступ к соответствующей базе данных. При нажатии кнопки «открыть», осуществится загрузка данных из хранилища данных в витрину, для текущей работы с ними. Далее пользователю необходимо выбрать закладку «Исходные данные».

При выборе закладки «Исходные данные», появляется меню, содержащее две опции «Настройка ГА» и «Данные по проектам». Выбор опции «Настройка ГА» приводит к появлению соответствующей экранной формы (рисунок 2.14).

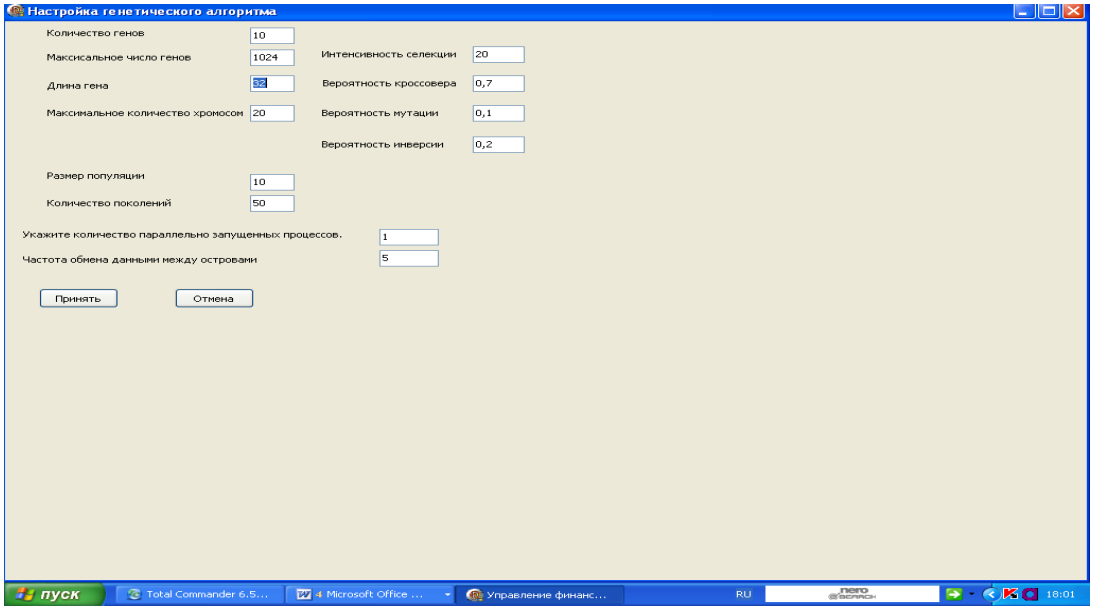


Рисунок 2.14. – Вид экранной формы «Настройка ГА»

Группа исходных данных для настройки ГА включает: количество генов; максимальное число генов; длина гена; максимальное количество хромосом; размер популяции; количество поколений; интенсивность селекции; вероятность кроссовера; вероятность мутации; вероятность инверсии. Для настройки островной модели параллельных вычислений вводятся количество запущенных параллельных процессов и частота обмена данными между островами.

Выбор опции «Данные по проектам» приводит к появлению соответствующей экранной формы (рисунок 2.15).

Результирующая таблица содержит:

- перечень финансовых инструментов, выбранных пользователем;

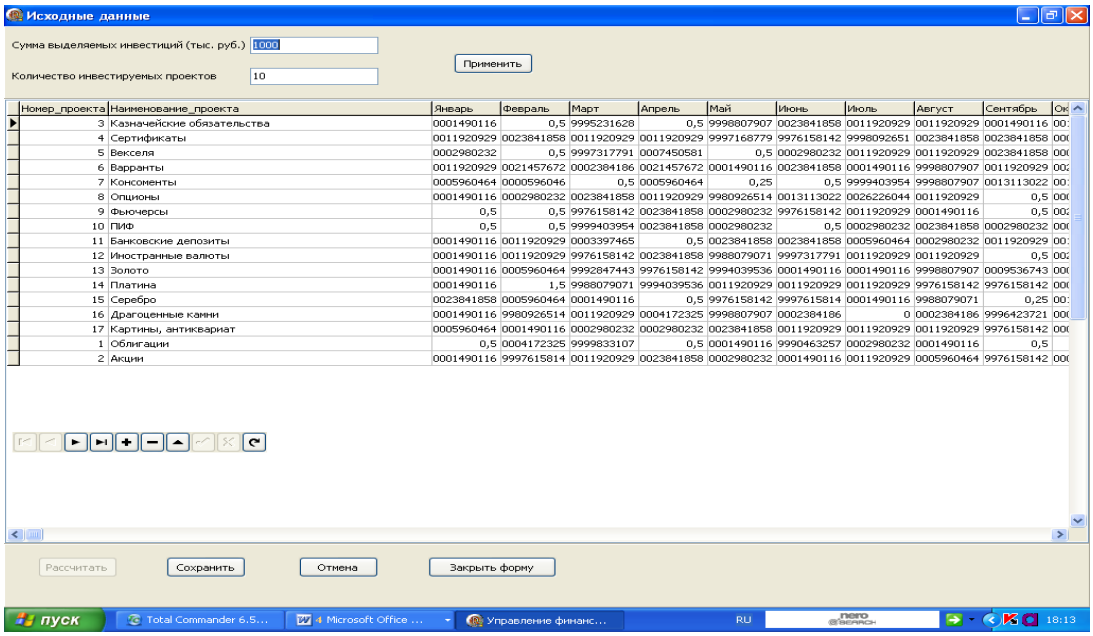


Рисунок 2.15. – Вид экранной формы «Данные по проектам»

Для ввода соответствующих значений необходимо нажать кнопку «Применить». После ее нажатия на экране появится экранная форма «Данные по проектам» (рисунок 2.15) с внесенными изменениями. После повторного отображения данного окна, на нем становится доступной кнопка «Рассчитать». Кроме того, на данной форме имеются кнопки «Отмена», «Сохранить» и «Закрыть форму», реализующие соответствующие действия.

Результаты работы ПК представляются в табличном и графическом видах. Табличная форма представления результатов приведена на рисунке 2.16.

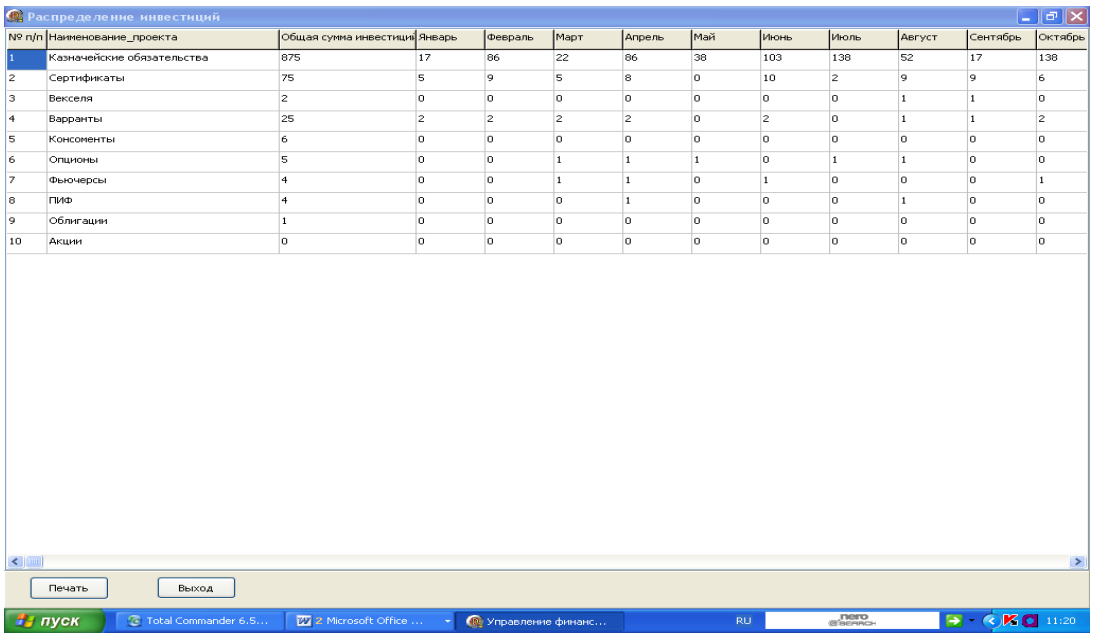


Рисунок 2.16. – Табличная форма представления результатов

- общее распределение суммы по финансовым инструментам;

- помесячное распределение сумм по финансовым инструментам.

На рисунке 2.17 приведена гистограмма распределения инвестиций на примере казначейских обязательств.

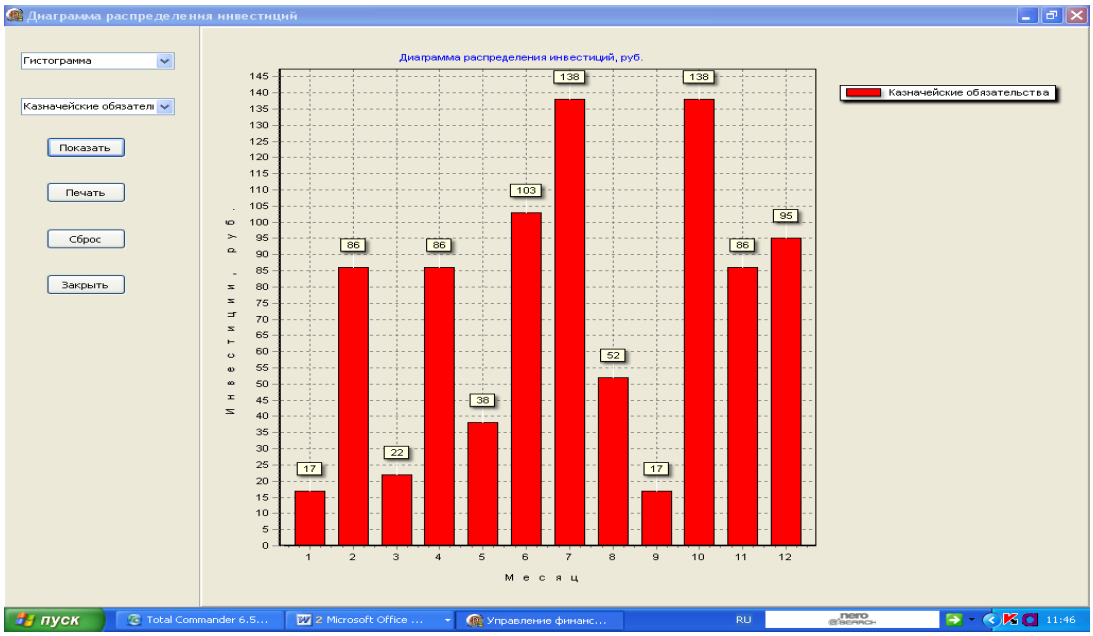


Рисунок 2.17. – Гистограмма распределения инвестиций

Обобщенная точечная диаграмма распределения инвестиций по всем финансовым инструментам данного ИП приведена на рисунке 2.18.

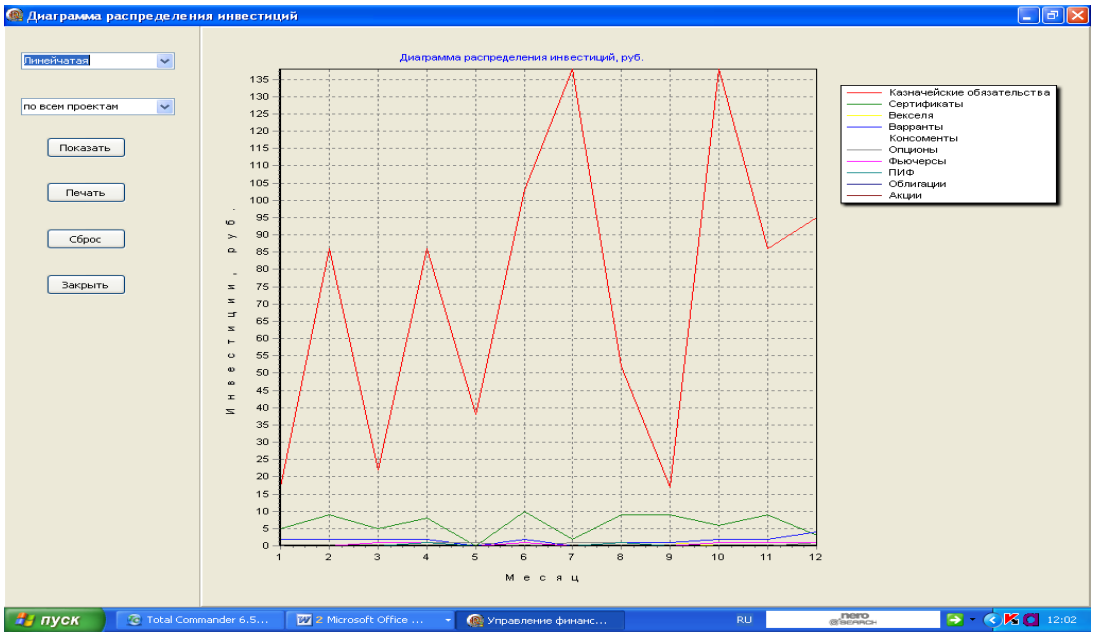


Рисунок 2.18. – Обобщенная диаграмма распределения инвестиций

ВЫВОДЫ

Анализ информационных процессов существующих моделей, методик и алгоритмов формирования ИП СЭО показал, что в качестве методологической основы для оценки включаемых в ИП активов с детерминированными данными целесообразно использовать классические модели ПИА Г. Марковица и У. Шарпа в силу доскональной изученности механизмов функционирования первой и значительного опыта практического использования второй.

В интересах повышения эффективности практического использования модели Г. Марковица теоретически доказана возможность и разработан алгоритм формирования единственного оптимального ИП в определенных условиях, основанный на геометрической интерпретации решаемой задачи. Сформулирована и решена задача определения долей ЦБ в оптимальном ИП, сформированном на основе модели Г. Марковица. Механизм ее решения, основанный на множителях Лагранжа, представлен в виде соответствующего алгоритма, доведенного до программной реализации.

Разработан алгоритм и соответствующее программное обеспечение, обеспечивающие определение долей ЦБ в ИП, сформированном на основе модели У. Шарпа.

Для оценки включаемых в ИП активов со стохастическими данными в качестве методологической основы для дальнейшего совершенствования, целесообразно использовать модель У. Шарпа, в которой процедура определения взаимной ковариации ЦБ, предложенная в модели Г. Марковица, заменена на более простую - сравнения доходности ЦБ с фондовым индексом (например, RTSI), что обеспечивает оперативное проведение расчетов.

На основе модели У. Шарпа разработана ее нейромодифицированная версия. Суть модификации заключается в придании классической модели У. Шарпа свойства оптимальности в упреждающие моменты времени на основе учета рыночных тенденций в виде значений «риск-эффектов». Это важно при прогнозировании доходностей ЦБ. Реализация данного свойства и модели в целом потребовала разработки алгоритма определения, хранения и примене-

ния «риск-эффектов» и его программной реализации. В качестве инструмента определения, хранения и применения «риск-эффектов» использована разновидность ИНС - многослойный персептрон.

Дальнейшим совершенствованием НМШ в плане повышения ее точности и достоверности явилась нейрокомитетная модель У. Шарпа. В отличие от НМШ, в НКМШ реализовано более точное, но и более сложное определение значений «риск-эффектов» множеством ИНС, называемым комитетом. Применение КИНС потребовало организации предварительной обработки временных рядов активов, основанной на их сегментации и кластеризации, с целью выделения типовых режимов работы РЦБ для обучения, верификации и непосредственного использования ИНС в условиях максимально приближенных к рыночным. Это позволило реализовать процесс параллельного обучения и верификации ИНС, а также выбор ИНС-обработчика, используемого для прогнозирования направления тренда «риск – эффекта», адекватного текущим условиям проведения оценки активов, что повышает оперативность и точность получаемых прогнозных значений неизвестных дискрет в стохастических данных ЦБ.

Модифицированный генетический алгоритм традиционно используется в качестве инструмента оценки реальных инвестиций. Его способность парирования основных трудностей применения большинства методов оптимизации нелинейных унимодальных функций, связанных с проблемами «проклятия размерности» и «застревания» в локальных экстремумах, определила целесообразность его включения в состав ИСППИД. Его модификация, состоявшая из двух этапов, проводилась в интересах повышения быстродействия. Суть первого этапа заключалась в развитии, лежащего в основе поиска решений, метода ветвей и границ, в части построения не всего дерева решений, а перспективных поддеревьев и реализации механизма поиска решений (обхода вершин дерева решений) в соответствии со стратегией – «снизу-вверх». На втором этапе использовалась островная модель параллельных вычислений.

3. МЕТОДЫ, МОДЕЛИ И АЛГОРИТМЫ ПРОГНОЗИРОВАНИЯ ЗНАЧЕНИЙ АКТИВОВ В ИНФОРМАЦИОННОМ ПРОЦЕССЕ ФОРМИРОВАНИЯ ИНВЕСТИЦИОННОГО ПОРТФЕЛЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЙ ОРГАНИЗАЦИИ

Методы, модели и алгоритмы прогнозирования значений активов в информационном процессе формирования ИП СЭО входят в состав методического обеспечения подсистемы прогнозирования значений параметров (ППЗП) ЦБ разработанной ИСППИД СЭО.

ППЗП решает две основные задачи: широкую (исследование многослойных персептронов и их комитетов: поиск их оптимальных структур и рациональных обучающих методов и алгоритмов) и узкую (формирование и обучение ИНС в интересах выявления закономерностей в информационных потоках данных и прогнозирования значений «риск – эффектов» для НМШ и НКМШ).

Обобщенная структурная схема ППЗП приведена на рисунке 3.1.

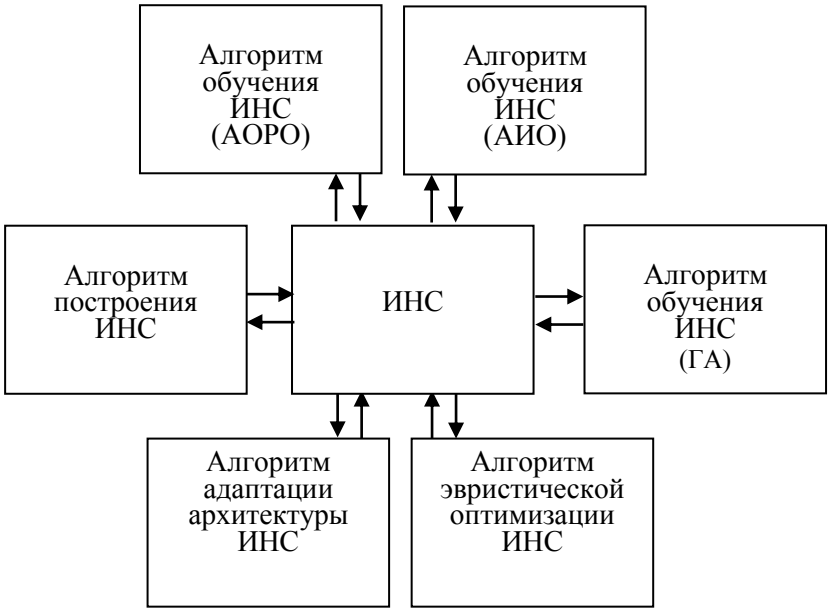


Рисунок 3.1. – Обобщенная структурная схема ППЗП

В состав ППЗП входят шесть алгоритмов: обратного распространения ошибки (АОРО); имитации отжига (АИО); построения ИНС; генетический алгоритм (ГА); адаптации архитектуры ИНС; эвристической оптимизации ИНС. Кроме того в состав ППЗП входит модуль ИНС, который предназначен для непосредственного формирования состава и структуры сети. Известно, что

существует несколько разновидностей ИНС, различающихся функциональным назначением, структурой и принципом действия. В данной работе исследуются многослойные персептроны, которые находят достаточно широкое распространение на практике.

Вопросы теоретического исследования и практической реализации ППЗП потребовали разработки программного комплекса нейросетевого прогнозирования временных рядов, характеристика которого приведена ниже.

3.1 Характеристика программного комплекса нейросетевого прогнозирования временных рядов

Основное функциональное назначение программного комплекса нейросетевого прогнозирования временных рядов (ПК НПВР) – обеспечить формирование и обучение ИНС в интересах прогнозирования значений «риск-эффектов» для НМШ и НКМШ. Достижение данной достаточно узкой цели оказалось невозможным без решения более широкой задачи – исследования ИНС. Процесс исследования ИНС заключается в поиске их оптимальных структур и рациональных обучающих методов и алгоритмов. Разработанный комплекс обеспечивает реализацию данного процесса.

Структурно-функциональная схема ПК НПВР приведена на рисунке 3.2.

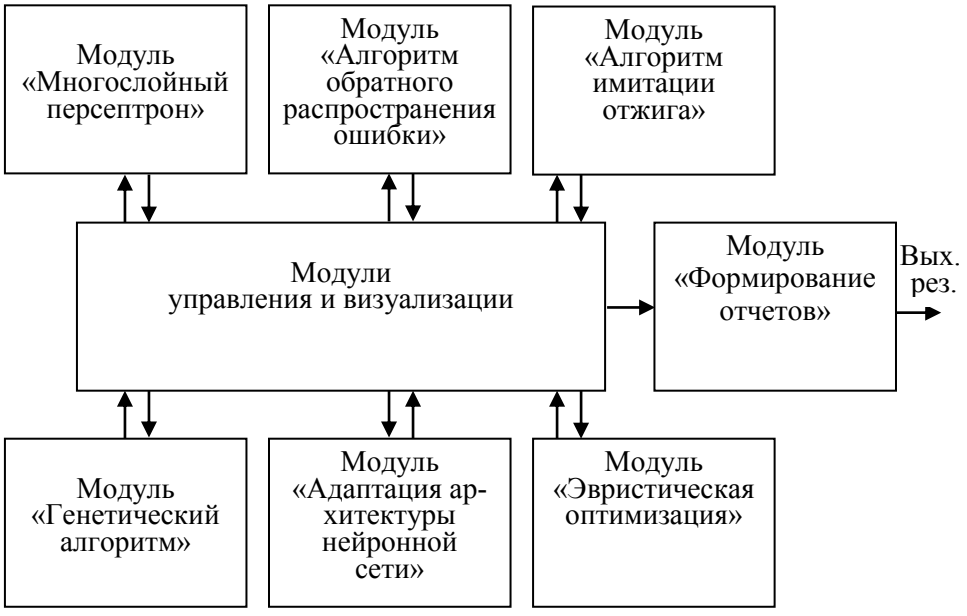


Рисунок 3.2. – Структурно-функциональная схема ПК НПВР

ПК НПВР включает следующие программные модули: «Многослойный персептрон»; «Алгоритм обратного распространения ошибки»; «Алгоритм имитации отжига»; «Эвристическая оптимизация»; «Адаптация архитектуры нейронной сети»; «Генетический алгоритм»; «Управление и визуализация»; «Формирование отчетов». Описание теоретических аспектов, заложенных в них решений, детализация алгоритмов и особенности программной реализации приведены ниже.

3.1.1 Теоретические аспекты, алгоритм и программная реализация многослойного персептрона

Выбор многослойного персептрона (МП) в качестве основного элемента системы прогнозирования значений параметров обусловлен хорошей изученностью его свойств и особенностей [111, 185, 196, 209, 259, 260], а также положительными практиками применения в области прогнозирования временных рядов [91, 102, 156, 160]. К основным его недостаткам следует отнести сложность теоретического анализа по причине распределенной формы нелинейности и высокой связности сети и затруднения процесса визуализации обучения сети из-за наличия скрытых нейронов. В процессе обучения определяется набор признаков входящего сигнала, которые обычно представляют скрытыми нейронами. Это приводит к усложнению процесса обучения по причине необходимости выполнения поиска в широкой области возможных функций [91], так как, выбор должен производиться среди альтернативных представлений входных образов [258].

На рисунке 3.3 показан структурный граф МП с двумя скрытыми слоями и одним выходным слоем.

Показанная на рисунке сеть является полносвязной. Это значит, что каждый нейрон в любом слое сети связан со всеми нейронами (узлами) предыдущего слоя. Сигнал передается по сети в прямом направлении, слева направо, от слоя к слою [173].

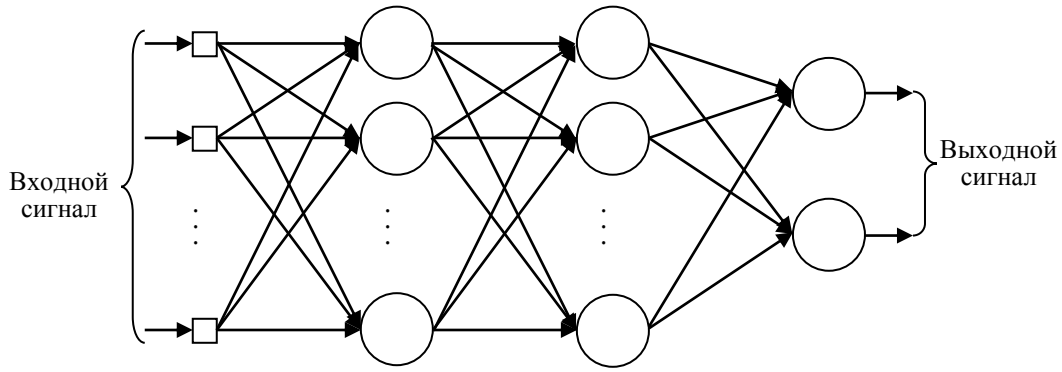


Рисунок 3.3. – Структурный граф многослойного персептрона

Функциональный сигнал это входной сигнал, поступающий в сеть и передаваемый вперед от нейрона к нейрону по всей сети. Такой сигнал достигает конца сети в виде выходного сигнала. Данный сигнал является функциональным по двум причинам. Во-первых, он предназначен для выполнения некоторой функции на выходе сети. Во-вторых, в каждом нейроне, через который передается этот сигнал, вычисляется некоторая функция с учетом весовых коэффициентов [173].

Выходные нейроны составляют выходной слой сети. Остальные нейроны относятся к скрытым слоям. Первый скрытый слой получает данные из входного слоя, составленного из сенсорных элементов (входных узлов). Результирующий сигнал первого скрытого слоя, в свою очередь, поступает на следующий скрытый слой, и т.д., до самого конца сети.

Любой скрытый или выходной нейрон многослойного персептрона может выполнять два типа вычислений [173].

1. Вычисление функционального сигнала на выходе нейрона, реализуемое в виде непрерывной нелинейной функции от входного сигнала и синаптических весов, связанных с данным нейроном [173].

2. Вычисление оценки вектора градиента (т.е. градиента поверхности ошибки по синаптическим весам, связанным с входами данного нейрона), необходимого для обратного прохода через сеть [173].

Один из важных теоретических вопросов применительно к МП заключается в определении минимального числа скрытых слоев, обеспечивающих аппроксимацию некоторого непрерывного отображения. Ответ содержится в

формулировке и доказательстве теоремы об универсальной аппроксимации нелинейного отображения. Она в определенном плане расширяет теорему Вейерштрасса [258]. Последняя утверждает, что на основе абсолютно и равномерно сходящегося ряда полиномов можно представить любую непрерывную функцию на замкнутом интервале действительной оси. Впервые интерес к данной теме возник в работе Розенблата [259]. Для доказательства теоремы об универсальной аппроксимации использовалась усовершенствованная теорема Колмогорова о суперпозиции [258]. В дальнейшем было показано, что многослойный персептрон с одним скрытым слоем, косинусоидальной пороговой функцией и линейным выходным слоем представляет собой частый случай "сети Фурье", обеспечивающей на выходе аппроксимацию заданной функции рядом Фурье [257]. Данная теорема формулировалась следующим образом.

Пусть $\varphi(\bullet)$ ограниченная, не постоянная монотонно возрастающая непрерывная функция. Пусть I_{mo} - m_o-мерный единичный гиперкуб $[0, 1]^{m0}$. Пусть пространство непрерывных на I_{mo} функций обозначается символом $C(I_{mo})$. Тогда для любой функции $fC(I_{mo})$ и $\varepsilon > 0$ существует такое целое число m_1 и множество действительных констант α_i , b_i и w_{ij}, где $i = 1, \ldots, m_1$, $j = 1, \ldots, m_o$, что

$$F(x_1,...,x_{m_0}) = \sum_{i=1}^{m_1} \alpha_i \varphi(\sum_{j=1}^{m_0} w_{ij} x_j + b_i) , \qquad (3.1)$$

является реализацией аппроксимации функции $\varphi(\bullet)$, т.е.

$$\left| F(x_1,...,x_{m_0}) - f(x_1,...,x_{m_0}) \right| < \varepsilon , \qquad (3.2)$$

для всех $x_1,...,x_{m_0}$ принадлежащих входному пространству [257].

Теоретическое значение данной теоремы велико. Данная теорема доказывает потенциальную возможность применения МП в интересах решения задач аппроксимации. Однако в интересах их практического использования необходимо обоснование ряда положений, таких как [257]:

1. Определение пределов ошибок аппроксимации.

2. Обоснование границ риска, возникающего при использовании мно-

гослойного персептрона.

3. Определение размера обучающего множества, необходимого для хорошего обобщения.

4. Влияние «проклятия размерности» на использование многослойного персептрона для решения задач аппроксимации.

5. Обоснование необходимости использования нескольких слоев для решения задач аппроксимации.

Результаты исследований аппроксимирующих свойств однослойного персептрона с единственным выходным нейроном и функцией активации в виде сигмоиды приведены в [257]. Обучение данной сети осуществлялось на основе алгоритма обратного распространения ошибок (АОРО). Затем сеть тестировалась на новых данных. Во время обучения сети предъявлялись выбранные точки аппроксимируемой функции *f*. В итоге была сформирована аппроксимирующая функция *F*, описываемая выражением (3.1). В случае предъявления сети ранее не использованных данных, функция *F* "оценивала" новые точки целевой функции (ЦФ), т.е. *F = f*. При этом гладкость ЦФ *f* выражалась в нотации преобразования Фурье. Значение предельной амплитуды функции *f* представлялось средним значением нормы вектора частоты, взвешенного амплитудными значениями распределения Фурье. Допустим, имеется функция $\hat{f}(w)$ - многомерное преобразование Фурье функции *f(x)*, $x \in \Re^{m_0}$, где *w* — вектор частоты. Тогда представление функции *f(x)*, на основе преобразования Фурье $\hat{f}(w)$, будет определяться в соответствии с выражением [257]:

$$f(x) = \int_{\Re^{m_0}} \tilde{f}(w) \exp(jw^T x) dw, \qquad (3.3)$$

где $j = \sqrt{-1}$.

Первый абсолютный момент распределения функции *f* с учетом преобразования Фурье $\hat{f}(w)$ с интегрируемой функцией $w\,\hat{f}(w)$ может быть определен следующим образом:

$$C_f = \int_{\Re^{m_0}} |\tilde{f}(w)| \times \| w \|^{0,5} dw, \qquad (3.4)$$

где ||*w*|| — Евклидова норма вектора *w*; |*f(w)*| — абсолютное значение функции *f(w)*.

Из выражения (3.4) следует, что первый абсолютный момент C_f представляет собой меру гладкости функции *f*. Тем самым его можно считать основой для расчета пределов ошибки, возникающей вследствие использования многослойного персептрона. В общем случае ошибка аппроксимации оценивается мерой вероятности μ для шара $B_r = \{x : \|x\| \le r\}$ радиуса $r > 0$. Поэтому утверждение для предела ошибки аппроксимации, можно сформулировать следующим образом [260].

Для заданной непрерывной функции *f(x)*, имеющей конечный первый момент C_f , и любого $m_1 > 1$ существует набор сигмоидальных функций *F(x)* вида:

$$\int_{B_r} (f(x) - F(x))^2 \mu(dx) \le \frac{C_f^{'}}{m_1}, \tag{3.5}$$

где $C_f^{'} = (2rC_f)^2$.

Эмпирический риск, возникающий в случае наблюдения функции *f(x)* на множестве значений $\{x_i\}_{i=1}^N$ входного вектора *x*, принадлежащего шару B_r, определяется в соответствии с выражением:

$$R = \frac{1}{N} \sum_{i=1}^{N} (f(x_i) - F(x_i))^2 \le \frac{C_f^{'}}{m_1}. \tag{3.6}$$

С учетом выражения (3.7), могут быть определены границы риска *R*, который возникает при применении многослойного персептрона с m_o входными узлами и m_1 скрытыми нейронами [257]:

$$R \le O\left(\frac{C_f^2}{m_1}\right) + O\left(\frac{m_0 m_1}{N} \log N\right). \tag{3.7}$$

Слагаемые в выражении (3.7) отражают противоречивые требования к размеру скрытого слоя и компромисс между ними. Это имеет важное практическое значение. В выражении (3.7) заключен второй важный результат, ка-

сающийся требуемого объема обучающей выборки. В нем показано, что точная оценка ЦФ не требует экспоненциально больших: обучающего набора и размерности входного пространства m_0, даже если при этом первый абсолютный центральный момент C_f остается конечным. Это подтверждает целесообразность использования МП в качестве универсального аппроксиматора. Значение разности между эмпирическим и оптимальным значениями аппроксимации можно считать ошибкой оценивания. Если ε_0 - СКО значения ошибки оценивания, то размер N обучающего набора, необходимый для качественного обобщения, должен иметь порядок $m_0 m_1 / \varepsilon_0$. Отсюда следует, что для качественной аппроксимации размер обучающей выборки должен быть больше отношения общего числа свободных параметров ИНС к СКО значения ошибки оценивания.

Суть третьего важного результата, содержащегося в (3.7) заключается в следующем. Если размер скрытого слоя выбирать в соответствии с выражением:

$$m_1 = C_f \left(\frac{N}{m_0 \log N}\right)^{1/2}, \tag{3.8}$$

то риск R ограничивается величиной $O(C_f \sqrt{m_0 (\log N / N)})$. Это значит, что скорость сходимости, как функция размера обучающего множества N, имеет порядок $(1/N)^{1/2}$, умноженный на логарифмический член. Для обычной гладкой функции (например, тригонометрической или полиномиальной) скорость сходимости отличается. Если s — мера гладкости, определяемая как число существующих производных, то для обычной гладкой функции минимаксная скорость сходимости общего риска R имеет порядок $(l/N)^{2s/(2s+m_0)}$. Зависимость скорости сходимости от размерности входного пространства m_0, называют "проклятием размерности" (впервые этот термин был введен Ричардом Белманом в 1961 году в работе, посвященной процессам адаптивного управления [237]). Тем самым применение общих гладких функций в данном случае ограничено. Следовательно, применение МП для решения задач аппроксимации имеет безусловное преимущество. Условием проявления

преимущества является конечность первого абсолютного момента C_f . В этом суть ограничения гладкости.

Для пояснения особенностей "проклятия размерности" рассмотрим пример, в котором *x* - m_0-мерный входной вектор, а множество *{(xi,di)},i = 1, 2,..., N,* задает обучающую выборку. Плотность дискретизации пропорциональна значению $N^1/m0$. Пусть *f(x)* — поверхность в m_0-мерном входном пространстве, проходящая около точек данных $\{x_i, d_i\}_{i=1}^{N}$. Если функция *f(x)* сложна и в большей части неопределенна, то для детализации исследования поверхности необходимо уменьшить шаг дискретизации точек зондирования. На практике, наличие "проклятия размерности" приводит к трудностям поиска обучающей выборки с повышенным числом точек зондирования. Поскольку увеличение плотности точек зондирования, приводит к экспоненциальному росту сложности и увеличению ресурсных затрат на их обработку. Кроме того, проблематичным становится поиск областей с равномерным распределением случайных точек.

Еще одна особенность заключается в следующем. Функция, определенная в пространстве большой размерности, более сложная, чем функция, определенная в пространстве меньшей размерности. Эту сложность трудно выявить. Для парирования "проклятия размерности" необходимы корректные априорные знания о функции, учитываемые при обучении. Поэтому, для корректной оценки сложных функций в пространствах высокой размерности необходимо повышать гладкость неизвестной функции путем увеличения размерности входных данных [260].

Для обоснования необходимости использования нескольких нейронных слоев для решения задач аппроксимации, необходимо отметить следующее. Несмотря на теоретическое обоснование возможности использования для аппроксимации однослойной ИНС, на практике это реализовать затруднительно. Причина заключается в том, что у ИНС с одним скрытым слоем нейроны могут взаимодействовать друг с другом на глобальном уровне. Это приводит к усложнению задачи повышения качества аппроксимации в одной точке без

ухудшения в другой. Поэтому необходимо использовать несколько скрытых слоев. Механизм функционирования нескольких скрытых слоев заключается в следующем. В первом скрытом слое извлекаются локальные признаки, а во втором – глобальные. Нейрон второго скрытого слоя "обобщает" выходные сигналы нейронов первого скрытого слоя, которые относятся к соответствующей области входного пространства. Он обучается глобальным признакам этой области. Для других областей он не активен (выходной сигнал равен нулю). Описанный механизм подобен сплайновому подходу к аппроксимации кривых [260].

Математическое обоснование целесообразности использования нескольких нейронных слоев в контексте обратных задач представлено в [260].

Для известной непрерывной вектор-функции $f : \Re^m \to \Re^M$, компактного подмножества $C \subseteq \Re^M$, содержащемся в пространстве образов функции *f*, и некоторого положительного числа $\varepsilon > 0$ требуется найти такую вектор-функцию $\varphi : \Re^M \to \Re^m$, которая удовлетворяет условию:

$$\|\varphi(f)(u) - u\| < \varepsilon \text{ для любого } u \in C.$$

Данная задача принадлежит области обратной динамики. В данной предметной области наблюдаемое состояние *x(n)* системы определяется ее текущим состоянием *u(n)* и предыдущим состоянием *x(n — 1)*:

$$x(n) = f(x(n\text{-}l), u(n)).$$

Предполагается, что функция *f* является обратимой. Это значит, что *u(n)* можно представить как функцию от *x(n)* для любого *x(n - 1)*. Функция *f* описывает прямую динамику, а функция φ - обратную. Требуется построить такую функцию φ, которая может быть реализована многослойным персептроном. В общем случае она должна быть разрывной. В [260] математически обосновано, что для построения функции φ одного скрытого слоя недостаточно, даже при использовании нейронной модели с разрывными активационными функциями. В тоже время, персептрон с двумя скрытыми слоями решает эту задачу для любых возможных *C, f* и ε.

Блок- схема алгоритма формирования МП представлена на рисунке 3.4.

Рисунок 3.4. – Блок-схема алгоритма формирования МП

Блоки 1, 12 обеспечивают пуск и остановку алгоритма формирования МП.

В блоке 2 реализован ввод исходных данных, таких как:

- начальные значения весовых коэффициентов межнейронных связей;

- вид функции активации. При проведении исследований использовались такие функции активации, как: линейная; пороговая; сигмоидальная; логистическая; гиперболический тангенс; радиально-базисная.

В блоке 3 задается число входов во входном слое.

Блок 4 предназначен для задания значений обучающей выборки.

Блок 5 используется для нормализации обучающей выборки.

В блоке 6 задается число скрытых слоев персептрона.

В блоке 7 задается число нейронов в скрытом слое.

Блок 8 обеспечивает проверку условия удовлетворенности пользователем заданного числа слоев.

Блок 9 обеспечивает проверку условия удовлетворенности пользователем заданного числа нейронов.

В блоке 10 в интересах исследования нового класса ИНС задается число разрывов связей нейронов в скрытых слоях. Детализация данного вида исследований приведена в описании подсистемы управления.

Блок 11 реализует вывод результатов. В качестве результата выступает сформированный МП.

Экранная форма исследования персептронов приведена на рисунке 3.5.

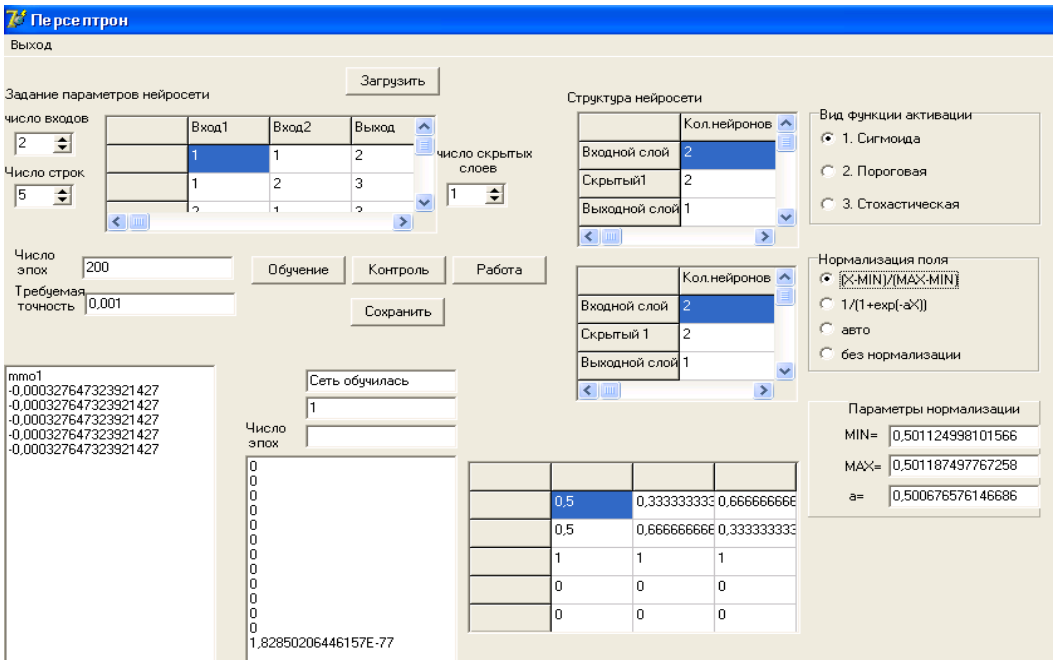


Рисунок 3.5. – Вид экранной формы исследования МП

3.1.2. Теоретические аспекты и программная реализация алгоритма обратного распространения ошибок

Алгоритм обратного распространения ошибки (АОРО), реализованный в соответствующем программном модуле, предназначен для обучения МП. Впервые данный метод был предложен Руммельхартом и Хинтоном в 1986 г. На рисунке 3.6 показаны потоки сигналов и ошибок, характерных для МП.

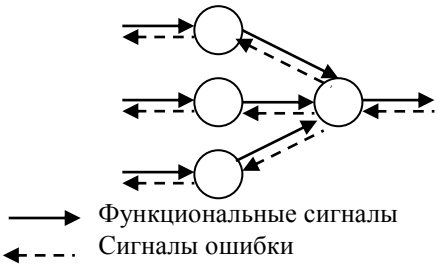


Рисунок 3.6. – Потоки сигналов и ошибок в МП

Сигнал ошибки распространяется с выхода в начало (от слоя к слою). Он получил свое название благодаря тому, что вычисляется каждым нейроном сети на основе функции ошибки, представленной в той или иной форме. В ходе проведения экспериментов было установлено, что причиной ошибок, которые возникают на выходе сети, являются ошибки нейронов скрытых слоев. Эти ошибки малоизученны. Выявлено, что чем сильнее синаптическая связь между нейронами скрытого слоя и выходными нейронами, тем значительнее ошибки первых влияет на ошибки вторых. Оценка ошибки нейронов скрытых слоев может быть получена как взвешенная сумма ошибок последующих слоев [173].

С целью упрощения пояснения механизма обучения ИНС на основе АОРО, рассмотрим персептрон с единственным скрытым слоем. Обозначим матрицу весовых коэффициентов между нейронами входного и скрытого слоев, как - *W*, а матрицу весов между нейронами скрытого и выходного слоев - как *V*. Для идентификации входов, будет использоваться индекс *i*, для элементов скрытого слоя - индекс *j*, а для элементов выходного слоя - индекс *k*.

Обучающая выборка имеет символьное представление в виде: $(X\alpha, Y\alpha)$, $\alpha = 1..p$. Малыми символами *y* будем обозначать активности нейронов, а малыми символами *x* - суммарные взвешенные входы нейронов.

Пошаговое описание АОРО представлено в таблице 3.1 [196].

Таблица 3.1 - Описание алгоритма обратного распространения ошибок

Шаг 0.	Начальные значения весов всех нейронов всех слоев *V(t=0)* и *W(t=0)* полагаются случайными числами.
Шаг 1.	Предъявление сети входного образа -$X\alpha$. В результате формируется выходной образ $y \neq Y\alpha$. При этом нейроны последовательно от слоя к слою функционируют по следующим формулам [196]: скрытый слой $$x_j = \sum_i W_{ij} X_i^{\alpha}; \quad y_j = f(x_j)$$ выходной слой $$x_k = \sum_j V_{jk} Y_j; \quad y_k = f(x_k)$$ Здесь *f(x)* - сигмоидальная функция, определяемая по формуле $$y = 1 \Big/ (1 + \exp(-\left[\sum_i W_i x_i - \Theta\right]))$$
Шаг 2.	Минимизация функционала квадратичной ошибки сети для данного входного образа, имеющего вид [196]: $$E = 1/2 \sum_k (y_k - Y_k^{\alpha})^2$$ В основе минимизации классический градиентный метод оптимизации. Его суть состоит в итерационном уточнении аргумента согласно формуле [196]: $$\mathrm{V_{jk}(t+1) = V_{jk}(t) - h \cdot \frac{\partial E}{\partial V_{jk}}}$$ Функция ошибки в явном виде не содержит зависимости от веса V_{jk}, поэтому воспользуемся формулами неявного дифференцирования сложной функции [196]: $\frac{\partial \mathrm{E}}{\partial \mathrm{V_{jk}}} = \delta_{\mathrm{k}} = (y_k - Y_k^{\alpha})^2$ $$\frac{\partial \mathrm{E}}{\partial y_k} = \delta_{\mathrm{k}} = (y_k - Y_k^{\alpha});$$ $$\frac{\partial \mathrm{E}}{\partial x_k} = \frac{\partial \mathrm{E}}{\partial y_k}\frac{\partial y_k}{\partial x_k} = \delta_{\mathrm{k}} \cdot y_k(1 - y_k);$$ $$\frac{\partial \mathrm{E}}{\partial \mathrm{V_{jk}}} = \frac{\partial \mathrm{E}}{\partial y_k} \cdot \frac{\partial y_k}{\partial x_k} \cdot \frac{\partial x_k}{\partial \mathrm{V_{jk}}} = \delta_{\mathrm{k}} \cdot y_k(1 - y_k) \cdot y_{\mathrm{j}}.$$ Здесь учтено полезное свойство сигмоидальной функции *f(x)*: ее производная выражается только через само значение функции, *f'(x)=f(1-f)*. Таким образом, все необходимые величины для подстройки весов выходного слоя *V* получены [196].
Шаг 3.	Подстройка весов скрытого слоя. Градиентный метод по-прежнему дает [196]: $W_{ij}(\mathrm{t}+1) = W_{ij}(\mathrm{t}) - \eta \cdot \frac{\partial \mathrm{E}}{\partial W_{ij}}$. Вычисления производных выполняются по тем же формулам, за исключением некоторого усложнения формулы для ошибки α_j.

	$$\frac{\partial \mathrm{E}}{\partial x_k} = \frac{\partial \mathrm{E}}{\partial y_k}\frac{\partial y_k}{\partial x_k} = \delta_{\mathrm{k}} \cdot y_k(1-y_k);$$ $$\frac{\partial \mathrm{E}}{\partial \mathrm{V_j}} = \delta_{\mathrm{j}} = \sum_k \frac{\partial \mathrm{E}}{\partial x_k} \cdot \frac{\partial x_k}{\partial y_{\mathrm{j}}} = \sum_k \delta_{\mathrm{k}} \cdot y_k(1-y_k) \cdot \mathrm{V_{jk}}$$ $$\frac{\partial \mathrm{E}}{\partial W_{ij}} = \frac{\partial \mathrm{E}}{\partial y_{\mathrm{j}}} \cdot \frac{\partial y_{\mathrm{j}}}{\partial x_{\mathrm{j}}} \cdot \frac{\partial x_{\mathrm{j}}}{\partial W_{ij}} = \delta_{\mathrm{j}} \cdot y_{\mathrm{j}}(1-y_{\mathrm{j}}) \cdot X^{\alpha}_{\mathrm{j}} =$$ $$= \left[\sum_k \delta_k \cdot y_k(1-y_{\mathrm{k}}) \cdot \mathrm{V_{jk}}\right] \cdot \left[y_{\mathrm{j}}(1-y_{\mathrm{j}}) \cdot X^{\alpha}_{\mathrm{j}}\right]$$ При вычислении δ_j частные производные берутся только по переменным последующего слоя. По полученным формулам модифицируются веса нейронов скрытого слоя. Если в нейронной сети имеется несколько скрытых слоев, процедура обратного распространения применяется последовательно для каждого из них, начиная со слоя, предшествующего выходному, и далее до слоя, следующего за входным. При этом формулы сохраняют свой вид с заменой элементов выходного слоя на элементы соответствующего скрытого слоя [196].
Шаг 4.	Повторение шагов 1-3 для всех обучающих векторов. Обучение завершается по достижении малой полной ошибки или максимально допустимого числа итераций, как и в методе обучения Розенблатта [196].

Из описания шагов 2-3 видно, что обучение представляет собой решение задачи оптимизации функционала ошибки градиентным методом [173].

Параметр η характеризует темп обучения МП. При его выборе необходим компромисс. С одной стороны его увеличение приводит к более быстрому обучению МП, а с другой к возрастанию ее неустойчивости (сходимости). Теоретически он выбирается достаточно малым для сходимости метода. Результаты экспериментов показали, что сходимость данного метода не высокая. Это обусловлено не совпадением локального направления градиента с направлением к минимуму и независимостью подстройки весов для каждой пары об образов обучающей выборки. Последнее означает, что улучшение функционирования на некоторой заданной паре может, приводить к ухудшению работы на предыдущих образах [173].

Для повышения темпа обучения МП рассмотрим детализацию шага 3 АОРО. Обратный проход начинается с выходного слоя предъявлением ему сигнала ошибки, передаваемого справа налево от слоя к слою с параллельным вычислением локального градиента для каждого нейрона. В данном

процессе изменяются синаптические веса по правилу (3.9):

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n) \tag{3.9}$$

где $\delta_j(n)$ - локальный градиент; $y_i(n)$ - входной сигнал нейрона.

Для нейрона, расположенного в выходном слое, локальный градиент равен соответствующему сигналу ошибки, умноженному на первую производную нелинейной функции активации. Затем соотношение (3.9) используется для вычисления изменений весов, связанных с выходным слоем нейронов. Зная локальные градиенты для всех нейронов выходного слоя, с помощью (3.10) можно вычислить локальные градиенты всех нейронов предыдущего слоя, а значит, и величину коррекции весов связей с этим слоем:

$$\delta_j(n) = \varphi_j'(v_j(n)) \sum \delta_k(n) w_{kj}(n) . \tag{3.10}$$

Такие вычисления проводятся для всех слоев в обратном направлении.

После предъявления очередного обучающего примера он "фиксируется" на все время прямого и обратного проходов [258 - 260].

Введем в (3.9) момент инерции (3.11):

$$\Delta w_{ji}(n) = \alpha \Delta w_{ji}(n-1) + \eta \delta_j(n) y_i(n) \tag{3.11}$$

где α - постоянная момента (положительное число). Как показано на рисунке 3.7, она является управляющим воздействием в контуре обратной связи для $\Delta w_{ji}(n)$, где z^{-1} - оператор единичной задержки.

Уравнение (3.11) называется обобщенным дельта-правилом. При $\alpha = 0$ оно вырождается в обычное дельта-правило [173].

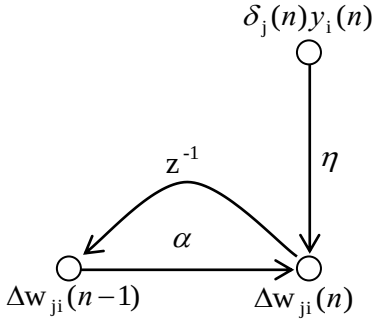


Рисунок 3.7. – Граф прохождения сигнала в МП

Для того чтобы оценить влияние последовательности представления обу-

чающих примеров на синаптические веса в зависимости от константы α, перепишем (3.11) в виде временного ряда с индексом *t*. Значение индекса *t* будет изменяться от нуля до текущего значения *n*. При этом выражение (3.11) можно рассматривать как разностное уравнение первого порядка для коррекции весов $\Delta w_{ji}(n)$. Решая это уравнение относительно $\Delta w_{ji}(n)$, получим временной ряд длины $n + 1$:

$$\Delta w_{ji}(n) = \eta \sum \alpha^{n-t} \delta_j(t) y_i(t). \qquad (3.12)$$

Поскольку $\delta_j(n) y_i(n) = -\partial E(n)/\partial w_{ji}(n)$, то выражение (3.12) можно переписать в эквивалентной форме (3.13):

$$\Delta w_{ji}(n) = -\eta \sum_{t=0}^{n} \alpha^{n-t} \frac{\partial E(t)}{\partial w_{ji}(t)}. \qquad (3.13)$$

Анализ выражения (3.13) позволяет сделать ряд выводов [259].

1. Текущее значение коррекции весов $\Delta w_{ji}(n)$ представляет собой сумму экспоненциально взвешенного временного ряда. Для того чтобы этот ряд сходился, постоянная момента должна находиться в диапазоне $0 < \alpha < 1$. Если константа α равна нулю, АОРО работает без момента. Следует также заметить, что константа α может быть и отрицательной, хотя эти значения не рекомендуется использовать на практике.

2. Если частная производная $\frac{\partial E(t)}{\partial w_{ji}(t)}$ не изменяет знак на нескольких последовательных итерациях, то экспоненциально взвешенная сумма $\Delta w_{ji}(n)$ возрастает по абсолютному значению, поэтому веса $w_{ji}(n)$ могут изменяться на очень большую величину. Включение момента в АОРО ведет к ускорению спуска в некотором постоянном направлении.

3. Если частная производная $\frac{\partial E(t)}{\partial w_{ji}(t)}$ на нескольких последовательных итерациях меняет знак, экспоненциально взвешенная сумма $\Delta w_{ji}(n)$ уменьшается по абсолютной величине, поэтому веса $w_{ji}(n)$ изменяются на небольшую

величину. Таким образом, добавление момента в АОРО ведет к стабилизирующему эффекту для направлений, изменяющих знак.

Включение момента в АОРО обеспечивает незначительную модификацию метода корректировки весов, оказывая положительное влияние на его работу. Кроме того, слагаемое момента может предотвратить нежелательную остановку алгоритма в точке какого-либо локального минимума на поверхности ошибок [173].

Результаты экспериментов показали, что параметр η целесообразно задавать для каждой конкретной связи η_{ji}. Это значит, что применяя различные параметры скорости обучения в разных областях сети, можно оптимизировать процедуру обучения. Кроме того, при реализации АОРО можно изменять как все синаптические веса сети, так и только часть из них, оставляя остальные на время адаптации фиксированными. В последнем случае сигнал ошибки распространяется по сети в обычном порядке, однако фиксированные синаптические веса будут оставаться неизменными. Этого можно добиться, установив для соответствующих синаптических весов $w_{ji}(n)$ параметр интенсивности обучения η_{ji} равным нулю.

Для эффективного обучения МП использовался ряд эмпирических приемов, в частности: выбор размера обучающего множества; применение перекрестной проверки; использование раннего останова, применение последовательного и пакетного режимов обучения, обоснование критерия останова обучения.

При выборе размера обучающего множества учитывалось важное эмпирическое правило. Размер обучающего множества *N* должен удовлетворять следующему соотношению [257]:

$$N = O(W/\varepsilon), \qquad (3.14)$$

где *W* - общее количество свободных параметров (т.е. синаптических весов и порогов) сети; ε - допустимая точность ошибки классификации; O(-) - порядок заключенной в скобки величины. Например, для ошибки в 10% коли-

чество примеров обучения должно в 10 раз превосходить количество свободных параметров сети. Выражение (3.14) получено из эмпирического правила Видроу для алгоритма LMS. Данное правило устанавливает, что время стабилизации процесса линейной адаптивной временной фильтрации примерно равно объему памяти линейного адаптивного фильтра в задаче фильтра на линии задержки с отводами, деленному на величину рассогласования [257]. Рассогласование в алгоритме LMS выступает в роли ошибки ε из выражения (3.14).

Суть использования перекрестной проверки состояла в следующем. Обучение методом обратного распространения заключается в кодировании отображения входа на выход (представленного множеством маркированных примеров) в синаптических весах и пороговых значениях МП. Предполагается, что на примерах из прошлого сеть будет обучена настолько хорошо, что сможет обобщить их на будущее. С такой точки зрения процесс обучения обеспечивает настройку параметров МП для заданного множества данных. Более того, проблему настройки МП можно рассматривать как задачу выбора наилучшей модели из множества структур-"кандидатов" с учетом определенного критерия. Решить такую задачу возможно на основе стандартного статистического подхода, получившего название перекрестной проверки [257]. В рамках этого подхода имеющиеся в наличии данные сначала случайным образом разбиваются на обучающее множество и тестовое множество. Обучающее множество, в свою очередь, разбивается на подмножество для оценивания (используется для выбора модели) и проверочное подмножество (используется для тестирования модели). Целесообразность применения проверочного множества отличного от оценочного, обусловлена необходимостью исключения переобученности МП.

Применение метода обучения с ранним остановом тесно связано с перекрестной проверкой. Обычно обучение МП методом обратного распространения происходит поэтапно - переходя от более простых к более сложным функциям отображения. Это объясняется тем фактом, что при нормальных условиях среднеквадратическая ошибка уменьшается по мере увеличения количества

эпох обучения: она начинается с довольно больших значений, стремительно уменьшается, а затем продолжает убывать все медленнее по мере продвижения сети к локальному минимуму на поверхности ошибок. Если главной целью является хорошая способность к обобщению, то по виду кривой довольно сложно определить момент, когда следует остановить процесс обучения. Если вовремя не остановить сеанс обучения, то существенно повышается вероятность излишнего переобучения. Наступление стадии излишнего переобучения определялось с помощью перекрестной проверки, в которой данные разбиты на два подмножества — оценивания и проверки. Множество оценивания использовалось для обычного обучения сети с небольшой модификацией: сеанс обучения периодически останавливался (через каждые несколько эпох), после чего сеть тестировалась на проверочном подмножестве. Более точно, периодический процесс оценивания-тестирования выполнялся следующим образом. После завершения этапа оценивания (обучения) синаптические веса и уровни порогов МП фиксировались, и МП переключалась в режим прямого прохода. Ошибка МП вычислялась для каждого примера из проверочного подмножества. После завершения тестирования наступал следующий этап процесса обучения, и все повторялось. Использовалась так называемая процедура обучения с ранним остановом. На рисунке 3.8 показаны зависимости среднеквадратических ошибок от числа эпох для тестирования и обучения [173].

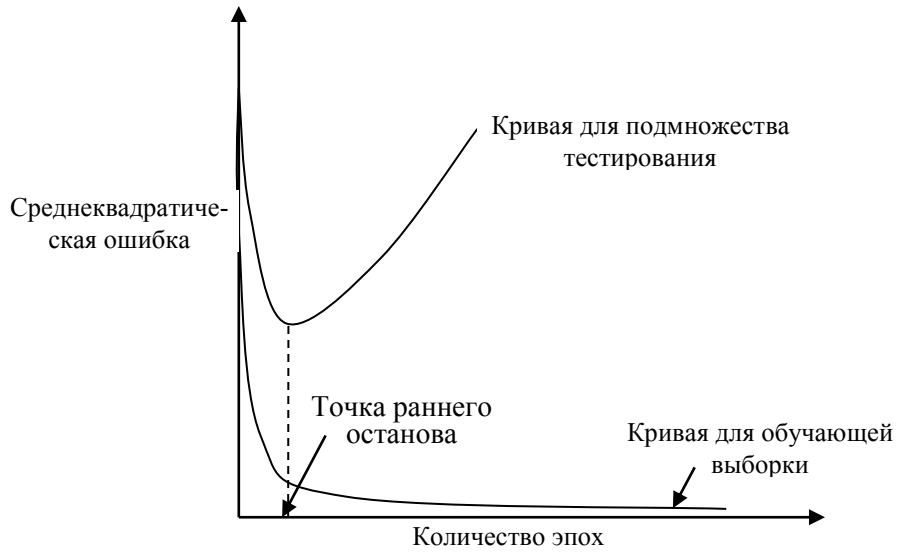


Рисунок 3.8. – Графики обучения МП с ранним остановом

Первая кривая относится к измерениям на подмножестве для оценивания, а вторая - к измерениям на проверочном подмножестве. Анализ данных функций показывает их различное поведение. Функция обучения на подмножестве оценивания монотонно убывает с увеличением количества эпох, в то время как кривая обучения на проверочном подмножестве монотонно убывает до некоторого минимума, после чего при продолжении обучения начинает возрастать. Если учитывать только кривую, построенную на подмножестве оценивания, то возникает соблазн продолжать обучение и после прохождения точки минимума на кривой, построенной на проверочном подмножестве. Однако на практике оказалось, что после прохождения этой точки сеть обучалась только посторонним шумам, содержащимся в обучающей выборке. Поэтому было принято решение об использовании точки минимума кривой тестирования в качестве критерия останова сеанса обучения. Когда ошибки на подмножествах оценивания и проверки не одинаково стремились к нулю (что характерно для данных, не содержащих шума), емкости сети оказывалось не достаточно для точного функционирования модели. В этом случае минимизировалась интегрированная квадратичная ошибка, что (приблизительно) эквивалентно минимизации обычной глобальной среднеквадратической ошибки при равномерном распределении входного сигнала [173].

Для повышения эффективности обучения МП использовалось комбинирование последовательного и пакетного режимов обучения. В рамках последовательного режима обучения корректировка весов проводилась после подачи каждого примера. Для примера рассмотрим эпоху, состоящую из N обучающих примеров, упорядоченных следующим образом: *((x1),d(1)),.., (x(N),d(N)).* Сети предъявлялся первый пример *(x(l),d(l))* этой эпохи, после чего выполнялись описанные выше прямые и обратные вычисления. В результате проводилась корректировка синаптических весов и уровней порогов в сети. После этого сети предъявлялась вторая пара *(x(2),d(2))* в эпохе, повторялись прямой и обратный проходы, приводящие к последующей коррекции

синаптических весов и уровня порога. Этот процесс повторялся, пока сеть не завершала обработку последнего примера (пары) данной эпохи - *(x(N), d(N)).* В пакетном режиме обучения корректировка весов проводилась после подачи в сеть всех примеров обучения (всей эпохи). Для каждой эпохи функция стоимости определялась как среднеквадратическая ошибка, представленная в составной форме (3.15) [173]:

$$E_{av}(n) = \frac{1}{2N}\sum_{n=1}^{N}\sum_{j\in C} e_j^2(n), \qquad (3.15)$$

где сигнал ошибки $e_j(n)$ соответствует нейрону j для примера обучения n. Ошибка $e_j(n)$ равна разности между $d_j(n)$ и $y_j(n)$ для j-го элемента вектора желаемых откликов $d(n)$ и соответствующего выходного нейрона сети. В выражении (3.16) внутреннее суммирование по j выполняется по всем нейронам выходного слоя сети, в то время как внешнее суммирование по n выполняется по всем образам данной эпохи. При заданном параметре скорости обучения η корректировка, применяемая к синаптическому весу w_{ji} связывающему нейроны i и j, определялась следующим дельта-правилом (3.16) [173]:

$$\Delta w_{ji}(n) = -\eta\frac{\partial E_{av}}{\partial w_{ji}} = -\frac{\eta}{N}\sum_{n=1}^{N} e_j(n)\frac{\partial e_j(n)}{\partial w_{ji}}. \qquad (3.16)$$

Согласно (3.16), в пакетном режиме корректировка веса $\Delta w_{ji}(n)$ выполняется только после прохождения по сети всего множества примеров.

Процесс комбинирования данных режимов обучения состоял в следующем. В случае наличия избыточных данных использовался последовательный режим обучения. Его применение позволяло использовать меньший объем внутреннего хранилища для каждой синаптической связи. Более того, путем предъявления обучающих примеров в случайном порядке (в процессе последовательной корректировки весов), достигался стохастический поиск в пространстве весов. Это, в свою очередь, минимизировало возможность остановки алгоритма в точках локальных минимумов. Во всех остальных случаях использовался пакетный режим обучения. Он обеспечивал более точную оценку вектора градиента при «обедненных» данных и позволял рас-

параллелить вычисления.

Для обоснования критерия прекращения (останова) обучения рассматривались уникальные свойства локального и глобального минимумов поверхности ошибок. Обозначим символом *w* вектор весов, обеспечивающий минимум, будь то локальный или глобальный. Необходимым условием минимума является достижение вектора градиента *g(w)* (т.е. вектора частных производных первого порядка) для поверхности ошибок в точке нулевых значений. На практике это означает, что АОРО сошелся, если Евклидова норма вектора градиента достигла достаточно малых значений. Недостатком этого критерия сходимости является то, что для сходимости обучения может потребоваться довольно много времени. Кроме того, необходимо постоянно вычислять вектор градиента *g(w)*. Другим уникальным свойством минимума является то, что функция стоимости (или мера ошибки) E_{av} *(w)* в точке w = w* стабилизируется. Отсюда можно вывести еще один критерий сходимости. Критерием сходимости АОРО, является достаточно малая абсолютная интенсивность изменений среднеквадратической ошибки в течение эпохи. Практические результаты показали, что интенсивность изменения среднеквадратической ошибки можно считать достаточно малой, если она лежит в пределах от 0,1-1% за эпоху. В процессе функционирования АОРО, использовался еще один критерий сходимости. После каждой итерации обучения сеть тестировалась на эффективность обобщения. Процесс обучения останавливался, когда эффективность обобщения становилась удовлетворительной или когда оказывалось, что пик эффективности уже пройден [173].

Блок-схема АОРО представлена на рисунке 3.9 [173].

Блоки 1,10 используются для пуска и остановки АОРО.

В блоке 2 реализован ввод исходных данных, таких как:

- обучающая выборка в нормализованном виде;
- значения весов;
- вектор выходных значений сигналов.

Блок 3 обеспечивает расчет выходного сигнала.

В блоке 4 рассчитывается значение среднеквадратичной ошибки.

Блок 5 реализует определение производных сложной функции (сигнала и функции активации).

Рисунок 3.9. – Блок-схема АОРО

В блоке 6 фиксируется значение производной.

Блок 7 обеспечивает проверку изменения знака производной. Если он не изменился, то управление подается в блок 9. Иначе управление подается в блок 8.

Блок 9 вывод результатов. В качестве результатов выступает набор весовых коэффициентов, при которых достигается значение минимума функции ошибки.

3.1.3 Теоретические аспекты и программная реализация алгоритма имитации отжига

Алгоритм имитации отжига (АИО), реализованный в соответствующем программном модуле, предназначен для оптимизации поиска глобального экстремума ЦФ. АИО представляет собой алгоритмический аналог физического процесса управляемого охлаждения. Предложенный Н. Метрополисом в 1953 г. [63] и доработанный многочисленными последователями, он в настоящее время считается одним из немногих алгоритмов, позволяющих оперативно находить глобальный минимум функции нескольких переменных. АИО основан на идее, заимствованной из статической механики. Он отражает поведение материального тела при отвердевании с применением процедуры отжига [63]. Для получения качественного материала из расплава, при отвердевании, его температура должна уменьшаться постепенно, вплоть до момента полной кристаллизации. Если процесс остывания протекает слишком быстро, образуются значительные нерегулярности структуры материала, которые вызывают внутренние напряжения. В результате общее энергетическое состояние тела, зависящее от его внутренней напряженности, остается на гораздо более высоком уровне, чем при медленном охлаждении [63]. Быстрая фиксация энергетического состояния тела на уровне выше нормального аналогична сходимости оптимизационного алгоритма к точке локального минимума. Энергия состояния тела соответствует ЦФ, а абсолютный минимум этой энергии – глобальному минимуму. В процессе медленного управляемого охлаждения, называемого отжигом, кристаллизация тела сопровождается глобальным уменьшением его энергии, однако допускаются

ситуации, в которых она может на какое–то время возрастать (в частности, при подогреве тела для предотвращения слишком быстрого его остывания). Благодаря допустимости кратковременного повышения энергетического уровня возможен выход из ловушек локальных минимумов, которые возникают при реализации процесса. Только понижение температуры тела до абсолютного нуля делает невозможным какое–либо самостоятельное повышение его энергетического уровня. В этом случае любые внутренние изменения ведут только к уменьшению общей энергии тела [63].

В реальных процессах кристаллизации твердых тел температура понижается ступенчатым образом. На каждом уровне она какое-то время поддерживается постоянной, что необходимо для обеспечения термического равновесия. На протяжении всего периода, когда температура остается выше абсолютного нуля, она может, как понижаться, так и повышаться. За счет удержания температуры процесса поблизости от значения, соответствующего непрерывно снижающемуся уровню термического равновесия, удаётся обходить ловушки локальных минимумов, что при достижении нулевой температуры позволяет получить и минимальный энергетический уровень [195, 196].

Блок-схема АОЖ, используемый для обучения МП приведена на рисунке 3.10.

Блоки 1, 14 используются для пуска и остановки алгоритма.

В блоке 2 реализован ввод исходных данных, таких как: значения входных параметров и соответствующего им отклика МП из обучающей или тестирующей выборок, либо реальных данных.

Блок 3 обеспечивает генерацию начальных значений весов нейронных связей и температуры T_{max}.

В блоке 4 рассчитывается значение энергии в соответствии с соотношением $E = (Y_1 - Y_1')^2 + (Y_2 - Y_2')^2 + (Y_3 - Y_3')^2 + ... + (Y_n - Y_n')^2$.

В блоке 5 реализована проверка условия наличия положительной температуры тела. Если данное условие выполнено, то управление подается в блок 6, иначе вычислительный процесс завершается и управление передается в блок 13.

Блок 6 обеспечивает генерацию значений приращений весов w’.

В блоке 7 рассчитывается обновленное значение энергии E’.

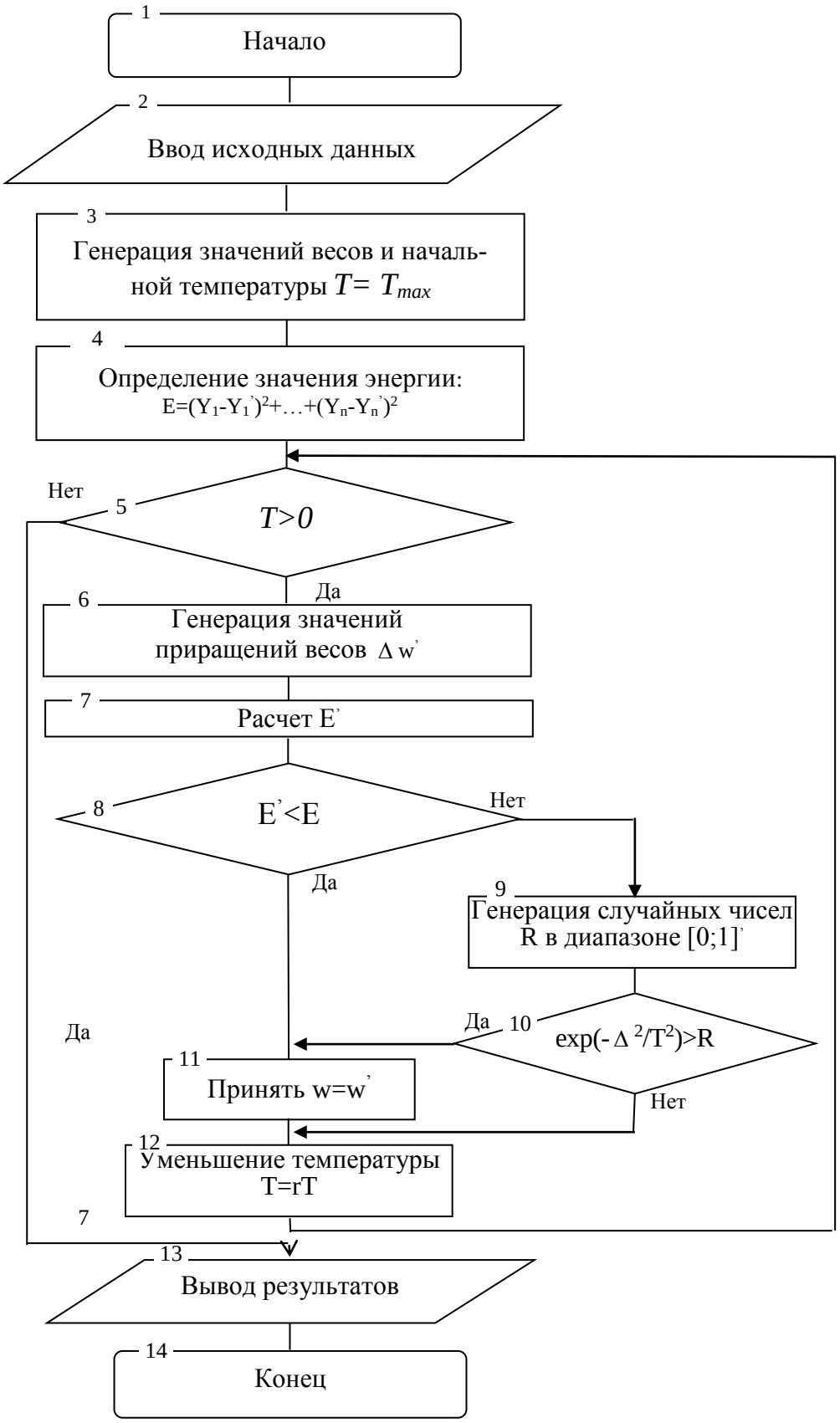


Рисунок 3.10. – Блок-схема алгоритма имитации отжига

В блок 8 проверяется условие превышения предыдущего значения энергии над текущим значением. Если это так, то управление передается в блок 11, в противном случае управление передается в блок 9.

В Блоке 9 реализована генерация случайных чисел в диапазоне от 0 до 1.

В блок 10 проверяется условие $\exp(-\Delta^2/T^2)>R$. Если это так, то управление передается в блок 11. В противном случае управление передается в блок 12.

В блоке 11 весам присваиваются новые значения.

Блок 12 обеспечивает снижение температуры.

В блоке 13 осуществляется вывод полученных результатов – оптимальных значений весов ИНС.

В АИО реализован стохастически управляемый метод повторных рестартов. Для оценки его эффективности был проведен сравнительный анализ функционирования АИО, ГА и алгоритма прямого перебора.

На рисунке 3.11 приведен вид экранной формы интерфейса исследования данных алгоритмов обучения МП.

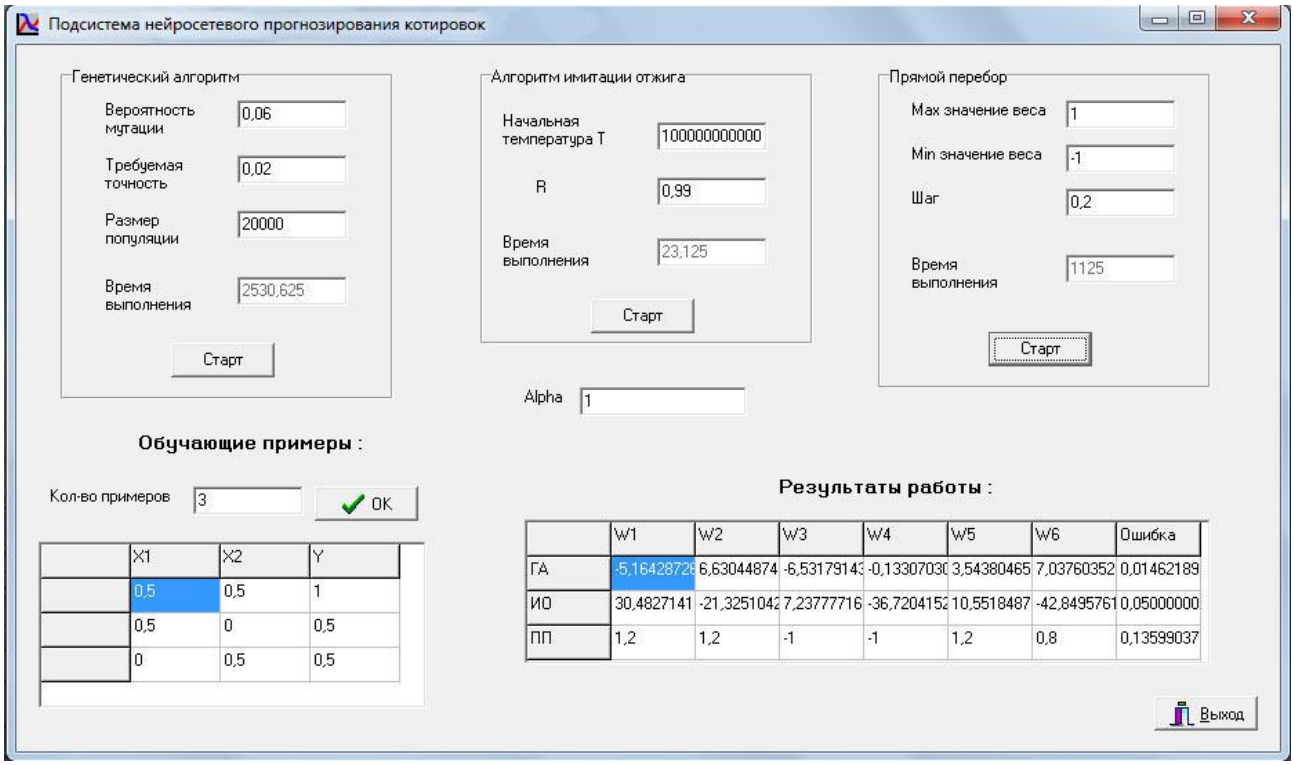


Рисунок 3.11. – Вид интерфейса исследования алгоритмов обучения МП

Представленный интерфейс имеет блочную структуру. В верхней части окна располагаются блоки заданий параметров трех алгоритмов обучения (ГА, АИО и прямого перебора). Кроме того, в каждом блоке имеется окно, в котором отображается время выполнения соответствующего алгоритма обучения, а также кнопка его запуска – «Старт». В нижней левой части экранной формы находится компонент задания обучающих примеров. В нижней правой части экранной формы находится таблица, в которой отображаются значения весов нейронных связей и ошибки соответствующих ИНС. В центре экранной формы имеется окно задания параметра α. Данный параметр определяет крутизну функции активации. При $\alpha \to \infty$, функция вырождается в пороговую. При $\alpha \to 0$, функция активации становится линейной функцией с пороговым значением - 0,5. При $\alpha \to 1$, данная функция является плавной сигмоидальной функцией (вариант, представляющий интерес в работе).

В таблицах 3.2 и 3.3 приведены экспериментальные значения основных параметров функционирования (времени работы и значений ошибки) для трех алгоритмов обучения.

Таблица 3.2 - Результаты работы ГА и АИО

Критерии	Генетический алгоритм									
Время работы (мсек)	1324	1184	5471	2334	1091	1177	2330	1111	2354	1123
Ошибка	0,016	0,014	0,019	0,019	0,014	0,018	0,01	0,007	0,01	0,012
	Алгоритм имитации отжига									
Время работы (мсек)	23	22	21	22	23	23	28	26	25	21
Ошибка	0,05	0,09	0,07	0,07	0,05	0,09	0,08	0,05	0,05	0,08

Таблица 3.3 - Результаты работы алгоритма прямого перебора

Критерии	Прямой перебор				
	1	2	3	4	5
Время работы (мсек)	1157	42960	39760	1673589	303858
Ошибка	0,13	0,14	0,095	0,099	0,071

В таблице 3.2 приведены результаты работы двух алгоритмов обучения, при постоянных исходных данных. В таблице 3.3 представлены резуль-

таты функционирования алгоритма прямого перебора для 5 вариантов исходных данных:

1) Max значение веса – 1, Min значение веса – -1, Шаг – 0,2;

2) Max значение веса – 1, Min значение веса – -1, Шаг – 0,1;

3) Max значение веса – 2, Min значение веса – -2, Шаг – 0,2;

4) Max значение веса – 2, Min значение веса – -2, Шаг – 0,1;

5) Max значение веса – 3, Min значение веса – -3, Шаг – 0,2.

Анализ полученных результатов показывает, что АИО является наиболее оперативным, но не самым точным из рассматриваемых методов обучения. Более высокую точность показал ГА, но он затрачивает намного больше времени на работу чем АИО. Теоретически, алгоритм прямого перебора является самым точным методом поиска глобального минимума функции. Однако, на практике, по причине недостаточной аппаратной мощности ПК, оказалось невозможным получить значение глобального минимума за приемлемое время.

3.1.4. Программная реализация генетического алгоритма

Модуль «Генетический алгоритм» реализует случайный поиск экстремума функции для нескольких аргументов. Он представляет собой симбиоз переборного и градиентного методов. Его основу составляет естественный отбор - главный механизм эволюции [27, 34, 65]. В нем реализуются генетические коды индивидуумов – ДНК (дезоксирибонуклеиновая кислота), приспособленность индивидуумов, хромосомы, скрещивание, мутации и др. Детальное описание ГА и его программная реализация приведены в разделе 2.6.

3.1.5 Теоретические аспекты и алгоритмы адаптации архитектуры персептрона

Модуль адаптации архитектуры МП обеспечивает рациональный выбор числа слоев сети, числа нейронов в слое и количество требуемых связей между слоями адекватно поставленной задаче.

Рассмотрим последовательность процесса обучения сети. Обучение ведется путем минимизации целевой функции *E(w)*, определяемой только на

обучающем подмножестве L. При этом значение целевой функции определяется, как [196]:

$$E(w) = \sum_{k=1}^{p} E(y_k(w), d_k)\,, \tag{3.17}$$

где p - количество обучающих пар (x_k, d_k), y_k — вектор реакции сети на возбуждение x_k.

Минимизация функции (3.17) обеспечивает достаточное соответствие выходных сигналов сети ожидаемым значениям из обучающих выборок [196].

Цель обучения состоит в таком подборе архитектуры и параметров МП, которые обеспечат минимальную погрешность распознавания тестового подмножества данных, не участвовавших в обучении. Эту погрешность будем называть погрешностью обобщения $E_G(w)$. Со статистической точки зрения погрешность обобщения зависит от уровня погрешности обучения $E_L(w)$ и от доверительного интервала ε. Она характеризуется неравенством (3.18) [196]:

$$E_G(w) =\leq E_L(w) + \varepsilon(\frac{p}{h}, E_L)\,. \tag{3.18}$$

Установлено [196], что значение ε функционально зависит от уровня погрешности обучения $E_L(w)$ и от отношения количества обучающих выборок p *к* фактическому значению h параметра, называемого мерой Вапника-Червоненкиса (МВЧ) и обозначаемого VC_d. Данная мера отражает уровень сложности МП и тесно связана с количеством содержащихся в ней весов. Значение ε уменьшается по мере возрастания отношения количества обучающих выборок к уровню сложности сети. Значение МВЧ функционально зависит от количества синаптических весов, связывающих нейроны между собой. Чем больше количество различных весов, тем больше сложность сети и соответственно значение меры VC_d. Верхняя и нижняя границы МВЧ определяются в соответствии с неравенством (3.19) [196]:

$$2\left[\frac{K}{2}\right]N \le VC_d \le 2N_w(1+\lg N_n)\,, \tag{3.19}$$

где [ ] - целая часть числа, N — размерность входного вектора, K -

количество нейронов скрытого слоя, N_w — общее количество весов МП, а N_n - общее количество нейронов МП [196].

Из выражения (3.20) следует, что нижняя граница диапазона приблизительно равна количеству весов, связывающих входной и скрытый слои, тогда как верхняя граница превышает двукратное суммарное количество всех весов МП. В связи с невозможностью точного определения меры VC_d в качестве ее приближенного значения используется общее количество весов МП [196].

Таким образом, на погрешность обобщения оказывает влияние отношение числа обучающих выборок к количеству весов сети. Небольшой объем обучающего подмножества при фиксированном количестве весов вызывает хорошую адаптацию сети к его элементам, однако не усиливает способности к обобщению, так как в процессе обучения наблюдается относительное превышение числа подбираемых параметров (весов) над количеством пар фактических и ожидаемых выходных сигналов сети. Эти параметры адаптируются с чрезмерной (а вследствие превышения числа параметров над объемом обучающего множества - и неконтролируемой) точностью к значениям конкретных выборок, а не к диапазонам, которые эти выборки должны представлять. Фактически задача аппроксимации подменяется в этом случае задачей приближенной интерполяции. В результате всякого рода нерегулярности обучающих данных и измерительные шумы могут восприниматься как существенные свойства процесса. Функция, воспроизводимая в точках обучения, будет хорошо восстанавливаться только при соответствующих этим точкам значениях. Даже минимальное отклонение от этих точек вызовет значительное увеличение погрешности, что будет восприниматься как ошибочное обобщение. Результаты численных экспериментов показали, что высокие показатели обобщения достигаются при превышении числа обучающих выборок МВЧ в несколько раз. Однако, при чрезмерном количестве нейронов и весов в МП, возникает эффект гиперразмерности, приводящий к значительным ошибкам [196]. Поэтому для оптимизации числа используемых нейронов использовался алгоритм каскадной корреляции Фальмана. Его суть заключается в динамиче-

ском добавлении нейронов в МП по результатам оценки ее обобщающих свойств после определенного количества циклов обучения. Достоинством данного алгоритма является частичное сохранение результатов обучения МП, полученных на предыдущих итерациях обучения. Реализация данного алгоритма заключалась в следующем. На начальном этапе обучения МП использовалось количество нейронов, заведомо недостаточное для решения задачи. Для обучения использовались вышеизложенные алгоритмы (АОРО, ГА или АИО) и обычные методы. Обучение происходило до тех пор, пока ошибка не переставала убывать и не выполнялись условия (3.20) [196]:

$$\left\{\begin{array}{l}\dfrac{E(t)-E(t-\delta)}{E(t_0)}<\Delta_r\\ t\geq t_0+\delta\end{array}\right\}, \qquad (3.20)$$

где t — время обучения; Δ_r - пороговое значение убыли ошибки; δ - минимальный интервал времени обучения между добавлениями новых нейронов; t_0 — момент времени последнего добавления. Когда выполнялись оба условия, добавлялся новый нейрон. Веса и порог нейрона инициализировались небольшими случайными числами. Обучение повторялось до тех пор, пока не выполнялось условие (3.19). Зависимость значения ошибки от времени обучения для алгоритма каскадной корреляции Фальмана приведена на рисунке 3.12 [173].

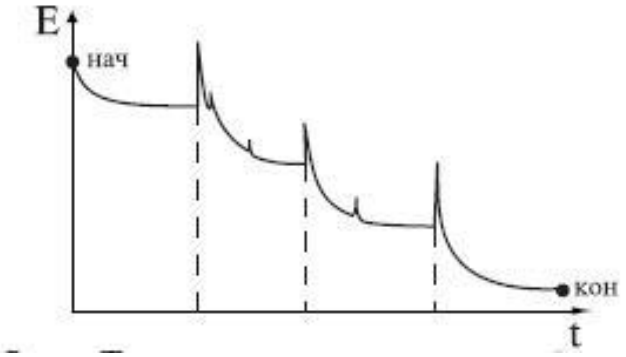


Рисунок 3.12. – Зависимость значения ошибки от времени обучения МП для алгоритма Фальмана

Моменты добавления новых нейронов отмечены пунктиром. После каждого добавления ошибка сначала резко возрастает, т.к. параметры нейро-

на случайны, а затем быстро сходится к меньшему значению [196].

В интересах рационального подбора архитектуры МП, наряду с алгоритмом каскадной корреляции Фальмана использовался способ минимизации структуры МП основанный на модификации ЦФ, позволяющей выявлять и исключать малоактивные скрытые нейроны. Его выбор обусловлен максимальной технологической приспособленностью к общему процессу обучения МП в данной работе. В данном способе учитывается, что если выходной сигнал какого-либо нейрона при любых обучающих выборках остается неизменным (на его выходе постоянно вырабатывается 1 или 0), то его присутствие в сети излишне. И напротив, при высокой активности нейрона считается, что его функционирование дает важную информацию [173]. При этом, модифицированная целевая функция, предложенная И. Шовеном [196], имела вид (3.21):

$$E(w) = E^{(0)}(w) + \mu \sum_{i=1}^{K} \sum_{j=1}^{p} e(\Delta_{ij}^2) \,. \qquad (3.21)$$

В этом выражении Δ_{ij} означает изменение значения выходного сигнала *i*-го нейрона для *j*-й обучающей выборки, а $e(\Delta_{ij}^2)$ - это корректирующий фактор целевой функции, зависящий от активности всех *K* скрытых нейронов для всех *j(j = 1, 2, ..., p)* обучающих выборок. Коэффициент *m* определяет степень относительного влияния корректирующего фактора на значение целевой функции. Вид корректирующей функции подбирался так, чтобы изменение целевой функции зависело от активности скрытого нейрона, причем при высокой его активности (т.е. частых изменениях значения выходного сигнала). Величина ΔE должна быть малой, а при низкой активности - большой [196]. Это достигалось применением функции *e*, удовлетворяющей отношению (3.22) [196]:

$$e' = \frac{\partial e(\Delta_i^2)}{\partial \Delta_i^2} = \frac{1}{(1 + \Delta_i^2)^n} \,. \qquad (3.22)$$

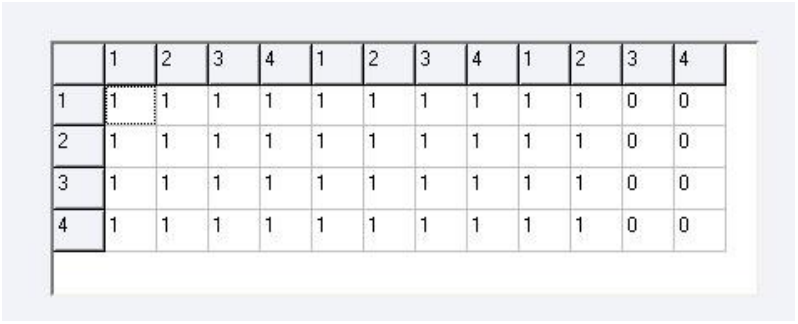
Индекс *n* позволяет управлять процессом штрафования за низкую ак-

тивность [196].

При n = 2 функция e принимает вид: $e = \frac{1}{1+\Delta_i^2}$. Малая активность нейронов карается сильнее, чем высокая, что в конечном итоге приводит к полному исключению пассивных нейронов из сети [173].

Оба подхода к оптимизации архитектуры МП, основанные как на учете чувствительности, так и на модификациях ЦФ, привели к минимизации количества весов и нейронов МП, уменьшив, таким образом, уровень ее сложности и улучшив соотношение между количеством обучающих выборок и МВЧ. Это обеспечило возрастание способности МП к обобщению [173].

При программной реализации алгоритма каскадной корреляции Фальмана в среде Delphi, в интересах задания межнейронных связей использован компонент StringGrid, позволяющий обрабатывать большие массивы данных (см. рисунок 3.13) [173].



	1	2	3	4	1	2	3	4	1	2	3	4
1	1	1	1	1	1	1	1	1	1	1	0	0
2	1	1	1	1	1	1	1	1	1	1	0	0
3	1	1	1	1	1	1	1	1	1	1	0	0
4	1	1	1	1	1	1	1	1	1	1	0	0

Рисунок 3.13. – Компонент StringGrid, используемый в интересах задания межнейронных связей

Кроме того, в совокупности с данным компонентом использован компонент «Hint», обеспечивающий отображение реквизитов (номер строки и столбца) на сфокусированном элементе матрицы. Последнее обеспечило удобство поиска необходимой связи с целью последующего изменения ее состояния.

3.1.6. Реализация эвристической оптимизации ИНС

Модуль «Эвристическая оптимизация» обеспечивает оптимизацию выбора ряда параметров действующих алгоритмов формирования и обучения МП эмпирическим путем.

Применительно к АОРО эвристический подход использовался для его модификации и уточнения ряда параметров. Проведенные вычислительные эксперименты показали, что как и любой градиентный алгоритм, АОРО "застревает" в локальных минимумах функции ошибки, т.к. градиент вблизи локального минимума стремится к нулю. Это приводило к большим ошибкам МП и его дальнейшей непригодности. Суть первого способа преодоления локального минимума заключалась в неоптимальном выборе шага дискретизации. Было обнаружено явление, оправдывающее неоптимальный выбор шага. Известно, что поверхность функции ошибок *E(P)* имеет множество долин, седел и локальных минимумов. Оказалось, что первый найденный минимум редко имел малую величину *E*. Чем больше нейронов и синапсов было в сети, тем менее чаще определялся глобальный минимум ЦФ. Выбор не оптимального шага приводил к тому, что он, в силу своей большой величины, позволял выходить из окрестности одного локального минимума и попадать в окрестность другого минимума. Для того, чтобы из-за большого шага дискретизации не «проскочить» другие локальные минимумы и возможно глобальный минимум, был разработан специальный алгоритм. Его идейную основу составил алгоритм «упругого распространения» - Rprop [196]. Данный алгоритм использует знаки частных производных для подстройки весовых коэффициентов. Для определения величины коррекции используется следующее правило (3.23):

$$\Delta_{ij}^{(t)} = \begin{Bmatrix} \eta^{+}\Delta_{ij}^{(t)}, \dfrac{\partial E^{(t)}}{\partial w_{ij}} \dfrac{\partial E^{(t-1)}}{\partial w_{ij}} > 0 \\ \eta^{-}\Delta_{ij}^{(t)}, \dfrac{\partial E^{(t)}}{\partial w_{ij}} \dfrac{\partial E^{(t-1)}}{\partial w_{ij}} < 0 \\ 0 < \eta^{-} < 1 < \eta^{+} \end{Bmatrix}, \qquad (3.23)$$

где $\Delta w_{ij}^{n} = -\eta \dfrac{\partial E}{\partial w_{ij}}$, а $\Delta w_{ij}^{n} = -\eta \cdot \delta_{j}^{n} \cdot y_{i}^{n-1}$.

Если на текущем шаге частная производная по соответствующему весу

w_{ij} изменила знак, то из этого следует, что последнее изменение было большим и алгоритм «проскочил» локальный минимум. Следовательно, величину изменения необходимо уменьшить на η^- и вернуть предыдущее значение весового коэффициента, то есть сделать «откат» - $\Delta w_{ij}(t) = \Delta w_{ij}(t) - \Delta_{ij}^{(t-1)}$. В отличие от этого действия в разработанном модифицированном алгоритме Rprop, «откат» заменяется «обратным ходом» с уменьшенным шагом дискретизации. Применительно к алгоритму имитации отжига это эквивалентно незначительному разогреву металла (вещества). Дальнейшие действия реализуются в соответствии с алгоритмом Rprop [173].

Если знак частной производной не изменился, то нужно увеличить величину коррекции на η^+ для достижения более быстрой сходимости. В результате проведенных экспериментов было установлено, что целесообразно выбирать η^+=1.2, а η^-=0.5. Начальные значения для всех Δ_{ij} устанавливались равными 0.1. Для вычисления значения коррекции весов использовалось следующее правило (3.24) [173]:

$$\Delta w_{ij}(t) = \begin{Bmatrix} -\Delta_{ij}^{(t)}, \dfrac{\partial E^{(t)}}{\partial w_{ij}} > 0 \\ +\Delta_{ij}^{(t)}, \dfrac{\partial E^{(t)}}{\partial w_{ij}} < 0 \\ 0, \dfrac{\partial E^{(t)}}{\partial w_{ij}} = 0 \end{Bmatrix}. \qquad (3.24)$$

Если производная положительна, т. е. ошибка возрастает, то весовой коэффициент уменьшался на величину коррекции, в противном случае – увеличивался.

Затем веса подстраивались в соответствии с выражением (3.25):

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t). \qquad (3.25)$$

Блок-схема модифицированного алгоритма Rprop приведена на рисунке 3.14.

Блоки 1,9 используются для пуска и остановки процесса обратного распространения ошибок.

В блоке 2 реализован ввод исходных данных.

Блок 3 обеспечивает инициализацию значений величин коррекции Δ_{ij} .

В блоке 4 предъявляются все примеры из выборки и вычисляются частные производные.

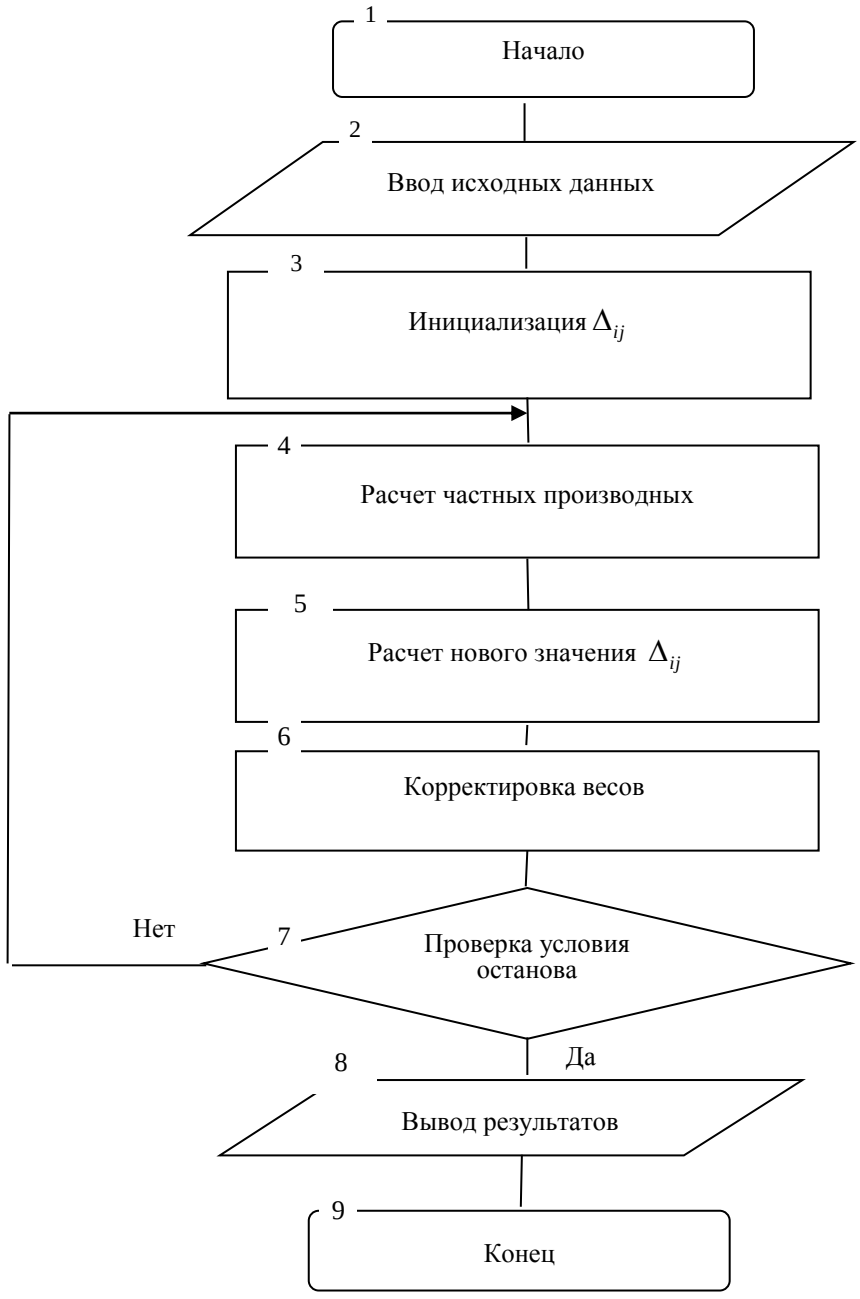


Рисунок 3.14. – Блок-схема модифицированного алгоритма Rprop

Блок 5 рассчитывает новые значения Δ_{ij} по формулам (3.23) и (3.24).

В блоке 6 реализуется корректировка весов в соответствии с выражением (3.25).

Блок 7 проверяет условие останова данного процесса. Если условие

останова не выполняется, то управление передается блоку 4. В противном случае управление передается блоку 8.

В блоке 8 реализован вывод полученных результатов.

Совокупность проведенных экспериментов (более 200), показала, что данный алгоритм сходится почти в 6 раз быстрее, чем стандартный алгоритм обратного распространения ошибок и на 30% быстрее, чем алгоритм Rprop.

Другой способ преодоления локальных минимумов заключался в обучении с шумом. Корректировка весов проводилась в соответствии с соотношением (3.26):

$$\Delta w = -\varepsilon \frac{\partial E}{\partial w} + n, \qquad (3.26)$$

где n — случайная величина, имеющая нулевое математическое ожидание и небольшую дисперсию. При выборе n использовалось нормальное распределение. Реализация данного способа приводила к увеличению времени обучения до 20 %. Однако вероятность определения глобального максимума существенно возрастала. После обучения МП модифицированным алгоритмом Rprop, из ста предъявленных тестов, МП ошибался лишь в двух. Экспериментальным путем удалось установить, что оптимальная величина n лежит в пределах 0,012-0,23.

Важными параметрами эвристической оптимизации являются границы нормированных значений входных и выходных сигналов МП. Известно, что общепринятый и описанный в литературе динамический диапазон нормированных входных и выходных сигналов МП лежит в пределах от 0 до 1 [196]. В ходе проведенных экспериментов было замечено, что применение нормированных значений в диапазоне ±½ уменьшает время сходимости алгоритма обратного распространения. Это обосновывается следующим образом. Так как величина коррекции веса $\Delta w_{pq,k}$ пропорциональна выходному уровню нейрона, порождающего $OUT_{p,j}$, то нулевой уровень ведет к тому, что вес не меняется. При двоичных входных векторах половина входов в среднем будет равна нулю, и веса, с которыми они связаны, не будут обучаться [196]. Более эффективным оказывается приведение входов к значениям ±½ и добавлении

смещения к функции активации, чтобы она также принимала значения ±½. Новая функция активации стала выглядеть следующим образом:

$$OUT = -½ + \frac{1}{1+e^{-NET}}.$$

На рисунке 3.15 показаны графики зависимостей величины достигаемого уровня ошибки *E* (ось ординат), от числа итераций *N* (ось абсцисс) для различных видов функции активации, полученные экспериментальным путем.

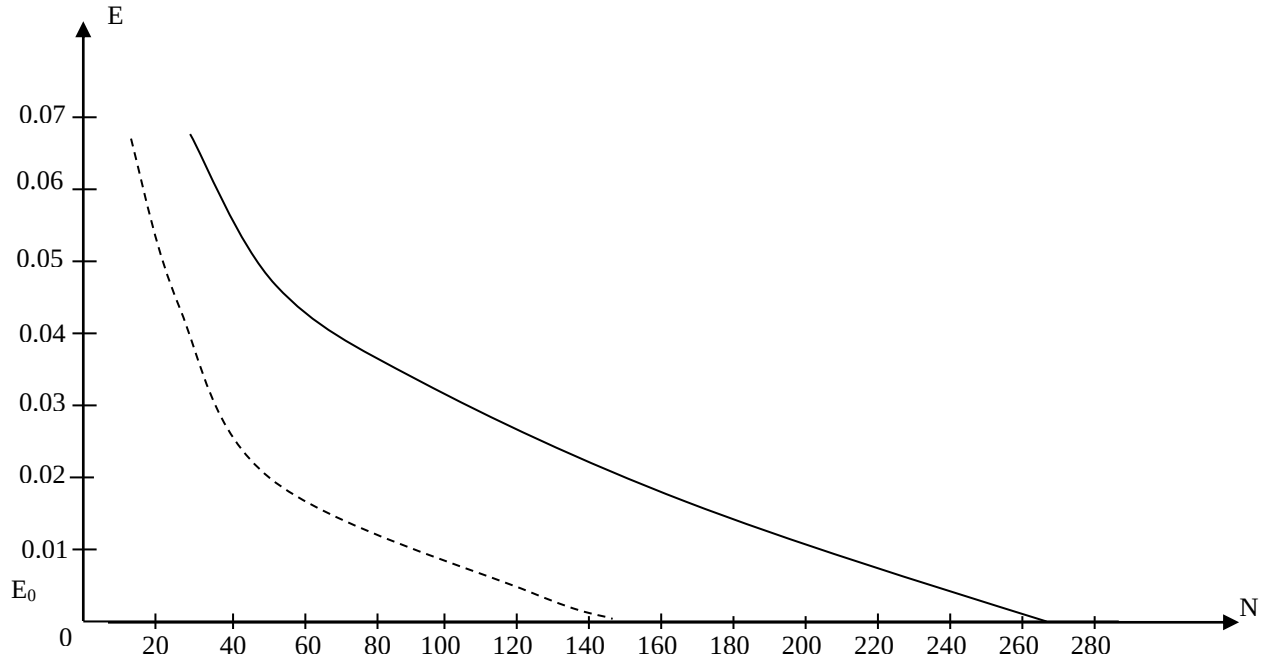


Рисунок 3.15. – Зависимости величины достигаемого уровня ошибки, от числа итераций

Из данного графика видно, что в результате использования функции активации нового вида (пунктирная кривая) сходимость (скорость достижения порогового значения) увеличилась почти в 2 раза. Последнее означает, что для достижения требуемой точности МП теперь требуется не 270 итераций, а лишь 150.

3.1.7. Реализация управления и визуализации многослойного персептрона

Модуль управления обеспечивает интерфейс взаимодействия ЛПР с подсистемами, общее проведение вычислений и информационный взаимообмен между соответствующими моделями и модулями.

Модуль визуализации (МВ) обеспечивает отображение исследуемого МП и проведение визуального моделирования путем исключения каких-либо

нейронов, слоев и связей между ними, то есть оперативной корректировки структуры сети. Впервые описание данного процесса приведено в [173, 178].

Известно, что в нервных системах реальных биологических объектов, откуда берут свое начало МП, некоторые связи между нейронами могут отсутствовать. Отдельные нейроны (и даже группы) могут не действовать. При этом нервная система способна эффективно функционировать, не «обращая внимания» на существующие «неисправности». Применительно к МП данный момент в настоящее время малоизучен. Представляет интерес учесть вышеизложенные особенности в рамках существующей технологии моделирования МП, чтобы восполнить этот пробел. Для этой цели в качестве инструмента совершенствования технологии моделирования МП для исследователей-неспециалистов в области программирования предлагается использовать МВ.

МВ обеспечивает визуальное представление матрицы размером $n \times m$, где n – число нейронов в слое, а m – количество слоев. Каждый элемент матрицы, представляющий собой факт наличия или отсутствия связи на k - м слое между i- м и j- м нейронами, может принимать значение либо «0» - связь отсутствует, либо «1» - связь присутствует [173].

В качестве примера моделируемого МП (см. рисунок 3.16) представлен МП, имеющий 4 входа, входной (0-й) слой, два скрытых слоя с 4-мя нейронами в каждом и в[0]одной слс[1]: 2-мя нейр[2]ами [173].[3]

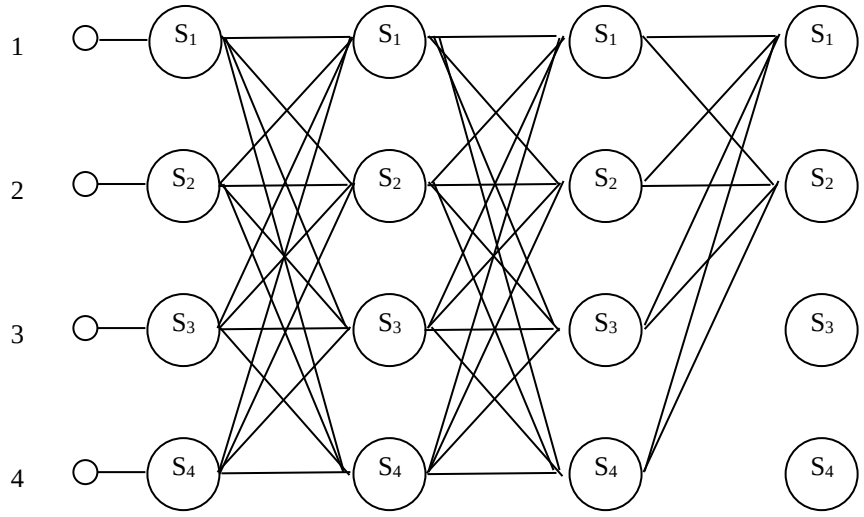


Рисунок 3.16 – Двухслойный персептрон

Матрица связей для данного МП представляется следующим образом [173]:

$$\begin{vmatrix} C_{11}^0 C_{12}^0 C_{13}^0 C_{14}^0 C_{11}^1 C_{12}^1 C_{13}^1 C_{14}^1 C_{11}^2 C_{12}^2 C_{13}^2 C_{14}^2 \\ C_{21}^0 C_{22}^0 C_{23}^0 C_{24}^0 C_{21}^1 C_{22}^1 C_{23}^1 C_{24}^1 C_{21}^2 C_{22}^2 C_{23}^2 C_{24}^2 \\ C_{31}^0 C_{32}^0 C_{33}^0 C_{34}^0 C_{31}^1 C_{32}^1 C_{33}^1 C_{34}^1 C_{31}^2 C_{32}^2 C_{33}^2 C_{34}^2 \\ C_{41}^0 C_{42}^0 C_{43}^0 C_{44}^0 C_{41}^1 C_{42}^1 C_{43}^1 C_{44}^1 C_{41}^2 C_{42}^2 C_{43}^2 C_{44}^2 \end{vmatrix}. \quad (3.27)$$

Рассмотрим элемент матрицы C_{11}^0. Верхний индекс указывает на номер слоя (в данном случае 0-й, т.е входной). В нижнем индексе указаны номера взаимодействующих нейронов (в данном случае оценивается взаимодействие 1-го нейрона 0-го слоя с 1-м нейроном 1-го скрытого слоя).

Если рассматривается случай, когда имеются взаимодействия между всеми нейронами МП, представленный на рисунке 3.15, то матрица связей с установленными значениями примет вид [173]:

$$\begin{vmatrix} 111111111100 \\ 111111111100 \\ 111111111100 \\ 111111111100 \end{vmatrix}. \quad (3.29)$$

Если рассматривается случай, когда в МП, представленной на рис. 3.15, отсутствует взаимодействие между 2 - м нейроном 1-слоя и 4-м нейроном 2-го слоя, то матрица связей с установленными значениями примет вид [173]:

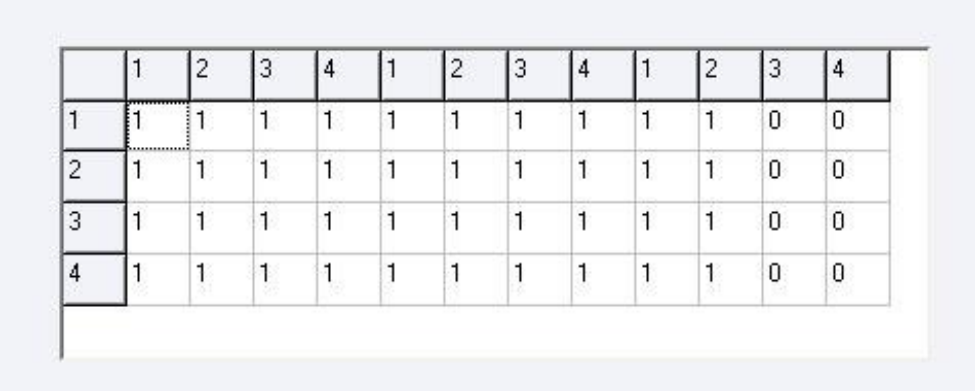
$$\begin{vmatrix} 111111111100 \\ 111111101100 \\ 111111111100 \\ 111111111100 \end{vmatrix}. \quad (3.30)$$

Матрица (3.30) используется при проведении расчетов в рамках последующих действий с МП, в частности при обучении, тестировании или эксплуатации. Значения установленных коэффициентов используются в качестве сомножителей при определении значений суммарных сигналов, поступающих на нейроны с различных входов (в качестве входов рассматриваются нейроны предыдущего слоя). Нулевые значения коэффициентов исключают из рассмотрения соответствующие входы, что эквивалентно разрыву связи между соответствующими нейронами.

Для программной реализации технологии визуального моделирования

в ПК НПВР использован стандартный компонент среды разработки Delphi 7.0 – StringGrid, представленный на рисунке 3.12. Его безусловным достоинством является возможность обработки больших массивов данных.

Визуальное представление моделируемого МП, приведенного на рисунке 3.16, с использованием этого компонента отображено на рисунке 3.17.



	1	2	3	4	1	2	3	4	1	2	3	4
1	1	1	1	1	1	1	1	1	1	1	0	0
2	1	1	1	1	1	1	1	1	1	1	0	0
3	1	1	1	1	1	1	1	1	1	1	0	0
4	1	1	1	1	1	1	1	1	1	1	0	0

Рисунок 3.17. – Визуальное представление моделируемого МП

С целью повышения наглядности отображения МП, совместно с компонентом StringGrid использован компонентом Hint. На его основе отображаются реквизиты (номер строки и столбца) на сфокусированном элементе матрицы, что обеспечивает удобство поиска необходимой связи с целью последующего изменения ее состояния. Применение МВ позволяет более гибко моделировать структуру МП и обеспечивает проведение исследований класса малоизученных (так называемых «изъянных») ИНС [87, 173].

3.2. Порядок использования и результаты апробации программного комплекса нейросетевого прогнозирования временных рядов

Для запуска программного комплекса необходимо выполнение файла ПК НСПВР.exe. После запуска появится главное окно программы (см. рисунок 3.18) [160].

ПК НПВР выполнен как SDI – приложение с набором функциональных вкладок: «Подготовка данных», «Обучение», «Прогнозирование». Рассмотрим функциональную составляющую и предназначение каждой из вкладок.

Вкладка «Подготовка данных» обеспечивает реализацию этапа формирования обучающей выборки. От полноты, состава, качества выборки обуче-

ния существенно зависят длительность обучения ИНС и качество получае-

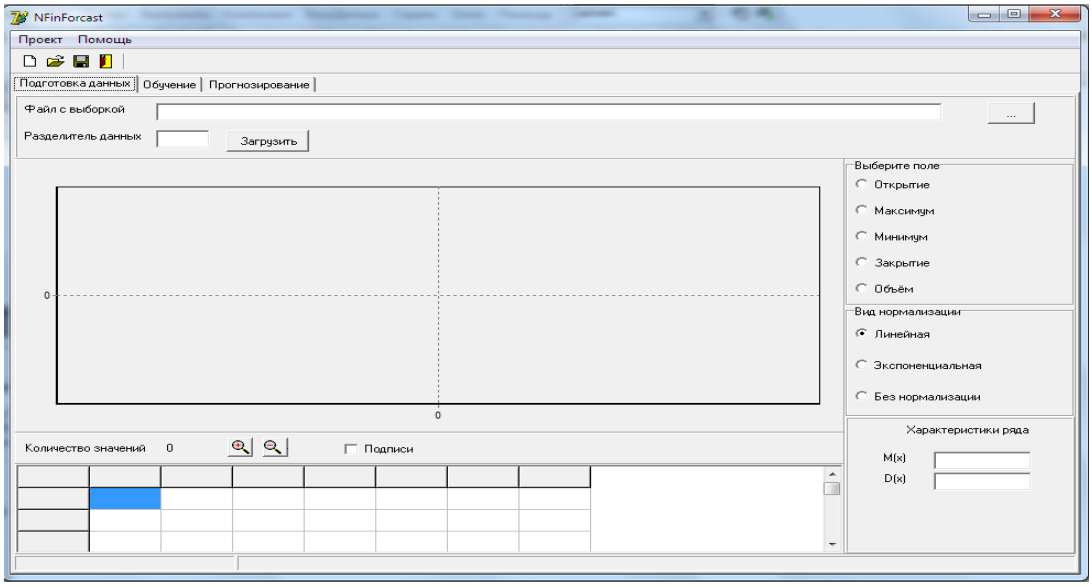


Рисунок 3.18. – Вид главной экранной формы ПК НПВР

мых прогнозов. Для ее загрузки необходимо выбрать текстовый файл, в котором она хранится. После загрузки обучающей выборки необходимо указать поля, используемые при обучении ИНС и способ их нормализации. В данной версии ПК НПВР предусмотрена линейная и экспоненциальная нормализация, а также ее отсутствие. На рисунке 3.19 представлен пример визуального отображения загруженной обучающей выборки с вычисленными значениями ее статистических характеристик – математического ожидания и дисперсии [160].

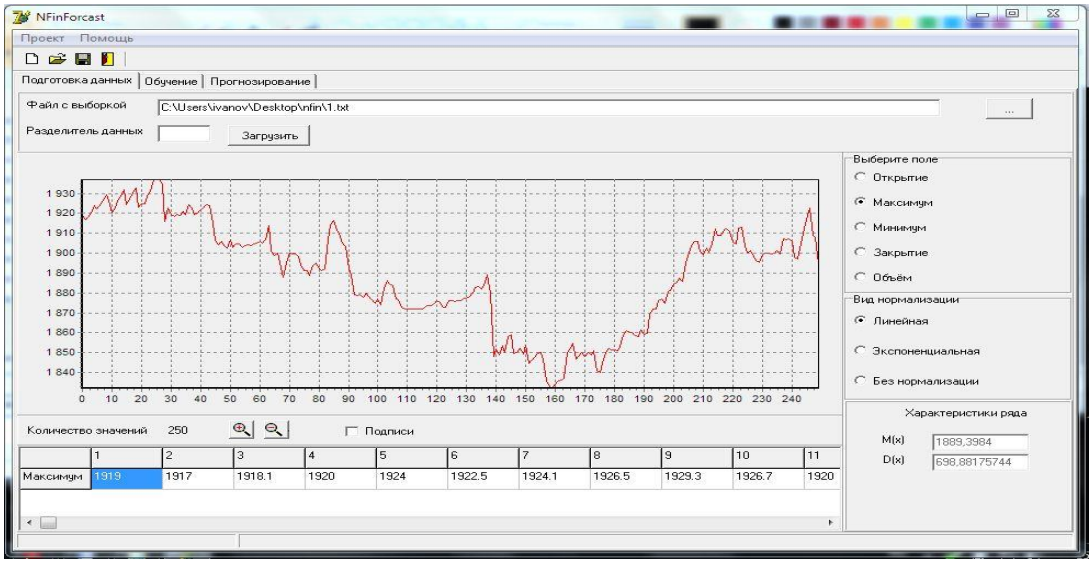


Рисунок 3.19. – Вид экранной формы ПК НПВР с загруженной обучающей выборкой

Вкладка «Обучение» реализует структурную настройку ИНС (см. рисунок 3.20).

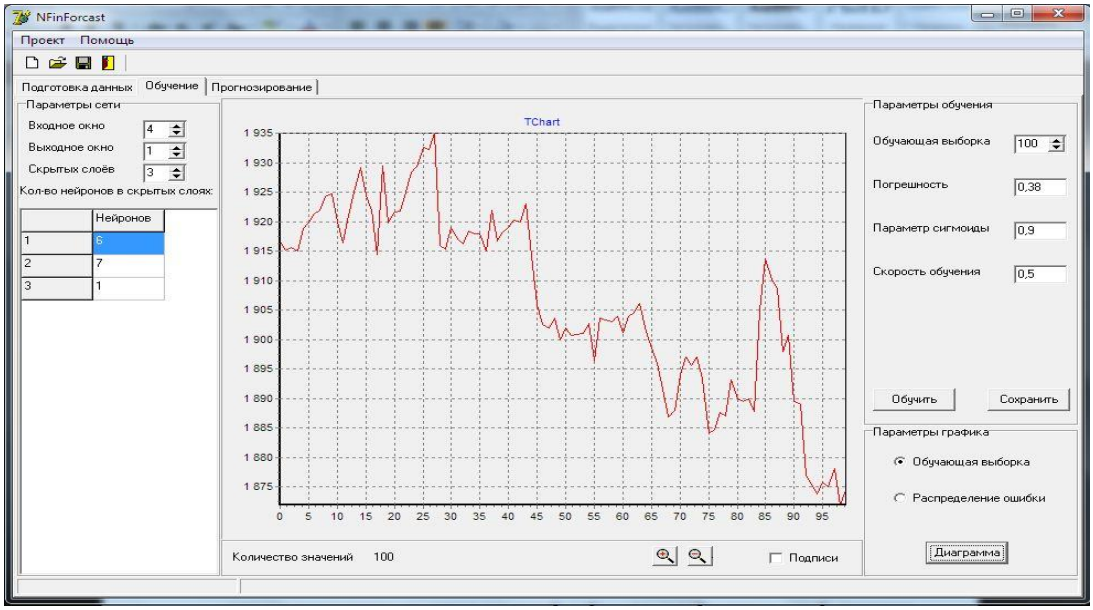


Рисунок 3.20. – Вид экранной формы ПК НПВР с открытой вкладкой «Обучение»

Она позволяет задать параметры обучения, в том числе: размер обучающей выборки, уровень погрешности, параметр сигмоиды, скорость обучения. В левой верхней части данной экранной формы расположена панель «Параметры сети», внутри которой можно задавать необходимую конфигурацию ИНС. Панель «Параметры графика», которая размещена в нижней правой части экранной формы, предоставляет возможность просмотра графика обучающей выборки и распределения ошибки после обучения. Кнопка «Обучить» запускает модуль «Алгоритм обратного распространения ошибок», внутри которого инициируется введённая конфигурация сети и параметры обучения. После этого происходит обучение сети. Результаты обучения ИНС (количество итераций и достигнутая погрешность) отображаются в окне информации - «Information» [160].

Чтобы сохранить обученную сеть необходимо нажать кнопку «Сохранить». При этом появится типизированное окно сохранения файла Windows, где следует указать место сохранения файла и нажать кнопку «Сохранить». Нейронная сеть будет сохранена в текстовый файл с расширением .txt.

Для дальнейшей работы с программным комплексом НПРВ в интере-

сах прогнозирования, необходимо перейти на вкладку «Прогнозирование» (см.

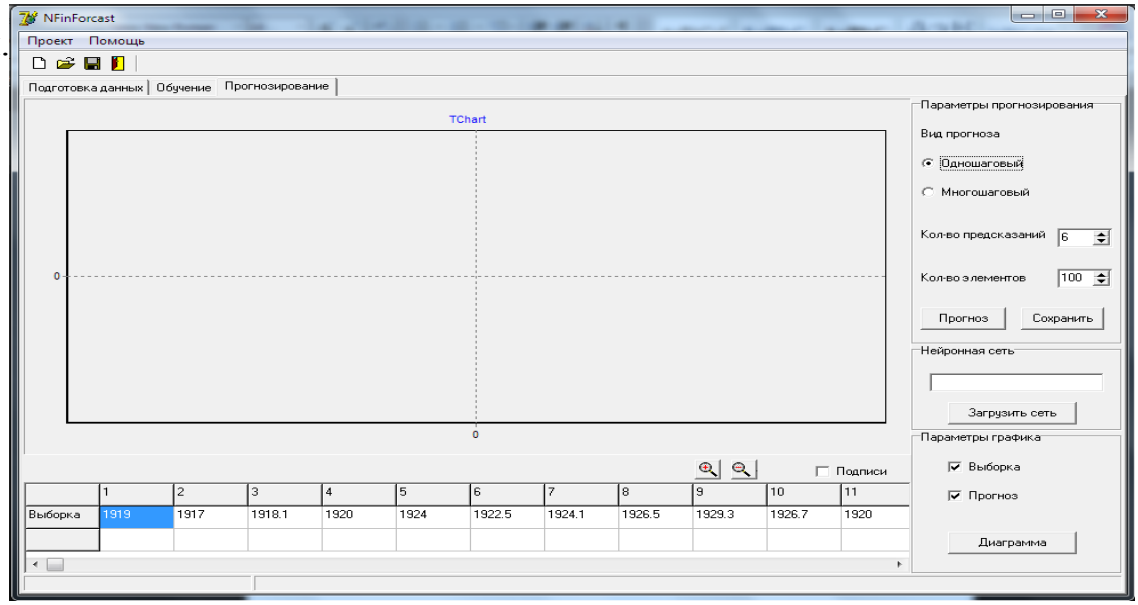


Рисунок 3.21. – Вид экранной формы ПК НПВР с открытой вкладкой «Прогнозирование»

На панели «Параметры прогнозирования» данной вкладки необходимо выбрать вид прогноза. Одношаговый прогноз является результатом единичного прогнозирования. При многошаговом прогнозе многократно реализуется последовательное применение прогноза и обучения. После этого задаётся количество предсказаний и нажимается кнопка «Прогноз». Пример экранной формы с результатами прогноза приведен на рисунке 3.22 [160].

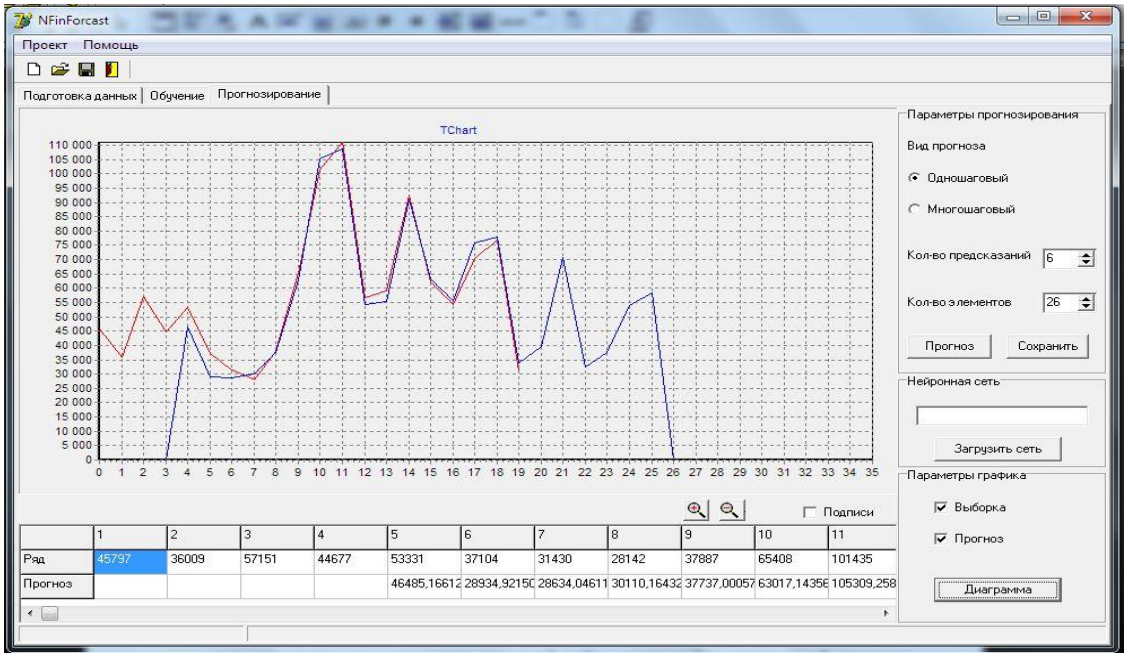


Рисунок 3.22. – Вид экранной формы ПК НПВР с результатами прогноза

Кнопка «Сохранить» записывает результаты прогноза в текстовый файл. Нажатие кнопки «Загрузить сеть» позволяет загрузить ранее обученную ИНС из файла и использовать ее для прогнозирования.

Для апробации работоспособности ПК НПВР была проведена серия много этапных экспериментов по прогнозированию курса акций Сбербанка РФ за период – с 17.05.16 по 01.06.16. Для обучения сети использовались различные наборы данных (10 минут, час, сутки). На первом этапе, для определенного вида представления исторических и прогнозируемых данных, формировались требуемая обучающая выборка и массив выходных и входных сигналов. Содержанием второго этапа являлись выбор структуры ИНС и настройка обучающих параметров. На третьем этапе осуществлялось обучение ИНС. При этом, в качестве основы, использовалась обучающая выборка, сформированная на первом этапе. Обучение останавливалось в случае, если ошибка обучения не превышала значения 0,01. На завершающем этапе ИНС тестировалась и оценивалось качество прогнозных значений котировок. После этого принималось решение о соответствии используемой архитектуры ИНС решаемой задаче. Было реализовано 23 опыта. Исследовались суточный, часовой и 10 -минутный наборы данных. В процессе проведения экспериментов установлено, что на точность прогнозных значений влияет протяженность исследуемого интервала. Чем он был больше, тем была выше точность прогноза. Данный факт можно объяснить тем, что ИНС на длинном интервале, за счет своей инерциальности успевает отрабатывать множество стохастических выбросов, представляющих собой «шум», и поэтому качество прогноза не страдает. Характерно, что 11 экспериментов показали относительно хороший результат прогноза. В первых 12 экспериментах наблюдались результаты, имеющие относительно высокие погрешности. Начиная с 13 эксперимента, ошибка ИНС уменьшилась до приемлемых значений (в среднем до 1,8%), за счет оптимизации структуры ИНС.

На рисунках 3.23 - 3.25 представлены результаты прогнозов трех показательных экспериментов. Параметры экспериментов приведены в таблице 3.4.

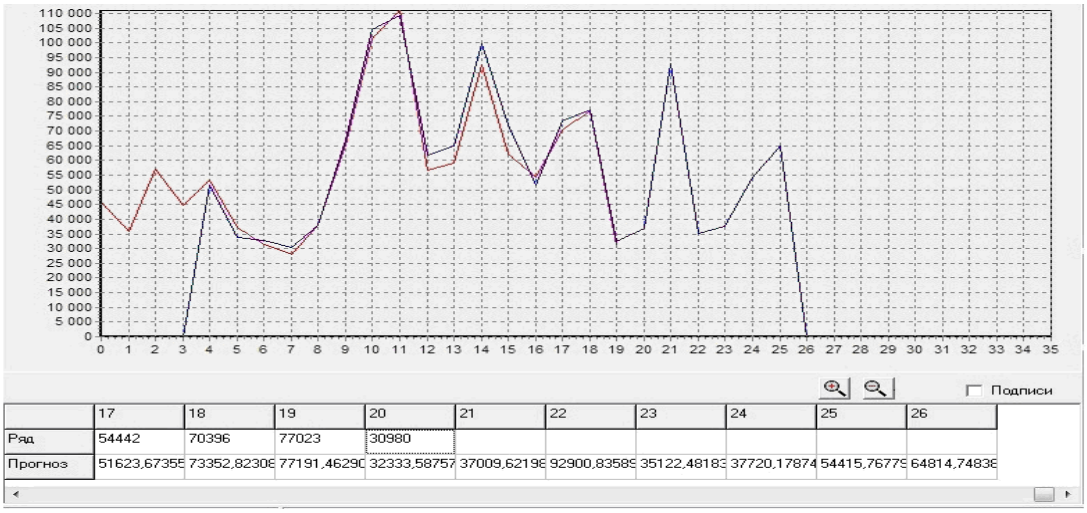


Рисунок 3.23. – Вид экранной формы с результатами прогноза для эксперимента №1

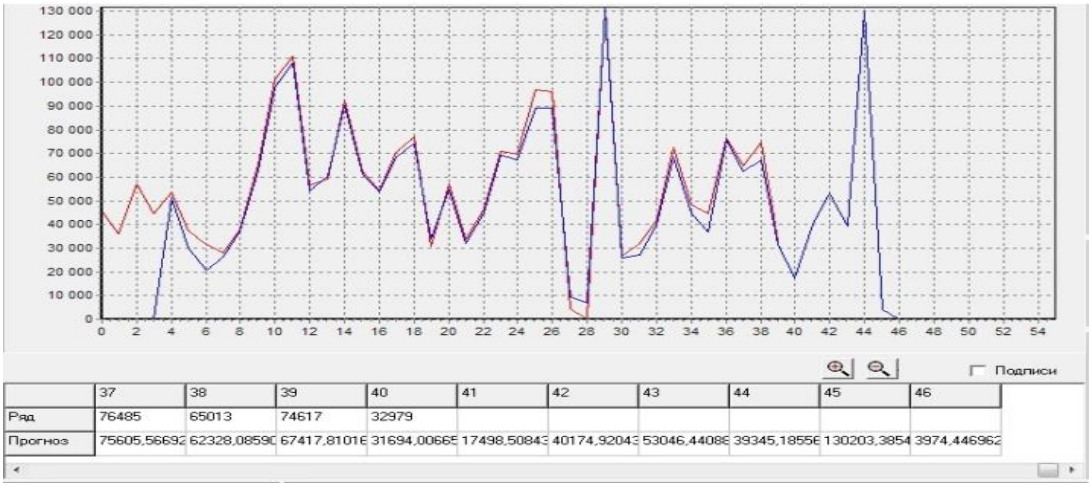


Рисунок 3.24. – Вид экранной формы с результатами прогноза для эксперимента №2

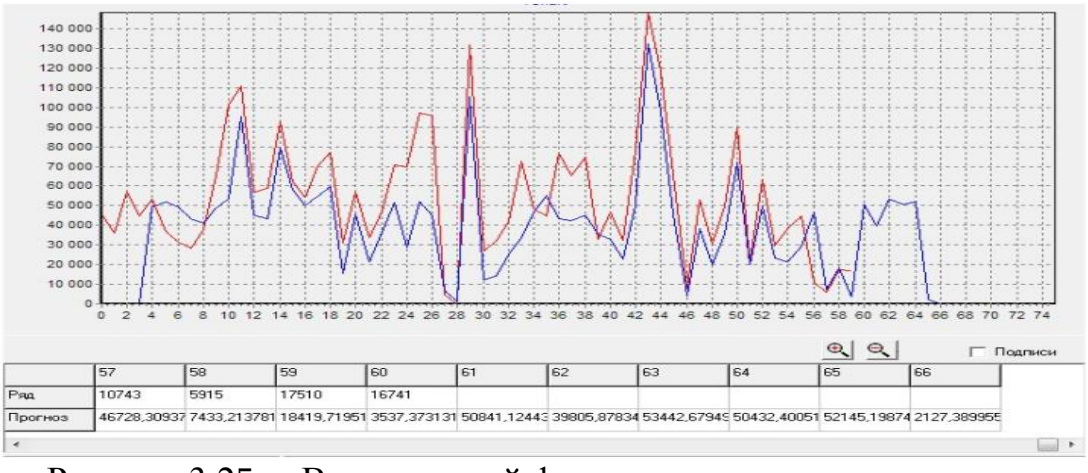


Рисунок 3.25. – Вид экранной формы с результатами прогноза для эксперимента №3

Таблица 3.4 - Параметры экспериментов исследования ИНС

№ эксперимента	Временной интервал	Число слоев ИНС	Число нейронов в слоях			
			1	2	3	4
1	10 минут	4	4	5	7	1
2	1 час	4	6	7	10	1
3	Сутки	3	5	10	1	-

Во всех экспериментах, в качестве критерия отбора входных данных используется поле «Объем» и линейная нормализация.

Визуальный анализ полученных результатов показывает, что точность прогнозных значений котировок зависит от структуры сети, в частности от числа используемых скрытых нейронных слоев. Наибольшая ошибка прогноза наблюдается в эксперименте №3, где используется один скрытый нейронный слой. Большей точностью обладает ИНС с двумя скрытыми нейронными слоями и увеличенным числом нейронов в них в эксперименте №2.

ВЫВОДЫ

Для исследования ИНС (оптимизации их структур и методов обучения) в интересах определения, хранения и применения «риск-эффектов» для обеспечения функционирования НМШ и НКМШ, в состав ИСППИД включен ПК НПВР. Он обладает следующими функциональными возможностями: загрузкой временных рядов; формированием выборок обучения, тестирования и контроля; загрузкой ранее обученных и протестированных ИНС для различных интервалов времени; формированием требуемой структуры ИНС; выбором метода обучения; обучения; тестирования; сохранения обученных ИНС.

В ПК НПВР генерируются и исследуются МП. К числу варьируемых параметров этого класса ИНС относятся: число входов и выходов; число скрытых слоев; количество нейронов в слоях; вид функции активации; вид нормализации. На этапе исследования ИНС, активируются и используются дополнительные функции: задание верхней и нижней границ меры Вапника - Червоненкиса; включение алгоритма Фальмана; включение процедуры выяв-

ления и исключения малоактивных нейронов на основе целевой функции Шовена.

В качестве методов обучения ИНС в ПК НПВР исследуются, модифицируются и используются: алгоритм обратного распространения ошибок; алгоритм имитации отжига; генетический алгоритм. Основное внимание уделено алгоритму обратного распространения ошибок. В интересах оптимизации его функционирования используется множество специальных процедур, в частности: введение момента инерции в дельта-правило настройки весов нейронов; эмпирическое задание размера обучающего множества; включение перекрестной проверки; включение раннего останова; включение последовательного, пакетного и комбинированного режимов обучения; включение дополнительного критерия останова; увеличение шага дискретизации; включение модифицированного алгоритма «упругого распространения»; включение обучения с шумом.

Перечень наиболее значимых модификаций алгоритмов обучения ИНС приведена в таблице 3.5 [173].

Таблица 3.5 - Характеристика модификаций алгоритмов обучения ИНС

Объекты и цели модификации	Суть модификации
АОРО. Улучшение обучения. Предотвращение переобучения.	1. Применение метода обучения с ранним остановом. Использование в качестве критерия останова - минимума кривой тестирования. 2. Введение второго критерия останова - малого изменения (0,1-1 %) среднеквадратической ошибки в течение эпохи.
АОРО. Повышение проходимости оврагов ЦФ.	1. Выбор увеличенного шага дискретизации. 2. Замена в алгоритме «упругого распространения» - Rprop «отката» на «обратный ход» с уменьшенным шагом дискретизации.
ИНС. Исследование новых видов.	Использование визуального контактора – коммутатора предназначенного для включения или исключения из исследований нейронов и связей между ними.

В состав ПК НПВР включен модуль визуализации, обеспечивающий отображение исследуемой ИНС и проведение визуального моделирования путем оперативной корректировки структуры сети. Применение данного модуля позволяет более гибко моделировать структуру ИНС и обеспечивает проведение исследований класса малоизученных (так называемых «изъянных») ИНС.

4. МЕТОДЫ, МОДЕЛИ И АЛГОРИТМЫ ТЕРМИНОЛОГИЧЕСКОГО ИНФОРМАЦИОННОГО ПОИСКА ПРАГМАТИЧЕСКОЙ ИНФОРМАЦИИ В РАМКАХ ПОРТФЕЛЬНОЙ ИНВЕСТИЦИОННОЙ ДЕЯТЕЛЬНОСТИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЙ ОРГАНИЗАЦИИ

4.1. Модель формализации текстовой информации

Реализация терминологического информационного поиска в ИСППИД СЭО предполагает формализацию текстов. Формализация обеспечивает представление текстовой информации в виде формальной системы [24, 75], базирующейся на определённых взаимосвязанных абстракциях – элементах (понятиях, аксиомах, теоремах, свойствах и др.), идеализации и искусственных символических языках (предикатах, моноидах и др.). Иерархическая схема уровней формализации текстовых данных представлена рисунке 4.1 [44].

Рисунок 4.1. – Иерархическая схема представления текстовых данных

Формализация позволяет систематизировать, уточнить и методологически прояснить содержание представления текстов, выяснить характер взаимосвязи между собой различных положений их обработки, выявить и сформулировать ещё не решённые проблемы [53]. Важное значение для формализации текста имеет модель, лежащая в ее основе [181]. В различные исторические

периоды разрабатывались разнообразные модели формализации текстовой информации. Результаты сравнительного анализа основных моделей формализации текстовой информации представлены в приложении Е [86].

Проведенный анализ моделей формализации текстовой информации показывает, что в интересах реализации терминологического информационного поиска необходимо использовать категоризацию лексики [240], обеспечивающую построение более простых и адекватных моделей представления текстовой информации. Она впервые возникла в моделях Ньютона и Лейбница [193], а дальнейшее развитие получила в лямбда-исчислении и семантической сети. В ее состав могут быть включены такие важные для терминологического информационного поиска категории, как тезаурус и терминологический портрет. Это позволяет дополнить иерархическую схему представления текстовых данных (см. рисунок 4.1) данными категориями и, на начальном этапе, сформировать математическую модель весовой распределенной информационной системы на тезаурусе, а затем на терминологическом портрете.

4.2. Математическая модель весовой распределенной информационной системы на тезаурусе

Рассмотрим ряд понятий необходимых для определения распределенной информационной системы, основанной на тезаурусе [207].

Определение 1. Тезаурусом будем называть конечное непустое множество $\{T\}$ слов *t*, отвечающих условиям [207]:

1) имеется непустое подмножество $\{T_0\} \subset \{T\}$, называемое множеством дескрипторов;

2) имеется симметричное, транзитивное рефлексивное отношение $R \subset T \times T$, такое, что:

$$\text{а) } t_1 \neq t_2 \wedge t_1 R t_2 \Rightarrow (t_1 \in T \setminus T_0) \vee (t_2 \in T \setminus T_0); \qquad (4.1)$$

$$\text{б) } t_1 \in T \setminus T_0 \Rightarrow (\exists t \in T_0)(t R t_1); \qquad (4.2)$$

где *R* - синонимическое отношение, а слова t_1 и t_2, отвечающие этому отношению, называются синонимическими дескрипторами;

3) имеется транзитивное и несимметричное отношение $K \subset T_0 \times T_0$, называемое обобщающим отношением.

В случае если два дескриптора t_1 и t_2 удовлетворяют отношению $t_1 K t_2$, то полагается, что дескриптор t_1 более общий, чем дескриптор t_2[207].

Элементы множества $\{T \setminus T_0\}$ называются множеством аскрипторов [207].

Определение 2. ИС с тезаурусом называется четверка *(T, D, M,* δ), где *T* – тезаурус с дескрипторным множеством T_0; *D* – коллекция документов; *M* – множество вопросов; $\delta : M \to 2^D$ - отображение, сопоставляющее каждому вопросу множество документов.

Описание любого документа $d \in D$ может быть представлено в виде [118]:

$$t(d) = \{t_1, t_2, \ldots, t_k\}, \tag{4.3}$$

и удовлетворяет отношению *K*.

Можно также считать, что каждый вопрос $m \in M$ представляется в форме, аналогичной описанию документов [207].

Множество описаний вопросов и документов частично упорядочено отношением включения $\underset{\sim}{<}$ следующим образом [207]: $\forall d_1, d_2 \in D$

$$t(d_1) \underset{\sim}{<} t(d_2) \Leftrightarrow (\forall t' \in t(d_1))(\exists t'' \in t(d_2))(t' K t'') \vee (t' = t''), \tag{4.4}$$

т.е. каждый дескриптор из $t(d_1)$ представляет собой обобщение дескриптора из $t(d_2)$ или идентичен дескриптору из $t(d_2)$.

Отношение $\underset{\sim}{<}$ позволяет сформулировать ответ *Q* на вопрос $m \in M$ в виде [118]:

$$Q = \delta(m) = \{d : d \in D \wedge m \underset{\sim}{<} \mathrm{t(d)}\}. \tag{4.5}$$

Пусть $S_1, \ldots, S_n$ - локальные ИС, где $S_j = (T_j, D_j, M_j, \delta_j), j = \overline{1,n}$, с составляющими, описанными в определении 1.

Объединим системы $S_1, \ldots, S_n$ в одну систему *S*, которую в дальнейшем будем называть распределенной ИС, базирующейся на глобальном тезаурусе $T = \underset{j}{U} T_j$ [207].

Пусть у нас имеется распределенная ИС *S=(T, D, M, δ)* с синонимическим отношением *R* и обобщающим отношением *K*. Тогда последовательность ИС можно определить следующим образом [118]:

$$S_j = (T_j, D_j, M_j, \delta_j), \tag{4.6}$$

где $T_j \subset T; D_j \subset D; M_j \subset M$; δ_j - сужение δ на $M_j; \forall j = \overline{1,n}$. Кроме того, отношения $R_j = R \cap (T_j \times T_j)$ и $K_j = K \cap (T_{0j} \times T_{0j})$; T_{0j} - множество дескрипторов системы S_j; отношение предпочтения $\underset{\sim j}{\leq} = \underset{\sim j}{\leq} \cap (M_j \times M_j)$;

$$\forall m \in M_j \qquad \delta(m) = \left\{d : d \in D_j \wedge m \underset{\sim j}{\leq} t(d)\right\}. \tag{4.7}$$

Множества $\{T_j\}$ и $\{D_j\}$ могут пересекаться.

Определим понятие подсистемы, которое позволит формализовать отношение включения одной системы в другую.

Определение 3. Пусть $S_1 = (T_1, D_1, M_1, \delta_1)$ и $S_2 = (T_2, D_2, M_2, \delta_2)$ - информационные системы. Система S_1 является подсистемой S_2, $S_1 \subset S_2$, если [118]:

$$1)\ (T_1 \subset T_2) \wedge (K_1 = K_2 \cap (T_{01} \times T_{02})) \wedge (R_1 = R_2 \cap (T_1 \times T_1)); \tag{4.8}$$

$$2)\ D_1 \subset D_2; \tag{4.9}$$

$$3)\ (M_1 \subset M_2) \wedge (\underset{\sim 1}{\leq} = \underset{\sim 2}{\leq} \cap (M_1 \times M_1)); \tag{4.10}$$

$$4)\ \delta_1(m) = \delta_2(m) \cap D_1, m \in M_1. \tag{4.11}$$

Поскольку распределенная ИС содержит набор документов локальных систем (так как локальная ИС - подсистема распределенной ИС), то общий ответ на некоторый вопрос к распределенной ИС может быть выражен как совокупность определенных операций, применительно к ответам локальных ИС.

Пусть имеется вопрос $\overline{m} = \{m\}$, который содержит дескриптор, а $\delta(\overline{m})$ - является ответом на него. Тогда ответ на локальный вопрос $\overline{m} = \{m\}$ в подсистеме S_j, согласно (4.5), будет описываться в виде [207]:

$$Q = \delta_j(\overline{m}) = \left\{d : d \in D_j \wedge m \underset{\sim j}{\leq} \mathrm{t(d)}\right\}. \tag{4.12}$$

Пусть, в распределенной ИС *S* ответ на глобальный вопрос

$\overline{m} = \{m\}$ представляет объединение ответов локальных ИС $\delta(\overline{m}) = \cup_j \delta_j(\overline{m})$.

Тогда, ответ на некоторый вопрос $\overline{m} = \{m_1, \ldots, m_k\}$ в распределенной ИС будет определяться в виде [207]:

$$\delta(\overline{m}) = \bigcap_{i=1}^{k} \delta(\overline{m}_i), \quad \overline{m}_i = \{m_i\}. \tag{4.13}$$

Можно показать верность выражения (4.13) применительно к последовательности ИС [207]:

$$\begin{aligned} \delta(\overline{m}) &= \bigcap_{i=1}^{k} \delta(\overline{m}_i) = \bigcap_{i=1}^{k} \bigcup_{j=1}^{n} \delta_j(\overline{m}_i) = \bigcap_{i=1}^{k} \bigcup_{j=1}^{n} \left\{ d : d \in D_j \wedge \overline{m_i} \underset{\sim}{\leq} \mathrm{t(d)} \right\} = \\ &= \bigcap_{i=1}^{k} \left\{ d : d \in \bigcup_{j=1}^{n} D_j \wedge \overline{m_i} \underset{\sim}{\leq} \mathrm{t(d)} \right\} = \bigcap_{i=1}^{k} \left\{ d : d \in D \wedge \overline{m_j} \underset{\sim}{\leq} \mathrm{t(d)} \right\} = \\ &= \left\{ d : d \in D \wedge \overline{m} \underset{\sim}{\leq} \mathrm{t(d)} \right\}. \end{aligned} \tag{4.14}$$

Определение 4. Распределенная ИС имеет свойство включения, когда набор вопросов частично упорядочен отношением $\underset{\sim}{\leq}$ и выполнено условие вида [207]:

$$\{\overline{m}_1, \overline{m}_2\} \subset M \wedge \overline{m}_1 \underset{\sim}{\leq} \overline{m}_2 \Rightarrow \delta(\overline{m}_1) \supset \delta(\overline{m}_2). \tag{4.15}$$

Данное свойство обеспечивает формулирование цепи ответов на ряд вопросов с уточнением ответов более детализированными вопросами [207].

В некоторых случаях при выполнении свойств аддитивности операции над ответами локальных систем могут быть упрощены. В частности:

1) если множества документов локальных ИС попарно не пересекаются, то глобальный ответ распределенной системы представляется как объединение локальных ответов,

$$(\vee_{1 \le k \ne l \le n} D_k \cap D_l = \varnothing) \wedge (\overline{m} = \{m_1, \ldots, m_S\}) \Rightarrow \delta(m) = \bigcup_{j=1}^{n} \delta_j(\overline{m}); \tag{4.16}$$

2) если все локальные ИС заданы на одном тезаурусе:

$$S_j = (T_j, D_j, M_j, \delta_j),\ j = \overline{1,n}, \tag{4.17}$$

а множества документов разнесены, то ответ на глобальный вопрос будет объединением локальных ответов, т.е. если $\overline{m} = \{m_1, \ldots, m_S\}$ - вопрос,

$$\underset{1\le i\le n}{\forall} T_i = T \text{, тогда } \delta(\overline{m}) = \bigcup_{j=1}^{n} \delta_j(\overline{m}). \qquad (4.18)$$

В плане дальнейшего обобщения представленной математической модели распределенной ИС [39], введем веса, которые будут описывать распределение информации на каждом слове. Информацию, заложенную в описании документа, будем рассматривать в виде некоторой единицы. Следовательно, вес w_i соответствует объему информации, попадающей на дескриптор t_i, т.е [207].

$$t(d) = \{< t_1, w_1 >, < t_2, w_2 >, \ldots, < t_k, w_k >\} \qquad (4.19)$$

и выполняются условия:

$$1)\ t_i \neq t_j \wedge t_i K t_j \Rightarrow (\forall w_i \in [0,1])(t_i, w_i) \not\subseteq t(d) \wedge \forall w_j \in [0,1])(t_j, w_j) \not\subseteq t(d); \qquad (4.20)$$

$$2)\ \sum_{i=1}^{k} w_i = 1. \qquad (4.21)$$

Рассмотрим подобие вопросов и описаний документов.

Определение 5. Пусть $t(d_1)$, $t(d_2)$ - описания документов:

$$t(d_1) = \{< t_{11}, w_{11} >, < t_{12}, w_{12} >, \ldots, < t_{1k_1}, w_{1k_1} >\}; \qquad 4.22)$$

$$t(d_2) = \{< t_{21}, w_{21} >, < t_{22}, w_{22} >, \ldots, < t_{2k_2}, w_{2k_2} >\}. \qquad (4.23)$$

Описание $t(d_1)$ подобно описанию $t(d_2)$ с точностью $\rho(0 \le \rho \le 1)$, если

$$t(d_1) \underset{\sim}{<} t(d_2) \wedge \forall w_{1i} \rho \le \sum_{j:[t_{2j} \in I(t_{1i})]} w_{2j}, \qquad (4.24)$$

$$< t_{1i}, w_{1i} > \in t(d_1)$$

где $I(t_{1i}) = \{t_{2j} : t_{1i} K t_{2j} \wedge 1 \le j \le k_1\}$.

Если $t(d_1)$ и $t(d_2)$ удовлетворяют условию (4.24), то запишем $t(d_1) \nabla_{\rho} t(d_2)$.

Рассмотрим пример.

Пример. Пусть $t_1 K t_4, t_1 K t_5, t_1 K t_6, t_2 K t_7, t_2 K t_8, t_3 K t_9$.

Положим

$$t(d_1) = \{< t_1, 0{,}5 >, < t_2, 0{,}3 >, < t_3, 0{,}2 >\};$$

$$t(d_2) = \{< t_4, 0{,}1 >, < t_5, 0{,}1 >, < t_6, 0{,}2 >, < t_7, 0{,}2 >, < t_8, 0{,}1 >, < t_9, 0{,}3 >\}.$$

Отношение $t(d_1) \nabla_{0,4} t(d_2)$ выполнено, если каждое неравенство в (4.25) верно:

$$\begin{cases} 0{,}5 \cdot 0{,}4 \le 0{,}1 + 0{,}1 + 0{,}2; \\ 0{,}3 \cdot 0{,}4 \le 0{,}2 + 0{,}1; \\ 0{,}2 \cdot 0{,}4 \le 0{,}3. \end{cases} \quad (4.25)$$

Рассмотрим ряд свойств отношения подобия, доказательство которых очевидно:

$$1.\ \text{Если } (t(d_1)\nabla_{\rho_1} t(d_2)) \wedge (\rho_1 \ge \rho_2), \text{ то } t(d_1)\nabla_{\rho_2} t(d_2). \quad (4.26)$$

$$2.\ \text{Если } (t(d_1)\nabla_{\rho_1} t(d_2)) \wedge (t(d_2)\nabla_{\rho_2} t(d_3)), \quad (4.27)$$

$$\text{то } (t(d_1)\nabla_{\rho} t(d_3)) \ (\rho \le \rho_1\rho_2). \quad (4.28)$$

Последнее свойство называется квазитранзитивностью [207].

Представляет интерес другая величина, характеризующая коэффициент подобия или меру корреляции пар векторов, удовлетворяющих отношению $\underset{\sim}{<}$.

Для векторов $t(d_1)$, $t(d_2)$ таких, что $t(d_1) \underset{\sim}{<} t(d_2)$, мера корреляции

$$\mu(t(d_1), t(d_2)) = \sum_{i:(t_{1i}, w_{1i}) \in t(d_1)} \min(w_{1i}, \sum_{j:[t_{2j} \in I(t_{1i})]} w_{2j}), \quad (4.29)$$

где $I(t_{1i}) = \{t_{2S} : t_{1i} K t_{2S} \wedge 1 \le i \le k_1\}$.

Фактически мера μ, оценивает пересекающуюся информацию, включенную в описании документов $t(d_1)$, $t(d_2)$.

Для описаний $t(d_1)$ и $t(d_2)$ в вышеприведенном примере, мера корреляции равна [207]:

$\mu(t(d_1), t(d_2)) = \min(0{,}5;\ 0{,}1 + 0{,}1 + 0{,}2) + \min(0{,}3;\ 0{,}2 + 0{,}1) + \min(0{,}2;\ 0{,}3) =$

$=0{,}4+0{,}3+0{,}2=0{,}9.$

Рассмотрим основные свойства меры корреляции [207]:

1. Мера корреляции неотрицательная величина, т.е.

$$\mu(t(d_1), t(d_2)) \ge 0; \quad (4.30)$$

2. Значение меры корреляции удовлетворяет неравенству:

$$0 \le \mu(t(d_1), t(d_2)) \le 1; \quad (4.31)$$

3. Если для каждого $(t_{1i}, w_{1i}) \in t(d_1)$ из определения подобия $w_{1i}\rho = \sum_{j:[t_{2j} \in I(t_{1i})]} w_{2j}$, то мера корреляции равна точности подобия, т.е.:

$$\mu(t(d_1),t(d_2))=\rho. \quad (4.32)$$

4. Если $t(d_1)\nabla_\rho t(d_2)$, то $\mu(t(d_1),t(d_2))\geq\rho$; (4.33)

5. Если $t(d_1)\underset{\sim}{<}t(d_2)\wedge\mu(t(d_1),t(d_2))=\rho_1$, (4.34)

то $t(d_1)\nabla_\rho t(d_2)\wedge\rho_2\leq\rho_1$.

Теперь возможно определить ИС на тезаурусе с весами.

Определение 6. ИС на тезаурусе с весами называется четверка *(T, D, M,* δ*),* где T – тезаурус с дескрипторным множеством $T_0\subset T$; D – набор документов; M – набор вопросов; δ- действие типа «отображение», $\delta: M\times[0,1]\to 2^{D\times[0,1]}$ которое определяет соответствие между парами (вопрос - точность подобия: документ - мера соответствия) [207].

Ответ ИС на вопрос $\overline{m}$ с требуемой точностью $\rho=c$ определяется в соответствии с выражением:

$$Q=\delta(\overline{m},c)=\left\{(d,\alpha): d\in D\wedge\overline{m}\nabla_c \mathrm{t(d)}\wedge\alpha=\mu(\overline{\mathrm{m}},\mathrm{t(d)})\right\}\subseteq D\times[0,1]. \quad (4.35)$$

В рамках ответа $N=\delta(\overline{m},c)$, выбор документов осуществляется в соответствии с отношением подобия ∇_c с точностью *c*. При этом, часть информации документа, соответствующая точному ответу на вопрос, определяется мерой корреляции $\mu(\overline{m},t(d))$ [207].

Одно из фундаментальных свойств ответа связано с точностью ρ и мерой корреляции μ вопроса к описанию документа, включенного в ответ. Представим это в виде соответствующей теоремы.

Теорема 1. Если документ $d\in D$ с мерой корреляции α включен в ответ, т.е. $(d,\alpha)\in\delta(\overline{m},c)$, то верно неравенство $\alpha\geq c$. Обратное утверждение неверно [207].

Представленная математическая модель весовой распределенной ИС, в сравнении с классической распределенной ИС, позволяет при формировании ответов на запросы учитывать неопределенности описания запросов и документов [207].

Дальнейшим развитием данной модели является математическая модель весовой обработки информации на терминологическом портрете.

4.3. Математическая модель весовой обработки информации на терминологическом портрете

Терминологический портрет (ТП) отличается от тезауруса, весовой взаимоувязкой терминов в нем и более высоким уровнем детализации рассматриваемой предметной области. Рассмотрим основные определения [119].

Определение 1. ТП будем называть конечное непустое множество $\{Z\}$ терминов *z*, отвечающих следующим условиям [119]:

1) имеется непустое подмножество $\{Z_0\} \subset \{Z\}$ - множество терминов;

2) имеется симметричное, транзитивное рефлексивное отношение $R \subset Z \times Z$, такое, что:

$$\text{а)}\ z_1 \neq z_2 \wedge z_1 R z_2 \Rightarrow (z_1 \in Z \setminus Z_0) \vee (z_2 \in Z \setminus Z_0); \quad (4.36)$$

$$\text{б)}\ z_1 \in Z \setminus Z_0 \Rightarrow (\exists z \in Z_0)(z R z_1); \quad (4.37)$$

при этом отношение *R* называется синонимическим отношением, а термины z_1 и z_2, отвечающие этому отношению, называются синонимическими терминами;

3) имеется транзитивное и несимметричное отношение $K \subset Z_0 \times Z_0$, называемое обобщающим отношением;

4) термины в ТП имеют веса в зависимости от их степени значимости. Вес g_i определяет степень значимости термина z_i, т. е:

$$Z(p) = \{\langle z_1, g_1\rangle, \langle z_2, g_2\rangle, \ldots, \langle z_k, g_k\rangle\}. \quad (4.38)$$

При этом сумма всех весов *(i: 1,...,N)* ТП равна *1*, т.е:

$$\sum_{i=1}^{N} g_i = 1. \quad (4.39)$$

На рисунке 4.2 приведен пример графической интерпретации фрагмента ТП «Экономический кризис» [119].

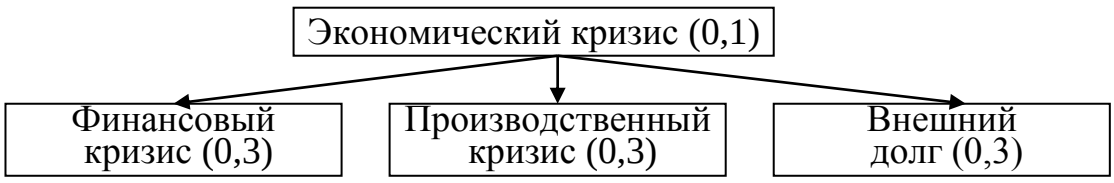


Рисунок 4.2. – Фрагмент терминологического портрета «Экономический кризис»

Данный пример показывает, что в рамках рассматриваемого ТП, термины «Финансовый кризис», «Производственный кризис» и «Внешний долг» по степени значимости – равнозначны и имеют веса равные 0,3. Однако, каждый из них, по степени значимости, выше обобщающего термина «Экономический кризис», который имеет вес равный 0,1. Формальная запись данного ТП представляется следующим образом: Экономический кризис (0.1): финансовый кризис (0.3) $\wedge$ производственный кризис (0.3) $\wedge$ внешний долг (0.3) [119].

Определение 2. Подсистемой весовой обработки информации на заданном ТП будем называть четверку [119]:

$$(Z, D, M, \delta), \tag{4.40}$$

где Z – ТП $Z_0 \subset T$; D – коллекция документов; M – множество запросов; δ - отображение, $\delta : M \times [0,1] \to 2^{D \times [0,1]}$ сопоставляет каждой паре (запрос, точность подобия) множество пар (документ, мера корреляции).

Остальные определения 3-6 и рассуждения, приведенные в [118, 207] приемлемы для данного случая. Отличие заключается в механизме распределения запросов по подсистемам поиска. Применительно к тезаурусу [207], нагрузка на информационные подсистемы поиска равномерная. Применительно к ТП, нагрузка на информационные подсистемы поиска избирательная, зависящая от весов терминов [119]. В данном случае, ответ подсистемы обработки информации Q на запрос $\overline{m} = z_i \cdot g_i$ с точностью $\rho = c$, будет определяться в соответствии с выражением (3.41) [119]:

$$\begin{aligned} &Q = \sum_{i=1}^{n} \delta_i (z_i \cdot g_i, c) = \\ &\sum_{i=1}^{n} (\{(d,\alpha) : d \in D \wedge z_i \cdot g_i \nabla_c \mathrm{t(d)} \wedge \alpha = \mu(z_i \cdot g_i, \mathrm{t(d)})\}) \subseteq D \times [0,1] \end{aligned}, \tag{4.41}$$

где ∇_c - отношение правдоподобия. При определении ответа $N = \delta(\overline{m}, c)$, отношение подобия ∇_c осуществляет выбор документов, точность подобия которых не менее с. Мера корреляции $\mu(\overline{m}, t(d))$ показывает, какая часть информации в документе соответствует ответу на вопрос. Если документ $d \in D$ с мерой корреляции α включен в ответ, т.е. $(d, \alpha) \in \delta(\overline{m}, c)$, то верно неравенство $\alpha \geq c$.

Один из возможных алгоритмов реализации механизма обработки терминологического портрета заключается в следующем. На начальном этапе информационные подсистемы отыскивают и обобщают информацию до определенного порогового значения $\alpha \geq c$, для термина с максимальным весом. Затем для более меньшего веса, и т. д, вплоть до термина с наименьшим весом.

Разработанная математическая модель подсистемы весовой обработки информации на заданном терминологическом портрете (4.41), позволяет при формировании ответов на запросы учитывать неопределенности описания запросов и документов, а также перераспределять нагрузку на локальные информационные подсистемы в зависимости от значимости обслуживаемых терминов [119].

Рассмотрим синтез структуры распределенной информационной системы.

4.4. Синтез структуры распределенной информационной системы

Пусть $Sj = (Tj, Dj, Mj, \delta j)$, $j = \overline{1,\nu}$ - локальные ИС, предназначенные для обеспечения функционирования сложной управляющей системы, состоящей из совокупности $K = \{1, \ldots, \varphi\}$ объектов. Информация с каждой локальной ИС поступает в центр сбора информации для дальнейшей передачи ее управляющему органу. Одна локальная ИС может обслуживать несколько объектов [39].

Распределенная ИС $S = (T, D, M, \delta)$ определяется через локальные составляющие [39]:

$$1)\ T = \bigcup_j T_j;\ R_j = R \cap (T_j \times T_j);\ K_j = K \cap (T_{0j} \times T_{0j});\qquad (4.42)$$

$$2)\ D = \bigcup_j D_j;\qquad (4.43)$$

$$3)\ (K \supset \bigcup_j K_j) \wedge (\underset{\sim}{<}_j = \underset{\sim}{<}_j \cap (M_j \times M_j));\qquad (4.44)$$

$$4)\ \forall m \in M_j \qquad \delta(m) = \{d : d \in D \ \wedge m \underset{\sim}{<} t(d)\}.\qquad (4.45)$$

Распределенная ИС выступает в роли центра сбора и обработки информации, т. е. обработки ответов на запросы управляющих органов [39].

Для формализации задачи введем обозначения: c_{jl} - стоимость сбора

информации об *l*-м объекте *j*-й локальной ИС; b_{jl} - стоимость передачи единицы информации об *1*-м объекте в центр из *j*-й локальной ИС; x_{jl} — булева переменная, равна *1*, если *l*-й объект обслуживается *j* –й локальной ИС, и равна *0* в противном случае [39].

Описание состояния каждого объекта представляется в виде своей информационной модели [39]:

$$S_j^l = (T_j^l, D_j^l, M_j^l, \delta_j^l), l \in K,\ j \in \{1,\ldots,\nu\}, \tag{4.46}$$

где T_j^l - тезаурус с дескрипторным множеством T_{0j}^l, описывающий состояние *l*-го объекта; D_j^l - коллекция возможных документов, которые требуются органам управления для принятия решения; M_j^l - множество допустимых запросов со стороны органов управления; $\delta_j^l : M_j^l \to 2^{D_j^l}$ отображение, сопоставляющее каждому вопросу множество документов.

Индекс *j* указывает, что информационная модель *1*-го объекта сформирована в *j* -й локальной ИС [39].

Если локальная ИС обслуживает несколько объектов, то она формально является распределенной системой "точечного" типа.

Информация об объектах представляется независимыми информационными моделями и вместе с тем, она сосредоточена в одном месте, в одной "точке". Таким образом, информационные модели объектов являются подсистемами глобальной информационной модели. Это дает возможность формулировать ответы на запросы в виде [39]:

$$\delta_j(m) = \left\{ d : d \in D_j \wedge m \underset{\sim}{<}_j t(d) \right\}, \tag{4.47}$$

где $m \in M_j$.

Правильность ответа гарантируется свойствами распределенной системы, представленной в п. 4.3.

Для запросов сложного типа, включающих в себя несколько дескрипторов, выражение (4.47) имеет вид:

$$\delta_j(m) = \bigcap_{i=1}^{k} \bigcup_{l=1}^{p} \delta_j^l(\overline{m}) = \bigcap_{i=1}^{k} \bigcup_{l=1}^{p} \left\{ d : d \in D_j^l \wedge m_j \underset{\sim}{<}_j^l t(d) \right\}, \quad (4.48)$$

где $\overline{m} = \{m_1, \ldots m_k\}$, $\overline{m}_i = \{m_i\}$; число объектов, обслуживаемых любой системой.

Объем передаваемой информации на запрос $\overline{m}$ из j-й локальной информационной системы равен [39]:

$$\alpha_{\overline{m}} = F(\delta_j(\overline{m})), \quad (4.49)$$

где F – оператор преобразования информации к виду, предназначенному для передачи в каналы связи.

Сформулируем задачу распределения объектов по локальным информационным подсистемам при множестве допустимых запросов в распределенной системе $M = \{m_1, \ldots m_r\}$, на которые ответы формируются последовательно без повторения запросов. Для удобства положим, что каждый запрос описывается одним дескриптором [39].

Требуется найти [39]:

$$\min_{\{x_{jl}\}} \sum_j \sum_l c_{jl} x_{jl} + \sum_i \sum_j \sum_l b_{jl} F(\delta_j^l(m_i)) x_{jl}, \quad (4.50)$$

при ограничениях:

$$\sum_j x_{jl} = 1, \forall l \in K; \quad (4.51)$$

$$\sum_l x_{jl} \geq (\leq) N_j, \forall j \in \{1, \ldots \nu\}. \quad (4.52)$$

Ограничение (4.51) требует обслуживания каждого объекта только одной информационной системой. Условие (4.52) ограничивает количество объектов, подлежащих обслуживанию локальными подсистемами, либо, напротив, требует, чтобы их было не меньше заданного числа [39].

Усложним требования к распределенной системе. Потребуем, чтобы органы управления получали ответы на запросы даже в случае, если любая локальная информационная система перестанет функционировать [39].

Иначе говоря, возникает потребность в дублировании информации об объектах [39].

Постановка задачи меняется только в части, касающейся изменения ограничения (4.51). Оно принимает вид [39]:

$$\sum_j x_{jl} = 2, \forall l \in K \qquad (4.53).$$

Модифицируем еще раз постановку задачи. Будем считать, что после закрепления объектов за локальными ИС в задаче (4.50)-(4.51) перераспределения объектов не происходит.

Однако органам управления необходимо иметь информацию о предыдущих состояниях объектов даже после прекращения функционирования любой из локальных подсистем. Это означает, что происходит дублирование информационных моделей об объектах постоянно в процессе их функционирования и информация локальных ИС перераспределяется между ними [39].

Упростим ситуацию, полагая, что перераспределение информации осуществляется один раз. Хотя в реальной ситуации информация о состоянии объектов передается по мере ее поступления на всем интервале времени функционирования локальных ИС, указанное ограничение не снижает общих рассуждений, так как полная постановка задачи потребует просто дополнительного суммирования по дискретным моментам времени [39].

Введем следующие обозначения: a_{js} - стоимость передачи единицы информации из *j*-й локальной ИС в *s*-ю; y_{jsl} - булева переменная, равна *1*, если информация об *l*-м объекте перераспределяется из *j*-й локальной ИС в *s*-ю [39].

Окончательно задача синтеза структуры распределенной ИС формулируется так: необходимо найти [39]:

$$\min_{\{x_{jl}, y_{isl}\}} \sum_j \sum_l c_{jl} x_{jl} + \sum_i \sum_j \sum_s b_{jl} F(\delta_j^l(m_i)) x_{jl} + \sum_j \sum_s \sum_l a_{js} [F(T_j^l, D_j^l, M_j^l, \delta_j^l) x_{jl}] y_{jsl}, \qquad (4.54)$$

при ограничениях:

$$\sum_j x_{jl} = 1, \forall l \in K; \qquad (4.55)$$

$$\sum_l x_{jl} \geq (\leq) N_j, \ \forall j \in \{1, \ldots \nu\} \qquad (4.56)$$

$$\sum_s y_{jsl} = 1, \forall l \in K_j = \{l : x_{jl} = 1\}, \forall j \in \{1, \ldots \nu\}. \qquad (4.57)$$

Ограничение (4.57) требует перераспределения информации об *l*-м объекте из *j*-й локальной ИС [39].

На основе задачи (4.54)-(4.57) можно сформулировать ряд задач, учитывающих те или иные требования относительно структуры распределенной ИС. Для этого необходимо сформулировать требования в виде ограничений и ввести их в описание задачи [39].

Разработанная модель учитывает распределение информации по дескрипторам, описывающим документ. Это позволяет органам управления оценивать неопределенность ответов на запросы.

Пусть документы в локальных ИС представляются в виде:

$$t(d_j^l) = \{< t_1^{l_j}, w_1 >, < t_2^{l_j}, w_2 >, \ldots, < t_k^{l_j}, w_k >\}, \qquad (4.58)$$

$$d_j^l \in D_j^l, \quad S_j^l = (T_j^l, D_j^l, M_j^l, \delta_j^l).$$

В реальных ситуациях при описании состояния объектов, разные локальные ИС, обеспеченные различными техническими средствами измерения и наблюдения за состоянием объектов, представляют не совпадающие между собой описания одних и тех же объектов. В формализованном виде это означает, что при запросе $\overline{m}$ органов управления о состоянии некоторого объекта *l* локальные ИС дадут ответы с точностью подобия ρ [39]:

$$\delta_j^l(\overline{m}, \rho) = \left\{(d_j^l, \alpha_j^l) : d_j^l \in D_j \wedge \overline{m} \nabla_\rho \mathrm{t}(\mathrm{d}_j^l) \wedge \alpha_j^l = \mu(\overline{m}, \mathrm{t}(\mathrm{d}_j^l))\right\}, \forall l \in K, \quad j = \overline{1, \nu}. \qquad (4.59)$$

Очевидно, что показатели α_j^l - могут быть использованы для уточнения структуры распределенной ИС, т. е. выбора такой структуры, которая бы обеспечивала получение ответов с максимальной мерой корреляции или с мерой корреляции не менее заданной [39]. Последнее требование эквивалентно условию передачи информации с отвечающей запросу содержательной частью не менее заданной. Формализация этого условия применительно к постановке задачи (4.54)-(4.57) состоит во введении дополнительных ограничений [39]:

$$\alpha_j^l = \mu(\overline{m}, \mathrm{t}(\mathrm{d}_j^l)) \geq \alpha_j^+, \forall l \in K, \quad \forall j : x_{jl} = 1. \qquad (4.60)$$

В интересах практической реализации терминологического поиска на основе разработанных моделей и синтезированной структуры распределен-

ной ИС, необходимо рассмотреть технологические аспекты решаемой задачи. Первый из них состоит в разработке алгоритма идентификации текстовой информации заданному ТП.

4.5. Алгоритм идентификации текстов

ТП может быть построен как для отдельного текста терминологический портрет текста (ТПТ), так и для некоторой предметной области. Владельцем терминологического портрета предметной области может быть как ЛПР, так и ИСППИД СЭО. Считая ЛПР и ИСППИД единым целым будем называть последний терминологическим портретом информационной системы (ТПИС). ТПИС содержится в специальном разделе витрины данных и является своеобразным фильтром, через который «пропускается» множество тематических текстов, поступающих из различных источников информации. Блок-схема алгоритма идентификации некоторого текста приведена на рисунке 4.3.

Блоки 1, 11 реализуют начало и окончание процесса идентификации текста.

Блок 2 обеспечивает ввод исходного текста.

В блоке 3 осуществляется формализация текста на основе соответсвующей модели - многоуровневой иерархической схемы текстовых данных (см. рис. 4.1) [44], реализованной в ряде ИС [80, 84, 94, 116, 170]. В ней используется 5 уровней иерархии: дискурс (связный текст); предложение; словосочетание; слово; морфема. Уровни с 6 по 8 предназначены для звукового анализа и в данной работе не рассматриваются.

В блоке 4 реализовано выделение терминов. Для выделения терминов и построения терминологического портрета текста используется специальный иерархический терминопостроитель, содержащий соответствующую систему моделей и алгоритмов терминологического информационного поиска. Его детализация приведена в п. 4.6.

Блок 5 используется для преобразования выделенных терминов к каноническому виду с целью определения их морфологических форм, таких

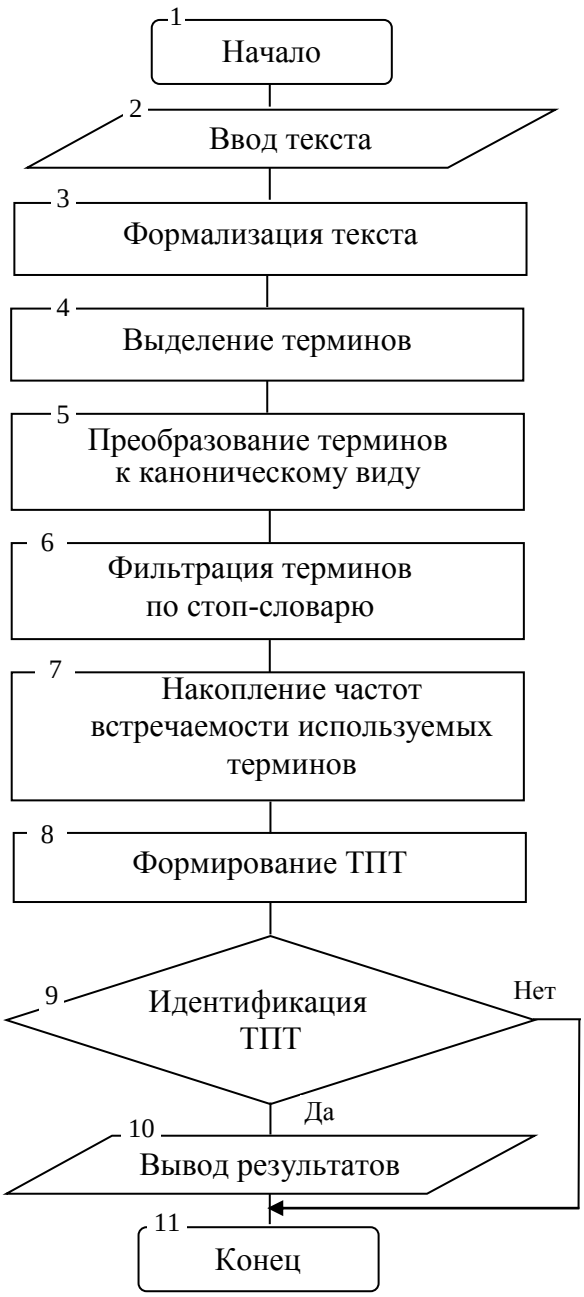


Рисунок 4.3. – Блок-схема алгоритма идентификации текста

как: существительное, местоимение, форма отрицания, нестандартная форма, различные характеристики прилагательного и причастия (сравнительная степень, краткая грамматическая форма, сложная грамматическая форма с дефисом), различные характеристики деепричастия (от глагола совершенного вида, от глагола несовершенного вида, повелительное наклонение), различные характеристики глагола (вид, время, возвратная форма и др.) и др. [73].

Блок 6 обеспечивает фильтрацию выделенных терминов по стоп – словарю для исключения из рассмотрения заведомо не перспективных

терминов (предлоги, частицы, союзы, местоимения и др).

В блоке 7 реализовано накопление и анализ частот встречаемости выявленных терминов с целью определения их весов.

Блок 8 формирует ТПТ, представляющий собой вектор весов информационных признаков $(v_1,\ldots,v_к)$, где v_i – вес i-го признака $i=1,\ldots,k$ в тексте, k –число признаков. Вес i-го информационного признака соответствует частоте встречаемости терминов из которых он состоит и определяется в соответствии с выражением (4.61) [73]:

$$v_i = \frac{n_i}{N}, \tag{4.61}$$

где n_i - число терминов i-го признака в данном тексте; N - общее количество терминов.

В блоке 9 реализована идентификация ТПТ. Для принятия решения о соответствии данного текста исследуемой предметной области - ТПИС, рассчитывается значение косинуса угла между весовыми векторами ТПТ и ТПИС. Порядок проведения расчетов следующий.

Рассмотрим в n – мерном пространстве два произвольных вектора AB и CD с координатами a_i, b_i, c_i, d_i:

$$\vec{AB} = \{b_1 - a_1, b_2 - a_2, \cdots, b_n - a_n\} \tag{4.62}$$

$$\vec{CD} = \{d_1 - c_1, d_2 - c_2, \cdots, d_n - c_n\}. \tag{4.63}$$

Координаты a_i, b_i, c_i, d_i являются координатами случайных величин A = $\{a_1, a_2, \cdots, a_n\}$, B = $\{b_1, b_2, \cdots, b_n\}$, C = $\{c_1, c_2, \cdots, c_n\}$ и D = $\{d_1, d_2, \cdots, d_n\}$. На всей области определения $[-\infty, \ +\infty]$ они имеют нормальное распределение.

Введем систему прямоугольных координат. Тогда случайные величины X и Y равные $A – B$ и $C – D$ соответственно, будут представлять собой те же вектора, которые проведены из начала координат [73]. Формально это записывается следующим образом:

$$x_i = b_i - a_i$$

$$y_i = d_i - c_i.$$

Известно [15, 224], что $\hat{r}_{X,Y}$ - выборочный коэффициент корреляции двух независимых случайных величин X и Y имеет вид (4.64):

$$\hat{r}_{X,Y} = \frac{\hat{\mathrm{cov}}\{\mathrm{X},\mathrm{Y}\}}{\hat{\sigma}_{\mathrm{X}} \cdot \hat{\sigma}_{\mathrm{Y}}} = \frac{\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{x_i} - \hat{\mathrm{m}}_{\mathrm{X}})(\mathrm{y_i} - \hat{\mathrm{m}}_{\mathrm{Y}})}{\frac{1}{\mathrm{n}}\sqrt{\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{x_i} - \hat{\mathrm{m}}_{\mathrm{X}})^2 \sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{y_i} - \hat{\mathrm{m}}_{\mathrm{Y}})^2}}, \quad (4.64)$$

где $\hat{\sigma}_X$ и $\hat{\sigma}_Y$ - оценки средних квадратических отклонений величин X и Y соответственно; $\hat{\mathrm{cov}}\{X,Y\}$ - оценка второго смешанного центрального момента случайной величины (X,Y) также называемого корреляционным моментом; $\hat{m}_x$ и $\hat{m}_y$ - оценки математических ожиданий величин X и Y [73].

Выражение (4.64) может быть записано следующим образом [58]:

$$\hat{r}_{X,Y} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{m}_X)(y_i - \hat{m}_Y)}{\sqrt{\sum_{i=1}^{n}(x_i - \hat{m}_X)^2 \sum_{i=1}^{n}(y_i - \hat{m}_Y)^2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(b_i - a_i - \hat{m}_{B-A})(d_i - c_i - \hat{m}_{D-C})}{\sqrt{\sum_{i=1}^{n}(b_i - a_i - \hat{m}_{B-A})^2 \sum_{i=1}^{n}(d_i - c_i - \hat{m}_{D-C})^2}} =$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n}(b_i - a_i - (\hat{m}_B - \hat{m}_A))(d_i - c_i - (\hat{m}_D - \hat{m}_C))}{\sqrt{\sum_{i=1}^{n}(b_i - a_i - (\hat{m}_B - \hat{m}_A))^2 \sum_{i=1}^{n}(d_i - c_i - (\hat{m}_D - \hat{m}_C))^2}}. \quad (4.65)$$

Математические ожидания $\hat{m}_A, \hat{m}_B, \hat{m}_C, \hat{m}_D$ равны нулю [58], поскольку величины A, B, C и D имеют нормальное распределение. Тогда (4.65) примет вид:

$$\hat{r}_{X,Y} = \frac{\frac{1}{n}\sum_{i=1}^{n}(b_i - a_i)(d_i - c_i)}{\sqrt{\sum_{i=1}^{n}(b_i - a_i)^2 \sum_{i=1}^{n}(d_i - c_i)^2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}x_i y_i}{\sqrt{\sum_{i=1}^{n}x_i^{\,2} \sum_{i=1}^{n}y_i^{\,2}}}. \quad (4.66)$$

Выражение (4.66) формально отражает тот факт, что коэффициент корреляции двух случайных векторов с координатами, нормально распределенными в пространстве, определяется, как значение величины косинуса угла между ними [73].

В случае не превышения рассчитанного значения коэффициента корреляции заданного порогового уровня, управление передается блоку 11. Иначе передача управления осуществляется в блок 10 [73].

Блок 10 выводит результат о принадлежности текста ТПИС.

4.6 Характеристика иерархического терминопостроителя

Иерархический текстовый терминопостроитель (ИТТ) предназначен для формирования ТП исследуемого текста. Он базируется на системе моделей и алгоритмов формализации и проведения различных видов анализа текста (морфологического, синтаксического и семантического). Впервые подобная система моделей и алгоритмов была обоснована в [73, 142] и прошла практическую апробацию [83, 84] в интересах рубрицирования текстов. В уточненном виде, применительно к данной работе, система моделей и алгоритмов ПТП приведена на рисунке 4.4.

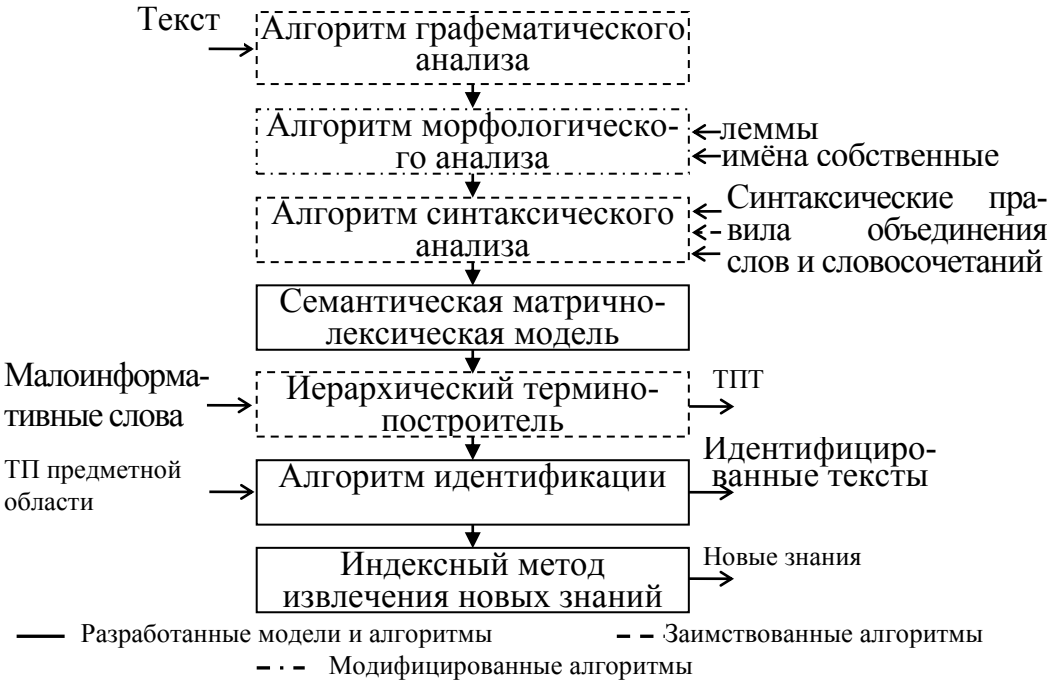


Рисунок 4.4. – Система моделей и алгоритмов ПТП

Как видно из схемы, на начальном этапе исходная текстовая информация проходит графематическую обработку, реализуемую соответствующим алгоритмом. В рамках данной обработки, текст разделяется на отдельные самостоятельные фразы [73].

В дальнейшем, каждое из предложений подвергается морфологическому анализу, который реализуется соответствующим алгоритмом. На его основе определяется морфология слов (суффикс, приставка, корень и др.) [73].

Алгоритм синтаксического анализа обеспечивает разбор предложений по частям речи [73].

Семантическая матрично-лексическая модель (СМЛМ) обеспечивает представление текста в виде сети с узлами – понятиями и смысловыми отношениями между ними.

Иерархический терминопостроитель формирует терминологический портрет текста (ТПТ) на основе анализа частот встречаемости понятий [159] и использования шинглов [137]. В [116] приведены результаты его использования для аннотирования текстов.

Алгоритм идентификации обеспечивает селекцию текстов в соответствии с заданным (эталонным) терминологическим портретом.

Индексный метод извлечения новых знаний (ИМИНЗ) основан на определении логических цепочек взаимосвязей терминов в массиве текстовых документов и в случае их наличия, извлечении новых знаний (дополнительной терминологической информации) в виде совокупности новых терминов.

4.7. Семантическая матрично-лексическая модель анализа текста

Целевое назначение данной модели заключается в выделении терминов заданного текста на основании смысла. Ее основу составляют лексемы.

Под лексемой (лексической единицей) будем понимать элементарную семантически значимую единицу языка, представляющую собой слово, устойчивое словосочетание или другую языковую конструкцию, способную обозначать предметы, явления, их признаки и др. [252]. Она имеет свойства.

Свойство 1. Равенство двух лексем возможно в том случае, когда число символов, их составляющих равно и имеет место позиционное совпадение символов [73].

Свойство 2. Число повторений лексемы в различных фразах текста представляет собой частоту лексемы. Уникальная лексема представляет собой лексему, у которой частотой встречаемости является единичной [73].

Свойство 3. Повторяющиеся лексемы составляют лексическое множество связей текста.

Свойство 3 показывает, что для текста может быть построена своя матри-

ца лексических связей (МЛС). Данная матрица строится следующим образом.

Пусть имеется текст, в котором присутствует множество лексем $\{L\} \equiv \{l_1,\ldots,l_i,\ldots,l_{Nk}\}$. Они упорядочены последовательностью появления в тексте. Общее число лексем связи в тексте составляет N_k. Выявлены частоты встречаемости лексем $F \equiv \{f_1,\ldots,f_i,\ldots,f_{N_k}\}$. При этом $f_i > 1$ при любом $i \le N_k$, поскольку уникальные лексемы удалены [73].

Предложениям текста P_n, $n \in \{1,\ldots,J_p\}$ соответствует множество входящих в них лексем $l_i \in M$ [73].

Введем двоичный параметр q_i^n, определяемый в соответствии с выражением:

$$q_i^n = \begin{cases} 1, \text{ если } l_i \in \Pi^n \\ 0, \text{если } l_i \notin \Pi^n \end{cases} \qquad (4.67)$$

Тогда P_n соответствует вектор: $pn = (q_1^n,\ldots,q_i^n,\ldots,q_{Jm}^n)$, а тексту МЛС:

$$(q_i^n) = \begin{pmatrix} & & & & \\ q_1^1 & \cdots & q_i^1 & \cdots & q_{N_i}^1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ q_1^n & \cdots & q_i^n & \cdots & q_{N_i}^n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ N & & N & & N \\ q_1^p & \cdots & q_i^p & \cdots & q_{N_i}^p \end{pmatrix}. \qquad (4.68)$$

МЛС является одним из инструментов представления текста в виде ориентированного графа, позволяющего проводить его дальнейший анализ. Он строится следующим образом.

Имеется текст $T_K(R,D)$ представленный в виде множества вершин-предложений R, и множества дуг, порождаемых лексическим множеством связей – D.

Пусть лексема m_i с частотой f_i представляет число дуг $D_{f_i}^2$, где D_b^a - число сочетаний из b по a.

Вершины r^m и r^b лексемы m_i из *M* связаны семантической (смысловой) дугой $D_i^{m,b}$, если $\mu_i^m = \mu_i^b = 1$.

Кроме того узлы R^m&R^b являются смежными. Другими словами дуга $D_i^{m,b}$ является инцидентной для вершин (r^m), (r^b). Она берет начало в r^m и заканчивается в r^b.

Количество дуг, являющихся инцидентными *i*-ой вершине, равно сумме элементов на *i*-ой строке в МЛС [73].

Построение для текста $T_K(R,D)$ матриц смежности и инцидентности представляется следующим образом.

Матрица *P(G)* с размерами $a \times b$ (где *a* - количество вершин, *b* - количество дуг графа *T*), а *(i,j)*-й элемент равен *1* в том случае, когда вершина o_i в графе *T* инцидентна ребру l_j, или в том случае, когда o_i является началом дуги l_j, и равен *-1*, в том случае, когда вершина o_i является концом дуги l_j (применительно к орграфам), и равен *0* для остальных случаев, соответствует матрице инцидентности *(0,1)* [73].

Матрица *C(T)* размером $a \times a$ (*a* - число вершин в *T*), *(i,j)*-й элемент c_{ij} которой равен *1*, если вершины o_i и o_j смежны, т.е. соединены дугой (или ребром) (o_i, o_j), и равен *0* в противном случае, соответствует матрице смежности - *(0,1)*.

Процедура формирования данных матриц из МЛС заключается в следующем. Построим следующую матрицу:

$$c(m,b) = \sum_{i=1}^{A_m} \mu_i^m * \mu_i^b, \qquad (4.69)$$

где *c(m,b)* - элемент матрицы смежности размерности $A_r \times D_g$, равный числу дуг, идущих из вершины r^m в вершину r^b.

Введем следующие обозначения: *j* - номер дуги, *j+(1,...,D_g)*; *m* и *b* - номера вершин. Смысл записи $D_i^{m,b}$ заключается в утверждении существо-

вания m_i-ой лексемы, порождающей j-ую дугу, инцидентную r^m, r^b вершинам. При этом элемент $p(m,b)$ матрицы инцидентности размерности $A_r \times D_g$ будет иметь вид:

$$p(m,j) = \begin{cases} 1, & \text{если } D_j^{m,b} \\ -1, & \text{если } D_j^{b,m} \\ 0 & \text{в других случаях} \end{cases}, \qquad (4.70)$$

где $D_g = \sum_{i=1}^{A_m} \frac{i!}{2*(f_b - 2)}$.

В том случае, если последовательность вершин от r^m до r^b частично упорядочена таким образом, что последующая вершина смежна с предыдущей, то имеет место путь из r^m в r^b, представляющий собой совокупность соответствующих дуг. При этом длина пути, есть число дуг в последовательности. Длина L между множествами лексем, входящими в r^m и r^b предложения есть путь между двумя вершинами. Путь от предложения, порождающего аксиоматическую лексему до предложения, порождающего произвольную терминальную лексему, есть полная ассоциация m_i-ой аксиоматической лексемы. Фактически локализованная тема - m_i, есть объединение всех ее полных ассоциаций без повторов.

Семантическое расстояние между двумя лексемами в пределах заданного текста (графа T_K) определяется как минимум из длин всех путей, соединяющих предложения областей существования этих двух лексем [73].

Важной характеристикой ориентированного графа текста является ассоциативная мощность. Она позволяет определить нагрузку дуги (ее длину) и, что очень важно, оценить важность ассоциации.

Ассоциативное графовое представление текста заключается в виде последовательности предложений-вершин, а фактически лексем, связанных между собой смыслом. В процессе определения наиболее важных понятий, содержащихся в тексте, необходимо соответствующее ассоциативное множе-

ство слов, привести к нормальной форме [73]. Для определения ассоциативного множества слов, целесообразно использовать меры связности графа с аксиоматическими лексемами [117].

На практике, наряду с представлением текста в виде графа, используют семантическую сеть. Она позволяет представлять смысл текста в более естественной форме – в виде совокупности связанных между собой понятий (слов и словосочетаний), несущих основную смысловую нагрузку и наиболее часто, встречающихся в тексте. Она, как и граф, состоит из вершин и рёбер [73].

Имена, приписываемые вершинам и рёбрам семантической сети, совпадают с именами соответствующих сущностей и отношений используемых в естественном языке. Ребро и связываемые им вершины образуют подграф семантической сети, несущий минимальную информацию - факт наличия связи определённого типа между соответствующими объектами [73].

Пример фрагмента семантической сети, описывающей кризисные состояния (предкризисное состояние, кризис, рецессия) внешней среды в сфере экономики, представлен на рисунке 4.5.

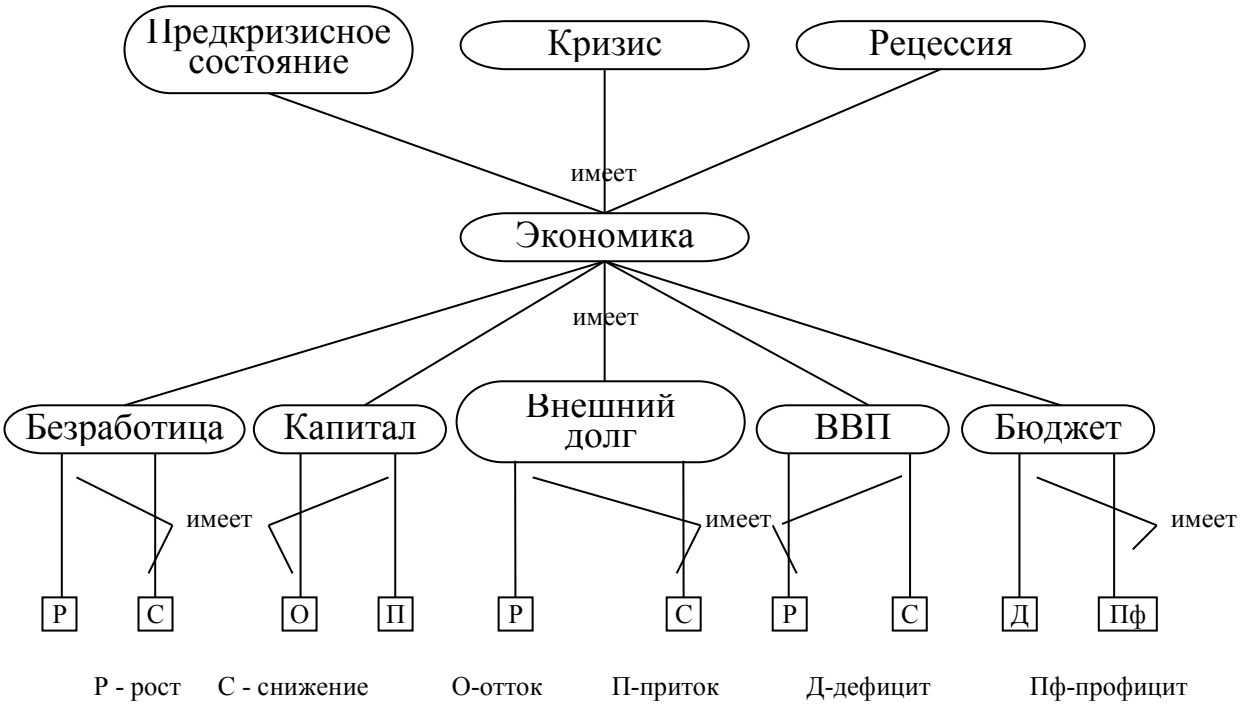


Рисунок 4.5. – Фрагмент семантической сети

Пример, представленный на рисунке 4.5, может быть проработан на любую глубину. При этом слова, используемые при обозначении объектов в семантической сети, могут быть продублированы в других названиях сколь угодно раз [73].

Воспользовавшись фрагментом вышеприведенной семантической сети, можно построить структуру «экономические показатели», которая в схематичном виде приведена на рисунке 4.6.

СОЗДАТЬ-СТРУКТУРУ (ИМЯ СТРУКТУРЫ = ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ
ЧИСЛО АТРИБУТОВ = 5
АТРИБУТ = БЕЗРАБОТИЦА
АТРИБУТ = КАПИТАЛ
АТРИБУТ = ВНЕШНИЙ ДОЛГ
АТРИБУТ = ВВП
АТРИБУТ = БЮДЖЕТ).

Рисунок 4.6. – Пример структуры «Экономические показатели»

Для реализации данной структуры целесообразно использовать аппарат фреймов [214]. Известно, что наибольшее распространение находят три типа фреймов: фреймы-структуры, фреймы-роли, фреймы-сценарии [214]. В [73] для построения модели семантического анализа русскоязычных текстов использовались фреймы-сценарии. В данной работе использованы фрейм-структуры в силу трех причин. Во–первых, они в максимальной степени соответствуют представлению информации по своему принципу построения. При заполнении их элементов-слотов определёнными значениями, фрейм-структуры превращается в описания конкретных фактов, событий, процессов [73, 82, 256]. Во-вторых, фрейм-структуры обладают объектно-ориентированными свойствами инкапсуляции, наследования и полиморфизмом объектов [214], что гармонично сочетается с общей технологией разработки СППИР, в частности, использованием объектно-ориентированной среды разработки Delphi 7.0. В-третьих, табличная форма представления фрейм-структур оптимально вписывается в общую концепцию построения БД СППИР (как в рамках хранилища данных, так и витрины данных).

Недостатками фреймового подхода к проведению семантического анализа текста, являются более жёсткое, чем при подходе, основанном на семан-

тической сети, выделение объектов, ситуаций и их свойств [73], а также возрастание сложности фреймовой модели (появление разнотипных вложенных фреймов) при увеличении в тексте числа иерархически взаимосвязанных разноплановых понятий, терминов, фактов, событий, процессов и др. [225]. Данные недостатки приводят к существенным трудностям формирования терминологических портретов.

Для парирования данных недостатков использовалась гибридная модель семантического анализа, содержащая как семантическую сеть, так и фреймы-структуры. При этом составляющие данной модели использовались избирательно. Так, в интересах семантического анализа газетных статей приоритет в использовании отдавался фреймам.

Поскольку газетные статьи являются одним из источников снижения неопределенности внешней среды для ЛПР, исследование последних актуально. Проведенные исследования показали, что для их семантического анализа целесообразно применять фреймы. Поскольку первейшей заботой журналиста является быстрая и не загружающая мозг передача информации, то, как правило, все характеристики, требующие применения медленных индуктивных или дедуктивных процессов [73], для семантического анализа которых необходимы семантические сети, не используются [225]. В данной работе, для обработки газетных подборок использовалась следующая технология. Для обработки заголовка статьи и одного - двух первых предложений вызывался центральный фрейм. Его ячейки заполнялись понятиями, содержащимися в выделенных лексемах. В зависимости от их значимости, некоторые понятия становились термами новых фреймов, ячейки которых заполнялись понятиями из лексем других фрагментов статьи. В процессе заполнения фреймовой структуры, трудности вызывали сложные местоимения. В интересах их устранения, часть сложных местоимений ограничивалась в использовании на основе стоп-словаря [67].

Анализ других источников информации (интернет, электронные журналы и др.), содержащих описание кризисного состояния внешней среды, показал целесообразность использования фреймов в совокупности с семантической сетью. Как оказалось, ещё одним важным достоинством использования

фреймов, явилась возможность с их помощью реализовать ответы на вопросы. В рамках данной работы развитие этого направления не рассматривалось.

Применительно к описанию кризисного состояния внешней среды, использование гибридной модели синтаксического анализа позволило создать базу знаний (БЗ) в виде набора терминологических портретов, характеризующих различные понятия данной предметной области. На основе созданной БЗ, объекты данной предметной области, идентифицировались и накапливались в соответствующей базе данных.

Ниже приведены несколько правил из БЗ:

1. ЕСЛИ ОБЪЕКТ. АТРИБУТ = “ЛИКВИДНОСТЬ“ И
ОБЪЕКТ. ЛИКВИДНОСТЬ = “ОТСУТСТВУЕТ”, ТО
ОБЪЕКТ. ИМЯ = “БАНКРОТ”, И
ОБЪЕКТ. АТРИБУТ = “ПОТЕНЦИАЛЬНЫЙ”
2. ЕСЛИ ОБЪЕКТ. АТРИБУТ = “ВВП“ И
ОБЪЕКТ. ВВП = “НИЗКИЙ” И
ОБЪЕКТ. АТРИБУТ = “ДОЛГ” И
ОБЪЕКТ. ДОЛГ = “ВНЕШНИЙ” И
ОБЪЕКТ. ДОЛГ = “ВЫСОКИЙ”, ТО
ОБЪЕКТ. ИМЯ = “КРИЗИС” И
ОБЪЕКТ. КРИЗИС = “ФИНАНСОВЫЙ”.

Подобные правила позволили добавлять фреймы, атрибуты во фреймы, менять значения атрибутов и др.

Применение разработанной СМЛМ анализа текстов в совокупности с базой знаний, позволили с определенной точностью идентифицировать в текстовой информации определенные понятия, их атрибуты и связи с другими объектами. В результате появилась потенциальная возможность извлечения дополнительных данных (новых знаний), содержащихся в текстах, что обусловило необходимость разработки соответствующей модели.

4.8. Индексный метод извлечения новых знаний

Индексный метод извлечения новых знаний (ИМИНЗ) базируется на определении логических цепочек взаимосвязей терминов в массиве текстовых документов и в случае их наличия, извлечении новых терминов [117].

Под новыми знаниями будем понимать новую дополнительную терминологическую информацию, извлекаемую из текста (текстов).

Схема, поясняющая суть ИМИНЗ, приведена на рисунке 4.7.

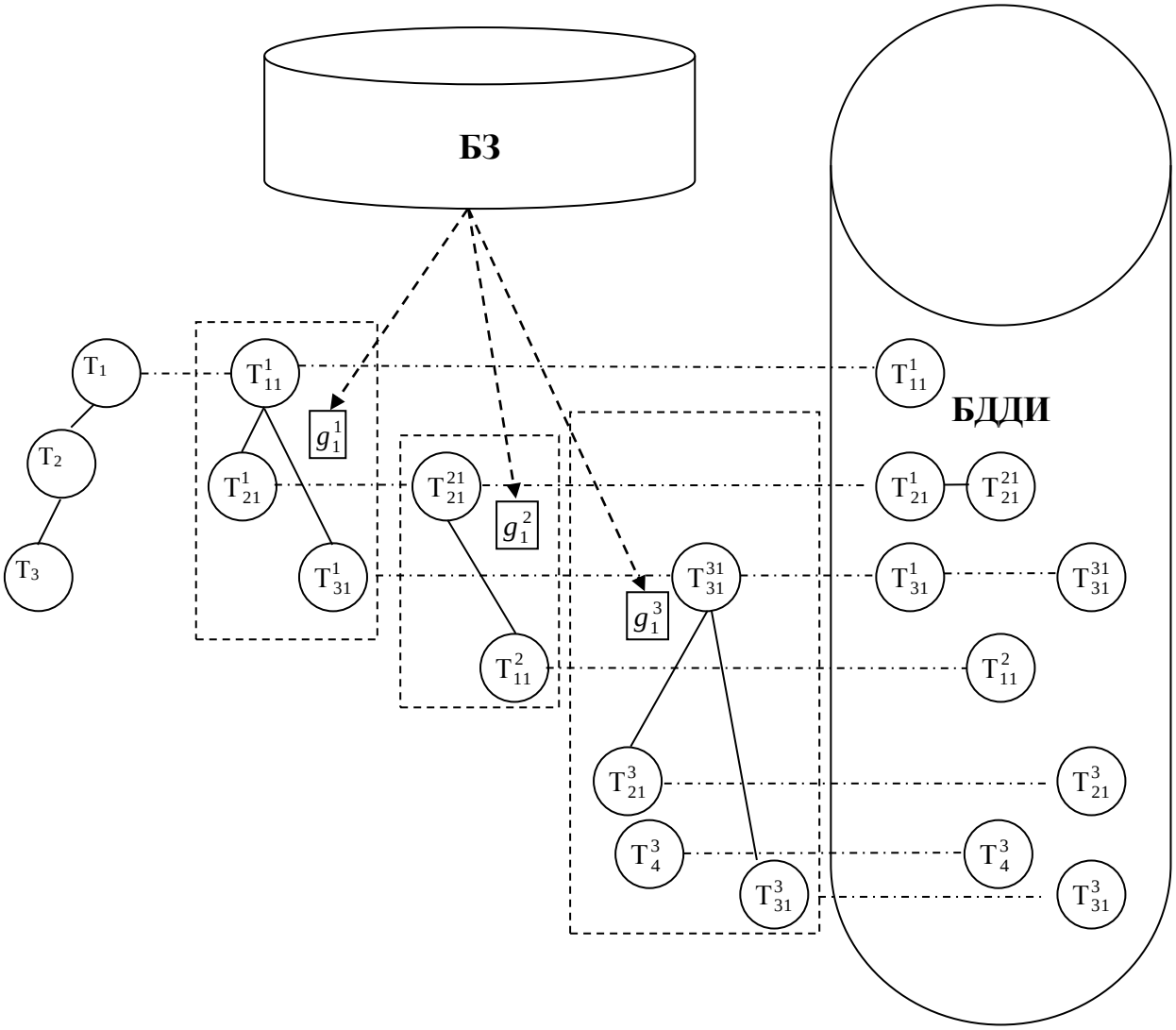


Рисунок 4.7. – Схема, поясняющая механизм ИМИНЗ

На вход модуля, реализующего ИМИНЗ, подается набор терминов *(T_1, T_2, T_3)*, для которых необходимо найти дополнительную информацию. Пользователем может быть сформирован набор правил (критериев) отбора новых терминов, содержащихся в БЗ, в виде пороговых значений частот их встречаемости в текстах (g_1^1, g_1^2, g_1^3). Кроме того, он может задавать синонимы искомых терминов (например, «жаргонизмы»), которые не содержатся в используемом словаре [43].

В результате проведенного семантического анализа установлено, что из множества текстов, принадлежащих заданной предметной области ТПИС, в трех текстах (отмечены прямоугольниками с пунктирными границами)

найдены восемь терминов ($T^1_{11}, T^1_{21}, T^1_{31}, T^{21}_{21}, T^2_{11}, T^{31}_{31}, T^3_{21}, T^3_{31}$), имеющих логические цепочки взаимосвязей (представлены на рисунке прямыми штрихпунктирными линиями) лишь с одним термином (T_1) из трех, поступивших на вход модуля, реализующего ИМИНЗ. Термины внутри цепочек имеют *n, m* – мерную двухуровневую индексацию (*n* – число текстов, связанных с данным термином, *m* – число взаимосвязанных терминов). Верхний *n* – мерный (в приведенной схеме – двойной) индекс указывает на тот факт, что термин текста за номером левого индекса логически связан с термином текста за номером правого индекса. Верхний одинарный индекс указывает на тот факт, что данный термин логически связан с одним из терминов только этого текста. Нижний *m* – мерный (в приведенной схеме – двойной) индекс указывает на тот факт, что термин текста за номером левого индекса логически связан с термином за номером правого индекса. В анализируемых текстах могут оказаться одиночные термины, не связанные с терминами других текстов. Для них используются верхние и нижние одиночные индексы. Все выявленные термины (связанные и не связанные) хранятся в базе данных дополнительной информации (БДДИ). К каждому из них, у пользователя имеется доступ.

Для записи каждого термина в БДДИ имеется три поля: два поля, в которых хранятся верхний и нижний индексы, и текстовое поле в котором хранится определение данного термина. Для хранения дублирующих терминов используется правило: «Если дублирующий термин одиночный (не содержит логических цепочек с другими терминами) и подобный термин уже содержится в БДДИ, то он исключается из БДДИ. В противном случае он записывается в БДДИ» [101].

Хранение верхнего и нижнего индексов терминов позволило программными средствами Delphi и СУБД Access реализовать процедуру восстановления межтекстовых терминологических взаимосвязей (логических цепочек). Данная процедура прошла апробацию [94] и показала себя аналогом подсистемы объяснения, используемой в экспертных системах. Однако для ее реализации не потребовалось применение специального математического аппарата (например, предикатов первого порядка и др.) и соответству-

ющих высокоуровневых языков программирования (Пролог, ЛИСП и др.).

Рассмотрим особенности программной реализации технологии терминологического информационного поиска в ИСППИД СЭО.

4.9. Программная реализация терминологического информационного поиска в информационной системе поддержки портфельной инвестиционной деятельности

В интересах повышения оперативности ТИП в ИСППИД СЭО, была разработана система виртуального распределенного терминологического информационного поиска (СВРТП), в дальнейшем с целью сокращения – система [170]. Она представляет собой пакет программ, предназначенный для распределенного поиска информации о терминах, вводимых пользователем, и извлечения связанных с ними новых дополнительных сведений в системах поддержки принятия решений, используемых в интересах снижения неопределенности внешней среды. Особенностью данной системы является возможность реализации ТИП на виртуальных машинах (на локальной виртуальной машине и в сети виртуальных машин).

Виртуальная машина представляет собой специальную программу, эмулирующую реальный ПК (BIOS, ЖМД, операционную систему и др.) либо их сеть. Она позволяет осуществить обмен данными между локальным ПК, виртуальной машиной и их сетью [35].

Система выступает в роли интеллектуального поискового агента (ИПА).

Интеллектуальность поискового агента обусловлена наличием процедур, реализующих механизм весовой генерации запросов, эквивалентный механизму пользовательских предпочтений. Разработанный ИПА обладает следующими важными свойствами:

- автономностью;
- интерактивностью;
- рациональностью;
- проактивностью.

Автономность характеризует тот факт, что ИПА самостоятельно вы-

полняет значительную часть своей работы. Во время выполнения работы ИПА не взаимодействует с пользователем и с другими агентами.

Интерактивность подразумевает умение ИПА взаимодействовать с пользователем [163]. Суть взаимодействия заключается в настройке моделей и механизмов ИПА пользователем, получении заданий от него и предоставлении ему полученных результатов.

Рациональность заключается в выполнении определенных действий для достижения поставленной цели и отсутствии противоположных действий, препятствующих ее достижению.

Проактивность предполагает наряду со сбором основной информации (по введенному термину) также сбор дополнительной информации.

ИПА реализует функции передаточного механизма между потребностями ЛПР и стандартными поисковыми системами, такими как [169]: Яндекс, Google, Rambler и др., кроме того, ИПА обладает специфическими функциями, такими как:

- построение ТП предметной области пользователя;
- построение ТПТ;
- селекция основной терминологической информации;
- селекция дополнительных терминов;
- выборка информации по дополнительным терминам;
- ведение терминологической БД;
- накопление в БД основной и дополнительной терминологической информации;
- формирование инсталляционных пакетов ИПА для локальных и виртуальных машин.

Структурно-функциональная схема ИПА приведена на рисунке 4.8.

Интерфейс пользователя предназначен для получения заданий от пользователя и предоставления ему полученных результатов. На начальном этапе функционирования системы на основе интерфейса формируется ТП предметной области пользователя путем ввода набора терминов и их распределе-

ния по различным уровням иерархии.

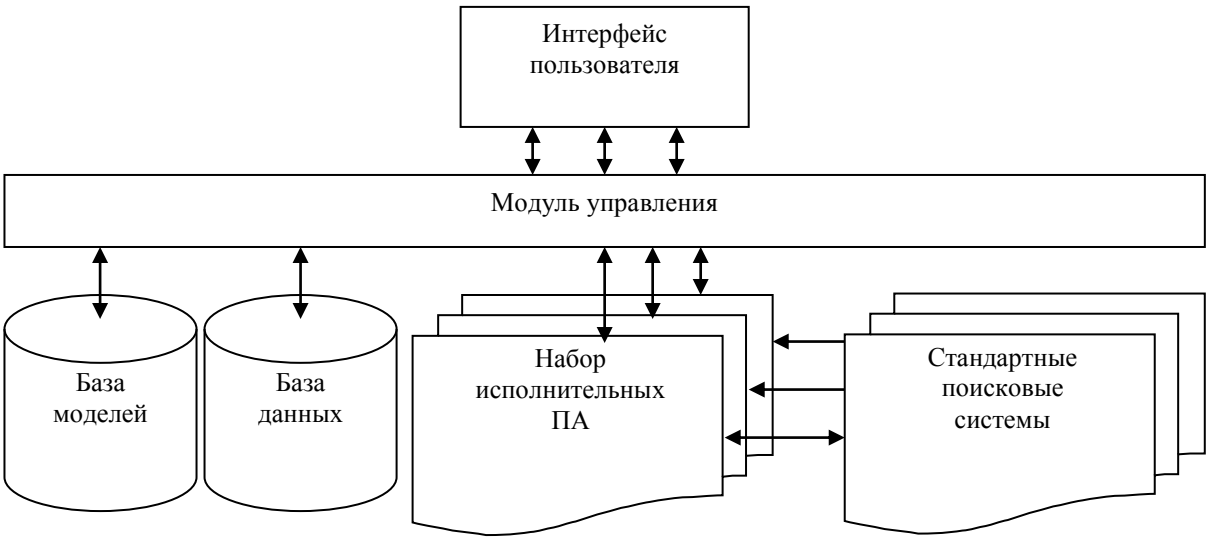


Рисунок 4.8. – Структурно-функциональная схема ИПА

Модуль управления занимает центральное место в системе. Он обеспечивает взаимодействие всех элементов системы и координацию их функционирования в соответствии с заданной логикой. После запуска системы он автоматически переходит в активное состояние и ожидает команды пользователя. Активное состояние модуля управления предполагает выполнение с определенной периодичностью набора ряда процедур, таких как:

- упорядочение записей в распределенной БД в соответствии с иерархией терминологического портрета предметной области;

- индексирование записей в распределенной БД по иерархическим сегментам ТП предметной области;

- очистка БД от информации, невостребованной пользователем в течение заданного временного интервала;

- опрос готовности дежурного исполнительного поискового агента.

По мере поступления команд от пользователя модуль управления выполняет набор следующих процедур:

- прием входного термина или их последовательности;

- идентификация иерархической принадлежности терминов;

- оценка загрузки канала дежурного исполнительного поискового агента;

- формирование пакета инсталляции дополнительных исполнительных

поисковых агентов для локального ПК;

- формирование пакета инсталляции дополнительных исполнительных поисковых агентов для локальной вычислительной сети;

- формирование пакета инсталляции дополнительных исполнительных поисковых агентов для виртуальных машин;

- оценка загрузки каналов дополнительных исполнительных поисковых агентов;

- распределение поисковой терминологической нагрузки среди дополнительных исполнительных поисковых агентов;

- репликация распределенной БД.

В каждом исполнительном ИПА имеется свой модуль управления, реализующий такие функции, как:

- инициализация терминологического портрета предметной области;

- прием последовательности терминов пользователя;

- выбор поисковой системы и генерация запросов;

- построение терминологических портретов текстов;

- идентификация поступивших текстов;

- селекция основных (искомых) терминов и информации по ним;

- селекция дополнительных (сопутствующих) терминов и информации по ним;

- занесение информации по основным и дополнительным терминам в БД.

База моделей ИПА обеспечивает требуемую логику его функционирования.

Состав базы моделей системы представлен на рисунке 4.9.

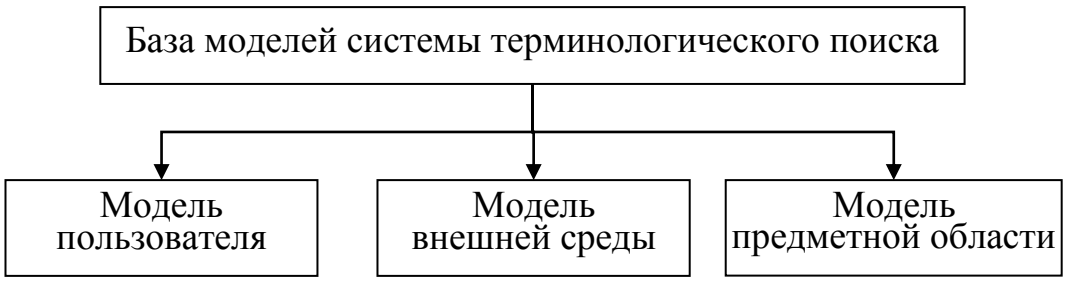


Рисунок 4.9. – Состав базы моделей системы

Модель пользователя отражает механизм взаимодействия последнего с

базой данных. Наличие данной модели позволяет повысить эффективность поиска и обеспечить реализацию упреждающего поиска (проактивность) информации и мониторинга потоков ее поступления. Кроме того, данная модель отображает постоянные информационные потребности пользователя (информационные запросы по отслеживанию предпочитаемых источников и запросы безотносительно источников) [23]. ЛПР может корректировать модель.

Модель внешней среды устанавливает порядок взаимодействия пользователя с окружающим информационным пространством. Элементами модели внешней среды являются:

- информационные ресурсы, включающие сектора деловой информации, юридической информации, информации для специалистов, а также массовой, потребительской информации;

- набор правил работы с информационными ресурсами;

- средства добычи информации (поисковые системы, поисковые машины, программы-роботы и др.);

- набор правил настройки средств добычи информации;

- набор сценариев информационного поиска.

Набор сценариев информационного поиска включает:

- поиск в локальном массиве электронных документов форматов: txt, html, pdf, doc;

- поиск на основе внешних конструкций;

- монопоиск и мультипоиск по ключевым словам;

- монопоиск и мультипоиск по образцу (шаблону).

Модель предметной области представляет собой иерархическую структуру терминов (терминологический портрет), связанную с областью исследований пользователя [170]. Терминологический портрет отличается от тезауруса, весовой взаимоувязкой терминов в нем и более высоким уровнем детализации рассматриваемой предметной области. Для построения терминологического портрета текста используется специальный иерархический терминопостроитель, содержащий систему моделей и алгоритмов терминологиче-

ского поиска, описанных в разделах 4.5 и 4.6 [182].

Для запуска системы виртуального распределенного терминологического поиска необходимо исполнить командный файл Термин_поиск.bat. В состав скрипта командного файла входит набор команд, инсталлирующих программные средства, составляющие окружение данной системы. Устанавливаются драйверы, специальные пакеты программ, программная среда Python и исполнительный файл системы – analyzer.py. Кроме того в состав набора команд могут быть включены драйверы, операционные системы и исполнительные файлы, инсталлирующие виртуальные машины и виртуальную вычислительную сеть. В процессе инсталляции на экране монитора появляется окно командной строки, в котором можно увидеть результаты инсталляции системы.

Если инсталляция системы прошла успешно, то на мониторе появляется главная экранная форма «Терминологический поиск». Внешний вид главной экранной формы «Терминологический поиск» представлен на рисунке 4.10.

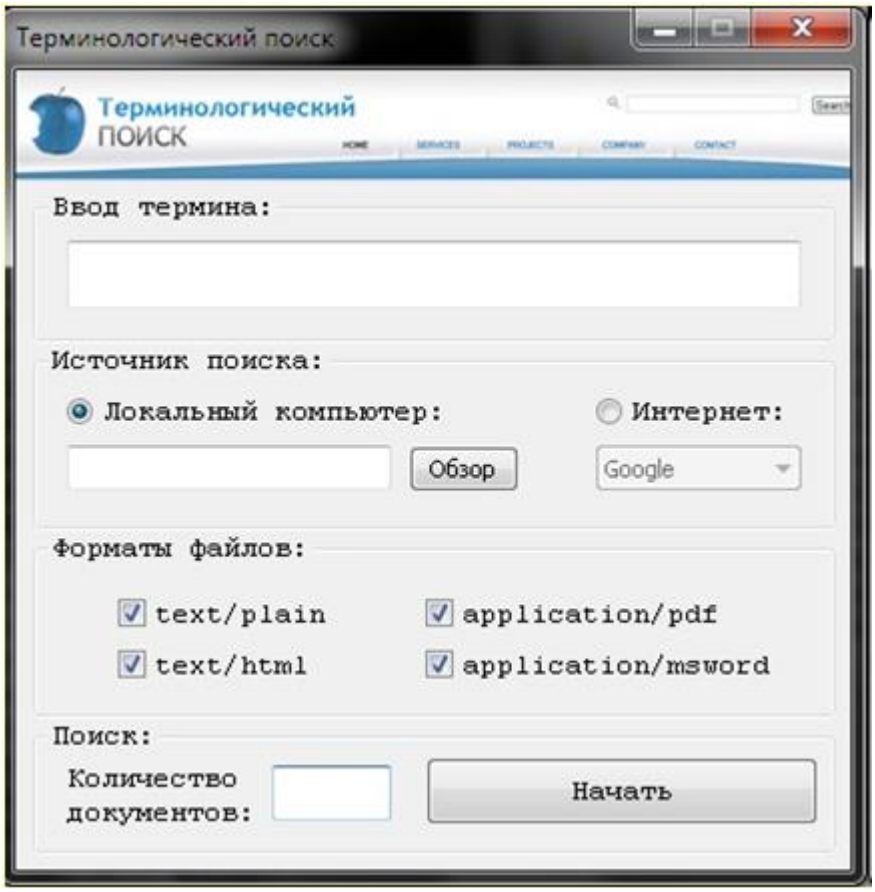


Рисунок 4.10. – Вид главной экранной формы системы «Терминологический поиск»

В верхней части главной экранной формы системы находится окно ввода терминов. С использованием данного окна пользователем может быть введен терминологической портрет предметной области в специальном формате: «облигация:3:акция:3:дюрация:4». Приведенная запись означает, что пользователем формируется терминологический портрет, в котором термины «облигация» и «акция» находятся на 3 уровне иерархии, а термин «дюрация» находится на 4 уровне иерархии. Для ввода последовательности терминов, определение которых необходимо найти, а также дополнительную информацию по ним, используется формат вида: «облигация; акция; дюрация».

В средней части окна находятся источники информации, в которых может проводиться поиск и возможные форматы искомых файлов. В разработанной версии программы, в качестве источников поиска, могут использоваться локальный компьютер, локальная вычислительная сеть, глобальная сеть Интернет и виртуальная машина. Если не активированы «Локальный компьютер» и «Интернет», то используется виртуальная машина. Среди форматов электронных документов, в массивах которых проводится поиск, в данной версии используются: txt, html, pdf, doc.

В нижней части окна находится область с инструментами поиска. В ней размещена кнопка «Начать» и окно «Количество документов». Нажатие на кнопку «Начать» инициирует процесс ТИП. В окне «Количество документов» отображается число обработанных документов.

После завершения процесса ТИП, на мониторе появляется экранная форма «Результаты поиска», внешний вид которой представлен на рисунке 4.11.

В левой части экранной формы «Результаты поиска» находится окно с определениями искомых терминов и указанием источников, содержащих данную информацию. Внизу окна указывается число найденных документов. В приведенном примере осуществлялся поиск определения термина «бюджети-

рование» в глобальной сети Интернет. Найдено 25 документов, содержащих

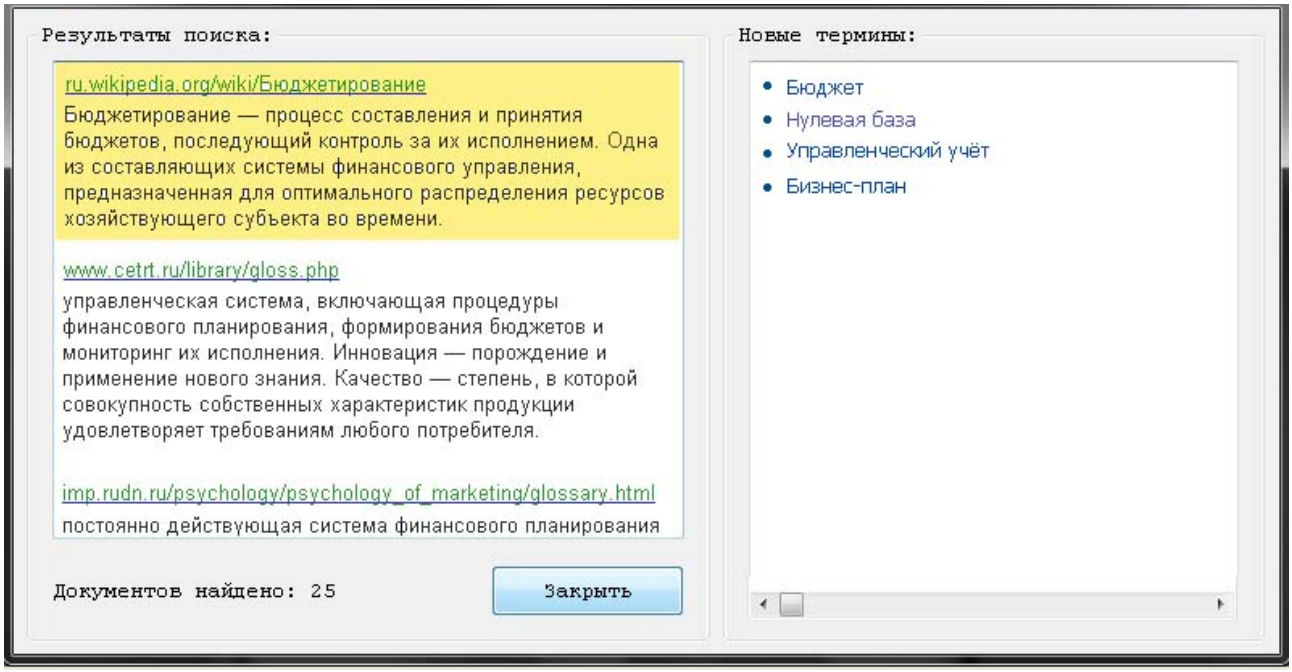


Рисунок 4.11. – Вид экранной формы «Результаты поиска»

данную информацию. Внизу окна указывается число найденных документов. В приведенном примере осуществлялся поиск определения термина «бюджетирование» в глобальной сети Интернет. Найдено 25 документов, содержащих данное определение в требуемом контексте. На экране отображены определения, содержащиеся в википедии, в одной из библиотек и в одном из глоссариев.

В левой части экранной формы «Результаты поиска» находится окно с перечнем дополнительных терминов (бюджет, нулевая база, управленческий учет, бизнес - план).

Нажатие кнопки «Закрыть» приводит к закрытию данной экранной формы и возврату на главную экранную форму «Терминологический поиск».

Выход из программы осуществляется путем нажатия соответствующей кнопки.

Наряду с внедрением данной системы в состав ИСППИД, отдельные ее элементы (алгоритм идентификации текстов, иерархический текстовый терминопостроитель, семантическая матрично - лексическая модель анализа текста) прошли апробацию в документальной информационно-поисковой си-

стеме «Научные труды» [94]. Документальная информационно-поисковая система (ДИПС) «Научные труды» представляет собой программную систему, предназначенную для ведения базы данных научных трудов профессорско-преподавательского состава кафедры, формирования отчетных материалов об их публикациях, а также реализации терминологического информационного поиска в распределенной базе данных и в сети Интернет. На рисунке 4.12 приведена экранная форма ДИПС «Запросы пользователя БД» с реализованным запросом по термину «Технология».

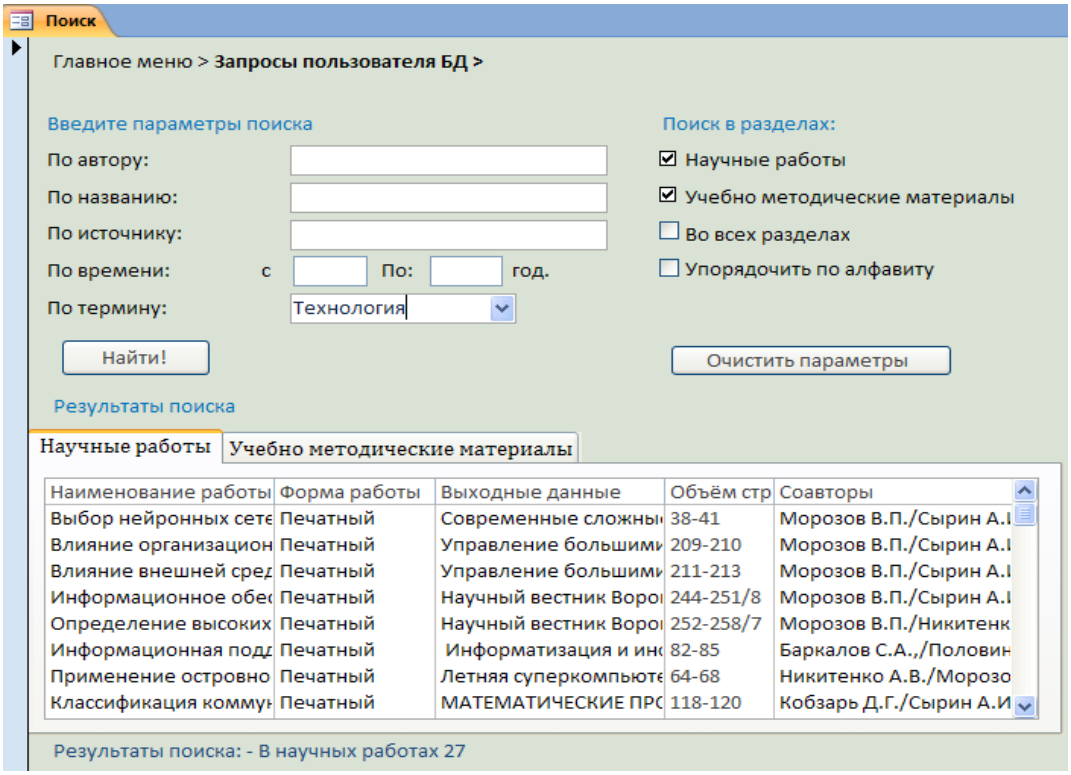


Рисунок 4.12. – Вид экранной формы «Запросы пользователя БД» с реализованным запросом по термину «Технология»

В нижней части данной экранной формы имеются закладки «Научные труды» и «Учебно-методические материалы», содержащие таблицы с результатами поиска. Полоса прокрутки, размещенная справа, позволяет просмотреть все найденные документы (в данном случае « Научные труды»). Двойной клик мыши по выделенному научному труду вызывает экранную форму «Просмотр». Она обеспечивает визуальный просмотр выбранного документа и позволяет увидеть, в каком контексте использовался данный термин.

ВЫВОДЫ

Для добычи полезной информации (новых знаний) в области ПИД СЭО, связанной с особенностями определения параметров ЦБ, стратегиями управления ИП и др., в составе ИСППИД функционирует подсистема ТИП. В нее входят: модель формализации текстовой информации; алгоритм графематического анализа; алгоритм морфологического анализа; алгоритм синтаксического анализа; семантическая матрично-лексическая модель; иерархический терминопостроитель; алгоритм идентификации; модель извлечения новых знаний.

Идейной основой для модели формализации текстовой информации служит категоризация лексики. Она обеспечивает построение более простых и адекватных моделей представления текстовой информации [73]. Первые упоминания о ней относятся к моделям Ньютона и Лейбница. Дальнейшее развитие категоризация лексики получила в лямбда-исчислении и семантической сети. В ее состав могут быть включены такие важные для ТИП поиска категории, как тезаурус и терминологический портрет. Терминологический портрет отличается от тезауруса, весовой взаимоувязкой терминов в нем и более высоким уровнем детализации рассматриваемой предметной области [182].

Иерархический текстовый терминопостроитель предназначен для формирования терминологического портрета исследуемого текста. Он базируется на системе моделей и алгоритмов формализации и проведения различных видов анализа текста (морфологического, синтаксического и семантического). В формализованном виде терминологический портрет представляет собой вектор весов информационных признаков. Вес каждого информационного признака соответствует частоте встречаемости терминов из которых он состоит [73]. Иерархический текстовый терминопостроитель реализует построение подобных векторов.

Семантическая матрично-лексическая модель анализа текста предназначена для выделения терминов заданного текста на основании смысла. Ее основу составляют лексемы - элементарные семантически значимые единицы

языка, представляющие собой слова, устойчивые словосочетания или другие языковые конструкции, способные обозначать предметы, явления, их признаки и др. Для заданного текста определяется матрица лексем. Она является одним из инструментов представления текста в виде ориентированного графа, позволяющего проводить его дальнейший анализ. В дальнейшем он преобразуется в семантическую сеть, которая позволяет представлять смысл текста в более естественной форме – в виде сети с узлами – понятиями и смысловыми отношениями между ними [73].

Индексный метод извлечения новых знаний базируется на определении логических цепочек взаимосвязей терминов в массиве текстовых документов и в случае их наличия, извлечении новых терминов [117]. Под новыми знаниями понимается новая дополнительная терминологическая информация, извлекаемая из текста (текстов). Логическая цепочка представляет собой набор сцепленных индексов понятий. Процедура восстановления соответствия «индекс-понятие» показала себя аналогом подсистемы объяснения, используемой в экспертных системах. Однако для ее реализации не потребовалось применение специального математического аппарата (например, предикатов первого порядка и др.) и соответствующих высокоуровневых языков программирования (Пролог, ЛИСП и др.).

В интересах повышения оперативности ТИП в ИСППИД, разработана система виртуального распределенного ТИП. Особенностью данной системы является возможность реализации ТИП на виртуальных машинах (на локальной виртуальной машине и в сети виртуальных машин).

5. МОДЕЛИ И АЛГОРИТМЫ ХРАНЕНИЯ И ОБРАБОТКИ ДАННЫХ ПОРТФЕЛЬНОЙ ИНВЕСТИЦИОННОЙ ДЕЯТЕЛЬНОСТИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЙ ОРГАНИЗАЦИИ

5.1. Состав и структура подсистемы ввода и хранения данных

Модели и алгоритмы хранения и обработки данных входят в состав подсистемы сбора, хранения и обработки и данных (ПСХОИ). Она используется в интересах информационного обеспечения основных функциональных подсистем ИСППИД: формирования ИП, прогнозирования значений параметров ЦБ и терминологического информационного поиска.

Структурно-функциональная схема ПСХОИ приведена на рисунке 5.1.

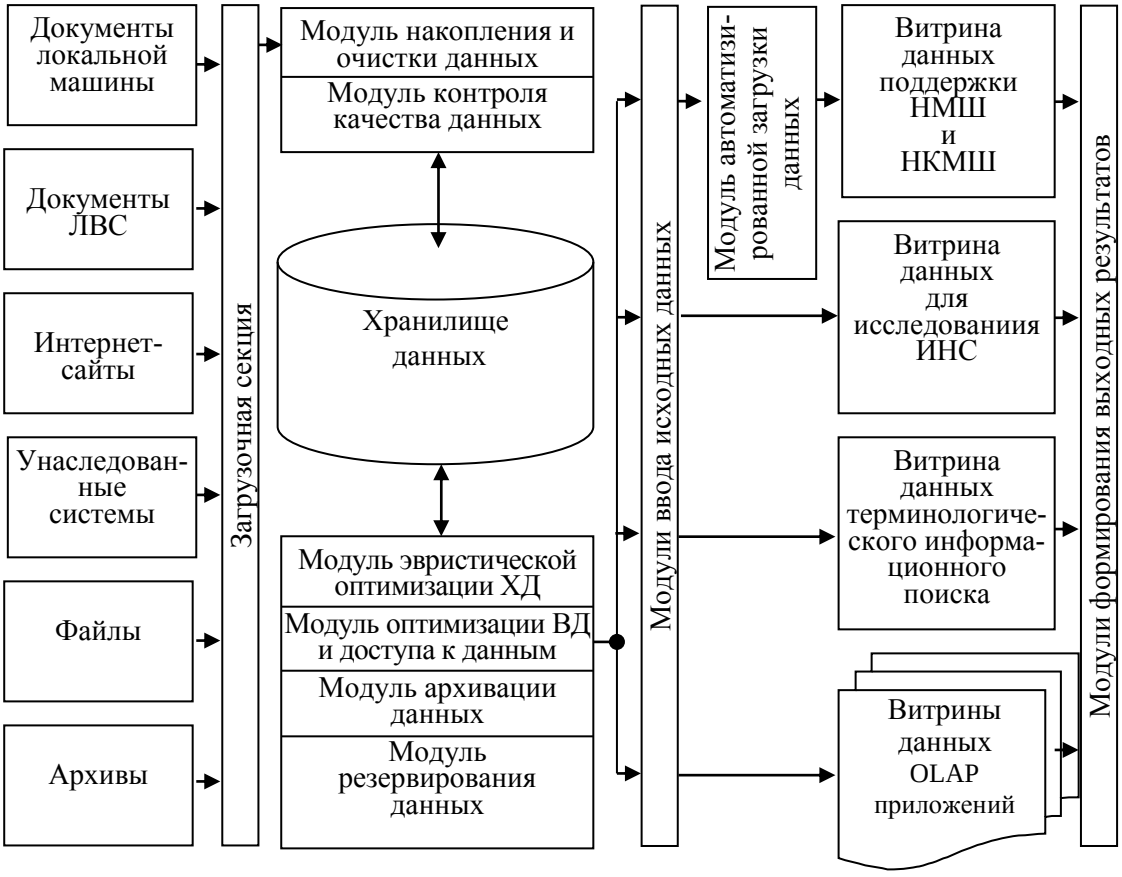


Рисунок 5.1. – Структурно-функциональная схема ПВХД

В состав ПСХОИ входит хранилище данных, загрузочная секция, набор витрин данных (поддержки НМШ и НКМШ, исследования ИНС, ТИП, других OLAP приложений) и программных модулей (эвристической оптимизации хранения данных, ввода данных, накопления и очистки данных, контроля качества

данных, оптимизации витрин данных (ВД) и доступа к данным, архивации данных, резервирования данных, формирования выходных результатов).

Хранилище данных (ХД) предназначено для долговременного хранения данных [95] и представляет собой БД, отличающуюся специфическим администрированием. Ее содержимое имеет ряд свойств [218]:

- предметная ориентация (характер данных, содержащихся в ХД, ориентирован не на имеющиеся приложения, а на задачи, решаемые в интересах информационной поддержки управленческих решений);

- нормализованный формат данных (единый стандарт);

- временная инвариантность (данные сохраняют свою истинность независимо от манипуляции с ними в любой момент времени);

- информационная устойчивость (данные, которые были загружены однажды, не должны меняться) [218].

Разработка ХД осуществлялась поэтапно [19, 50, 70, 80].

На начальном этапе проводились концептуальное и логическое проектирование ХД, независящие от деталей его воплощения (от используемой СУБД). В основу концептуального проектирования была положена объектно-реляционная модель данных [70]. При этом использовался стандарт IDEF0 [77].

На его основе представим систему в виде совокупности взаимосвязанных работ или функций. Для функций и работ реализован независимый анализ. На этой основе прозрачно моделируется логика и взаимосвязи процессов функциональных подсистем ИСППИД [78, 79]. Графическое представление бизнес-процессов составляет основу методологии IDEF0. Оно реализуется в виде специальных диаграмм, которые взаимосвязаны между собой в определенном порядке [19]. Подобная диаграмма есть единица представления системы. Для своего размещения она требует отдельный лист. При разработке информационных моделей ХД и ВД использовались модели «сущность-связь» (ER-модели) [82]. Модель "сущность-связь" представляет собой высокоуровневую концептуальную модель данных. Данная модель – это набор концепций, которые описывают структуру БД и связывают с ней транзакции обновления и извлечения данных. На основании

модели «сущность-связь» были построены таблицы для занесения данных. Таблицы строились с помощью программного средства Database Desktop в среде визуального программирования Delphi 7.0 с использованием Paradox 7. Процесс их проектирования проводился с использованием инструмента BP-Win [70]. На его основе из логической модели данных, была сформирована БД физического уровня. В рамках ее реализации выполнялась следующая последовательность действий [212]:

- формирование набора таблиц в соответствии с логической моделью;
- выбор методов доступа к хранимым данным и требуемых видов структур, обеспечивающих высокий уровень оперативности обработки запросов;
- разработка системы защиты хранимых данных.

Такой подход к реализации ХД позволил реализовать [212]:

- мониторинг за появлением избыточных данных;
- непротиворечивость хранимых данных;
- повысить объем хранимых полезных данных;
- возможность совместного использования данных;
- мониторинг и поддержку целостного характера данных;
- относительно высокую безопасность;
- применение стандартных шаблонов хранения данных;
- повышение производительности обработки хранимых данных при масштабировании системы;
- уменьшение времени доступа к хранимым данным;
- снижение затрат на сопровождение БД;
- улучшение качества функционирования службы восстановления хранимых данных и резервного копирования.

При разработке ХД учитывались следующие особенности его эксплуатации:

- использование для хранения текстовых и цифровых данных;
- обеспечение возможности варьирования требований к запрашиваемой информации (изменение структуры запросов и др.);
- гибкость модели хранения данных и приемлемая оперативность.

5.2. Алгоритм эвристической оптимизации хранилища данных

Модуль эвристической оптимизации ХД обеспечил формирование его рациональной структуры. С этой целью в рамках данного модуля был разработан специальный алгоритм эвристической оптимизации, блок-схема которого представлена на рисунке 5.2 [173].

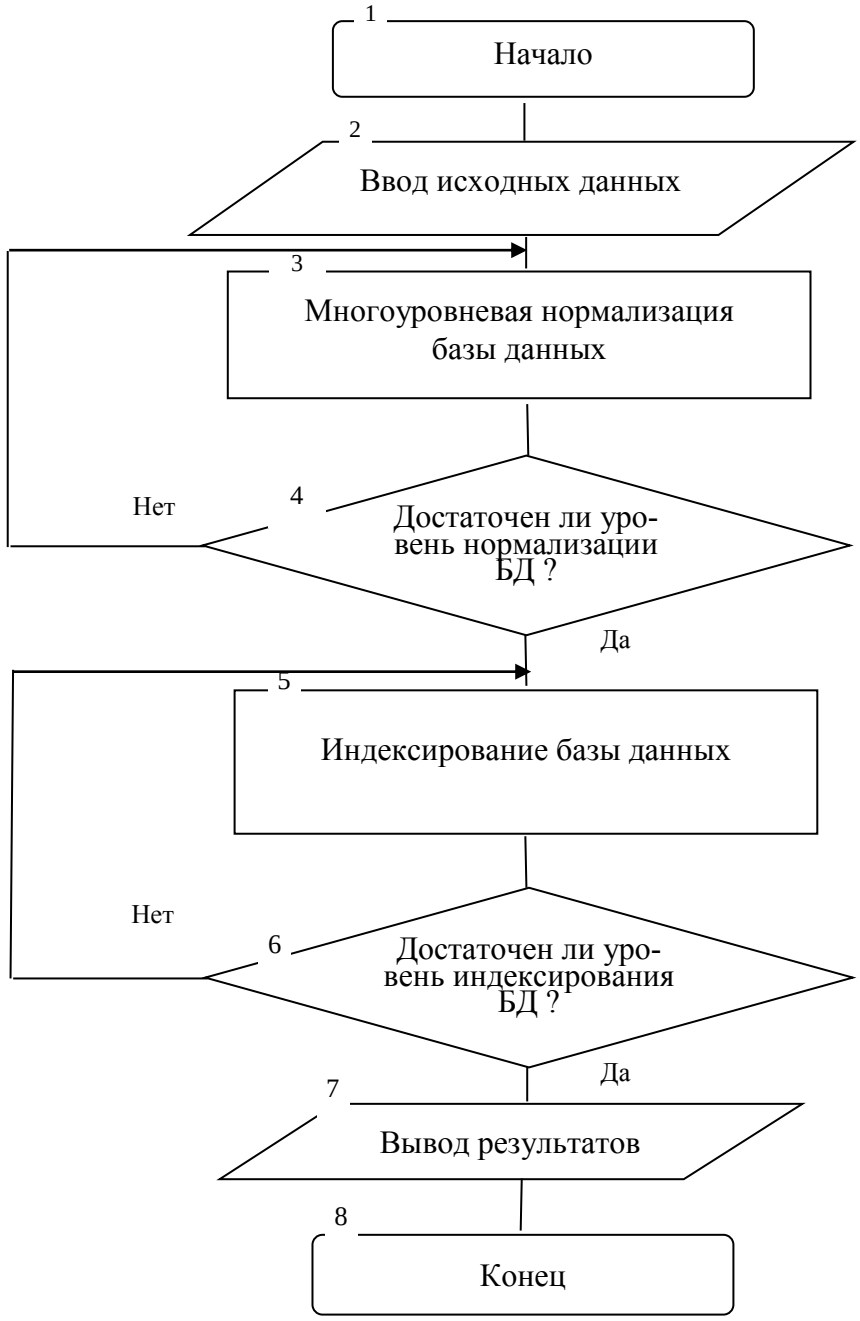


Рисунок 5.2. – Блок-схема алгоритма эвристической оптимизации структуры БД

Блоки 1 и 8 используются для пуска и остановки алгоритма эвристической оптимизации структуры базы данных.

В блоке 2 реализован ввод исходных данных, таких как число уровней нормализации и число индексируемых полей.

В блоке 3 реализована многоуровневая нормализация ХД. Нормализация - это формальный метод анализа отношений на основе их первичного ключа и существующих функциональных зависимостей. В ходе нормализации формат отношений становится все более ограниченным (строгим) и менее восприимчивым к аномалиям обновления. В теории реляционных баз могут использоваться 5 нормальных форм. Каждая нормальная форма имеет ограничения. Отношение находится в определенной нормальной форме, если оно удовлетворяет набору ограничений этой формы. С переводом структуры отношений базы данных в формы более высокого порядка происходит удаление из таблиц избыточной неключевой информации [230].

Первая нормальная форма (1НФ) - отношение, в котором на пересечении каждой строки и каждого столбца содержится одно и только одно значение.

Вторая нормальная форма (2НФ) - отношение, представленное в *1НФ* в котором каждый атрибут, который не входит в состав первичного ключа, полностью от него зависим.

Третья нормальная форма (3НФ) - отношение, находящееся в *1НФ* и *2НФ.* Его атрибуты полностью входят в первичный ключ, имея при этом, транзитивную функциональную зависимость от него.

Четвертая нормальная форма (4НФ) – отношение, не содержащее сложных мультизначных зависимостей, имеющее представление в нормальной форме Бойса-Кодда.

Пятая нормальная форма (5НФ) – отношение, которое не имеет зависимости взаимосвязи.

Для обеспечения высокой функциональности при максимуме нормализации [33], была реализована ее 3НФ.

Блок 4 обеспечивает проверку достаточности уровня нормализации ХД. Если он достаточен, то управление подается в блок 5. Иначе управление подается в блок 3.

В блоке 5 реализовано индексирование ХД. Индекс представляет собой структуру ускоряющую обнаружение файловых записей. Это обеспечивает повышение оперативности реализации пользовательских запросов. По своей сути индекс в ХД является аналогом книжного предметного указателя. С его помощью можно избежать длительный последовательный или пошаговый просмотр файла для поиска требуемой информации. Индекс БД упорядочен. Каждый его элемент включает наименование искомого объекта и один или несколько идентификаторов (указателей) в месте его локализации. Объем памяти, который занимает индекс гораздо меньше объема памяти таблицы. Следовательно, перебор значений в нем более оперативен, чем поиск и считывание данных в отношении. Учитывая, что индексные значения при хранении упорядочены, поиск нужной строки ускоряется. Индексы позволяют реализовать выборку строк отношений в заданном интервале значений. Одно отношение может включать ряд индексов. Для разных отношений, содержащих идентичные атрибуты, возможно сформировать мультииндекс, в котором одному общему атрибуту может соответствовать несколько ссылок. При этом каждая ссылка будет указывать на строку с таким значением в разных отношениях.

Блок 6 обеспечивает проверку уровня достаточности индексирования ХД. Если он достаточен, то управление передается в блок 7. В противном случае управление передается в блок 5.

Блок 7 обеспечивает результат - оптимизированное ХД.

В качестве примера в таблице 5.1 приведены данные (временной ряд доходностей), хранящиеся в ХД, предназначенные для обучения искусственной нейронной сети. В качестве данных приведены значения котировок акций ОАО «Газпром», имевших место в период с 25.03.2017 по 04.04.2017 г.

Таблица 5.1 – Данные, хранящиеся в ХД для обучения ИНС

Дата	25.03.17	28.03.17	29.03.17	30.03.17	31.03.17	01.04.17	04.04.17
Цены открытия	224,87	223,01	225,9	225,51	229,38	229,1	235,98
Норм. цена открытия	0,23	0,15	0,27	0,26	0,42	0,40	0,69
Цены закрытия	222,81	225,5	224,01	229,32	229,09	235,1	237,7
Норм. цена закрытия	0,11	0,22	0,16	0,38	0,37	0,62	0,73
Прогнозная цена	223,01	225,9	225,51	229,38	229,1	235,98	238,46
Норм. прогноз	0,14	0,26	0,25	0,41	0,40	0,68	0,79

В строках данной таблицы используются значения таких характеристик, как:

- «Дата» – дата, на которую приведены значения акций;

- «Цена откр.» – цена за акцию на момент открытия торгов;

- «Цена закр.» – цена за акцию на момент закрытия торгов;

- «Прогнозная цена» – цена открытия торгов на следующий день, значение которой необходимо спрогнозировать;

- «Норм. цена откр.», «норм. цена закр.», «норм. прогноз» – нормализованные значения.

Определение рациональной структуры ХД проводилось экспериментальным (эвристическим) путем.

Для различных вариантов структур ХД для подсистемы исследования ИНС формировалось множество запросов на поиск определенных атрибутов информации (по одному полю и по множеству полей). При этом время обработки каждого одноаспектного или многоаспектного запроса фиксировалось.

После завершения реализации множества запросов для соответствующего варианта структуры ХД оценивалось среднее время обработки множества запросов. Затем полученные значения сравнивались, и выбирался тот вариант структуры ХД, для которого значение среднего времени обработки множества запросов было минимальным.

Полученные значения усредненного времени реализации запросов приведены в таблице 5.2 [173].

Таблица 5.2 – Значения усредненного времени реализации запросов

Количество записей в ХД:10000					
Значения среднего времени выполнения запросов, *ms*					
Поля не индексированы					
1НФ					
Цена открытия	Нормальная цена открытия	Цены закрытия	Нормальная цена закрытия	Прогнозная цена	Нормальный прогноз
700	790	775	770	680	700
3НФ					
Цена открытия	Нормальная цена открытия	Цены закрытия	Нормальная цена закрытия	Прогнозная цена	Нормальный прогноз
400	380	390	370	410	415
Все поля проиндексированы					
1НФ					
Цена открытия	Нормальная цена открытия	Цены закрытия	Нормальная цена закрытия	Прогнозная цена	Нормальный прогноз
518	500	490	530	515	530
3НФ					
Цена открытия	Нормальная цена открытия	Цены закрытия	Нормальная цена закрытия	Прогнозная цена	Нормальный прогноз
180	150	160	130	160	150

Анализ полученных данных показывает, что время обработки запросов для ХД в 1НФ гораздо больше, чем время обработки запросов для ХД в 3НФ. Таким образом, можно сделать вывод о том, что оптимальной структурой ХД для подсистемы исследования ИНС является структура ХД, приведенная к 3НФ, все поля которой проиндексированы.

Анализ проведенных экспериментов показал, что среднее время обработки многоаспектных запросов меньше среднего времени обработки одноаспектных запросов. Следовательно, в СУБД MS Access целесообразнее осуществлять многоаспектный поиск [173].

5.3. Особенности реализации загрузочной секции

Загрузочная секция (ЗС) является буфером, предназначенным для временного хранения данных перед их загрузкой в ХД. В технологическом плане она представляет собой набор временных промежуточных таблиц, в которых находятся данные в необработанном формате. В интересах оперативной пересылки дан-

ных для их хранения использован внутренний формат системы. Данные, содержащиеся в ЗС, становятся доступными для пользователя, после их пересылки в ХД. Информация в ЗС поступает из различных источников (локальная машина, ЛВС, интернет-сайты, унаследованные системы, файлы, архивы).

Для сопряжения БД унаследованных систем «DB2» с текущей СУБД «Access 2007» использовалась специальная служба словаря информационных ресурсов (Information Resource Dictionary System — IRDS). Она реализует стандартный интерфейс для словарей данных СУБД. Это обеспечивает их более высокую доступность и упрощает взаимодействие. Данная служба является стандартом, разработанным Международной организацией стандартизации (International Organization for Standardization — ISO) [52]. В нее входят определения таблиц, которые включают словарь данных. Кроме того в ней содержится перечень операций, обеспечивающих доступ к данным таблицам. Они (операции) исключают противоречия при доступе к словарю данных и реализуют определения данных в разнотипных словарях.

В схематичном виде реализация данного подхода представлена на рисунке 5.3.

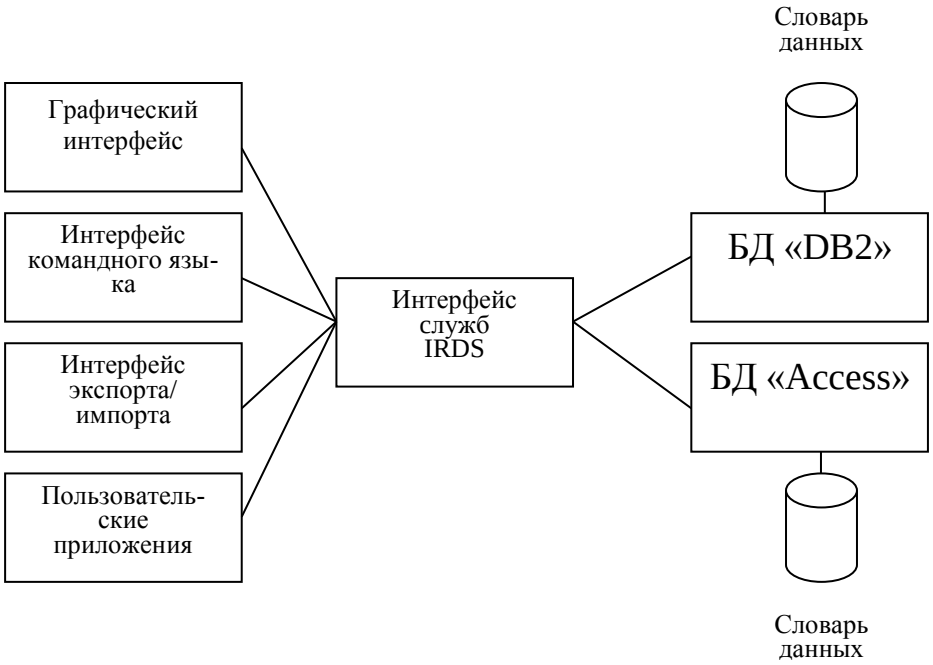


Рисунок 5.3. – Структурная схема обеспечения сопряжения

Информация, которая хранится в IRDS - словаре СУБД системы DB2, путем выполнения функций IRDS службы, передается в IRDS – словарь СУБД Access 2007, или передается в приложение, связанное с DB2. Ряд процедур данной службы обеспечивает организацию ввода и хранения новых типов данных, как для DB2, так и для Access 2007 [212].

Если пользователю требуется сформировать в какой либо СУБД новые виды отчетов, то набор функций системы IRDS позволяет это сделать достаточно оперативно, путем организации для этих целей соответствующего расширения. В технологическом плане, основу IRDS - службы составляет интерфейс служб, включающий совокупность функций, доступных для вызова пользователем в интересах организации доступа к IRDS - словарю. Данный интерфейс откликается на вызов следующих интерфейсов со стороны пользователя [212]:

- командный интерфейс в виде командной строки;
- интерфейс графических примитив;
- командные файлы импорта/экспорта;
- интерфейсы прикладных программ.

В разработанном прототипе ИСППИД СЭО применена СУБД Microsoft Access 2007 и реализована система запросов по некоторому образцу на основе языка QBE [212]. Для организации доступа к информации в БД в языке QBE реализован визуальный метод. В его основу положены шаблоны запросов [62]. Используется QBE следующим образом. Задаются образцы шаблонов запросов, в которых конкретизируются их параметры исходя из необходимости получения соответствующего ответа. Несмотря на свою простоту, язык QBE имеет хорошие возможности для работы с данными. В ИСППИД на языке QBE реализованы сложные запросы. Они используются для выборки информации, содержащейся в разных таблицах. Кроме того, на их основе формируется сложный набор полей, присутствующих в результирующей таблице.

5.4 Контроль качества и оптимизация хранения актуальных данных

Учет особенностей технологии хранения данных позволил при выборе инструмента использования ХД, отдать предпочтение витрине данных. В витрине

данных (ВД) хранятся текущие данные. Они представляет собой интегрированный, статичный, хронологически упорядоченный набор данных из разных источников. ВД базируется на нормализованной реляционной модели хранения данных. В состав ПСХОИ входит несколько витрин данных (поддержки НМШ, исследования ИНС, терминологического информационного поиска, OLAP приложений). Применение нескольких ВД обусловлено необходимостью поддержки различных форматов данных в соответствующих приложениях. ВД оптимизированы под специфические требования соответствующих приложений. ВД OLAP (on-line analytical processing - оперативной аналитической обработки) приложений являются перспективными и в разработанном прототипе ИСППИД не используются.

Модуль накопления и очистки данных обеспечивает проверку данных, поступающих в хранилище из различных источников. Необходимость этих действий обусловлена причиной существования данных с ошибками («замусоренных» данных), возникающих в результате ввода данных операторами ПК, различных форматов используемых словарей и отсутствием, системы контроля вводимых данных.

Модуль контроля качества данных предназначен для сохранения целостности и непротиворечивости последних, а также устранения их избыточности.

Модуль оптимизации ВД и доступа к данным предназначен для повышения оперативности реализации запросов пользователя и размещения актуальных данных на ВД. С целью повышения оперативности реализации запросов пользователя, был пересмотрен стандартный метод индексирования записей в БД и разработан модифицированный алгоритм обобщенного индексного дерева. Индексирование данных призвано повысить оперативность всех операций, используемых для поиска требуемой информации [203, 212]. Реализация процесса индексирования выявила следующую особенность, состоящую в следующем. В случае поиска записей для корректировки на основе условий, наложенных на индексируемый столбец, наблюдалось повышение оперативности поиска. Если же осуществлялось добавление некоторой записи, то оперативность падала. В виду того, что в процессе реализации сессий, выполнялось

множество разнородных операций с преобладанием добавления записей, а также индексной модификации, то в целом производительность падала. Для парирования данного недостатка, в интересах индексирования использовались P*-деревья. Последние, в основном, используются для индексирования пространственно-распределенных объектов [212]. Для их адаптации к текущим условиям были проведены исследования. Исследовались алгоритм обобщенного индексного дерева (АОИД) [42] и модифицированный алгоритм обобщенного индексного дерева (МАОИД). Суть модификации состояла в следующем. Основная функция «деления узла» была дополнена процедурами поиска неперекрывающихся множеств и деления непересекающихся узлов. Просмотр узловых записей базировался на процедуре деления прямоугольников и построении соответствующих списков, охватывающих непересекающиеся узлы. В теле главного цикла функции «деления узла», реализована процедура анализа возможных комбинаций деления. На основе этого была обеспечена итеративная сходимость алгоритма. Если существовало несколько вариантов узлового деления без пересечений, то предпочтение отдавалось тому, у которого была наименьшая площадь окружающего (щих) прямоугольника (ов). В случае отсутствия условий для реализации модифицированного алгоритма, подключался традиционный. Реализация вышеописанных процедур проводилась на основе стандартных библиотечных процедур Delphi 7.0.

Проведенные эксперименты показали [85], что по оперативности метод классических P-деревьев уступает предложенному в пределах 26%. Поэтому дальнейшее сравнение проводилось лишь с ними. На основе результатов экспериментов, построены зависимости быстродействия подобных индексных структур от их структуры.

В ходе экспериментов создавались три группы различных по размеру объектов: большие, малые и объекты со случайными размерами. Каждая группа включала набор экземпляров, формируемых по случайному закону, в соответствии с рядом: 5, 50, 500, 5000, 50000, 500000. Общее число исследуемых пакетов данных составило 50 [212].

Каждый вычислительный эксперимент включал следующие процедуры [212]:

- загрузка в БД множеств экземпляров объектов;

- формирование запросов к БД, реализующих пересечение ряда записей;

- отработка сформированного множества запросов;

- оценка длительности обработки индивидуальных запросов и их совокупностей различными алгоритмами;

- формирование результирующих отчетов.

Полученные результаты оценки длительности обработки запросов различными алгоритмами представлены в таблице 5.3 [212].

Таблица 5.3 - Результаты оценка длительности обработки запросов (*ms*) различными алгоритмами

Число записей	АОИД	МАОИД	Временной выигрыш, %
5	4,4	4,28	2,7
50	37,8	35,4	6,3
500	901,8	705,4	21,8
5000	1908,6	1307,4	31,5
50000	7132,1	3166,4	55,6
500000	30315,2	12108,3	60,1

Полученные результаты демонстрируют превосходство предложенного алгоритма в сравнении с традиционным более чем на 25% (по усредненному параметру). С увеличением числа записей, данное превосходство возрастает более чем на 50%. На основании вышеизложенного можно сделать вывод о перспективности предложенного алгоритма.

В модуле оптимизации ВД и доступа к данным реализован алгоритм взаимообмена данными между ВД и ХД, названный алгоритмом оптимизации доступа, который обеспечивает:

- оперативный обмен данными между ХД и ВД;

- доступ к актуальным данным за минимальное время;

- длительное хранение наборов данных включенных в ХД с заданным уровнем гарантии.

Для обеспечения эффективного взаимодействия ВД и ХД разработан

алгоритм оптимизации доступа, блок-схема которого представлена на рисунке 5.4 [173].

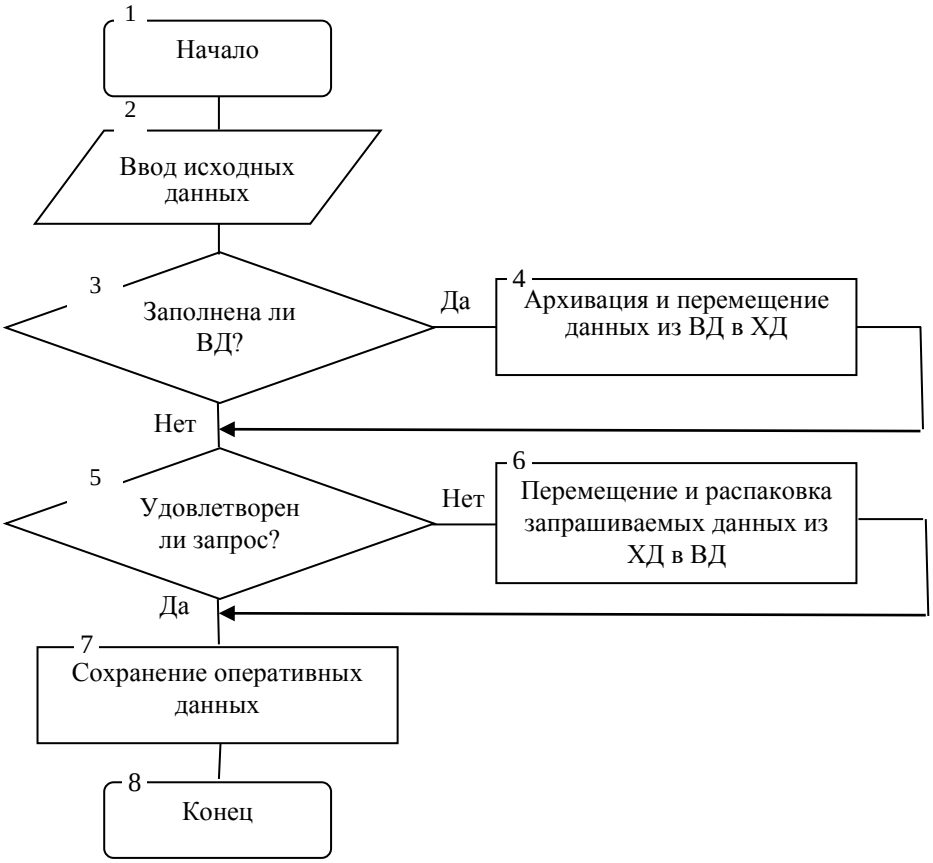


Рисунок 5.4 - Блок-схема алгоритма оптимизации доступа

Блоки 1,8 используются для пуска и остановки процесса обмена данными.

В блоке 2 реализован ввод исходных данных, таких как значение верхнего уровня объема данных ВД и значение нижнего уровня объема данных ВД. Запуск и остановка процедуры перемещения данных реализуется в соответствии со следующими критериями [218]:

- в случае, когда число записей в ВД превышает допустимый уровень, осуществляется перемещение некоторой части данных из ВД в ХД;

- в случае, когда при перемещении данных из ВД в ХД, объем числа оставшихся записей, станет меньше порогового значения, данный процесс завершается.

В блоке 3 реализована проверка загруженности ВД. В том случае, если ВД не заполнена, то управление передается в блок 5. В противном случае, передача управления осуществляется в блок 4 [218].

Блок 4 реализует архивацию и перемещение данных из ВД в ХД.

В блоке 5 реализован поиск данных, запрашиваемых пользователем. Если данные размещены в ВД, то реализуется их непосредственная загрузка, а затем дальнейшая обработка. Если данные размещены в ХД, то передача управления осуществляется в блок 6 [218].

В блоке 6 реализован поиск данных, запрашиваемых пользователем, в ХД. При наличии в ХД, осуществляется их перенос в ВД. В процессе переноса, данные распаковываются и подготавливаются для последующей обработки [218].

В блоке 7 реализована запись в ВД полученных данных [218].

Большое значение при архивации данных имеет определение последовательности их переноса. В ХД переносятся не все данные - об этом свидетельствует наличие нижнего допустимого уровня хранения данных в ВД. Перемещение производится в соответствии с рангом популярности данных: очередь на перенос формируется в очередности, обратной популярности данных, которую имеют данные с минимальным рангом. Мерой популярности данных являлась вероятность их запроса, определяемая эмпирическим путем. Ранг данных увеличивается, если увеличивается вероятность поступления соответствующего запроса на них [218].

Для оценки вероятности поступления запроса использовано предположение о том, что запросы формируют поток событий, подчиняющийся статистике простого пуассоновского потока [1]. Для определения его интенсивности, в виде число событий, происходящих за некоторый временной интервал, используется следующее выражение, определяемое эмпирическим путем [173, 218]:

$$\lambda^j = \frac{M^j}{t_M^j - t_O^j}, \qquad (5.1)$$

где M^j – число поступивших запросов на j - е множество данных, t_O^j - момент времени загрузки архивных данных, t_M^j - момент времени поступления M^j -го запроса. Для вероятности запроса на j - е множество данных с учетом плотности потока поступивших заявок λ^j в $t > t_M^j$ момент времени ис-

пользуется выражение [1, 218]:

$$p^j = 1 - \exp\left\{-\lambda^j\left(t - t_M^j\right)\right\}. \quad (5.2)$$

Значения полученных оценок позволяют вычислить ранг данного набора исходных данных. Применительно к множеству с *K* наборами данных суть ранжирования заключается в селекции исходных данных на основании следующего правила [173, 218]:

$$p^j < p^\nu, \text{ при } j < \nu \text{ для } j \in 1, K;\ \nu \in 1, K. \quad (5.3)$$

Результатом проведенной селекции данных является искомый набор с индексом *j*, который затем необходимо переместить в ХД. При таком подходе расчетные затраты, связанные с повторным извлечением данных из ХД, минимизируются [218]. При реализации данного алгоритма сортировка проводится при потенциально большом количестве кандидатов на перенос. В случае, когда количество переносимых файлов (наборов данных) невелико, реализуется простой последовательный перенос файлов.

Предложенный подход реализован в виде процедур, обеспечивающих:

1) контроль ресурсов ВД;

2) расчет ранга популярности данных;

3) автоматический перенос данных, инициируемых процедурами контроля состояния ВД.

5.5. Реализация архивации данных

Модуль архивации данных используется в интересах архивации и распаковки данных, соответственно при их передаче и приеме в ХД [171]. Обобщенная схема классификация методов сжатия информации приведена на рис. 5.5.

Подробное обоснование методов архивации данных приведено в [173].

Анализ эффективности архиваторов, базирующихся на методах сжатия данных, представленных на рисунке 5.5 и используемых на практике, показал следующее. Оценку прошли такие архиваторы, как [20, 173]: 7-Zip; ACE; ARJ; ARJZ; CABARC; Imp; JAR; PKZIP; RAR; WinZip; 1Zip.

Рисунок 5.5. – Классификация методов сжатия данных

Выбор этих архиваторов обусловлен следующими причинами. Во-первых, они наиболее рациональные в плане использования методов Зива – Лемпела. Во-вторых, они нашли достаточно высокую популярность. В-третьих, они оказали большое влияние в части развития словарных алгоритмов [173].

Результаты сравнительного анализа вышеперечисленных архиваторов представлены в таблице 5.4. Оценивалась степень сжатия файлов, время кодирования (архивации) и декодирования (разархивации). При этом использовался специальный сертифицированный тестовый набор исходных данных CalgCC [20, 173].

Коэффициент сжатия оценивался с помощью набора файлов, получившего название Calgary Compression Corpus2 (CalgCC). В состав CalgCC входят файлы различных типов данных (текстовых, графических, цифровых, смешанных) [20, 173].

Таблица 5.4 - Сравнительная характеристика архиваторов по степени сжатия файлов, времени кодирования и декодирования специального набора

	ARJ	PKZIP	ACE	RAR	CABARC	7-Zip
Bb	3.07	3.15	3.39	3.37	3.42	3.60
Bk1	2.42	2.44	2.77	2.82	2.90	2.93
Bk2	2.91	2.91	3.35	3.31	3.50	3.58
Go	1.47	1.48	1.54	1.52	1.71	1.85
Ns	2.55	2.60	3.01	3.05	3.05	3.12
O1	2.04	2.05	2.15	2.17	2.22	2.23
O2	3.00	3.03	3.37	3.38	3.52	3.94
Pr1	2.86	2.84	2.90	2.94	2.97	3.05
Pr2	2.73	2.76	2.84	2.86	2.93	3.02
Pc	9.31	9.78	10.55	10.38	10.65	11.73
Prc	2.94	2.91	3.04	3.02	3.03	3.12
Prl	4.36	4.41	4.43	4.54	4.61	4.74
Prp	4.33	4.31	4.52	4.56	4.61	4.72
Tr	4.64	4.78	5.18	5.22	5.31	5.54
Итого	**3.41**	**3.51**	**3.70**	**3.81**	**3.91**	**4.12**
$T_{код}$	**2.62**	**1.66**	**2.70**	**2.53**	**3.75**	**1.1**
$T_{дек}$	**2.51**	**1.92**	**2.81**	**2.95**	**3.97**	**1.1**

Набор состоит из 14 файлов, большая часть которых представляет собой тексты на английском языке или языках программирования [20].

При проведении исследований применялся тот алгоритм, который обеспечивал наилучшее сжатие [20, 173].

Полученные результаты показали следующее. Время архивации у ARJ и PKZIP в 4,5 раза ниже, чем у RAR и ACE. Последние более оперативны в сравнении с CABARC и 7-Zip (примерно на 30 %). Размер словаря, применительно к ARJ и PKZIP на порядок меньше, чем в остальных архиваторах [20].

За единицу измерения для $T_{код}$ и $T_{дек}$ принято время, затрачиваемое на сжатие полного тестового пакета CalgCC архиватором 7-Zip. Проведенное сжатие соответствовало скорости кодирования 2 Гб/с. При этом использовался ПК с процессором Pentium IV 3,0 ГГц, объеме ОЗУ 2ГГб, частоте шины 333 МГц, объеме жесткого диска 320 ГГб [173].

Дальнейшая оценка возможностей архиваторов с целью выбора наиболее эффективного проводилась на выборке данных, непосредственно используемых при исследовании НМШ. В ее состав входили: файлы с числовыми

данными; файлы с растровыми изображениями используемых графиков; диаграммы и гистограммы. Общий объем выборки включал 18 файлов. Их число файлов было определено экспериментально. При меньших значениях числа файлов имел место достаточно большой разброс значений коэффициента сжатия, а при больших значениях числа файлов наступало так называемое насыщение, при котором разброс значений коэффициента сжатия имел незначительные приращения [20, 173].

Среди исследуемых архиваторов рассматривались PKZIP, RAR и 7-Zip. В настоящее время они распространены и являются представителями определенных диапазонов значений коэффициентов сжатия данных (табл. 5.4).

В табл. 5.5 приведены результаты сравнения архиваторов по степени сжатия файлов на реальном наборе данных, состоящем из 18 файлов фиксированного размера 1639139 байт [173].

Таблица 5.5 - Результаты сравнения архиваторов по степени сжатия файлов на реальном наборе данных

	PKZIP	RAR	7-Zip
I1.txt	2.91	3.14	3.56
I2.txt	2.44	3.07	3.43
I3.txt	2.23	2.94	3.19
I4.txt	2.13	2.80	2.79
I5.txt	2.09	2.67	3.41
R1.doc	2.02	2.10	2.17
R2.doc	2.01	2.06	2.17
R3.doc	2.03	2.06	2.21
R4.doc	2.01	2.02	2.16
R5.doc	1.91	2.04	2.30
F1.bmp	2.93	3.02	3.16
F2.bmp	4.41	4.54	4.75
F3.bmp	4.38	4.58	4.74
F4.bmp	4.77	5.22	5.55
G1.jpeg	4.02	4.26	4.76
G1.jpeg	4.16	4.45	4.97
G1.jpeg	4.24	4.28	5.18
G1.jpeg	5.00	5.59	5.96
Итого	**3.07**	**3.37**	**3.68**
$T_{код}$	**1.62**	**1.13**	**2.56**
$T_{дек}$	**1.83**	**1.24**	**2.67**

Анализ полученных данных показывает, что они незначительно отличаются от данных, приведенных в таблице 5.4 (на десятые доличто повышает

степень доверия к полученным результатам.

Максимальные характеристики продемонстрировал архиватор 7-Zip. Ему незначительно уступает RAR [173].

В результате анализа временных характеристик при проведении экспериментов было установлено, что ценой увеличения степени сжатия является падение скорости в среднем в 2 и более раза [173].

По результатам проведенных экспериментов, для архивации данных при пересылке последних из оперативного хранилища в долговременное хранилище, был выбран алгоритм 7-Zip .

5.6 Реализация резервного копирования

Модуль резервирования данных обеспечивает резервное копирование данных из ХД и ВД на внешние носители информации - жесткие магнитные диски. Значимость резервирования данных трудно переоценить. Информация, получаемая в результате функционирования ИСППИД, стоит дороже компьютеров, на которых она хранится. Как показывает практика, восстанавливать данную информацию (как, впрочем, и любую) достаточно трудно. Поэтому сохранение этой информации – одна из важных и трудоемких задач.

Потеря информации может происходить по многим причинам, начиная от сбоя программ до выхода из строя оборудования. Поэтому эксплуатировать ИСППИД без резервных копий хранимой в ней информации нецелесообразно.

Основными проблемами при резервном копировании являются [173]:

- сокращение временного интервала резервного копирования (окна резервирования);

- возрастание числа интерфейсных приложений, поддерживаемых резервным копированием;

- трудности резервирования открытых активных файлов.

Детали первой проблемы состоят в следующем. В прошлом, резервное копирование реализовывалось в период, когда приложения не работали, т.е в нерабочее время (ночь, время обеда). Работа прекращалась, и затем запускались приложения резервного копирования. Функционирование ИСППИД

предусматривает практически постоянную ее работу. Следовательно, существенно сокращается окно резервирования.

Анализ второй проблемы показывает следующее.

На рисунке 5.6 в схематичном виде показано увеличение числа приложений, в рамках которых необходимо проведение, как резервного копирования данных, так и их восстановления [173].

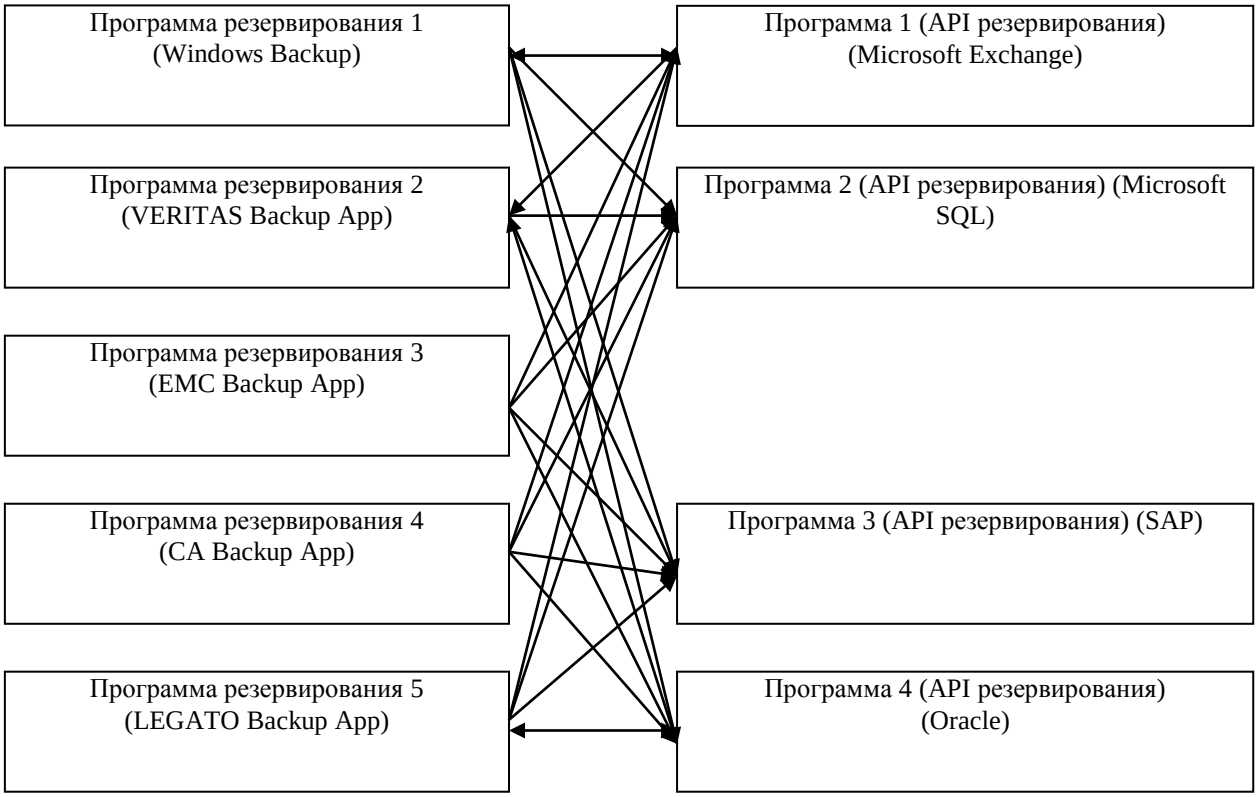


Рисунок 5.6. – Рост числа API для резервного копирования

Функция их возрастания носит экспоненциальный характер. В большинстве случаев пользователь стремится использовать дублирование резервного копирования несколькими программными продуктами, разработанными различными производителями. При этом у каждого поставщика используется свой специфический программный код, который он пытается защитить путем лицензирования [227]. Это приводит к возникновению ряда трудностей. Во-первых, таких лицензий достаточно много. Во-вторых, их мониторинг достаточно затратный. В-третьих, возможность использования их программных продуктов предполагает обучение персонала, определенную инфраструктуру и проведение других мероприятий [173].

Для третьей проблемы характерны значительные временные затраты,

расходуемые на резервное копирование. Так, при скорости записи информации на жесткий магнитный диск (ЖМД) в пределах 10 Гбайт/мин, время, затрачиваемое на резервное копирование 100 Гбайт ЖМД, составит 10 мин. За это время есть потенциальная возможность внесения изменений в наборы данных, что приведет к нарушению целостности резервной копии.

Для решения третьей проблемы имеется несколько путей [186]:

1. Возможно заблокировать работу с ЖМД до завершения резервного копирования. Этот путь применялся ранее в ночное время. В настоящее время его использование проблематично. Одна из причин заключается в постоянном времени работы приложений (нет времени на резервное копирование). Вторая состоит в больших объемах резервируемой информации, что приводит к существенному уменьшению времени резервного копирования, а порой к его отсутствию.

2. Возможно запретить пропускать открытые файлы в процессе резервирования данных. Однако могут возникнуть ситуации, когда при работе важных приложений, ряд данных может быть не записан на резервный диск;

3. Возможна программная селекция процедур ввода-вывода при резервировании и функционировании инициированных приложений. Проблема заключается в том, что при реализации такого подхода используются наборы команд, не представленных в официальном описании операционной системы (ОС). Это, так называемые, недокументированные возможности операционной системы. При выпуске обновлений для одной и той же ОС, такие наборы команд могут изменяться. Поэтому, ранее разработанные процедуры программной селекции, могут не работать или, что опаснее, работать неправильно.

Существует несколько подходов резервирования активных файлов при сохранении целостности резервной копии [173].

Суть первого из них заключается в создании нескольких областей хранения данных (первичной и вторичной). В процессе резервирования активные файлы записываются во вторичную область хранения. Таким образом, копируются все файлы. Для данного подхода характерна выборочная работа.

Если есть разрешение на запись в файл подкачки, то активные файлы не обрабатываются, а помещаются во вторичную область хранения. Таким образом, реализуется целостное резервирование данных. Обработка данных вторичной области хранения осуществляется лишь после завершения сессии резервного копирования. Причем эти данные записываются поверх данных первичной области хранения.

Реализация второго подхода состоит в возможности одновременного резервирования и работы приложений. Механизм заключается в следующем. При открытии файла для резервирования реализуется возможность работы с ним остальных приложений, т.е в него разрешена запись. В интересах разделения старых и новых данных, последние записываются во вторичную область хранения. В случае их запроса приложениями, дальнейшая их обработка осуществляется драйверами ОС. Схема участия драйверов в данном процессе приведена на рисунке 5.7 [173].

В случае поступления запроса от приложения резервирования, реализуется извлечение данных из хранилища. При открытии приложением файла, ОС организует отправку набора команд на считывание метаданных. Затем она организует отправку запросов на чтение и запись логических блоков в место хранения этого файла. С помощью драйвера фильтрации ОС реализует перехват файловых операций и перенаправление запросных вызовов. Для реализации вышеизложенного подхода у компании Microsoft имеется соответствующий пакет IFS. Он позволяет с помощью набора команд реализовать вышеизложенную последовательность действий в полном объеме.

В рамках второго подхода предлагается проводить резервирование на уровне образа. Это предполагает создание драйвера поддерживающего фильтрацию на низком уровне (драйвер фильтрации диска на рисунке 5.7).

При этом взаимодействие между драйверами осуществляется следующим образом. На уровне файловой системы реализованы операции ввода-вывода (показаны стрелкой 1) Операции ввода-вывода, выполняемые драйвером NTFS на уровне блоков ЖМД, показаны стрелкой 2. Набор драйверов,

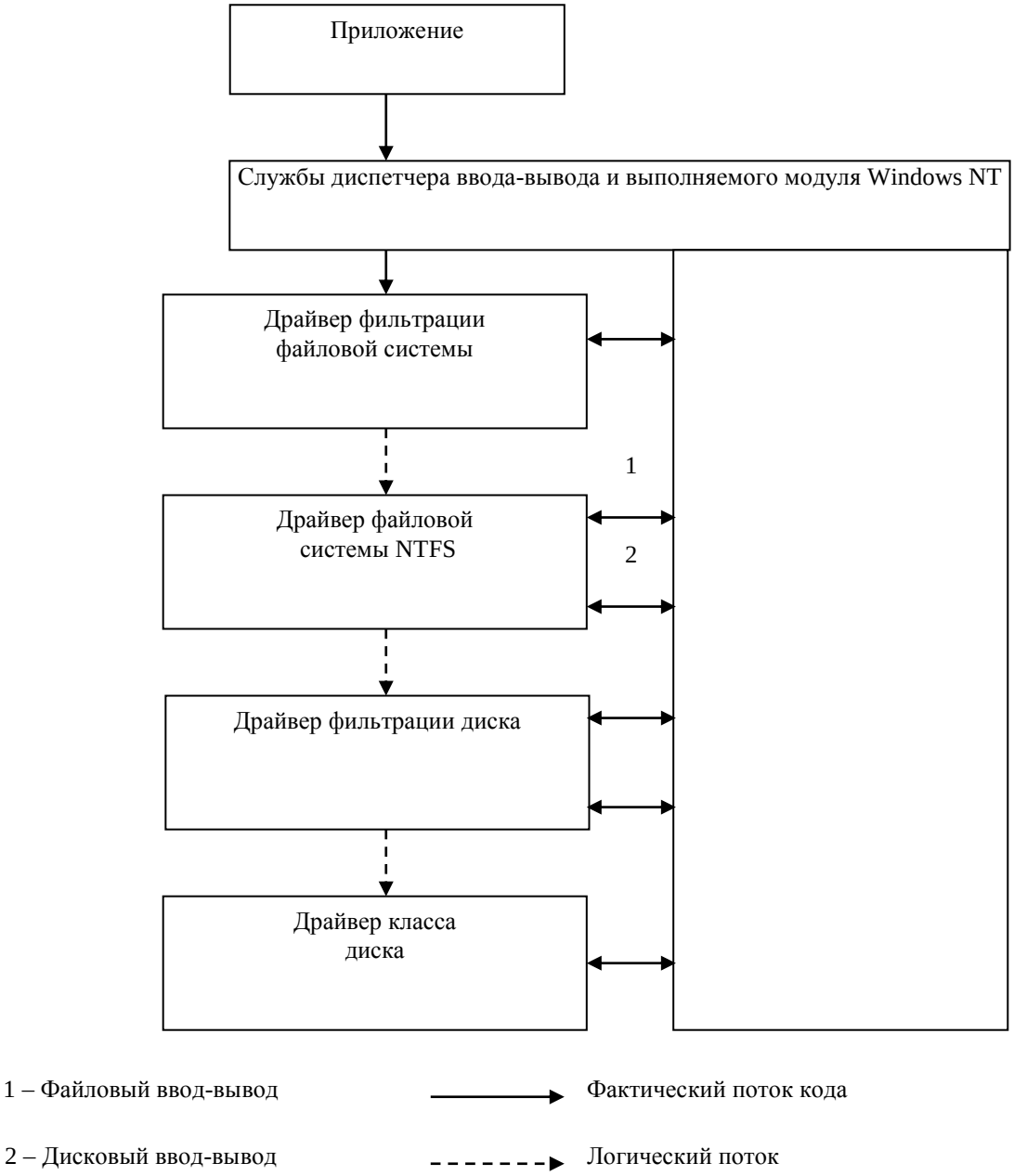


Рисунок 5.7. – Драйверы фильтрации Widows NT

реализующих данный подход, входят в состав специальной библиотеки Windows Driver Development Kit (DDK) [173].

Для третьего подхода характерно формирование моментальных снимков данных. Они копируются на резервный диск независимо от того, работают приложения копируемого диска или нет. Это достигается за счет использования оригинального тома. Формирование снимков обеспечивают различные программы и технические средства (аппаратные решения), например программные средства Microsoft Windows Server 2008 [173].

Для выбора типа резервирования, необходимо рассмотреть их классификацию. В схематичном виде она представлена на рисунке 5.8 [173].

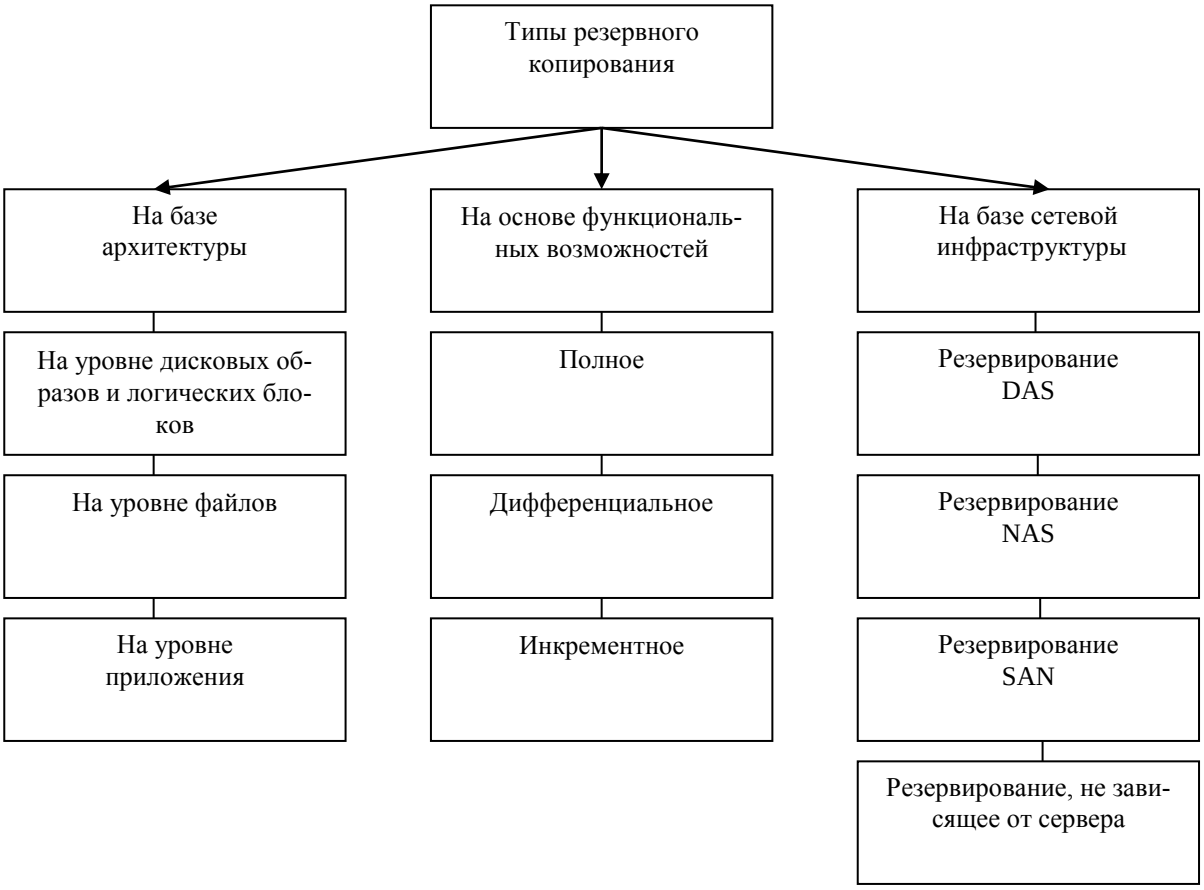


Рисунок 5.8. – Классификация резервного копирования

С учетом архитектурного признака резервное копирование подразделяется на три класса, различающихся уровнем доступа: дисковые образы и логические блоки; файлы; приложения [186].

В случае уровня дисковых образов и логических блоков в процессе резервирования участвуют блоки данных. В процессе реализации данной технологии резервного копирования взаимодействие работающих приложений с копируемыми данными не допускается. Механизм проведения данного вида резервирования состоит в следующем. Независимо от внутренней структуры ЖМД приложение резервирования получает к нему доступ (ЖМД). Затем оно на уровне блоков логического диска выполняет операции типа «чтение-запись». Достоинством данного типа резервирования является высокая оперативность резервирования и восстановления. Это имеет большое значение в си-

туациях с критическими сбоями в ПК. Данный тип резервирования имеет ряд недостатков. Во-первых, отсутствует доступ ОС и приложений к диску во время проведения резервирования. Во-вторых, избыточное копирование логических блоков с резервной копии (включаются неиспользуемые блоки). Для парирования второго недостатка ряд приложений резервирования включают логику, которая обнаруживает и пропускает логические блоки неиспользованные пользователем. В результате формируются разреженные копии диска [186].

На файловом уровне доступа функционирование программ резервирования обеспечивается операционной и файловой системами. Достоинствами данного типа резервирования являются. Во-первых, высокая эффективность, как файла, так и их набора. Во-вторых, при резервировании не прекращается доступ к данным, как ОС, так и приложений. Недостатками данного типа резервирования являются. Во-первых, более длительное время резервирования, по сравнению с вышеизложенным типом. Во-вторых, при копировании большого числа файлов малой размерности, существенно возрастает нагрузка на ОС и файловую систему. В-третьих, не все файлы закрываются после проведения процедуры резервирования.

Два вышеизложенных подхода к резервированию имеют общий недостаток. Реализация резервного копирования возможна лишь с правами доступа администратора. Поэтому файлы метаданных, используемые в интересах восстановления данных, должны быть корректно настроены, что обеспечивается дополнительной нагрузкой API на ОС и файловую систему. При этом немаловажную роль играет возможность корректного применения возможностей предоставляемых API [186].

Резервное копирование на уровне приложения проводится с полным участием API приложения. Для данного вида резервирования копия включает набор файлов и объектов. Последние являются своего рода «слепком» формирующим состояние системы в течение определенного времени. Основной недостаток данного вида резервирования связан с тесной связью операций резервного копирования и восстановления данных с разработанным прило-

жением. Как правило, если выходит новое приложение резервного копирования, то изменятся его API. Поэтому пользователь вынужден переходить на новый пакет API (новой версии). При установке приложения используется чистый диск, в котором не инициализирована файловая система. Кроме того, на чистый диск может быть записан большой по объему файл с собственными метаданными самого приложения. Для восстановления подобных файлов в ОС (Windows XP и Windows Server 2003), используется ряд функций NTFS. Восстановление файла происходит поблочно, с последующей маркировкий его конца специальной функцией SetFileValidData API Win32.

Дальнейший анализ рисунка 5.8 показывает, что по своим функциональным возможностям, резервирование может быть полным; дифференциальным и инкрементным [173].

Для полного резервирования характерно копирование на носитель резервной копии полного набора файлов или объектов. Кроме того, копируются метаданные, которые с ними связанны. Достоинством данного способа является использование единственного набора носителей (дисков) для восстановления данных. Это важно в случае отказа работы системы. Недостаток данного способа заключается в более длительном времени копирования. Это связано с необходимостью копирования всех данных. Данный способ резервирования наиболее часто используется на уровнях образа диска или блоков [186].

В рамках дифференциального резервирования упаковываются все изменения, произошедшие после последнего резервирования, которое проводилось в полном объеме. Копии дифференциального резервного копирования создаются, либо на уровне образа, либо на файловом уровне. Поэтому множество изменений на уровне образа будет представлено набором дисковых блоков, которые изменились. Для множества изменений на файловом уровне, это будет соответственно набор измененных файлов. Достоинством дифференциального резервирования является более низкая временная затратность по сравнению с полнообъемым резервным копированием. Его недостаток состоит в увеличении временных затрат на восстановление после сбоя. Это свя-

занно с тем, что в результате сбоя необходимо будет выполнить две группы операций, связанных с восстановлением данных. В рамках первой группы операций необходимо восстановление данных из полной резервной копии. В рамках второй группы операций необходимо восстановление данных из дифференциальной резервной копии. Целесообразность проведения дифференциального резервирования на уровне файлов возникает в тех ситуациях, когда приложениями создается множество малоразмерных файлов, а по завершении формирования полномасштабной резервной копии, изменяются лишь некоторые из них. Следует отметить, что дифференциальное резервирование не эффективно при работе с БД. В этом случае ЖМД практически постоянно взаимодействует с СУБД. Они достаточно часто вносят незначительные изменения в большие файлы БД. Следовательно, при таком резервировании на файловом уровне создается копия полнообъемного файла. В качестве примера такой программы можно рассматривать Microsoft Exchange. Она практически постоянно вносит незначительные изменения в большие файлы БД. Если в подсистемах хранения данных применяются более старшие модели резервного копирования, то дифференциальное резервирование может использоваться достаточно широко, в том числе и на уровне образа. При этом файлы приложений БД, также могут резервироваться. Эффективность такого подхода связана с тем, что при нем хранится большой объем метаданных, позволяющих оперативно найти дисковые блоки, которые подверглись изменениям с момента резервного копирования. Следовательно, реализуется резервирование (копирование) дисковых блоков подвергшихся изменениям. Остальные неизмененные блоки не копируются. Несмотря на высокую эффективность резервирования в рамках применения старших моделей хранения данных во внутренних подсистемах, актуальность применения API остается. Ее достоинство заключается в том, что она позволяет начать резервное копирование в определенный момент времени, а после его завершения продолжить операции ввода-вывода данных. Тем не менее, метод использования старшей модели в подсистемах хранения бывает полезен, и позволяет сократить набор лишних операций ввода-вывода данных при резервировании [173,

186].

В рамках инкрементного резервирования архивации подлежат только изменения, произошедшие с момента последнего, либо полного, либо дифференциального резервирования. Достоинством данного вида резервирования является уменьшение временных затрат. Это связано с тем, что файлы, которые не претерпевали изменений после создания последней полномасштабной или дополнительной резервных копий, на резервный диск не копируются. Недостаток данного метода состоит в повышенной длительности процедуры восстановления после произошедшего сбоя. Поскольку в ее реализации участвует набор, включающий несколько носителей. Это может быть последняя полномасштабная резервная копия и ряд дополнительных резервных копий. Когда старшие модели не используются в подсистеме хранения, дополнительное резервирование активируется при внесении изменений или дополнений в различные файловые наборы. В случае применения старших моделей в подсистемах хранения, возможно применение дополнительного блочного резервного копирования. Поскольку в этой ситуации доступен большой объем метаданных, которые могут быть использованы для идентификации блоков, претерпевших изменения.

На основе сетевой платформы имеется возможность реализации 4 типов резервного копирования: DAS, NAS, SAN и резервирование, которое не зависит от сервера [173].

Наиболее ранним резервированием является DAS. Оно возникло во времена непосредственного подключения аппаратных средств хранения к серверной станции. Данный тип резервирования остается популярным и настоящее время, применительно к серверам Windows. Безусловным преимуществом этого типа резервирования является его простота реализации. В данном случае, серверное приложение считывает информацию с определенного тома ЖМД и помещает их на резервный носитель. Недостатками резервирования DAS являются следующие. Во-первых, это достаточно затратный вид резервирования, поскольку требует несколько накопителей для резервного копирования (один на сервер). Во-вторых, одновременная эксплуатация единого ЖМД несколь-

кими серверами вызывает проблемы. В-третьих, в связи с тем, что ряд данных размещенных на нескольких серверах могут дублироваться и не синхронизироваться, при копировании похожих данных на разные носители, возможно их перепутывание. При этом ресурсы сервера будут постоянно заняты, поскольку при «чтении-записи» данных между ЖМД и накопителем он должен обрабатывать соответствующие запросы [186].

В рамках резервного копирования NAS имеется возможность применения одного внешнего накопителя для резервирования данных размещенных на нескольких серверах. Это обеспечивает более низкие материальные затраты. Тем не менее, данный вид резервирования не лишен недостатков. Во-первых, необходимость сегментации ЛВС с целью перенаправления резервируемых потоков в единый сегмент сети, приводит к уменьшению пропускной способности ЛВС. Резко увеличивается время технологической работы узлов, т.е существенно уменьшается время обслуживания запросов и транзакций пользователя. Во-вторых, возрастает объем данных предназначенных для хранения на сервере. Это требует увеличения времени резервирования таких данных [186].

Технология резервирования SAN базируется на положении о том, что сетевое хранение данных способно обеспечить требуемую пропускную способность между устройствами хранения. В зависимости от топологии она должна быть способна предоставить оперативную связь между рядом пар устройств. Однако известно, что применение топологии кольца с числом объектов превышающих 30 не способно реализовать вышеизложенное требование. Это связано с тем, что в кольцевой структуре сети, общая пропускная способность равномерно распределяется между всеми объектами. К достоинству данной технологии можно отнести возможность удаленного размещения внешнего накопителя от сервера с резервируемыми данными. Как правило, подобные накопители оснащаются оперативным интерфейсом обработки SCSI. Однако это ограничивает варианты их подключения (только к шине SCSI). Поэтому возникают сложности при совместном использовании одного накопителя рядом серверов. Данная проблема разрешается путем использо-

вания компонент Fibre Channel. Основная задача состоит в организации системы разрешений, обеспечивающих организацию корректного доступа к внешнему ЖМД. Одним из примеров подобной организации является метод зонирования. Он обеспечивает получение доступа одним сервером к внешнему ЖМД в заданный временной период (сессию). При этом имеются проблемы обеспечения требований зонирования для серверов. Их суть в корректном обеспечении переключения ряда внешних накопителей. Другим примером организации корректного доступа является применение интерфейсных команд SCSI, таких как Reserve и Release. Еще одним методом является применение специального серверного программного обеспечения, обеспечивающего совместный доступ внешнего ЖМД и сервера. Это значительно уменьшает стоимость резервирования. Примером внешнего накопителя может служить ЖМД Tivoli фирмы IBM.

Разновидностью реализации технологии SAN, может быть резервирование копирования без локальной сети. В данном случае данные передаются средствами SAN вне пределов ЛВС. При этом снижается нагрузка на ЛВС и возрастает ее пропускная способность, в том числе и для решения задач резервного копирования данных. Резервирование без ЛВС экономит использование ресурсов за счет совместного применения накопительных устройств. За счет одновременного проведения резервного копирования несколькими устройствами, возникает меньше ошибок и оперативнее проводится восстановление данных. Возникает возможность более эффективного ресурсного планирования. Более того, за счет высокой скорости передачи данных, процесс резервирования и восстановления данных отличается более высокой оперативностью. Вышеизложенный вид резервирования определяют как резервное копирование без сервера. Иногда его называют сторонним копированием [186].

Функциональная схема резервирования данных в ИСППИД показана на рисунке 5.9 [173].

В техническом исполнении резервное копирование реализовано на базе единого сервера под управлением операционной системы Windows Server

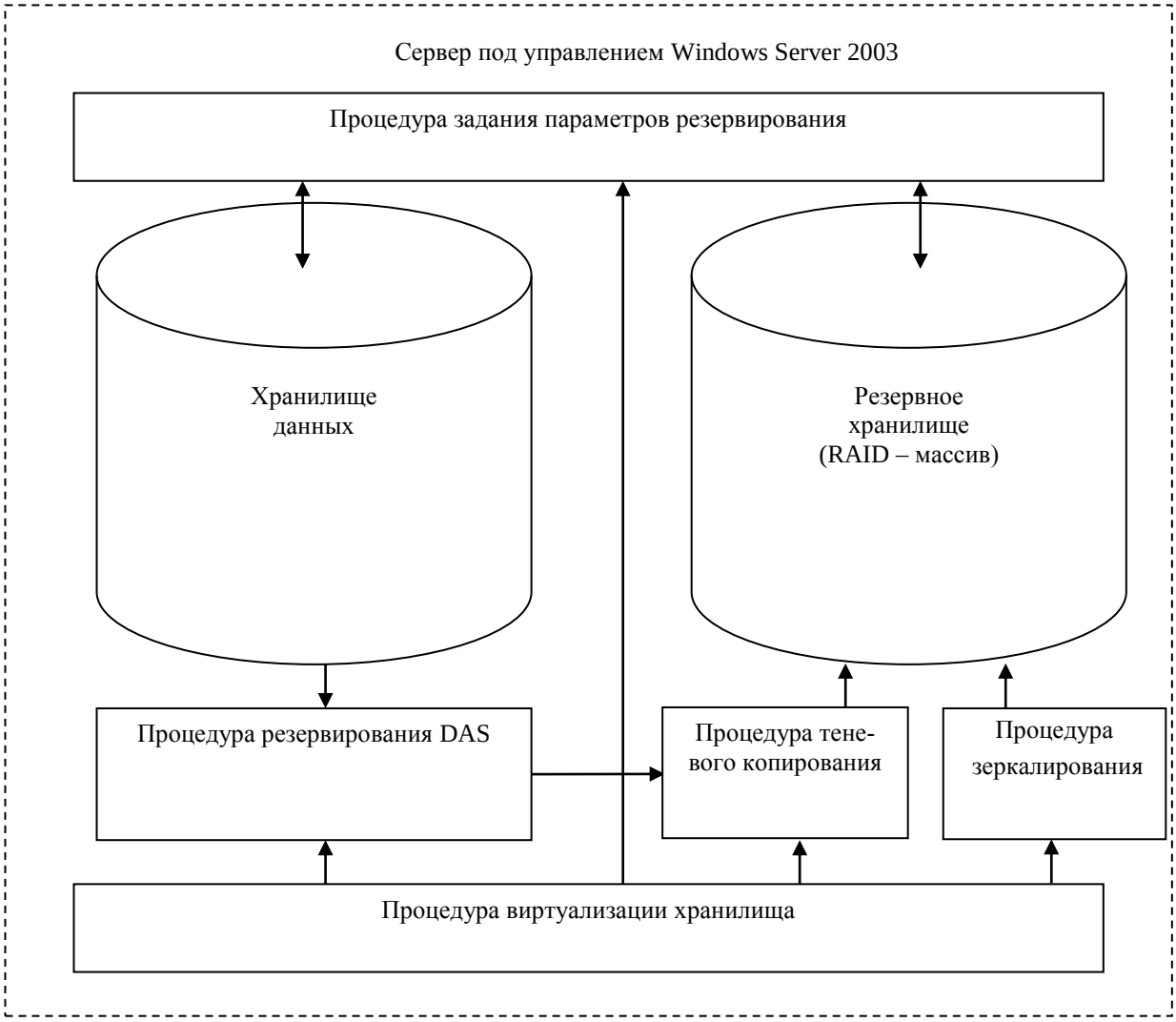


Рисунок 5.9. – Функциональная схема резервного копирования данных

2003. В состав соответствующего модуля входят следующие процедуры [173]:

- задания параметров резервирования;
- резервирования DAS;
- теневого копирования;
- зеркалирования;
- виртуализации хранилища.

Резервное хранилище включает 4 ЖМД емкостью 2000 ГГб каждый.

В рамках процедуры задания параметров резервирования системный администратор вводит значения требуемых параметров резервирования, таких как [173]:

- количество месячных резервных копий;
- количество еженедельных резервных копий;

- количество ежедневных резервных копий;

- наиболее целесообразный временной интервал резервирования данных;

- частота формирования теневых копий;

- размер места, отведенного для теневых копий.

Процедура резервирования DAS (direct-attached-storage) используется для реализации соответствующего типа резервирования. Выбор данного типа обусловлен его простотой реализации и относительно невысокой стоимостью. Поскольку резервное хранилище реализовано на жестких магнитных дисках, а не на накопителях на магнитных лентах, данный тип резервного копирования оказался практически лишенным недостатков [173].

Процедура теневого копирования реализует получение так называемых «Теневых копий» [222]. Системный администратор включает и настраивает службу Volume Shadow Copy, входящую в состав используемой операционной системы Windows Server 2003. Чтобы разрешить на сервере теневое копирование, вызывается окно свойств тома, осуществляется переход на вкладку «Теневые копии (Shadow Copies)», выбирается том и активизируется опция «Включить (Enable)». В результате пользователь имеет возможность доступа к различным версиям документа в соответствии с привязкой к определенному моменту времени. Само теневое копирование производится «на лету», независимо от того, открыт файл или нет. В этом режиме Windows Server 2003 копирует все файлы в общих папках и помечает их атрибутами с текущей датой и временем. Пока разрешено теневое копирование тома, Windows Server 2003 будет ежедневно создавать по две копии файлов. Этот процесс продолжится до исчерпания на диске места, отведенного под теневые копии. Вот почему в процедуре задания параметров резервирования важно настроить частоту копирования и размер места, отведенного для теневых копий [173].

Теневое копирование томов позволяет создавать библиотеки версий файлов, тем самым обеспечивая доступ к сохраненным версиям удаленных или поврежденных файлов, что избавляет системных администраторов от

одной из наиболее обременительных задач - восстановления одиночных файлов, ошибочно удаленных пользователями [173].

Обеспечение надежности и устойчивое восстановление данных по причине сбоев достигается за счет хранения данных в разных секторах и постоянном их обновлении. Периодическое обновление данных приводит к тому, что если они меняются в одной реплике, то автоматически меняются и в других. На этом принципе построены зеркальное отражение и репликация. Зеркалирование – технология, обеспечивающая копирование данных с одного носителя на другой в реальном времени. Она реализована в процедуре зеркалирования и обеспечивает систему хранения с резервированием всех данных. Технология зеркалирования тесно связана с формированием массивов RAID (Redundant Array of Independent Disks). RAID представляется как избыточный массив некоторого числа независимых дисков. Ключевым понятие в технологии RAID является чередование. Оно включает определение размера единицы «ввода-вывода» (аналог кластера). На практике эта величина лежит в пределах от 512 байт до 4 или 8 Мбайт. После инициализации размера осуществляется их (единиц) физическое дисковое размещение. Логические блоки для высокоуровневых кластеров формируются аналогично. Последовательность их размещения следующая: на диске 1 находится первая единица, на диске 2 – вторая и т. д [186]. Реализация массивов RAID может быть аппаратной и программной. Использована аппаратная реализация массива RAID. Аппаратный RAID - устройство RAID, созданное и поддерживаемое на аппаратном уровне. Для операционной системы такое устройство выглядит как один физический жесткий диск - отдельные диски, составляющие данное устройство, системе невидны. В силу удобства и надежности, несмотря на финансовые затраты, данное решение было реализовано.

Процедура виртуализации хранилища обеспечивает предоставление единого интерфейса для управления коммутаторами связной архитектуры. Данный интерфейс не зависит от производителя коммутатора, его модели и возможностей. Последовательность действий системного администратора

при проведении резервного копирования заключается в следующем [173]:

- формирование тома (возможна необходимость управления RAID-массивами или дисками);

- инициализация механизма формирования моментальных снимков (возможна перенастройка зонирования);

- формирование моментального снимка;

- инициализация тома с моментальным снимком для сервера резервирования;

- реализация резервирования;

- возврат моментального снимка текущих данных используемых в системе в набор свободных ресурсов хранилища (возможна дополнительная настройка коммутатора).

Реализация вышеизложенной последовательности действий осуществляется так: выполнение части операций администратором вручную путем набора команд в командной строке; запуск программного обеспечения (путем активации исполнительного файла); набор дополнительных операций в командной строке и т.д [173].

Процедура виртуализации хранилища позволяет полностью автоматизировать эти задачи.

5.7. Реализация ввода данных

Модуль ввода данных входит в состав каждой ВД и обеспечивает загрузку исходных данных, требуемых для работы соответствующего приложения. Применительно к НМХ это: число временных интервалов и их протяженности, значения исходного временного ряда и др. Данный модуль также обеспечивает загрузку данных из различных источников, выбираемых пользователем, в ХД.

Модуль автоматизированной загрузки данных предназначен для оперативного ввода больших объемов данных (курсов акций, котировок валют и др.) в автоматическом режиме. В качестве примера на рисунке 5.10 приведена блок-схема алгоритма автоматизированного ввода информации о котировках акций.

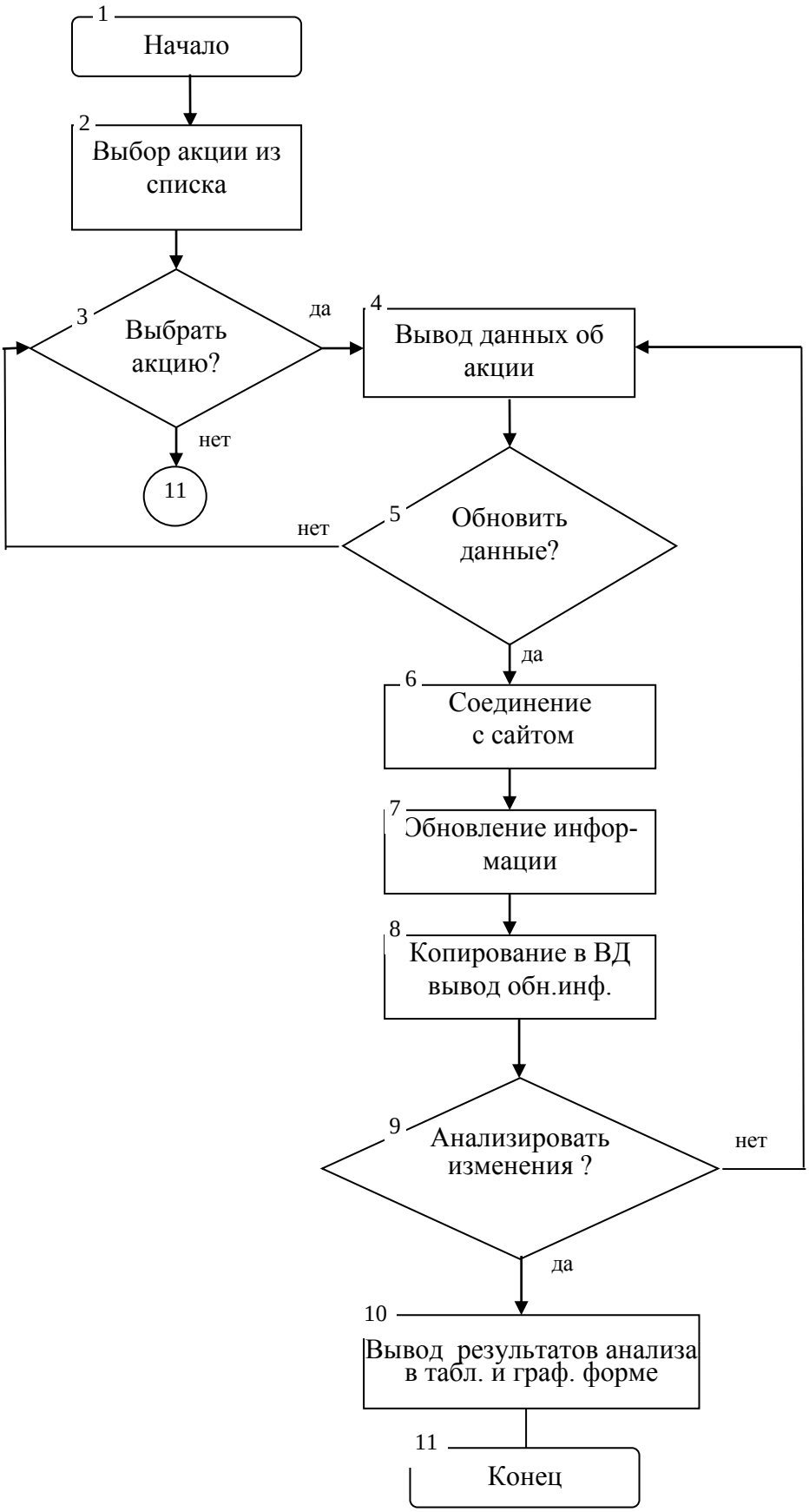


Рисунок 5.10. – Блок-схема алгоритма автоматизированного ввода информации о котировках акций

Блоки 1,11 обеспечивают пуск и остановку данного алгоритма.

В блоке 2 реализован выбор требуемой акции из списка.

В блоке 3 если ЛПР намерен работать с выбранной акцией, то управление передается в блок 4. В противном случае, в блок 11.

Блок 4 обеспечивает вывод характеристик выбранной акции.

В блоке 5 если ЛПР намерен обновить информацию по выбранной акции, то управление передается в блок 6. В противном случае, в блок 3.

Блок 6 предназначен для соединения с одним из сайтов, предоставляющих информацию о данном типе акции.

В блоке 7 реализовано обновление характеристик выбранной акции.

Блок 8 обеспечивает запись обновленной информации в базу данных.

В блоке 9 если ЛПР анализирует информацию об акции, то управление передается в блок 10. В противном случае, в блок 4.

Структурно – функциональная схема модуля автоматизированного ввода данных приведена на рисунке 5.11.

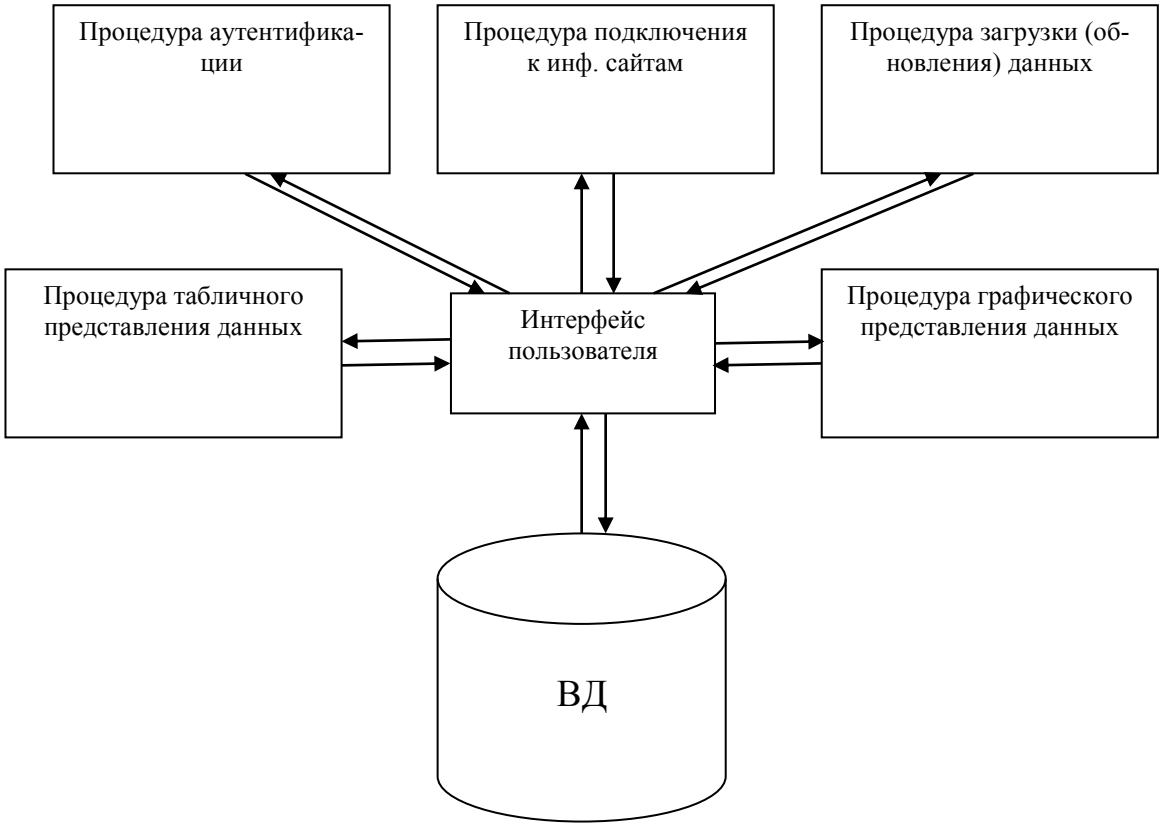


Рисунок 5.11. – Структурно-функциональная схема модуля автоматизированного ввода данных

Интерфейс пользователя это набор процедур, обеспечивающих взаимодействие ЛПР с программным модулем.

Процедура аутентификации обеспечивает проверку пароля.

Процедура подключения к информационным сайтам обеспечивает выход пользователя на сайты, содержащие информацию о вводимых данных.

Процедура загрузки (обновления) данных обеспечивает первичный ввод исходных данных либо их обновление по желанию ЛПР. Данные получают либо из сайта, либо из ВД. процедура включает форму выбора номенклатуры вводимой ЦБ из предложенного списка, а также таблицу статистики цен, показывающую данные об изменении котировок.

Процедура табличного представления данных обеспечивает формирование таблицы, в которой размещены исходные данные о ЦБ.

Процедура графического представления данных реализует отображение введенных данных (характеристики акций) в графическом виде.

Внешний вид окна главной экранной формы модуля автоматизированного ввода данных с открытой вкладкой «Статистика цен», приведен на рисунке 5.12.

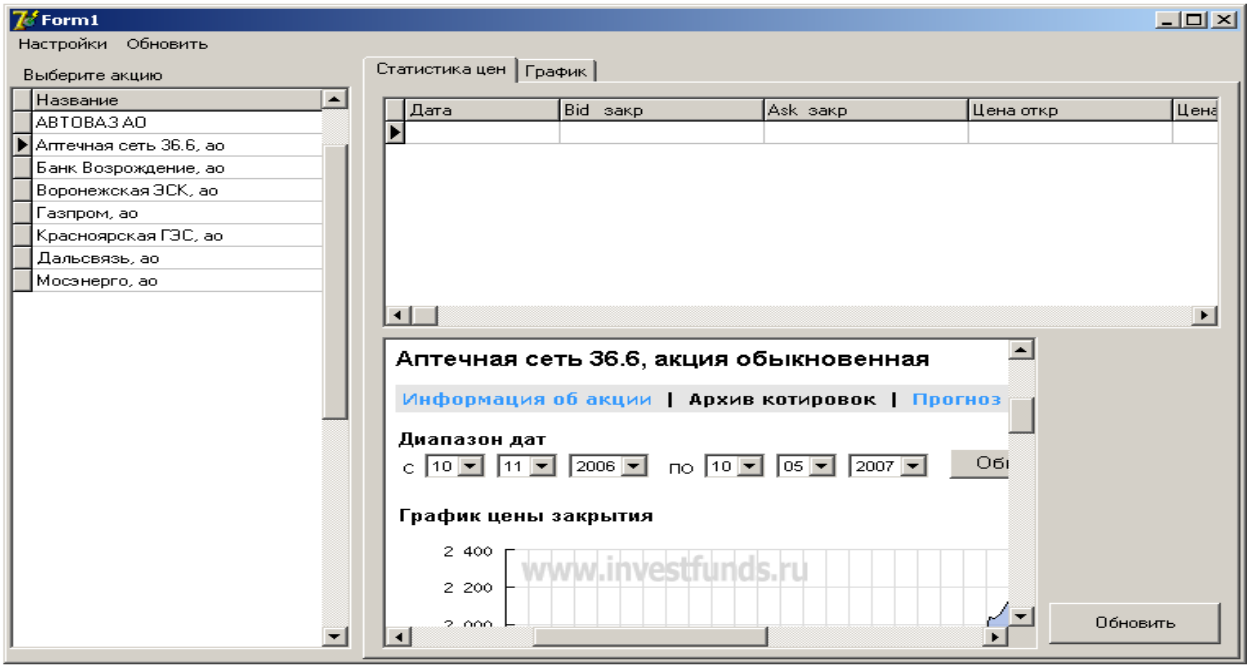


Рисунок 5.12. – Вид главной формы автоматизированного ввода данных

ВЫВОДЫ

Подсистема хранения и обработки данных разработана в интересах информационного обеспечения основных функциональных подсистем ИСППИД: анализа инвестиций, исследования временных рядов и терминологического информационного поиска. В ее состав входят: хранилище данных, загрузочная секция, набор витрин данных (поддержки НМШ и НКМШ, исследования ИНС, терминологического информационного поиска, других OLAP приложений) и программных модулей (эвристической оптимизации хранения данных, ввода данных, накопления и очистки данных, контроля качества данных, оптимизации витрин данных (ВД) и доступа к данным, архивации данных, резервирования данных, формирования выходных результатов).

Хранилище данных предназначено для долговременного хранения данных. Оно соответствует его классическому пониманию [95] и представляет собой специальным образом администрируемую базу данных. Для эвристической оптимизации ХД разработан соответствующий алгоритм, базирующийся на его нормализации (до 3-й нормальной формы) и индексировании хранимых данных. Апробация разработанного алгоритма показала его эффективность и показала, что среднее время обработки многоаспектных запросов меньше среднего времени обработки одноаспектных запросов. Следовательно, в СУБД MS Access целесообразнее осуществлять многоаспектный поиск.

Витрина данных обеспечивает работу с актуальными данными. Она представляет собой интегрированный, статичный, поддерживающий хронологию набор данных из разных источников, используемый для поддержки принятия инвестиционных решений. ВД базируется на нормализованной реляционной модели хранения данных. В состав ПСХОИ входит несколько витрин данных (поддержки НМШ и НКМШ, исследования ИНС, терминологического информационного поиска, OLAP приложений). Применение нескольких ВД обусловлено необходимостью поддержки различных форматов данных в соответствующих приложениях. Для обмена данными между хра-

нилищем и витриной реализован соответствующий алгоритм, основанный на эмпирической оценке закона их ранговой популярности, повышающий надежность хранения данных и эффективность манипулирования ими. Кроме того, для повышения оперативности обработки запросов разработан модифицированный алгоритм индексации данных, на основе их предварительной обработки методом «деления узла». Контрольные примеры подтвердили высокую эффективность разработанных алгоритмов.

Технологическое обеспечение ПСХОИ включает архивацию данных; резервное копирование; автоматизированный ввод данных. По результатам проведенных экспериментов, для архивации данных при пересылке последних из оперативного хранилища в долговременное хранилище, был выбран алгоритм 7-Zip. Для реализации резервирования выбрана технология DAS (direct-attached-storage) по причине простоты реализации и невысокой стоимости. В состав данной подсистемы вошли модули: задания параметров резервирования; резервирования DAS; теневого копирования; зеркалирования; виртуализации хранилища. Автоматизированный ввод данных реализован на таких процедурах, как: аутентификации; подключения к информационным сайтам; загрузки (обновления) данных; табличного представления данных; графического представления данных.

6. МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ВНУТРИСИСТЕМНОГО СИНТЕЗА, ОЦЕНКА ЭФФЕКТИВНОСТИ И РАЗВИТИЕ ИНФОРМАЦИОННОЙ СИСТЕМЫ ПОДДЕРЖКИ ПОРТФЕЛЬНОЙ ИНВЕСТИЦИОННОЙ ДЕЯТЕЛЬНОСТИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЙ ОРГАНИЗАЦИИ

6.1 Характеристика прототипа информационной системы поддержки портфельной инвестиционной деятельности

Для первичной апробации качества синтезированного облика ИСППИД разработан соответствующий прототип, представляющий собой первоначальный вариант ИСППИД с ограниченным функционалом, призванный проверить работоспособность, заложенных в ней идей, на практике. В перспективе, прототип доводится до полнофункционального лабораторного образца, подготовленного для дальнейшего промышленного изготовления и внедрения. Разрабатываемый прототип строится как открытая модульная система, реализующая ее основной функционал. В дальнейшем она доводится до полномасштабного образца, путем разработки и встраивания модулей, реализующих дополнительный или новый функционал.

Структурно-функциональная схема разработанного прототипа приведена на рисунке 6.1 [150].

В состав прототипа входят подсистемы формирования ИП, прогнозирования значений параметров ЦБ, терминологического информационного поиска, а также сбора, хранения и обработки информации. Кроме того, в его состав включены несколько отдельных модулей, таких как: системного управления, ввода данных, автоматизированного ввода данных и формирования выходной информации [150].

Подсистема формирования ИП СЭО обеспечивает определение характеристик ЦБ, являющихся кандидатами для включения в состав ИП, которые могут содержать детерминированные и стохастические значения параметров, расчет основных характеристик комплектуемых ИП и выбор оптимальных кандидатов по заданным критериям [150].

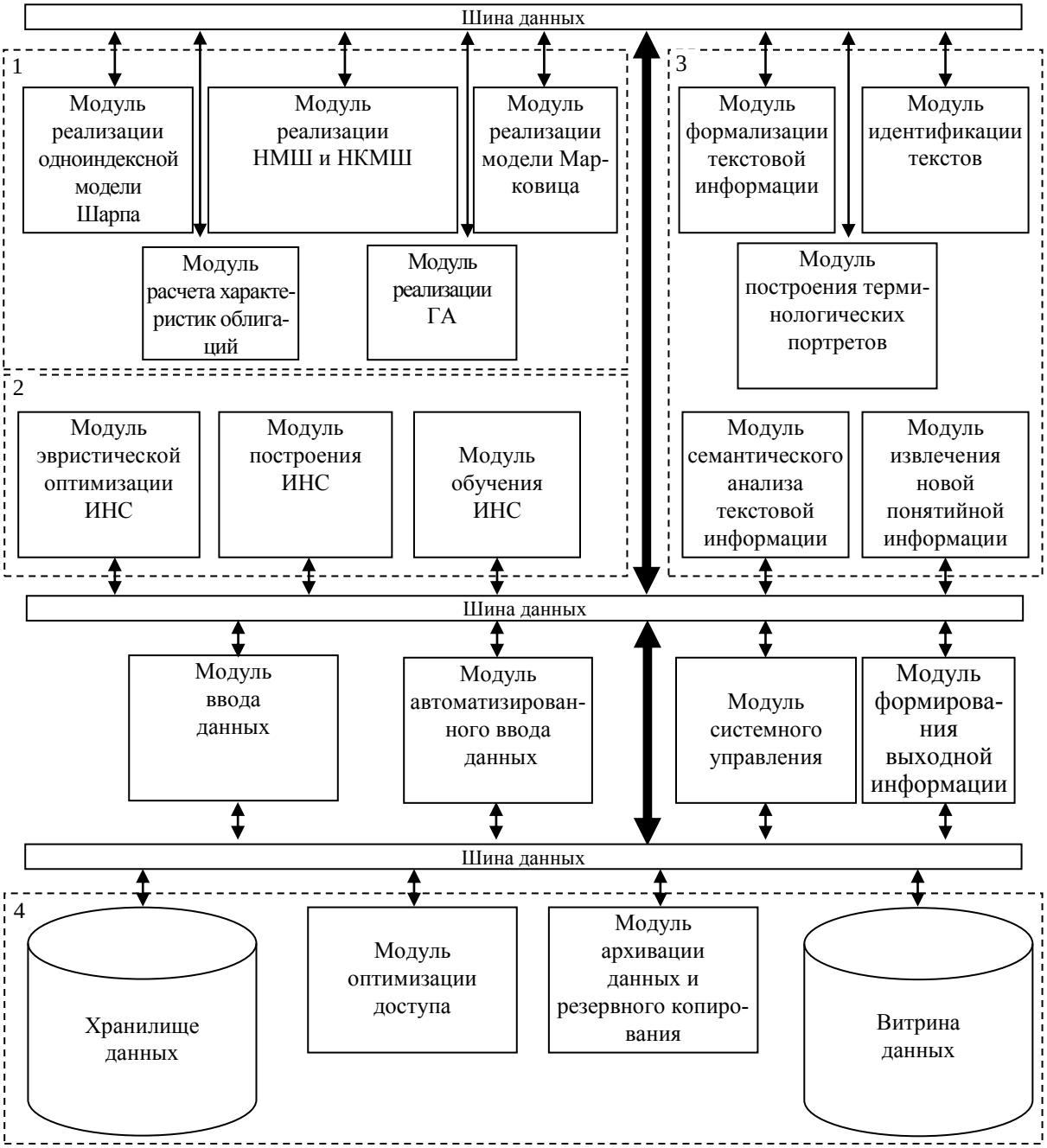


Рисунок 6.1 – Структурно-функциональная схема прототипа ИСППИД СЭО

Подсистема прогнозирования значений параметров ЦБ предназначена для обеспечения функционирования НМШ [132] и НКМШ [133], которые используются для оптимизации ИП применительно к активам, содержащим динамические стохастические данные. Функциональные модули подсистемы обеспечивают проведение расчета значений недостающих данных (дискрет) в динамически изменяющихся стохастических данных на базе математической аппроксимации функций, которая реализована в НМШ и НКМШ [132, 133].

Подсистема терминологического информационного поиска позволяет реализовать добычу прагматической (полезной) информации об особенностях и свойствах применения активов в разных областях экономической деятельности с целью выбора из них в наибольшей степени целесообразных в интересах размещения в ИП СЭО. Кроме того, данная подсистема обеспечивает поиск информации о перспективном и существующем методическом и информационном обеспечении процесса оптимизации ИП СЭО. Источниками информации выступают интернет-ресурсы и различные локальные БД [150].

Функциональным назначением подсистемы сбора, хранения и обработки информации является информационное обеспечение процессов функционирования подсистем ИСППИД СЭО. Ее состав включает витрину данных, хранилище данных и набор сервисных модулей [101].

Модуль ввода данных реализует загрузку различных групп наборов данных (общих расчетных данных, данных для настройки ГА, ИНС, ОМПВ) из разных источников (текстовых файлов, клавиатуры, баз данных).

Назначение модуля автоматизированного ввода данных заключается в вводе определенных множеств данных о характеристиках ЦБ (их минимума, максимума, значений открытия и закрытия, объема продаж) из интернет-источников (например, cbonds.ru, finam.ru и др.), их преобразования в требуемый формат описания и загрузку в ВД / ХД [150].

Модуль системного управления обеспечивает синхронизацию функционирования подсистем и модулей прототипа в соответствии с логической схемой его работы. В процессе управления участвуют, как стандартные процедуры общего программного обеспечения (операционной системы, СУБД и др.), так и специальные процедуры и функции (подсистемные, модульные и интерфейсные) в рамках сформированной логической схемы внутрисистемной работы и интерфейсного взаимодействия с ЛПР [150].

Модуль инициализации выходной информации предназначен для формирования и компоновки выходных информационных пакетов, содержащих опи-

сание состояний основных структурных компонентов (подсистем и модулей), динамических параметров ИП, рекомендаций по управлению ИП и структурными компонентами, представляемых в виде, выбираемом ЛПР (текстовом, графическом, табличном).

Разработанный прототип ИСППИД может работать в различных режимах. Логика изменения состояний системы следующая: «предварительное→исполнительное». Это означает, что ИСППИД сперва настраивается, а затем переводится в один из рабочих режимов. В данной версии прототипа предусмотрено пять режимов работы: «Настройка»; «Тестирование»; «Ручной»; «Автоматизированный»; «Автоматический». Два первых режима характерны для предварительного состояния функционирования системы, а оставшиеся три для ее исполнительного состояния.

Режим «Настройка» используется на начальном этапе функционирования системы и состоит в ее конфигурировании (выборе используемых компонентов) и задании начальных параметров настройки выбранных функциональных модулей. В подсистеме формирования ИП СЭО выбираются используемые модели (Марковица, Шарпа, НМШ, НКМШ, ГА) или модуль расчета характеристик облигаций, путем выбора соответствующих опций в меню подсистемы. В подсистеме прогнозирования значений параметров ЦБ выбираются соответствующие значения внутренних параметров, представленных в меню. Если выбраны НМШ или НКМШ, то вызывается меню ПК НПВР и осуществляется выбор опций из его интерфейса (например: «Инициализация МП», «Сформированный МП», «Визуализация структуры», «Задание эвристических параметров оптимизации», «Задание режима обучения» и др.). Если выбран ГА, то вызывается меню данного программного модуля и осуществляется его настройка. В интерфейсе подсистемы терминологического информационного поиска выбираются свои опции («Интернет», «Локальные БД», «Настройка ТПИС», «Настройка ИПА», «Настройка виртуализации» и др.). В подсистеме сбора, хранения и обработки информации осуществляется настройка на вид обрабатываемых дан-

ных («Детерминированные», «Стохастические», «Комбинированные»), используемый источник поступления («Хранилище данных», «Интерактивный ввод», «Ввод с сайта» и др.), оптимизацию доступа («Ранговая оптимизация»).

В режиме «Тестирование» проверяется работоспособность всех подсистем и модулей ИСППИД СЭО с выдачей полученных результатов ЛПР. В случае каких-либо отклонений система переводится в ручной режим работы и осуществляется ее донастройка.

Режим работы «Ручной» по своему функциональному назначению близок к режиму «Настройка», однако он реализуется на исполнительном этапе функционирования ИСППИД. Близость заключается в том, что все управление системой ЛПР берет на себя (управляет вводом-выводом данных, выбирает модели, выбирает форму представления результатов, может организовать индивидуальную работу с конкретной подсистемой или модулем, разбираясь самостоятельно в особенностях ее функционирования (например, он может просмотреть промежуточные данные, ввести новые расчетные формулы и др.), проводит корректировку выдаваемых промежуточных и финальных сообщений). Данный режим работы предназначен для системных разработчиков и высококомпетентных ЛПР.

Режим работы «Автоматизированный» является основным исполнительным режимом, используемым ЛПР после того как система прошла настройку и тестирование на множестве исторических данных и показала хорошие результаты по точности и оперативности. Особенностью данного режима является его диалоговый характер. После проведения определенных действий система предлагает ЛПР некоторые рекомендации:

- на этапе пересмотра ИП она предлагает рекомендации по изменению его структуры: какие ЦБ продать, какие купить, какие оставить и др.;

- в случае ухудшения точности или оперативности предлагается переход на другие модели;

- в рамках терминологического поиска представляется объяснение своих

действий (почему выбран тот или иной термин, в чем его суть и новизна, из какого источника он получен) и др.

Режим работы «Автоматический» является перспективным исполнительным режимом и применяется в том случае, когда система находится практически в идеальном состоянии, РЦБ показывает стабильное развитие (низкая волатильность ЦБ), используется консервативная стратегия управления ИП (применяются ЦБ с низким риском) и др. В данном режиме система является фактическим аналогом торговой системы. В силу вышеизложенных особенностей, данный режим используется редко.

Для использования разработанного прототипа требуются следующие технические и программные средства:

- ПК, включающий следующее множество конфигураций процессора: Celeron/AMD/Pentium/CoreDuo:*i3/i5/i7*;

- используемые операционные системы MS Windows NT-2000-XP-7;

- СУБД: Access 2007, MYSQL.

Для установки прототипа и его эксплуатации требуется следующая минимальная конфигурация оборудования: процессор Pentium III (3,0 ГГц); объем ОЗУ – 4096 МБ; объем свободного пространства на жестком магнитном диске 1 ТБ (с учетом резервирования 5 ТБ); оптическая мышь; графическая карта SVGA с объемом видеопамяти 1024 Кб; SVGA (LSD) – цветной монитор.

Для эффективной эксплуатации прототипа требуется следующая конфигурация оборудования: процессор CoreDuo *i7*; объем ОЗУ – 8192 МБ; объем свободного пространства на жестком магнитном диске 2 ТБ (с учетом резервирования 10 ТБ); оптическая мышь; графическая карта SVGA с объемом видеопамяти 2048 Кб; LED (LSD) – цветной монитор с экранной диагональю более 21дюйма.

После разработки прототипа ИСППИД СЭО представляет интерес оценка его эффективности.

6.2. Методика интегральной оценки эффективности информационной системы поддержки портфельной инвестиционной деятельности в составе социально-экономической организации

Известно, что в технологическом плане оценка эффективности системы в составе СЭО заключается в определении частных показателей эффективности и их свертке в единый интегральный показатель. В [150] проведен сравнительный анализ различных видов сверток (аддитивных, дихотомических, метрических, мультипликативных и оценочных). Их применение зависит от решаемых задач на различных уровнях иерархии.

Для комплексной оценки эффективности ИСППИД в составе СЭО сформирована иерархическая система частных показателей, представленная в графическом виде на рисунке 6.2 [150].

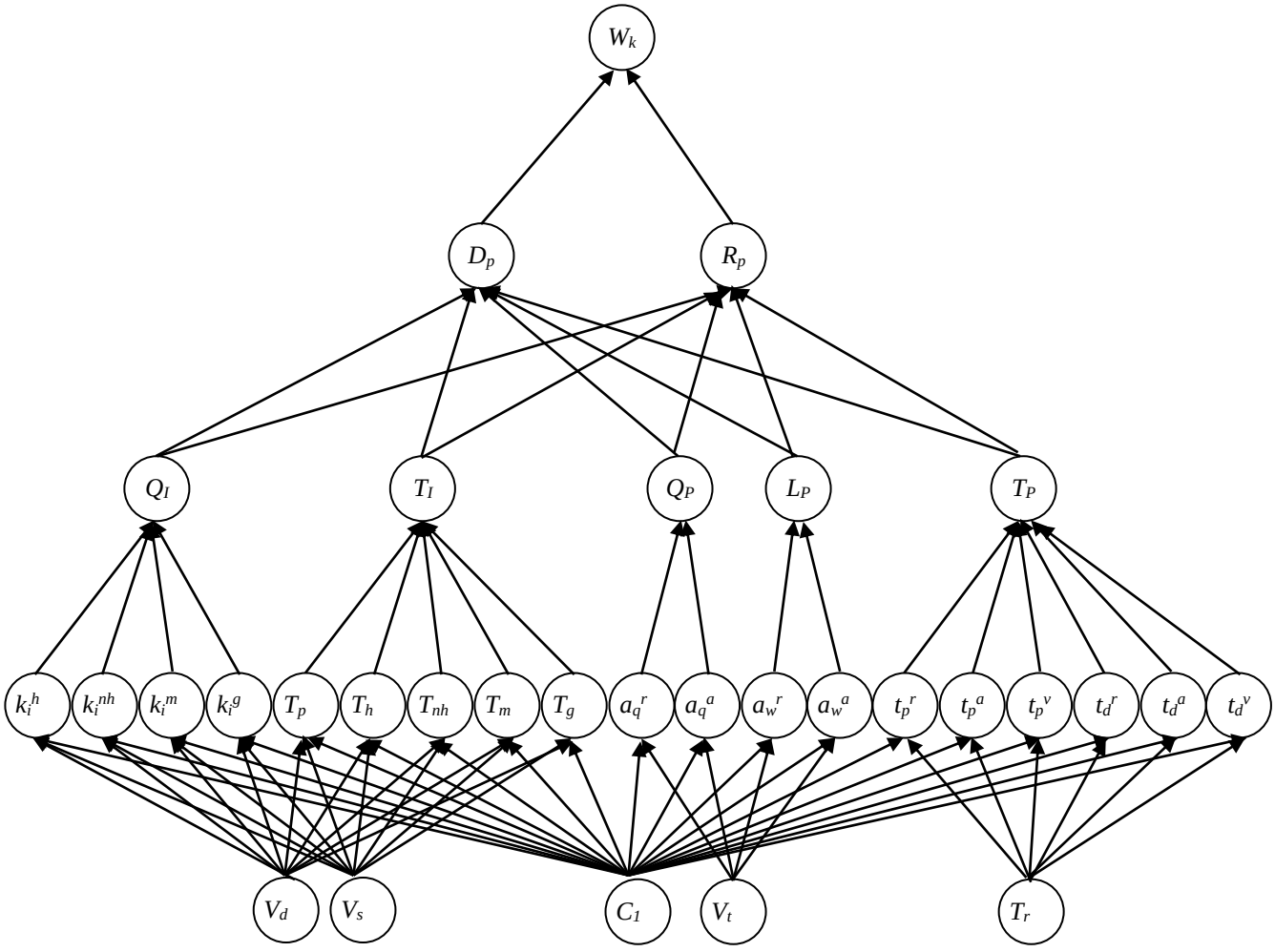


Рисунок 6.2. – Показатели эффективности ИСППИД СЭО

Она включает интегральные, интегрально-системные, системные, ин-

формационно-системные и информационные показатели, которые приведены в таблице 1.1 первой главы диссертационной работы.

На первом иерархическом уровне присутствуют такие информационные частные показатели, как: V_d - объем данных о ЦБ близких к детерминированным; V_s - объем данных о ЦБ, носящих стохастический характер; V_t – объем обрабатываемых текстовых данных; T_r - время оперативной загрузки данных, C_1 - объем выделенных денежных средств на разработку системы.

На втором уровне присутствуют интегральные информационно-системные показатели, являющиеся частными для следующего уровня иерархии: k_i^h - коэффициент отклонения средней доходности ИП Шарпа; k_i^{nh} - коэффициент отклонения средней доходности ИП НКМШ [10]; k_i^m - коэффициент отклонения средней доходности ИП Марковица; k_i^g - коэффициент отклонения средней доходности ИП МГА; T_p - усредненное время управления ИП ЛПР; T_h - усредненное время управления ИП Шарпа; T_{nh} - усредненное время управления ИП НКМШ; T_m - усредненное время управления ИП Марковица; T_g - усредненное время управления ИП МГА; a_q^r - коэффициент идентификационной точности ручного способа; a_q^a - коэффициент идентификационной точности на основе ПТП; a_w^r - коэффициент идентификационной полноты ручного способа; a_w^a - коэффициент идентификационной полноты на основе ПТП; t_p^r - время построения ТПТ на основе ручного способа, с; t_p^a - время построения ТПТ на основе ПТП, с; t_p^v - время построения ТПТ на основе ПТП с использованием виртуального пространства, с; t_d^r - время сбора новой информации на основе ручного способа, с; t_d^a - время сбора новой информации с использованием ПТП, с; t_d^v - время сбора новой информации на основе ПТП с использованием виртуального пространства, с [150].

На третьем уровне используются следующие системные показатели эффективности: Q_I – инвестиционная точность; T_I – инвестиционная оперативность; Q_P – понятийная точность; L_P – понятийная полнота; T_P – понятий-

ная оперативность. Значения данных системных показателей характеризуют качество функционирования подсистем ИСППИД. Для определения общего системного показателя ИСППИД используется соответствующий коэффициент качества - η, рассчитываемый в соответствии с выражением мультипликативной интегральной свертки [150]:

$$\eta = g_1 Q_I g_2 T_I g_3 Q_p g_4 L_p g_5 T_p, \tag{6.1}$$

где $g_1 \ldots g_5$ – веса соответствующих частных показателей и $(\sum_{i=1}^{5} g_i = 1)$.

Выбор мультипликативной интегральной свертки обусловлен простотой ее реализации и высокой чувствительностью к изменениям значений составляющих частных показателей [150].

Интегральными показателями четвертого уровня эффективности являются портфельная доходность – D_p и портфельный риск – R_p . Портфельный риск в данной задаче является ограничением. Частными интегрально-системными показателями выступают: η – коэффициент качества использования ИСППИД; C_2 – объем выделяемого инвестиционного капитала. Значение портфельной доходности рассчитывается по формуле [150]:

$$D_p = \mu C_2 \left(\frac{2\eta}{1+\eta} - R_p \right), \tag{6.2}$$

где μ - значение допустимого выигрыша или проигрыша на РЦБ, которое определяется статистическим путем.

Интегральным показателем вклада ИСППИД в ПИД СЭО и, в конечном счете, в чистую прибыль СЭО является показатель пятого уровня – долевой коэффициент чистой прибыли (ДКЧП) – W_k, рассчитываемый по формуле [150]:

$$W_k = \frac{\Delta\Theta}{C_0} = \frac{D_p - R_p}{C_1 + C_2 + C_3}, \tag{6.3}$$

где $\Delta\Theta$ – значение доли прибыли СЭО, получаемой на основе ПИД, руб.; C_0 - общие затраты ПИД СЭО, включающие: затраты на разработку ИСППИД - C_1, руб., финансовый капитал на формирование ИП - C_2, руб., компенсационные затраты транзакционных издержек ПИД и оплаты труда

ЛПР – C_3, руб. Переменная C_2 учтена уже один раз в выражении (6.2), поэтому в (6.3) не учитывается.

6.3. Методика и результаты оценки эффективности функционирования подсистемы формирования инвестиционного портфеля социально-экономической организации

Методика оценки качества функционирования основных элементов подсистемы формирования ИП в схематичном виде представлена на рисунке 6.3.

Рисунок 6.3. – Методика оценки качества функционирования основных элементов подсистемы формирования ИП

Она предназначена для исследования точностных и временных характеристик основных элементов подсистемы формирования ИП СЭО, таких как: одноиндексная модель Шарпа; нейрокомитетная одноиндексная модель Шарпа; обобщенный алгоритм реализации модели Марковица; модифицированный генетический алгоритм распределения инвестиций. В ее состав включены модуль формирования исходных данных и три оценочных модуля: измерения

временных параметров, оценки точностных характеристик, сравнительного анализа эффективности функционирования объектов исследований.

Модуль формирования исходных данных использовался в интересах компоновки наборов данных предназначенных для тестирования объектов исследований. В интересах получения объективной картины о возможностях оцениваемых моделей и алгоритмов, наборы данных были едины для всех объектов исследований. В качестве исходных данных использовались параметры облигаций, за истекшие периоды времени. Источником информации являлся интернет – сайт export.rbc.ru. Он позволяет экспортировать данные из нескольких десятков источников в трех форматах: Browser; Excel; MetaStock. В работе использовался формат данных Excel. Пример экспортируемых данных приведен в таблице 6.1.

Таблица 6.1 – Пример экспортируемых данных

TICKER	DATE	OPEN	HIGH	LOW	CLOSE	VOL	WAPRICE
USD000000TOD	2015-04-01	58.15	58.47	57.79	57.88	1271685000	58.2405
USD000000TOD	2015-04-02	57.401	57.401	56.47	56.649	1363006000	56.9229
USD000000TOD	2015-04-03	56.506	57.11	56.082	56.7	1370867000	56.7985
USD000000TOD	2015-04-06	56.515	56.651	55.61	55.7	958926000	56.2992

где TICKER - символ ценной бумаги; HIGH – наибольшая цена; DATE - дата; OPEN - цена при открытии; CLOSE - цена при закрытии; LOW - наименьшая цена; VOL - объем сделок, заключенных за данный период времени, в бумагах; WAPRICE – средневзвешенная цена.

Значения индекса РТС экспортировались с того же сайта export.rbc.ru.

Пример экспортируемых данных индекса РТС приведен в таблице 6.2.

Таблица 6.2 – Пример экспортируемых данных индекса РТС

TICKER	DATE	OPEN	HIGH	LOW	CLOSE	NUMTRADES
RTSI	2016-03-25 11:00:00	859.05	864.3	859.05	863.07	48099621
RTSI	2016-03-25 12:00:00	863.08	864.4	861.65	863.66	19871183
RTSI	2016-03-25 13:00:00	863.66	866.32	863.54	865.48	12986322
RTSI	2016-03-25 14:00:00	865.52	868.82	865.09	868.13	13936163

где TICKER – символ индекса; DATE - дата; OPEN – значение индекса

при открытии; HIGH - максимальное значение индекса; LOW - минимальное значение индекса; CLOSE - значение индекса при закрытии; NUMTRADES – количество сделок.

Использование параметров финансовых инструментов за прошлые периоды времени обусловлено целесообразностью исследования разработанных моделей и алгоритмов в условиях детерминированной внешней среды. В таких условиях возможно оптимально настроить и в максимальной степени реализовать потенциал, заложенный в объектах исследований. Для расчета других параметров облигаций, таких как: курс облигации (рыночная цена); цена облигации с постоянной купонной ставкой, имеющей *m* купонных выплат в год; стоимость облигации с нулевым купоном; купонная доходность; текущая доходность облигации; эффективная доходность к погашению; доходность за период владения; доходность государственных облигаций; доходность муниципальных облигаций; доходность корпоративных облигаций; доходность еврооблигаций; дюрация Маколея; предельная величина дюрации; модифицированная дюрация; выпуклость облигации; историческую волатильность доходность, дюрация, выпуклость и др. использовался разработанный программный модуль, включенный в состав ИСППИД СЭО [171]. Сформированные наборы данных записывались в тестовую базу данных. Применение тестовой БД решало несколько задач [172]:

1. Обеспечение локализации исходных данных.

2. Апробация эвристических алгоритмов оптимизации БД (нормализация и индексирование).

3. Исследование эффективности алгоритма обмена данными на основе их ранговой популярности.

Для оценки эффективности функционирования нейрокомитетной модели Шарпа использовалась отдельная группа исходных данных. Последние предназначены для инициализации многослойных персептронов, обученных в различные моменты времени. Технологически они размещены в текстовых файлах. Исходные данные для инициализации ИНС включают: число скрытых слоев; размер

входного окна (число обучающих дискрет); количество нейронов, входящих в соответствующие слои; значения весов обученной сети; значение индекса РТС, соответствующее моменту времени обучения ИНС.

Модуль измерения временных параметров обеспечивает определение и расчет общих временных издержек каждого из объектов исследований. Для измерения временных характеристик функционирования, разработанных моделей и алгоритмов, использовался встроенный таймер. Для оценки времени, затрачиваемого на формирование инвестиционного портфеля пользователем, использовался специальный программный модуль «Таймер-секундомер».

Модуль оценки точностных характеристик используется для фиксации значений результирующих характеристик инвестиционных портфелей, рассчитанных пользователем, а также на основе моделей и алгоритмов, входящих в состав ИСППИД. В качестве опорных (эталонных) значений с которыми сравнивались остальные принимались результаты полученные пользователем. Поскольку в роли пользователя выступала группа экспертов, степень доверия результатам, полученных ими, была максимальной.

Модуль сравнительного анализа эффективности функционирования объектов исследований формирует интегральную картину качества разработанных моделей и алгоритмов.

Методика проведения экспериментов заключалась в следующем. Формировался набор ЦБ, содержащий 5 типов облигаций и 5 типов акций. Каждый тип включал 100 экземпляров. Цена одной облигации (акции) составила в среднем 1000 руб. Сумма, выделяемая на формирование портфеля составила 100000 руб. Величина значения уровня риска составила 7%. При формировании портфеля решалась прямая задача, заключающаяся в обеспечении его максимальной доходности при заданном уровне риска. Для получения усредненных результатов оценки точности и оперативности, сформированные портфели объединялись в группы, в соответствии с рядом: 1-5-10-15-20-25-30-35-40-45-50. Формирование портфелей ручным способом ограничилось 20 по причине утомляемости экспертов.

На основе проведенных экспериментов в соответствии с разработанной методикой, получены следующие результаты, представленные в таблице 6.3.

Таблица 6.3 – Результаты оценки качества подсистемы формирования ИП

M	k_i^h	k_i^{nh}	k_i^m	k_i^g	T_p	T_h	T_{nh}	T_m	T_g
1	0,991	0,991	0,912	0,982	1220	2,8	2,2	2,8	2,8
5	0,987	0,994	0,893	0,980	6908	4,4	64,3	73,1	21,4
10	0,968	0,974	0,877	0,978	152700	9,1	118,4	141,2	35,1
15	0,967	0,961	0,843	0,977	-	13,4	81,1	181,4	58,4
20	0,960	0,959	0,847	0,975	-	46,7	94,6	258,3	77,3
25	0,957	0,977	0,875	0,973	-	69,9	141,4	314,2	117,1
30	0,95	0,981	0,853	0,972	-	100,4	244,2	374,4	158,4
35	0,941	0,901	0,844	0,971	-	141,3	250	451	216,7
40	0,936	0,878	0,840	0,970	-	175,8	264,6	479	283,4
45	0,924	0,871	0,836	0,969	-	212,4	342,8	527,6	318,8
50	0,901	0,858	0,801	0,968	-	296,3	426,7	576,3	407,2

В таблице 6.3 использованы следующие обозначения:

M – количество сформированных ИП;

k_i^h - коэффициент отклонения средней доходности ИП Шарпа;

k_i^{nh} - коэффициент отклонения средней доходности ИП НКМШ;

k_i^m - коэффициент отклонения средней доходности ИП Марковица;

k_i^g - коэффициент отклонения средней доходности ИП МГА;

T_p - усредненное время управления ИП ЛПР;

T_h - усредненное время управления ИП Шарпа;

T_{nh} - усредненное время управления ИП НКМШ;

T_m - усредненное время управления ИП Марковица;

T_g - усредненное время управления ИП МГА.

Значения коэффициентов отклонения средней доходности ИП на основе соответствующих моделей и алгоритмов рассчитываются в соответствии с выражениями (6.4 – 6.7):

$$k_i^h = \frac{\sum_{i=1}^{M} D_i^h}{\sum_{i=1}^{M} D_i^p}, \qquad (6.4)$$

$$k_i^{nh} = \frac{\sum_{i=1}^{M} D_i^{nh}}{\sum_{i=1}^{M} D_i^{p}}, \quad (6.5)$$

$$k_i^{m} = \frac{\sum_{i=1}^{M} D_i^{m}}{\sum_{i=1}^{M} D_i^{p}}, \quad (6.6)$$

$$k_i^{g} = \frac{\sum_{i=1}^{M} D_i^{g}}{\sum_{i=1}^{M} D_i^{p}}, \quad (6.7)$$

где D_i^p - доходность i портфеля, формируемого пользователем; $D_i^h, D_i^{nh}, D_i^m, D_i^g$ - доходности i портфелей, формируемых на основе модели Шарпа, нейрокомитетной модели Шарпа, алгоритма Марковица, модифицированного генетического алгоритма соответственно.

Значения усредненных времен управления ИП ЛПР, на основе модели Шарпа, нейрокомитетной модели Шарпа, алгоритма Марковица, модифицированного генетического алгоритма рассчитываются в соответствии с выражениями (6.8 – 6.12):

$$T_p = \sum_{i=1}^{M} t_i^p , \quad (6.8)$$

$$T_h = \sum_{i=1}^{M} t_i^h , \quad (6.9)$$

$$T_{nh} = \sum_{i=1}^{M} t_i^{nh} , \quad (6.10)$$

$$T_m = \sum_{i=1}^{M} t_i^m , \quad (6.11)$$

$$T_g = \sum_{i=1}^{M} t_i^g , \quad (6.12)$$

где t_i^p, t_i^h, t_i^{nh}, t_i^m, t_i^g - время формирования i портфеля ЛПР, на основе модели Шарпа, на основе нейрокомитетной модели Шарпа, на основе алгоритма Марковица, на основе модифицированного генетического алгоритма со-

ответственно.

На рисунке 6.4 приведены зависимости коэффициентов отклонений средних доходностей ИП от их числа (портфелей), сформированных на основе моделей и алгоритмов подсистемы формирования ИП ИСППИД, Прямая 1 соответствует зависимости отклонения средней доходности ИП, сформированного на основе модифицированного генетического алгоритма. Кривые 2, 3 представляют зависимости отклонения средней доходности ИП, сформированных на основе модели Шарпа и нейрокомитетной модели Шарпа, соответственно. Кривая 4 отражает зависимости отклонения средней доходности ИП, сформированного на основе алгоритма Марковица.

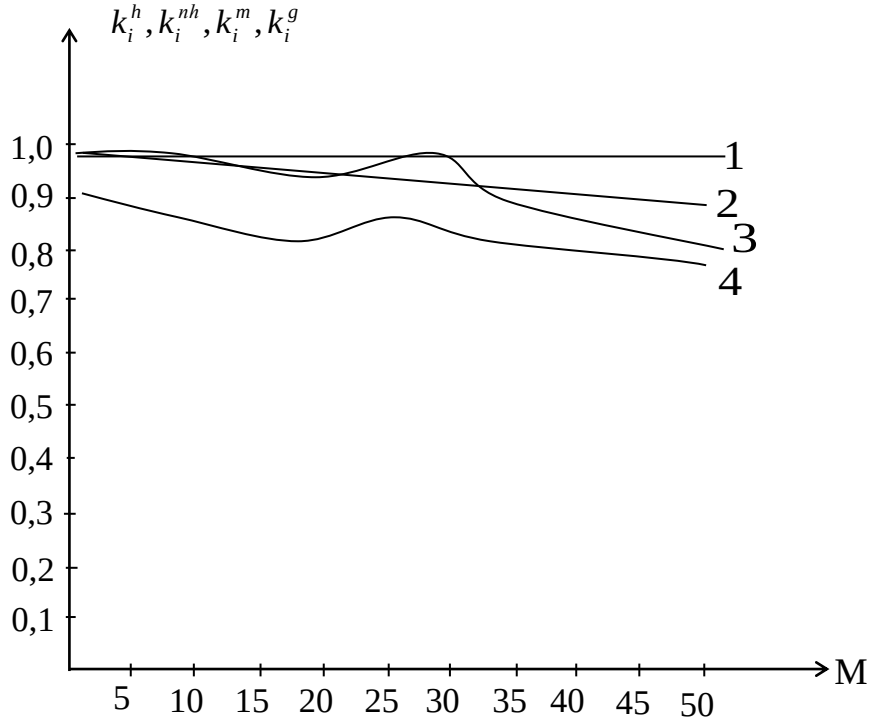


Рисунок 6.4. – Зависимости коэффициентов
отклонений средних доходностей ИП от их числа

Как отмечалось выше в качестве опорных (эталонных) значений доходностей с которыми сравнивались остальные принимались результаты полученные пользователем. Поэтому они на данном рисунке не представлены.

Зависимость результатов полученных на основе МГА (прямая 1) описывается линейной убывающей функцией. На интервалах [13; 31] и [44; 62] значения отклонений минимальны (достигается максимальная точность). Это

говорит об удачно подобранных настройках МГА. Наименее точные результаты показал алгоритм Марковица (кривая 4). Низкая точность обусловлена отсутствием полной информации о сделках по всем ЦБ на исследуемых временных интервалах. Промежуточное положение по точностным возможностям заняли модель Шарпа и нейрокомитетная модель Шарпа (кривые 2, 3). Результаты модели Шарпа линейны, что показывает ее относительно точную настройку на заданных наборах данных. Несмотря на полученные нестабильные результаты, нейрокомитетная модель Шарпа заслуживает внимания (кривая 3). Анализ ее результатов показал, что там, где были найдены априорные данные максимально приближенные к текущим ситуациям (интервалы [1;13] и контрольная точка [38]) значения полученных точностей были максимальны и превысили ее значения для МГА. Последнее показывает, что при хорошей настройке ИНС и существовании адекватной априорной информации текущим оценочным ситуациям, применение данной модели является перспективным.

На рисунке 6.5 представлены зависимости значений усредненного времени управления ИП на основе модели Шарпа (кривая 2), нейрокомитетной модели Шарпа (кривая 3), алгоритма Марковица (4) и МГА (кривая 1), от их (портфелей) числа M.

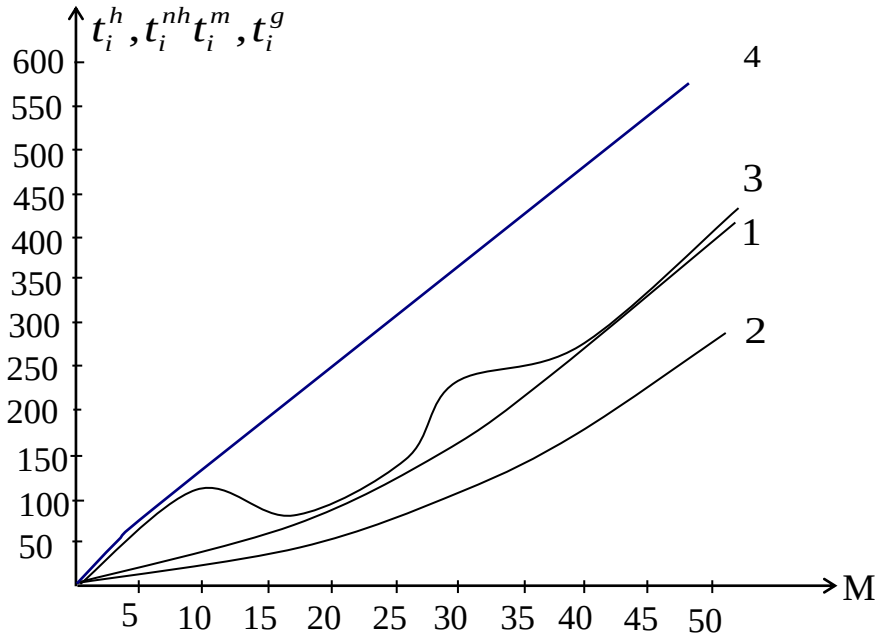


Рисунок 6.5 – Зависимости времени формирования ИП от их числа

Зависимость усредненного времени управления ИП ЛПР от их (портфелей) числа на данном рисунке не представлена. В данном масштабе, в силу значительных величин, ее невозможно отобразить. Анализ представленных графиков показывает следующее. Минимальные временные издержки показала модель Шарпа (кривая 2). Это связано с тем, что при проведении расчетов, доходности ЦБ сравниваются с константой (значением индекса РТС), а не между собой. Поэтому число ресурсоемких операций резко падает. Максимальными временными издержками обладает алгоритм Марко вица (кривая 4). Это обусловлено расчетом прямой и обратной ковариационных матриц облигационных доходностей на которые затрачиваются значительные временные ресурсы. Промежуточное положение по временным затратам занимают модифицированный генетический алгоритм (кривая 1) и нейрокомитетная модель Шарпа (кривая 3). Их взаимный анализ показывает, что МГА демонстрирует более высокую стабильность, чем НКМШ. Нестабильное поведение последней объясняется тем, что в контрольных точках (10 и 30 портфелей) НКМШ были предоставлены некорректные данные по причине несоответствия адекватности текущих и выбранных априорных условий. В результате НКМШ дообучалась, а это потребовало дополнительный временной ресурс.

Совместный анализ точностно-временных параметров показывает, что для заданных групп исходных данных наиболее эффективной является модель Шарпа. Она более оперативна и обладает приемлемой точностью. Наименьшая эффективность характерна для алгоритма Марковица. Он показал самые низкие точность и оперативность. Отдельного внимания заслуживает нейрокомитетная модель Шарпа. Несмотря на свое среднее положение среди исследуемых объектов в данной группе данных, она обладает весьма положительными качествами. На отдельных участках (на интервалах [10; 25] и [35; 50] предоставления корректных априорных данных) формирования портфелей она показала максимальную точность при приемлемой оперативности, что указывает на перспективность ее использования.

6.4. Методика оценки эффективности функционирования подсистемы терминологического информационного поиска

В качестве количественных показателей оценки эффективности функционирования подсистемы терминологического информационного поиска (ПТИП) ИСППИД, будем использовать интегральный идентификационный коэффициент и время сбора новой информации.

Интегральный идентификационный коэффициент определяется по формуле:

$$k_{in} = \frac{2a_q \cdot a_w}{a_q + a_w}, \tag{6.13}$$

где a_q и a_w - идентификационные коэффициенты точности и полноты соответствия элементов терминологических портретов текстов (ТПТ), терминологическому портрету информационной системы (ТПИС).

Идентификационная точность a_q это вероятность правильного отнесения понятия ТПТ к понятию ТПИС. Ее оценка определяется как:

$$a_q = \frac{\sum_{i=1}^{S} a_i}{\sum_{i=1}^{S} (a_i + \overline{a}_i)}, \tag{6.14}$$

где a_i - мощность множества терминов $\{A_i\}$, удовлетворяющих *i*-му термину S_i ТПП и фактически к нему причисленных; $\overline{a}_i$ - мощность множества терминов $\{\overline{A}_i\}$, неудовлетворяющих *i*-му термину S_i ТПП, но фактически к нему причисленных; *S* – число используемых терминов.

Идентификационная полнота a_w это вероятность того, что если исследуемый термин ТПТ соответствует термину ТПИС, то система отнесет его только к этому термину и после этого не остановится, а обработает все термины. Ее оценка определяется как:

$$a_w = \frac{\sum_{i=1}^{S} a_i}{\sum_{i=1}^{S} (a_i + b_i)}, \qquad (6.17)$$

где b_i – мощность множества терминов $\{B_i\}$ удовлетворяющих i-му термину S_i ТПП, но фактически к нему не причисленных.

Время сбора новой информации t_0 характеризует оперативность функционирования подсистемы терминологического информационного поиска и представляет собой временной интервал обработки одного или множества текстов.

Методика определения интегрального идентификационного коэффициента и времени сбора новой информации в схематичном виде приведена на рисунке 6.6.

Рисунок 6.6. – Методика оценки качества функционирования основных элементов подсистемы терминологического ТПИС ИСППИД

В соответствии с данной методикой реализовано формирование терминологического портрета информационной системы с использованием ИСППИД, оценка его качества ЛПР и последующий анализ множества текстов автоматизированным и ручным способами. Суть проводимого анализа заключалась в выделении терминов, формировании терминологических портретов текстов и определении количественных показателей оценки эффективности. Эксперты в области лингвистического анализа привлекались в интересах сравнительной оценки качества составленных терминологических портретов пользователем и системой, а также для оценки качества используемых словарей. Для оценки времени, затрачиваемого на проведение анализа системой, использован встроенный таймер. Для оценки времени, затрачиваемого на проведение анализа пользователем, использован внешний программный модуль «Таймер-секундомер». Анализируемые тексты группировались в наборы (пачки) в соответствии с рядом: 1-5-10-50-100. Проведение анализа ручным способом ограничилось обработкой первых трех членов ряда, по причине быстрой утомляемости пользователя.

При сравнении показателей качества функционирования подсистемы терминологического информационного поиска ИСППИД в качестве опорных принимались значения полученные ЛПР. Причин было несколько. Во-первых, в роли ЛПР выступил специалист, имеющий опыт работы в данной предметной области. Во-вторых, для оценки полученных результатов привлекались эксперты – лингвисты, которые оказывали практическую помощь ЛПР в процессе построения терминологических портретов в ходе экспериментов.

Оценка качества функционирования подсистемы терминологического информационного поиска ИСППИД проводилась для терминологического портрета «Облигационный портфель». Его структура приведена на рисунке 1.7 в первой главе диссертации. Исследовался уровень «Характеристики» ветви «Формирование». Его выбор обусловлен наличием информации по облигациям в кафедральной БД, что позволило проверить работоспособность подсистемы терминологического информационного поиска на локальной БД.

В качестве источников информации использовался Интернет и кафедральная локальная БД документальной информационно-поисковой системы «Научные труды» [94]. Исследовались понятия (определения), связанные с характеристиками облигаций. Применительно к подсистеме терминологического информационного поиска, наряду с обычным ее режимом работы, исследовался виртуальный режим для локального персонального компьютера. Для поиска информации в сети Интернет использовался браузер «Rambler». Результаты проведенных экспериментов представлены в таблице 6.4.

Таблица 6.4 – Результаты анализа качества ПТИП ИСППИД

N	a_q^r	a_w^r	a_q^a	a_w^a	t_p^r	t_d^r	t_p^a	t_d^a	t_p^v	t_d^v
1	1	1	0,87	0,87	220	80	1,2	0,8	0,8	0,3
5	0,99	0,99	0,89	0,89	490	110	18,3	7,1	8,4	3,6
10	0,97	0,97	0,83	0,83	2700	1100	57,4	11,2	15,1	5,2
50	0,95	0,95	0,8	0,8	-	-	264	56,4	98	25,4
100	0,92	0,92	0,85	0,85	-	-	518	82,3	153	49

В таблице 6.4 использован следующий ряд обозначений:

N – количество текстов;

a_q^r - коэффициент идентификационной точности ручного способа;

a_q^a - коэффициент идентификационной точности на основе ПТИП;

a_w^r - коэффициент идентификационной полноты ручного способа;

a_w^a - коэффициент идентификационной полноты на основе ПТИП;

t_p^r - время построения ТПТ на основе ручного способа, с;

t_p^a - время построения ТПТ на основе ПТИП, с;

t_p^v - время построения ТПТ на основе ПТИП с использованием виртуального пространства, с;

t_d^r - время сбора новой информации на основе ручного способа, с;

t_d^a - время сбора новой информации с использованием ПТИП, с;

t_d^v - время сбора новой информации на основе ПТИП с использованием виртуального пространства, с.

На рисунке 6.7 приведены зависимости идентификационной точности и полноты текстов от их числа, проведенных ЛПР (прямая 1) и ПТИП ИСППИД (прямые 2, 3).

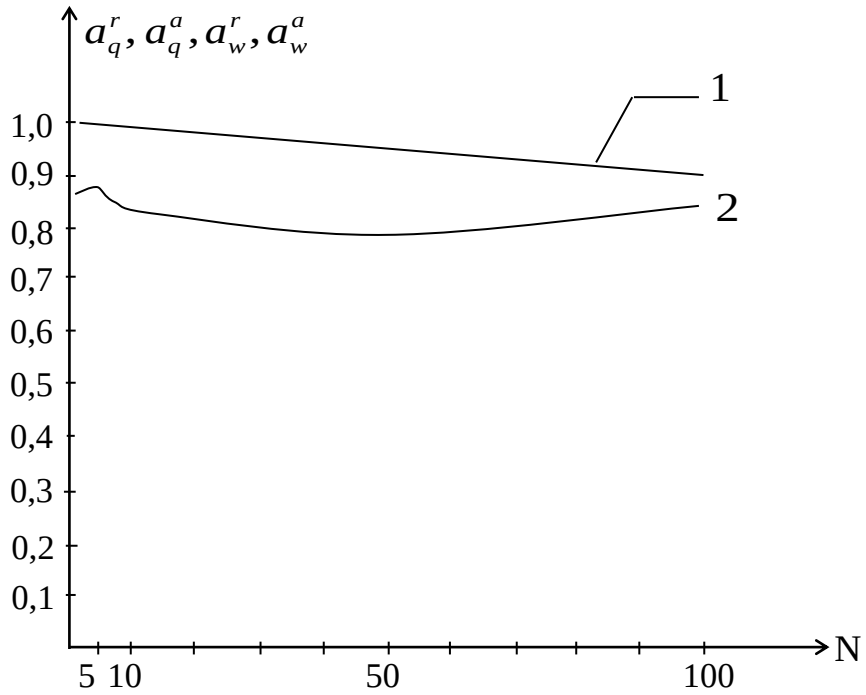


Рисунок 6.7. – Зависимости идентификационной точности и полноты текстов от их числа

Зависимость результатов полученных ЛПР (прямая 1) описывается линейной убывающей функцией. Линейность зависимости связана с высокой квалификацией ЛПР и дополнительно оказанной ему помощи экспертом-лингвистом. Убывающий характер зависимости определяется усилением усталости ЛПР при увеличении числа обрабатываемых текстов.

Зависимости результатов полученных с использованием ПТИП ИСППИД и ПТИП ИСППИД в виртуальном пространстве совпадают. Они представлены общим графиком (кривая 2). Данная зависимость носит усредненный характер. При проведении экспериментов наблюдались колебания результатов по точности и полноте идентификации. Они обусловлены несколькими причинами. Во-первых, сложность анализируемых текстов различна. Тексты в которых преобладали простые предложения обрабатывались оперативнее и точнее. Значительно затрудняли анализ текстов сложноподчиненные предложения, причастные и деепричастные обороты. Во-вторых, существенное влияние оказывала

авторская манера написания текста и логика его изложения. Встречались тексты со своеобразными манерой и логикой.

На рисунке 6.8 приведены зависимости суммарного времени текстовой обработки от числа текстов, проведенных ЛПР (прямая 1), ПТИП ИСППИД (прямая 2) и ПТИП ИСППИД в виртуальном пространстве (прямая 3).

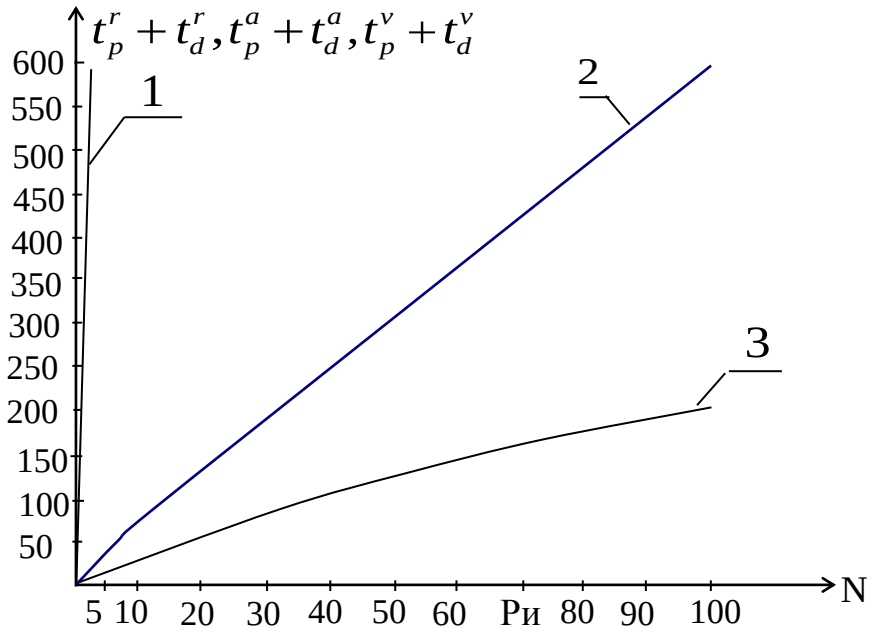


Рисунок 6.8. – Зависимости времени текстовой обработки от их числа

Анализ представленных зависимостей показал следующее. Наибольшие временные затраты при текстовой обработке характерны для ЛПР (прямая 1). При текстовой обработке нескольких текстов с использованием ПТИП ИСППИД (прямая 2) оперативность возрастает более чем на два порядка, а при обработке нескольких десятков текстов выигрыш составляет около четырех порядков. При дальнейшем увеличении количества обрабатываемых текстов данная тенденция сохраняется. Выигрыш в оперативности обработки сотни текстов с использованием ПТИП ИСППИД в виртуальном пространстве (прямая 3) в сравнении с обычной обработкой составляет 3 раза. Это связано с общей обработкой информации (доступ к БД) в оперативной памяти ПК. При дальнейшем увеличении числа обрабатываемых текстов данная тенденция также сохраняется.

Оценка качества функционирования подсистемы терминологического

информационного поиска ИСППИД показала, что, несмотря на более низкие точностные характеристики, по сравнению с ручным способом, текстовая обработка большого числа текстов с использованием ПТИП ИСППИД в виртуальном пространстве является эффективной и перспективной, в силу своей высокой оперативности.

Наряду с оценкой эффективности функционирования прототипа ИСППИД СЭО, важное значение имеет технология его внутрисистемного синтеза.

6.5. Предложения по содержанию внутрисистемного синтеза и перспективного развития информационных систем поддержки портфельной инвестиционной деятельности

В настоящее время имеется определенный опыт разработки технологии внутрисистемного синтеза информационных систем интегрированного менеджмента [97, 98, 112, 122, 138], достаточно близких по идеологии проектирования и конструктивному исполнению к ИСППИД СЭО, что показано в [101, 126]. Это обусловливает целесообразность применения накопленного опыта [101, 114, 126, 184] для систем данного класса.

Основные этапы внутрисистемного синтеза ИСППИД СЭО в схематичном виде приведены на рисунке 6.9.

На начальном этапе внутрисистемного синтеза ИСППИД СЭО определяются общие концептуальные положения, связанные с разработкой проекта создания системы. Они включают описание объекта управления, субъекта управления, процесса управления и окружение проекта [17].

Объектом управления в данном случае является ИСППИД СЭО, облик которой обоснован на стадии внешнесистемного синтеза, а завершающей стадией разработки является приемка ее в эксплуатацию [114].

Субъектом управления является проектная группа (ПГ), представляющая собой команду участников, задействованных в проекте. В состав ПГ входят: заказчик, главный конструктор, проектные менеджеры, непосредственные исполнители и др. На стадиях выработки и принятия решений между ними наблюдается тесное взаимодействие.

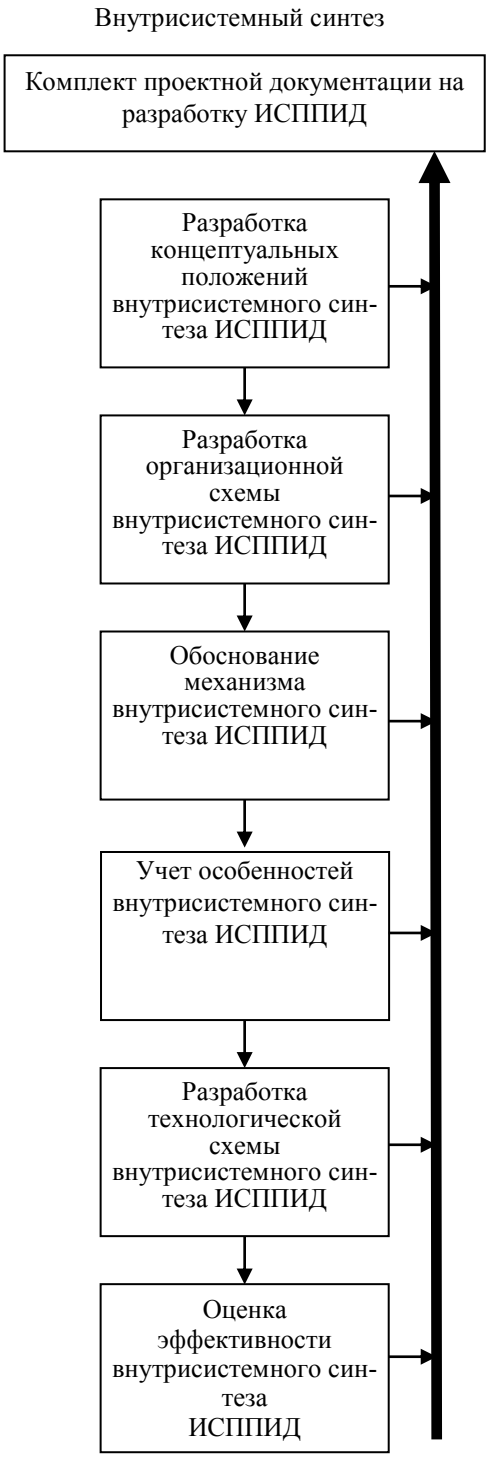


Рисунок 6.9. – Этапы внутрисистемного синтеза ИСППИД СЭО

В процессе проектного управления (ПУ) осуществляется целенаправленное воздействие субъекта на управляемый объект на основе принимаемых и реализуемых решений, которые учитывают различные технологические, организационные, технические, экономические и др. ограничения [114].

Совокупность различных воздействий внутренней и внешней сред, которые могут быть нейтральными, ускоряющими или замедляющими дости-

жение проектных целей, является окружением проекта [114].

Вторым этапом внутрисистемного синтеза ИСППИД СЭО является разработка первичной иерархической организационной схемы ПУ, предполагающей целеполагание, проектирование и реализацию объекта управления [101, 125, 126]. В схематичном виде она представлена на рисунке 6.10.

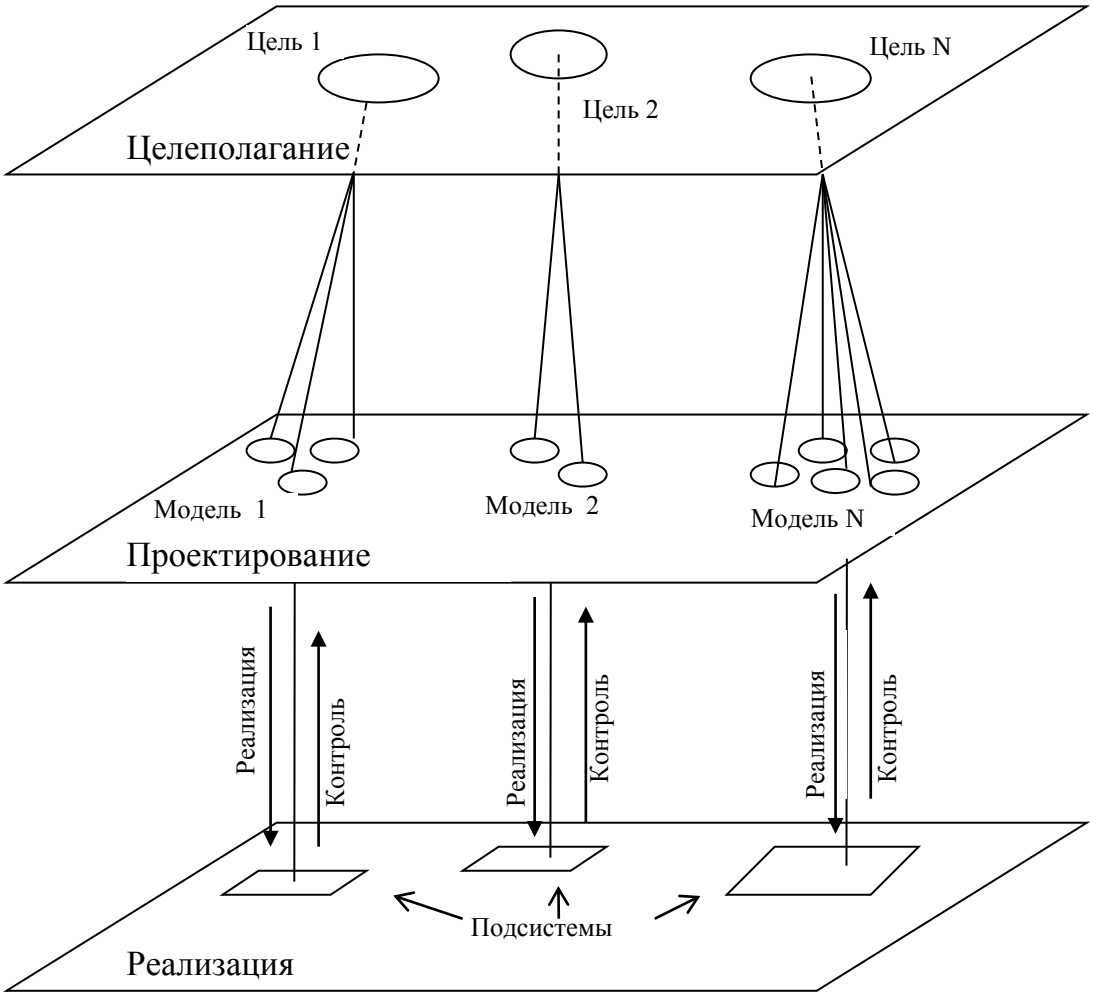


Рисунок 6.10. –Первичная организационная схема проектного управления ИСППИД СЭО

Из рисунка видно, что содержанием уровня целеполагания является определенный набор основных проектных идей, представляющих соответствующие цели. Они отражаются на уровне проектирования в содержании определенных моделей, представляющих собой детализированные планы действий, касающихся различных аспектов создания ИСППИД СЭО. В формальном виде подобные планы представляются в соответствующей документации (в аванпроекте, в технических предложениях, в техническом задании и др.)

[101].

На уровне проектирования решается несколько важных проектных задач:

- обоснования требуемых ресурсов для выполнения проекта;

- разработки организационной структуры проекта;

- определения коммуникационной схемы между проектными элементами;

- разработки схем организации и взаимодействия с внутренней средой (организация команд разработчиков, решение кадровых вопросов и др.);

- разработки сценариев взаимодействия проектных элементов с внешней средой;

- организации процесса финансирования проекта;

- организации материального, технического и других видов обеспечения проекта и др. [101].

Основным содержанием проектирования является разработка, анализ, уточнение и утверждение модели деятельности субъектов управления, включающей плановый комплекс мероприятий, направленных на реализацию конкретных целей.

Модели уровня проектирования находят отражение на уровне реализации, целевое назначение которого заключается в разработке конкретных подсистем ИСППИД СЭО и межподсистемного интерфейсного взаимодействия.

Следует отметить, что для проектирования систем, предназначенных для управления техническими средствами, представленных уровней вполне достаточно. Однако при проектировании ИСППИД СЭО, которые включают в своем составе программные комплексы, требуется еще один уровень иерархии, реализующий разработку программных решений (уровень программирования). Уточненная четырехуровневая схема организационного проектирования ИСППИД СЭО, представлена на рисунке 6.11 [101].

Уровень программирования предназначен для разработки программных модулей и организации их тестирования, проектирования БД, баз моделей, хранилищ и витрин данных, разработки интерфейсных экранных форм и др. [101].

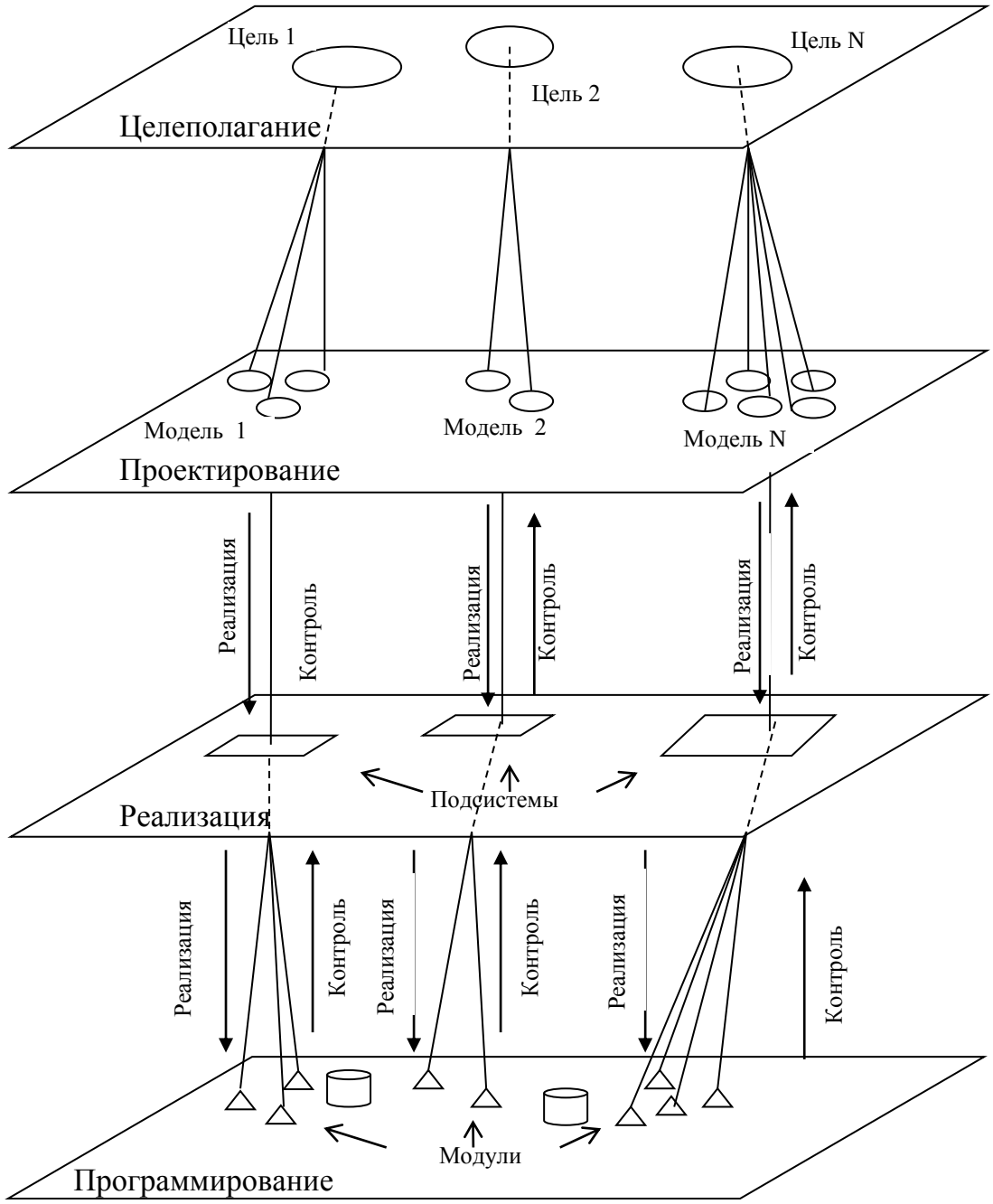


Рисунок 6.11. – Уточненная организационная схема проектного управления ИСППИД СЭО

Успешная реализация проекта невозможна без достижения определенного компромисса между существующими ресурсными возможностями и требованиями, выдвигаемыми проектными решениями. Реализация контроля на всех иерархических уровнях также является необходимым условием успешного завершения проектирования. Необходимость контроля обусловлена целесообразностью постоянного сравнения текущих показателей с теми,

которые запланированы в целях, моделях, подсистемах, модулях, программах и др. для оперативной фиксации возникших рассогласований и принятия мер по их устранению [101].

В результате реализации организационной схемы внутрисистемного синтеза ИСППИД СЭО формируется модель, которая отражает реальное проектное состояние. Ранее поставленные цели, сравниваются с результатами этой модели и на их основе, определяется проектная эффективность, как реальная степень их (целей) достижения.

Содержанием третьего этапа внутрисистемного синтеза ИСППИД СЭО является разработка механизма проектного управления (МПУ), представленного в схематичном виде на рисунке 6.12 [101].

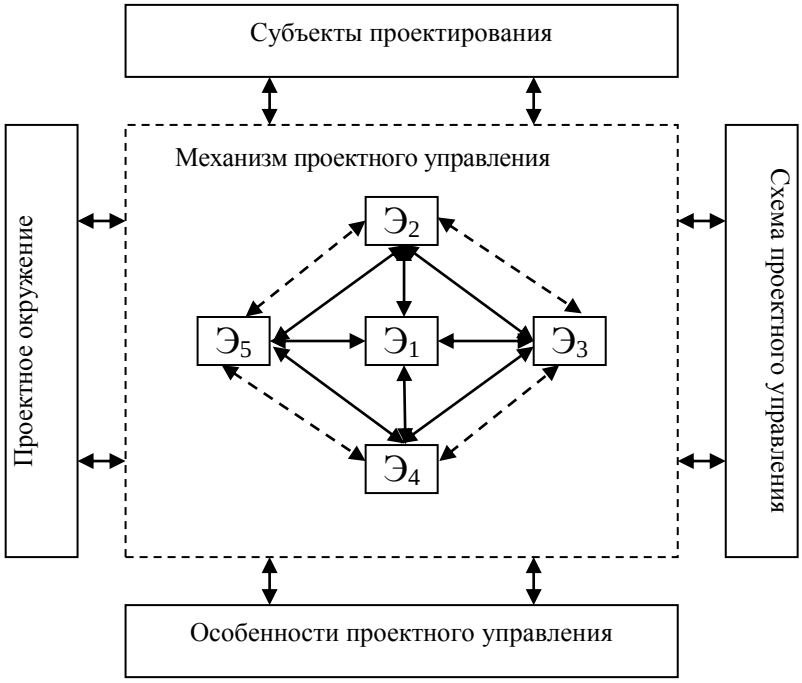


Рисунок 6.12. – Механизм мультиэтапного проектного управления ИСППИД СЭО

Содержанием МПУ является выполнение пяти этапов. На первом этапе $Э_1$ путем изучения, сформированного в результате внешнесистемного синтеза, технического проекта (конструкторской документации), осуществляется уяснение облика ИСППИД СЭО (состав, структура, режимы работы, внутрисистемные решения и др.). В рамках второго этапа $Э_2$ формируется первичный облик ПГ (состав, распределение функций, организация контроля, гра-

фик выполнения работ). Содержанием третьего этапа $Э_3$ является вербальное моделирование функционирования ПГ (взаимодействие, критерии оценки деятельности и др.). На четвертом этапе $Э_4$ осуществляется декомпозиция и анализ деятельности ПГ по функциональным направлениям (системотехника, алгоритмизация, программирование, тестирование, администрирование и др.) с учетом сохранения ее основных свойств. На завершающем этапе $Э_5$ реализуется объединение декомпозиционных элементов на критериальной основе и защита (обоснование) проектного решения – уточненный состав рабочей ПГ.

Особенностью МПУ является параллельное выполнение работ на всех этапах, путем проведения совместных совещаний, телеконференций и других телекоммуникационных взаимодействий, обусловливающее его мультиэтапный характер (отражено на рисунке 6.11 сплошными стрелками). Эффективность последнего в плане повышения оперативности согласования проектных решений очевидна. Для сравнения, на рисунке 6.11 пунктирными стрелками представлена последовательная стратегия реализации МПУ.

Практически на всех этапах проведения внутрисистемного синтеза ИСППИД СЭО важно учитывать ряд особенностей, влияющих на проведение различного вида оценок и принятие решений. Их учет составляет содержание четвертого этапа внутрисистемного синтеза ИСППИД СЭО. К подобным особенностям следует отнести субъективность, неопределенность, конфликтность, неполная структурированность, значимость априорной информации.

Субъективность обусловлена тем фактом, что основным участником данного процесса управления является человек, отличающийся индивидуальным видением некоторых явлений и процессов, обладающий определенными интересами и предпочтениями, которые могут отличаться от объективной реальности создания ИСППИД СЭО. Основным способом парирования субъективности является коллегиальность принятия решений.

Фактор неопределенности присутствует практически на всех этапах внутрисистемного синтеза ИСППИД СЭО. Это может привести к неопти-

мальному проектному управлению, серьезным ошибкам, неоправданному расходу ресурсов и др. Для снижения неопределенности целесообразно использовать различные виды моделирования, проведение натурных экспериментов, привлечение множества различных источников информации в интересах придания ей объективного характера, при разработке проектных решений привлекать экспертов и обеспечивать повышение уровня компетентности ПГ.

Наличие конфликтности обусловлено противоречиями, которые могут возникнуть внутри ПГ. Для снижения уровня конфликтности необходимо создавать доброжелательную рабочую обстановку в ПГ, разрабатывать объективную и прозрачную систему критериев оценки приемки узловых этапов проекта, обеспечивать своевременное вскрытие конфликтов и использовать существующие арсеналы их преодоления и др.

Наличие неполной структурированности в данной предметной области обусловлено тем, что некоторые важные свойства (особенно перспективные) ИСППИД СЭО (например, поведение в конкурентной среде) не до конца изучены, что приводит к невозможности их формализованного представления. Это может приводить к существенным ошибкам и значительно снижать эффективность применения системы. В интересах разрешения данной проблемы необходимо применение сценарного подхода исследования системы, проведение имитационного моделирования, использование математического аппарата теории нечетких множеств и др.

Значимость априорной информации состоит в том, что в интересах повышения оперативности разработки ИСППИД СЭО, целесообразно использовать предыдущий опыт подобных разработок т.е готовых решений. Аналогичным образом необходимо поступать и в рамках ПУ – использовать имеющийся опыт управления ПГ (например использовать методы Дельфи, экспертных оценок и др.).

В рамках пятого этапа внутрисистемного синтеза разрабатывается технологическая схема ПУ создания ИСППИД СЭО, которая в схематичном

виде приведена на рисунке 6.13 [101, 115, 162, 184, 253].

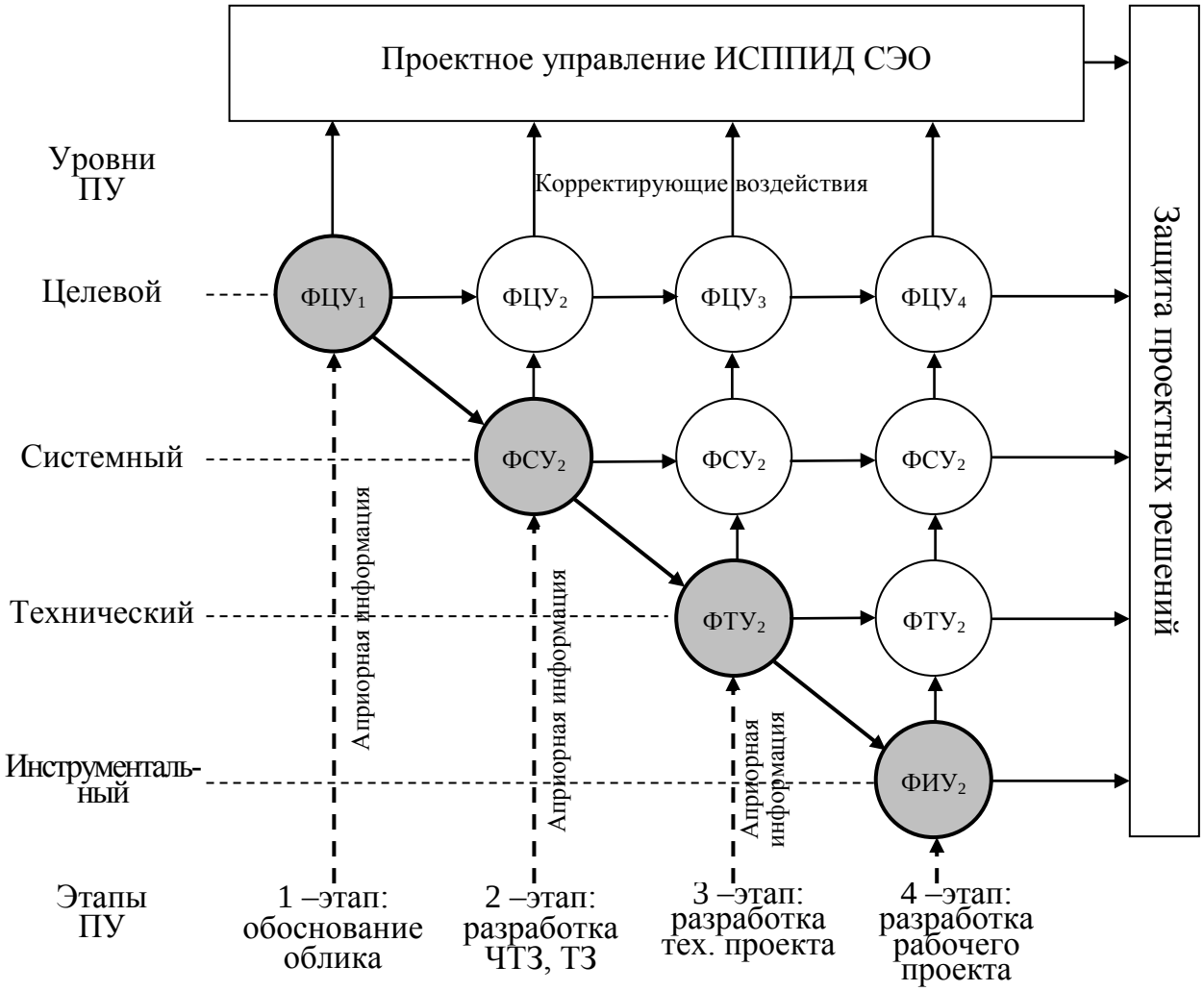


Рисунок 6.13. – Технологическая схема проектного управления ИСППИД СЭО

Отличительной особенностью предлагаемой технологии, от существующей линейной, заретушированной на рисунке 6.13, является применение многоуровневого параллельного итеративного процесса поиска оптимального проектного решения, учитывающего установленные ограничения [101].

При этом ПУ осуществляется на следующих четырех уровнях управления: инструментальном, техническом, системном, целевом. Самым нижним (четвертым) уровнем управления является инструментальный, на котором ИСППИД СЭО представляется в виде набора программных средств (например, модулей), имеющих единый интерфейс управления. На третьем (техническом) уровне управления система представляется набором требуемых алго-

ритмов, проектов БД, а также множеством сетевых и технических решений. На втором (системном) уровне рассматривается более высокое обобщение ИСППИД СЭО в виде множества подсистем и элементов, а также отношений (связей) между компонентами. Для первого (целевого) ПУ характерно цельное представление системы. На каждом уровне управления выполняется множество соответствующих ему функций в виде функционала с характерными обозначениями (функционал целевого уровня (ФЦУ); функционал системного уровня (ФСУ); функционал технического уровня (ФТУ); функционал инструментального уровня (ФИУ)).

Во времени процесс ПУ включает четыре этапа. Результатом первого этапа, наряду со сформированным обликом ИСППИД СЭО, представляется обоснованный первичный облик ПГ (состав, функциональная спецификация, организация контроля, ориентировочные временные границы выполнения работ и др.). В рамках второго этапа разрабатывается множество частных технических заданий (ЧТЗ), формируется общее техническое задание (ТЗ) и уточняется механизм функционирования ПГ (внутреннее и внешнее взаимодействие, критерии оценки деятельности и др.). На третьем этапе реализуется техническое проектирование (обоснование различных видов обеспечения и др.), а также проводится уточнение и детализация функциональных направлений деятельности ПГ (системотехники, алгоритмизации, программирования, тестирования, администрирования и др.). Завершающим этапом процесса ПУ является проведение рабочего (эскизного) проектирования системы и уточнение состава, структуры и временного режима (план-график выполнения работ) функционирования рабочей ПГ.

Итеративный характер процесса ПУ представлен наличием корректирующих воздействий, отображаемых на рисунке 6.13 в виде соответствующих стрелок. Кроме того, практически между всеми узлами ПУ имеются информационные связи и отдельно выделены информационные потоки с априорной информацией, которые демонстрируют возможность использования предыдущего опыта для решения текущих задач [101].

Завершается внутрисистемный синтез этапом оценки качества его проведения. Перечень критериев оценки качества внутрисистемного синтеза ИСППИД СЭО представлен на рисунке 6.14 [130].

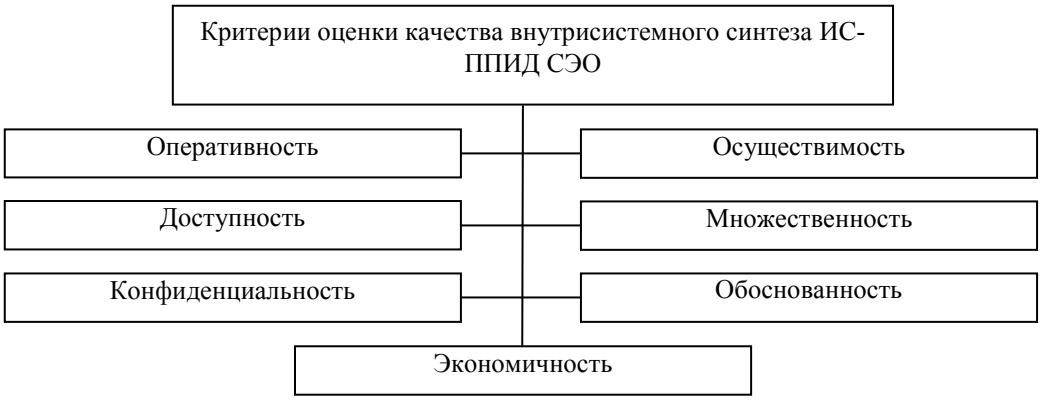


Рисунок 6.14. – Критерии оценки качества внутрисистемного синтеза ИСППИД СЭО

Одним из важных критериев оценки качества внутрисистемного синтеза ИСППИД СЭО является оперативность. Процесс формирования ПГ не должен затягиваться, но и не должен проводиться быстро и поверхностно. Последнее может негативно повлиять на конечный результат. Обычно временной интервал формирования ПГ составляет несколько недель. В данном процессе важно учитывать предыдущий опыт выбора команд (учитывать старые связи). Руководитель проекта, как правило, старается этим воспользоваться. Это положительно сказывается на слаженности ПГ, поскольку в команде работают проверенные люди, особенности которых хорошо известны.

Суть критерия осуществимости состоит в том, что при выборе ПГ должны учитываться реальные возможности членов группы (соответствующий уровень компетенций и практических навыков) и им должны ставиться реальные (достижимые) цели и задачи.

На основании критерия доступности, членам ПГ должен быть обеспечен своевременный и беспрепятственный доступ к требуемой информации, касающейся их функциональной деятельности и выделенных полномочий. Задержка доступа и/или предоставление некачественной информации может привести к затягиванию проекта и негативно сказаться на конечном резуль-

тате.

Наличие критерия множественности указывает на тот факт, что при обсуждении стратегий развития объекта ПУ и формирования ПГ должны обсуждаться различные модели и точки зрения, а выбираться должна одна (модель и точка зрения) на критериальной основе и механизме коллегиального принятия решений.

Существование критерия конфиденциальности показывает, что информация, которая содержится в ИСППИД СЭО, а также связанная с ее разработкой и формированием ПУ, должна быть доступна ограниченному круг лиц, оговоренному заказчиком. В случае получения доступа к информации неуполномоченного лица, ее конфиденциальность утрачивается.

Суть критерия обоснованности заключается в необходимости проведения объективной оценки идей и стратегий создания объекта ПУ и формирования ПГ. Данный критерий оказывает влияние на степень доверия к полученным результатам со стороны ЛПР и заказчика. Его достижению способствуют использование количественных и качественных показателей при оценке промежуточных и финальных результатов, проведение численных и полунатурных экспериментов, подтверждающих их (оценок) достоверность, применение системы меню, подсказок и объяснения полученных результатов в ИСППИД СЭО.

Важным критерием создания объекта ПУ и формирования ПГ является экономичность. Он предполагает обоснование оптимальных объемов ресурсов (материальных, финансовых, людских и др.), вкладываемых непосредственно в ИСППИД СЭО и затрачиваемых на ПУ по его разработке.

Особенностью предложенного внутрисистемного синтеза ИСППИД СЭО является применение стратегии мультиэтапного проектного управления ИСППИД СЭО. Ее достоинством является более высокая оперативность ПУ, за счет возможности распараллеливания выполняемых работ, и учета предыдущего опыта ПУ подобных систем.

Результатом внутрисистемного синтеза ИСППИД СЭО является ком-

плект проектной документации, включающий сетевой план-график выполнения соответствующих научно исследовательских и опытно – конструкторских работ по ее разработке, а также сформированная ПГ, обеспечивающая их реализацию.

Опытная эксплуатация отдельных подсистем и разработанного прототипа ИСППИД СЭО в целом послужила основой разработки предложений по повышению эффективности их функционирования по ряду направлений.

1. В части совершенствования подсистем:

а) применительно к ППЗА в рамках перспективных исследований целесообразно расширить возможности разработанной НКМШ в части поиска и учета статистики поведения активов в условиях стратегий принятия решений инвесторами РЦБ на повышение и понижение (так называемых «бычьего» и «медвежьего» трендов). С этой целью потребуется применение гибридных ИНС (сетей Кохонена и МП) и использовать роевые алгоритмы их обучения;

б) для ПТИП целесообразно:

- расширить перечень возможных форматов документов, используемых для поиска терминологической информации. В текущей версии используются форматы: txt, html, pdf, doc. Необходимо внедрение форматов djvu, docx и др;

- расширить интерфейс взаимодействия подсистемы с поисковыми машинами. В настоящее время в прототипе используется поисковая машина Rambler. Требуется подключение поисковых машин Google, Yandex и др. и оценка возможности их комплексирования;

- внедрить визуализацию терминологических портретов системы (в текущей версии не реализована).

в) опыт эксплуатации ПСОХИ показал, что при загрузке данных из ХД в ВД во время сбоев возможна потеря работоспособности «падение» ВД. Парированием данного недостатка явилось копирование (создание нескольких резервных копий) ВД и, в случае сбоя, организация перехода на резервную ВД. Это приводит к временным потерям и неэффективному использованию

дискового пространства жесткого магнитного диска и оперативной памяти. Перспективным решением данной проблемы является переход на технологию распределенных БД и использование механизма репликации локальных БД. Альтернативными решениями является организация локального сетевого обмена данными между ХД и ВД или переход на облачные технологии.

2. В части совершенствования всей системы:

а) в настоящее время, на основе использования разработанного ИМИНЗ, система способна объяснять свои действия в рамках терминологического информационного поиска. Необходимо распространить данную возможность на подсистему формирования ИП с целью обеспечения полной прозрачности и, тем самым, максимального доверия к системе, на самом ответственном этапе ее функционирования – формировании ИП. Реализация данной возможности потребует разработки подсистемы логического вывода из области информационной технологии экспертных систем;

б) в текущей версии ИСППИД реконфигурация системы, в зависимости от вида поступающих данных (детерминированных или стохастических), осуществляется ЛПР в автоматизированном режиме работы, путем выбора для использования соответствующей модели формирования ИП. Для расширения возможностей системы целесообразно данную функцию реализовать в автоматическом режиме с выдачей ЛПР информации о данном факте;

в) тестирование прототипа ИСППИД на массивах исторических данных позволило выявить в предкризисных состояниях РЦБ следующую закономерность: возникновение аномальных отклонений параметров законов распределения исследуемых активов в рамках функционирования НКМШ (на этапах сегментации и кластеризации исследуемого временного ряда ЦБ). В связи с этим, предлагается реализовать выдачу соответствующих предупреждений в автоматизированном и автоматическом режимах работы системы путем встраивания специальных процедур в ППЗА;

г) в настоящее время проверка работоспособности ИСППИД проведена применительно к фондовому рынку. В перспективе целесообразно провести ис-

следование системы применительно к финансовому рынку и уточнить ее состав и структуру;

д) проведение перспективных доработок, расширяющих возможности ИСППИД в данной предметной области, наиболее значимыми из которых являются:

- наращивание функционала системы в части внедрения рисковых финансовых активов (акций, опционов и др.);

- разработка инструментов диверсификации рисков с использованием недвижимости, безрисковых активов и др.;

- разработка инструментов хеджирования рисков на основе фьючерсов;

- разработка и внедрение модели арбитражного ценообразования и др.

ВЫВОДЫ

Для первичной апробации функционирования ИСППИД СЭО в соответствии с синтезированным обликом, разработан прототип, представляющий собой первоначальный вариант системы с ограниченным функционалом, призванный проверить работоспособность, заложенных в ней идей, на практике. В состав прототипа вошли подсистемы формирования ИП, прогнозирования значений параметров ЦБ, терминологического информационного поиска, а также сбора, хранения и обработки информации. Кроме того, в его состав включены несколько отдельных модулей, таких как: системного управления, ввода данных, автоматизированного ввода данных и формирования выходной информации. В данной версии прототипа предусмотрено пять режимов работы: «Настройка»; «Тестирование»; «Ручной»; «Автоматизированный»; «Автоматический». Два первых режима характерны для предварительного состояния функционирования системы, а оставшиеся три для ее исполнительного состояния.

В интересах комплексной оценки эффективности ИСППИД в составе СЭО сформирована иерархическая система интегральных, интегрально-системных, системных, информационно-системных и информационных показателей. Интегральным показателем вклада ИСППИД в ПИД СЭО и, в конеч-

ном счете, в чистую прибыль СЭО является показатель пятого уровня – долевой коэффициент чистой прибыли. В качестве информационных частных показателей первого уровня использованы объемы данных о ЦБ близких к детерминированным и, носящих стохастический характер, объем обрабатываемых текстовых данных, время оперативной загрузки данных и объем денежных средств, выделенных на разработку системы. На остальных уровнях представлены интегрально-системные, системные и информационно-системные показатели, связанные с параметрами функционирования соответствующих элементов, модулей и подсистем ИСППИД СЭО.

Оценка эффективности функционирования ИСППИД СЭО и ее основных подсистем проводилась на основе разработанных методик, включающих процедуры измерения временных параметров, оценки точностных характеристик и сравнительного анализа полученных результатов объектов исследований.

Анализ точностно-временных параметров подсистемы формирования ИП позволил обратить внимание на нейрокомитетную модель Шарпа. Несмотря на свое среднее положение по полученным характеристикам среди исследуемых объектов, она обладает положительными качествами. На отдельных участках (на интервалах [10; 25] и [35; 50] где представлены корректные априорные данные) формирования ИП она для стохастических данных о ЦБ показала максимальную точность при приемлемой оперативности, что указывает на перспективность ее использования.

Оценка эффективности функционирования подсистемы терминологического поиска ИСППИД показала, что, несмотря на более низкие точностные характеристики, по сравнению с ручным способом, обработка большого числа текстов с использованием ПТП ИСППИД в виртуальном пространстве является эффективной и перспективной, в силу своей высокой оперативности (выигрыш составляет более четырех порядков) при приемлемой точности. При проведении экспериментов наблюдались колебания результатов по точности и полноте идентификации. Они обусловлены несколькими причинами. Во-первых, сложность анализируемых текстов различна. Тексты в которых преоб-

ладали простые предложения обрабатывались оперативнее и точнее. Значительно затрудняли анализ текстов сложноподчиненные предложения, причастные и деепричастные обороты. Во-вторых, существенное влияние оказывала авторская манера написания текста и логика его изложения. Встречались тексты со своеобразными манерой и логикой.

Содержанием внутрисистемного синтеза ИСППИД СЭО, используемого в интересах формирования ПГ и уточнения ее облика, является последовательность выполнения следующих этапов: разработки концептуальных положений; разработки организационной схемы; обоснования механизма проектного управления; учета особенностей реализации; разработки технологической схемы реализации; оценки эффективности.

Концептуальные положения включают 5 категорий: окружение проекта, состав участников, организационную схема проектирования, механизм проектирования и учет специфических особенностей. В состав участников проектирования включены: субъект управления; объект управления; процесс управления. Схема проектирования базируется на 4 уровнях: целеполагания (множество целей), проектирования (множество моделей), реализации (множеством подсистем) и программирования (множество программных модулей). Система контроля, реализуемая на каждом уровне, используется в интересах сравнения достигнутых показателей с запланированными, заложенными в моделях, с последующей фиксацией возникших изменений. Таким образом, формируется модель, отражающая фактическое состояние проекта. Данная модель сравнивается с ранее поставленными целями, исходя из чего, определяется эффективность проекта, как степень достижения этих целей [115].

Организационная схема внутрисистемного синтеза ИСППИД базируется на многоуровневом параллельном итеративном процессе поиска рационального решения, который учитывает заданные условия и ограничения. Среди уровней проектирования выделены: целевой; системный; технический; инструментальный. На целевом уровне представляется создание (облик) и развитие ИСППИД как единой (цельной) системы. На системном уровне ИСППИД рассматривает-

ся как совокупность подсистем, элементов и взаимосвязей между ними. В рамках технического уровня ИСППИД позиционируется как совокупность алгоритмов, проектов баз данных, а также набора технических и сетевых решений. Для инструментального уровня характерно представление ИСППИД в виде совокупности программных модулей, управляемых с единого пользовательского интерфейса [184].

Механизм проектного управления основан на системном анализе. Он реализует: формулирование проблемы, определение целей, разработку и исследование альтернативных путей (способов) достижения поставленных целей, сравнение альтернатив и выбор оптимального способа достижения общей цели. В рамках данного механизма реализован метод априорных решений. Его суть заключается в использовании существующих информационных систем (подсистем и модулей), обеспечивающих своими функциональными возможностями достижение определенных подцелей в общем дереве целей. Достоинством данного метода является использование готовых (отработанных) решений, что обеспечивает сокращение сроков реализации проекта и уменьшение материальных затрат, а недостатком – сложности сопряжения существующих систем с разрабатываемыми. Примером использования данного метода, является применение при проектировании иерархического терминопостроителя ИСППИД, готового решения - ядра формализации и анализа текстовой информации.

Среди предложений по улучшению эффективности функционирования ИСППИД и ее подсистем выделяются следующие:

- для расширения диапазона поиска источников терминологической информации ПТИП требуется внедрение файловых форматов djvu, docx и др;

- для повышения эффективности поисковых возможностей ПТИП требуется подключение поисковых машин Google, Yandex и их совместное использование;

- в интересах совершенствования интерфейсных возможностей ПТИП целесообразно внедрение визуализации ТПИС;

- с целью повышения стабильности функционирования ВД ПСОХИ целесообразен переход на технологию распределенных БД с совместным использованием механизма репликации локальных БД, либо организация локального сетевого обмена данными между ХД и ВД, либо переход на облачные технологии;

- с целью расширения возможностей ППЗА в части учета стратегий принятия решений инвесторов на повышение и понижение, целесообразно применение гибридных ИНС (сетей Кохонена и МП) и использование роевых алгоритмов их обучения;

- для обеспечения максимального доверия к системе, на этапе формирования ИП, целесообразна разработка подсистемы логического вывода из области информационной технологии экспертных систем;

- в интересах расширения возможностей системы в данной предметной области предлагается: внедрение рисковых финансовых активов; применение для диверсификации рисков недвижимости, безрисковых активов и др.; реализация хеджирования рисков на основе фьючерсов; реализация модели арбитражного ценообразования и др.

ЗАКЛЮЧЕНИЕ

В ходе решения задач поставленных в диссертационной работе получены следующие основные результаты:

1. Обоснована модель снижения неопределенности внешней среды при формировании инвестиционного портфеля, отличающаяся учетом влияния уровня компетентности инвестора СЭО на основные показатели формируемого ИП.

2. Разработана модель определения состава ИСППИД, отличающаяся выявлением Парето-оптимальных комплектов подсистем и модулей, путем многокритериального оценивания их исходного инвариантного множества на основе метода дихотомического программирования.

3. Предложена стохастическая модель учета влияния неопределенности внешней среды на состав ИСППИД, основанная на увеличении числа дополнительных операций для достижения поставленной цели и позволяющая обосновать оптимальный по Парето ее набор компонент по критериям выделяемого финансового ресурса и достижимой точности.

4. Обоснован механизм учета априорной информации в дереве целей, базирующийся на применении готовых решений, обеспечивающих более оперативное достижение промежуточных целей и, в конечном итоге, глобальной цели.

5. Предложен метод оптимизации функциональной структуры ИСППИД, базирующийся на графоаналитической оценке и весовой обработке взаимосвязей между структурными элементами дерева целей, отличающийся построением и обработкой матриц весовых коэффициентов «цели-средства-данные», в отличие от традиционных «цели-средства», что позволяет построить адаптивную структуру системы, функционирующую на основе реконфигурации используемых моделей, методов и алгоритмов в зависимости от вида обрабатываемых данных.

6. Сформулирована постановка задачи синтеза структуры распределенной информационной системы, учитывающая распределение информации

по дескрипторам, описывающих документы, позволяющая органам управления оценивать неопределенность ответов на запросы.

7. Разработана нейромодифицированная одноиндексная модель Шарпа формирования ИП, учитывающая ситуативные рыночные изменения курсов ЦБ при прогнозировании их доходностей, на основе запоминания риск-эффектов с помощью ИНС.

8. Предложена нейрокомитетная модель У. Шарпа аналитического учета предыстории изменения доходностей активов при прогнозировании их значений в текущей ситуации в виде риск-эффектов с использованием ИНС для получения более точных результатов прогноза.

9. Разработан программный комплекс нейросетевого прогнозирования временных рядов, обеспечивающий оптимизацию структур ИНС, исследование и совершенствование численных методов их обучения, в интересах более точного и оперативного запоминания и воспроизведения риск-эффектов.

10. Предложен визуальный контактор, основанный на визуальной технологии моделирования, обеспечивающий отображение исследуемой ИНС и оперативную корректировку ее структуры без вникания в механизмы ее функционирования. Его применение позволяет более гибко моделировать структуру ИНС и обеспечивает проведение исследований малоизученных ИНС с разрушенными связями и нейронами.

11. Разработан модифицированный генетический алгоритм распределения инвестиций, отличающийся учетом построения перспективных поддеревьев решений, применением стратегии поиска решений «снизу-вверх» и использованием островной модели параллельных вычислений для повышения оперативности проведения расчетов за счет реализации параллельной обработки информации.

12. Предложена семантическая матрично-лексическая модель анализа текста, основанная на лексемах, представляющая текст в виде семантической сети с узлами – основными понятиями и дугами – семантическими отношениями между ними, позволяющая выделять из текста новые знания.

13. Разработан метод извлечения новых знаний, путем отбора новых терминов, содержащихся в тексте, на основе пороговых значений частот их встречаемости, позволяющий реализовать объяснение своих действий на основе восстановления логических цепочек взаимосвязей терминов в массиве текстовых документов.

14. Построена математическая модель весовой распределенной информационной системы на тезаурусе, учитывающая неопределенности в формируемых запросах и найденных документах.

15. Сформирована математическая модель весовой распределенной информационной системы на терминологическом портрете, позволяющая перераспределять поисковую нагрузку на подсистемы в зависимости от их загрузки и значимости искомых терминов.

16. Реализован алгоритм обмена данными между хранилищем и витриной, основанный на эмпирической оценке закона их ранговой популярности, повышающий надежность хранения данных и эффективность манипулирования ими.

17. Разработан модифицированный алгоритм индексации данных, на основе их предварительной обработки методом «деления узла», обеспечивающий повышение оперативности обработки запросов.

18. Разработаны концепция и технология внутрисистемного синтеза ИСППИД, основанные на мультиэтапном механизме проектирования, реализующем параллельное итеративное выполнение основных этапов проекта и применение априорной информации управления проектированием аналогов, позволяющие повысить оперативность проведения проектных работ и формирования проектных групп.

19. Разработано специальное математическое и программное обеспечение прототипа проблемно-ориентированной информационной системы поддержки портфельной инвестиционной деятельности, показавшее работоспособность и эффективность включенных в его состав методов, моделей и алгоритмов. Все разработанные программные продукты зарегистрированы в

фонде алгоритмов и программ, а также в центре информационных технологий и систем органов исполнительной власти.

20. Разработаны методики численной оценки эффективности функционирования основных подсистем ИСППИД, позволившие на основе вычислительных экспериментов выявить в них наиболее перспективные элементы (методы, модели и алгоритмы) для решении задач инвестиционного анализа на РЦБ в условиях стохастических данных об активах.

21. Обоснованы предложения по методологии внутрисистемного синтеза и перспективного развития ИСППИД, основанные на точностных и временных показателях разработанного прототипа, учитывающие закономерности изменения эффективности формируемых ИП в ПИД СЭО от точности, полноты, достоверности и оперативности информации предоставляемой ею в условиях стохастических данных об активах.

22. В рамках синтеза проектных решений предложен новый способ снижения неопределенности внешней среды – системной информированности. Его суть заключается в том, что в роли информатора в проекте выступает ИСППИД СЭО. Она реализует поиск, обработку и получение информации из различных источников (Интернет, научная литература, СМИ и др.) о возможном значении неизвестного параметра или способах его расчета. В отличие от остальных участников проекта, ИСППИД СЭО не обладает свойством манипулирования (не ищет выгоды для себя, и по этой причине не скрывает истинного положения дел).

СПИСОК СОКРАЩЕНИЙ

АИО - алгоритм имитации отжига

АОРО - алгоритм обратного распространения ошибки

БД - база данных

БДДИ - база данных дополнительной информации

БЗ - база знаний

БМ - база моделей

ВД - витрина данных

ГА - генетический алгоритм

ДИПС - документальная информационно-поисковая система

ДЦ - дерево целей

ЗС - загрузочная секция

ИМ - инвестиционный менеджмент

ИМИНЗ - индексный метод извлечения новых знаний

ИНС - искусственная нейронная сеть

ИО - инвестиционное обеспечение

ИП - инвестиционный портфель

ИПА - интеллектуальный поисковый агент

ИС - информационная система

ИСИМ - информационная система интегрированного менеджмента

ИСППИД - информационная система поддержки портфельной оптимизации

ИТТ - иерархический текстовый терминопостроитель

ИУ - информационное управление

ИФП - информационный процесс

КИНС - комитет искусственных нейронных сетей

КПИ - коэффициент прагматической информативности

КСЕИ - коэффициент семантической информативности

КСИИ - коэффициент синтаксической информативности

ЛПР - лицо, принимающее решения

МВ - модуль визуализации

МВЧ - мера Вапника-Червоненкиса

МГА - модифицированный генетический алгоритм

МЛС - матрица лексических связей

ММ - модель Марковица

МО - методическое обеспечение

МП - многослойный персептрон

МПУ - механизм проектного управления

МСА - метод системного анализа

НВС - неопределенность внешней среды

НКМШ - нейрокомитетная модель Шарпа

НМШ - нейромодифицированная модель Шарпа

НПВР - нейросетевое прогнозирование временных рядов

ОМПВ - островная модель параллельных вычислений

ОМШ - одноиндексная модель Шарпа

ПГ - проектная группа

ПИА - портфельный инвестиционный анализ

ПИД - портфельная инвестиционная деятельность

ПИП - портфельный инвестиционный процесс

ПИР - портфельное инвестиционное решение

ПИС - портфельная инвестиционная среда

ПИУР - портфельные инвестиционные управленческие решения

ПК - программный комплекс

ППЗА - подсистема прогнозирования значений активов

ПРО - подсистема расчета и оптимизации

ПСОХИ - подсистема сбора, обработки и хранения информации

ПТИП - подсистема терминологического информационного поиска

ПУ - проектное управление

ПФИ - портфельные финансовые инвестиции

РЦБ - рынок ценных бумаг

СА - системный анализ

СВРТП - система виртуального распределенного терминологического поиска

СМЛМ - семантическая матрично-лексическая модель

СМО - специальное математическое обеспечение

СМПО - специальное математическое и программное обеспечение

СППР - система поддержки принятия решений

СС - семантическая сеть

СУБД - система управления базой данных

СЭО - социально-экономическая организация

ТА - технический анализ

ТЗ - техническое задание

ТП - терминологический портрет

ТПО - теория портфельной оптимизации

ТПТ - терминологический портрет текста

ТПИС - терминологический портрет информационной системы

ФА - фундаментальный анализ

ФИУ - функционал инструментального уровня

ФСУ - функционал системного уровня

ФТУ - функционал технического уровня

ФЦУ - функционал целевого уровня

ХД - хранилище данных

ЦБ - ценная бумага

ЦФ - целевая функция

ЧТЗ - частное техническое задание

СПИСОК ЛИТЕРАТУРЫ

1. Айвазян, С.А. Прикладная статистика. Основы моделирования и первичная обработка данных / С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин – М.: Финансы и статистика, 1983. – 472 с.
2. Александер, В.С. Методы принятия инвестиционных решений на финансовых рынках. Дисс. к.э.н. Москва. 2006.
3. Анискин, Ю.Н. Управление инвестициями / Ю.Н. Анискин. – СПб.: «Омега-Л». 2002. – 167 с.
4. Арбузов, В.О. Имитационное моделирование микроструктуры фондового рынка на основе высокочастотной и трансакционной информации. Дисс. к.э.н. Пермь. 2016.
5. Аскинадзи, В.М. Инвестиционные стратегии на рынке ценных бумаг [текст] / В.М. Аскинадзи. – М.: ООО «Маркет ДС Корпорейшн», 2004. – 106 с.
6. Балаев, А.И. Составление портфелей ценных бумаг на основе прогнозирования совместной функции распределения доходностей. Автореф. дисс. к.ф-м.н. Москва. 2014.
7. Баркалов, С.А. Информационная поддержка принятия инвестиционных решений при формировании портфеля ценных бумаг / С.А. Баркалов, В.П. Морозов, А.В. Никитенко, А.И. Половинкина, А.И. Сырин // Информатизация и информационная безопасность правоохранительных органов: матер. XXII Всерос. науч. конфер. -М.: Академия управления МВД России, 2013. – С. 82-85.
8. Басовский, Л.Е. Экономическая оценка инвестиций / Е.Н. Басовская, Л.Е. Басовский. – М.: ИНФРА-М, 2007. – 241 с.
9. Батяева, Т.А. Рынок ценных бумаг / Т.А. Батяева, Т.И. Столяров. – М.: Инфра - М, 2006. – 408 с.
10. Бланк, И.А. Инвестиционный менеджмент / И.А. Бланк. – Киев: МП «ИТЕМ», ЛТД, 1995. – 315 с.
11. Бланк, И.Т. Управление инвестициями предприятия / И.Т. Бланк. – СПб.: «Ника-Центр». 2003. – 275 с.

12. Блауберг, И. В. Становление и сущность системного подхода / И.В. Блауберг, Э.Г. Юдин. – М., 1973. – 216 с.
13. Блех, Ю.А. Инвестиционные расчеты. Модели и методы оценки инвестиционных проектов / Ю.А. Блех, У.Р.Гетце. – Калининград: «Янтарный сказ», 1997. – 437 с.
14. Бочаров, В.П. Инвестиции / В.П. Бочаров. – СПб.: «Питер». 2002. – 288 с.
15. Бронштейн, И.Н. Справочник по математике для инженеров и учащихся вузов / И.Н. Бронштейн, К.А. Семендяев. – М.: Наука, 1986. – 534 с.
16. Бурков, В.Н. Задачи дихотомической оптимизации / В.Н. Бурков, И.В. Буркова // Системные проблемы качества, математического моделирования, информационных и электронных технологий – 2003: Международная науч.-тех. конф. –М.: Радио и связь. – С. 23-28.
17. Бурков, В.Н. Как управлять проектами / В.Н. Бурков, Д.А. Новиков - М.: СИНТЕГ- ГЕО, 1997. -188 с.
18. Валдайцев, В.Г. Инвестиции / В.Г. Валдайцев, П.П. Воробьев и др. – М.: ТК Велби, Изд-во Проспект, 2005. – 440 с.
19. Васютин, С.В. Базы данных. Интеллектуальная обработка информации / С.В. Васютин, А.Ф. Гореев, В.В. Корнеев - 2-е изд. - М.: Издатель Молгачёва С.В., Нолидж, 2001.-496 с.
20. Ватолин, Д. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео / Д.Ватолин, А. Ратушняк, М. Смирнов, В. Юкин. - М.: ДИАЛОГ-МИФИ, 2003. - 384 с.
21. Вахрин, П.С. Инвестиции / П.С. Вахрин. – СПб.: «Дашков и К», 2003. – 384 с.
22. Владимирова, Л.П. Прогнозирование и планирование в условиях рынка / Л.П. Владимирова. - М: Издательско-торговая корпорация «Дашков и К», 2008. – 400 с.
23. Волков, Г.Г. Компьютерные информационные технологии / Г.Г. Волков, О.Ю. Глинский. - БГЭУ: Бобруйск, 2010. – 86 с.
24. Волкова, В.Н. Основы теории систем и системного анализа / В.Н. Волкова, А.А. Денисов. Изд. 2-е. – СПб.: СПб.ГТУ, 1999. – 230 с.
25. Володин, С.Н. Эффективность методов технического анализа при

сверхкраткосрочных операциях на фондовом рынке. Дисс. к.э.н. Москва. 2013.

26. Воробьев, П.В. Рынок ценных бумаг: учеб. / П.В. Воробьев, В.А. Лялин. М.: Изд-во Проспект, 2009. - 384с.
27. Вороновский, Г.К. Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности / Г.К. Вороновский, К.В. Махотило, С.Н.Петрашев, С.А. Сергеев. – М.: ОСНОВА, 1997. – 112 с.
28. Галиев, Д.Р. Формирование эффективного портфеля ценных бумаг на основе альтернативных мер риска и доходности. Автореф. дисс. к.э.н. Уфа. 2013.
29. Герасимов, Е.С. Многомерные динамические сетевые модели управления инвестиционным портфелем. Автореф. дисс. к.ф-м.н. Томск. 2005.
30. Гитман, Л.Д. Основы инвестирования: Пер. с англ. / М.Д. Джонк. – М.: «Дело», 1997. – 1015 с.
31. Гладков, Л.А. Генетические алгоритмы / Л.А. Гладков. – М.: Физматлит, 2006. – 268 с.
32. Головачев, С.С. Прогнозирование доходности на фондовом и валютном рынках на основе моделей искусственных нейронных сетей. Дисс. к.э.н. Москва. 2014.
33. Грибачев, К Г. Delphi и Model Driven Architecture. Разработка приложений баз данных / К.Г. Грибачев. - СПб.: Питер, 2004. — 348 с.
34. Гудман, Э.Д. Эволюционные вычисления и генетические алгоритмы. Обозрение прикладной и промышленной математики / Э.Д. Гудман, А.П. Коваленко. - Т. 3. – Вып. 5. – М.: ТВП, 1986. – 760 с.
35. Гультяев, А. Виртуальные машины – несколько компьютеров в одном / А. Гультяев. – СПб.: Питер, 2006. – 224 с.
36. Давнис, В.В. Модифицированный вариант модели Шарпа, его свойства и стратегии управления инвестиционным портфелем / В.В. Давнис, С. Е. Касаткин, Е. А. Ратушная // Современная экономика: Проблемы и решения. – Воронеж, 2010. – № 9. – С. 135 – 145.
37. Данилюк, Е.Ю. Обработка данных финансового рынка и принятие ре-

шения о структуре Европейского опциона. Дисс. к.ф.-м.н. Томск. 2014.
38. Деева, А.К. Инвестиции / А.К. Деева. – М.: «Экзамен». 2004. – 320 с.
39. Дмитриев, А.К. П.А. Основы теории построения и контроля сложных систем / Дмитриев А.К. П.А. Мальцев. – Л.: Энергоатомиздат, 1988. – 192 с.
40. Дружинин, В.В. Проблемы системотологии (проблемы теории сложных систем) / В.В. Дружинин, Д.С. Конторов. – М.: Сов, радио, 1976. – 237 с.
41. Дункан, Д.У. Основополагающие идеи в менеджменте. Уроки основоположников менеджмента и управленческой практики. М.: Дело, 1996. 342 с.
42. Евстигнеев, В.А. Теория графов: алгоритмы обработки деревьев/ В.А.Евстигнеев, В.Н.Касьянов. - Новосибирск: Наука, 1994. -361 с.
43. Зализняк, А.А. Грамматический словарь русского языка: словоизменение / А.А. Зализняк. М.:Рус.яз., 1987. - 180 с.
44. Звегинцев, В. А. Предложение и его отношение к языку и речи / В.А. Звегинцев. – М.: Московский университет, 1976. – 250 с.
45. Зуев, Д.В. Аналитическая оценка структурированных производных финансовых инструментов. Дисс. к.э.н. Москва. 2016.
46. Ивасенко, А.Г. Информационные технологии в экономике и управлении: учебное пособие / А.Г. Ивасенко, А.Ю. Гридасов, В.А. Павленко. – 2-е изд., стер. – М.: КНОРУС, 2007. – 160 с.
47. Избачков, Ю.С. Информационные системы / Ю.С. Избачков, В.Н. Петров. – СПб.: Питер, 2006. – 656 с.
48. Информационная технология. Комплекс стандартов и руководящих документов на автоматизированные системы. - М.: Комитет стандартизации и метрологии СССР, 1991. - 120 с.
49. Карпиков, Е. / Е. Карпиков, А. Федоров. Основные постулаты классической теории портфельных инвестиций // http://speculator-fin.ru/page-id-121.html
50. Карпова, Т.С. Базы данных: модели разработки, реализация: учеб. пособие / Т.С. Карпова. – СПб.: Питер, 2001. - 304 с.
51. Киселевич, Ю.В. Формирование и моделирование инвестиционных портфелей с учетом особенностей фондового рынка РФ. Автореф.

дисс. к.э.н. Москва. 2006.

52. Клименко, С.В. Электронные документы в корпоративных сетях / С.В. Клименко, И.В. Крохин, В.М. Кущ, Ю.Л. Лагутин. - М.: Анкей, 1998. – 272 с.
53. Клини, С. Основания интуиционистской математики с точки зрения теории рекурсивных функций / С. Клини, Р. Весли. – М.: Издательство «НАУКА», 1978. – 271 с.
54. Ковалёв, В.В. Введение в финансовый менеджмент / В.В. Ковалев.- М.: Финансы и статистика, 2003.- 168 с.
55. Ковалёв, В.В. Инвестиции: учеб. / С.В. Валдайцев, П.П. Воробьев [и др.]; под ред. В.В. Ковалева, В.В. Иванова, В.А. Лялина. - М.: ТК Велби, Изд-во Проспект, 2005. - 440 с.
56. Колтынок, Б.Н. Инвестиционные проекты / Б.Н. Колтынок. – М.: «Дело». 2002. – 622 с.
57. Кормен, Т. Алгоритмы: построение и анализ / Т. Кормен, Ч. Лейзерсон, Р. Ривест. М.: МЦНМО: БИНОМ. Лаборатория знаний, 2004. – 348 с.
58. Корн, Г. Справочник по математике / Г. Корн, Т. Корн. М.: Издательство «Наука», 1973. – 832 с.
59. Корчагин, Ю.А. Инвестиции: теория и практика / Ю.А. Корчагин, И.П. Маличенко. – Ростов н/Д: Феникс, 2008. – 509 с.
60. Космачев, С.Н. Автоматизированные информационные системы / С.Н. Космачев. – М.: Издательский центр «Академия», 2009. – 97 с.
61. Костров, А.В. Системный анализ и принятие решений: учеб. пособие / А.В.Костров. - Владимир, 1995. - 66 с.
62. Крейнес, М.Г. Смысловой поиск и индексирование текстовой информации в электронных библиотеках: информационная технология "ключи от текста"[Электронный ресурс] / М.Г. Крейнес 167 //Электронные библиотеки. - 2001. – Т.2, вып. 3. Режим доступа: http://www.mmascience.ru/ktt/start.htm.
63. Круг, П.Г. Нейронные сети и нейрокомпьютеры: Учебное пособие по курсу «Микропроцессоры» / П.Г. Круг– М.: Издательство МЭИ, 2002. – 176 с.

64. Крюков, П.А. Финансовые инвестиции на международном валютном рынке. Автореф. дисс. к.э.н. Томск. 2015.
65. Курейчик, В.М. Генетические алгоритмы / В.М. Курейчик. - Таганрог, 1998. - 135 с.
66. Куреленкова, Ю.В. Система поддержки принятия решений при формировании портфеля ценных бумаг на основе комплексных мер риска. Автореф. дисс. к.т.н. Уфа. 2009.
67. Леонтьева, Н.Н. Русский общесемантический словарь (РОСС): структура, наполнение / Н.Н. Леонтьева // НТИ. Сер.2.-1997.-№ 12.-С.5-20.
68. Лойко, В.И. Информационные системы и технологии в экономике / В.И. Лойко и др. - М.: Финансы и статистика, 2005. - 416 с.
69. Лотоцкий, В.А. Идентификация структур и параметров систем управления / В.А. Лотоцкий // Измерения. Контроль. Автоматизация. 1991. – № 3–4. – С.30–38.
70. Маклаков, С.В. BPWin и ERWin. CASE – средства разработки информационных систем / С.В. Маклаков. – М.: ДИАЛОГ-МИФИ, 2000. – 256 с.
71. Малюгин, В.И. «Рынок ценных бумаг: Количественные методы анализа»: учеб. пособие /В.И. Малюгин. – М.: «Дело», 2003. – 320 с.70
72. Маренков, Р.Р. Основы управления инвестициями / Р.Р. Маренков. – М.: «Едиториал». 2003. – 480с.
73. Марлов, А.В. Логико-лингвистический интегратор русскоязычных текстов для информационной системы в строительстве. Автореф. и дисс. к.т.н. Воронеж. 2004.
74. Матвеев, Л.А. Компьютерная поддержка решений / Л.А. Матвеев. – СПб.: «Специальная литература», 1998. – 472 с.
75. Месарович, М. Общая теория систем: математические основы / М. Месарович, И. Такахара. - М.: Мир, 1978. – 248 с.
76. Месарович, М. Теория иерархических многоуровневых систем / Д. Мако, М. Месарович, И. Такахара. М.: Мир, 1973. – 344 с.
77. Методология IDF0. Функциональное моделирование. - М.: Метатехнология, 1993. – 117 с.

78. Методология IDFX1. Информационное моделирование. – М.: Метатехнология, 1993. – 120 с.
79. Методология динамического моделирования IDF0/CPN/ Материалы 6 семинара «Информационные технологии в проектировании систем и управлении бизнесом». – М.: Метатехнология, 1994. – 13 с.
80. Мешалкин, В.П. Введение в базы данных: Учеб. пособие/ В.П.Мешалкин, В.Н.Фефелов, М.А.Пермяков. -М., 1993. -127 с.
81. Микова, Е.С. Моментум эффект в динамике цен акций российского рынка. Дисс. к.э.н. Москва. 2014.
82. Мински, М. Фреймы для представления знаний /М. Мински. - М.: Энергия, 1979. – 180 с.
83. Миркин, Я.М. Российский рынок ценных бумаг: влияние фундаментальных факторов, приоритеты и механизм развития. Дисс. к.э.н. Москва. 2003.
84. Моисеев, Н.Н. Математические задачи системного анализа / Н.Н. Моисеев. – М.: Наука, 1981. – 488 с.
85. Морозов, В.П. Автоматизированное прогнозирование объемов авиаперевозок воздушно-транспортной системы региона с учетом сезонных волн: алгоритм и программное средство / Б.Г. Лютиков, В.П. Морозов, П.В. Селезнев // Гос. фонд алгор. и прогр. Рег.№50200700047 от 12.01.2007 г.
86. Морозов, В.П. Анализ лингвистических подходов к формализации языка для разработки процессоров сбора данных / В.П. Морозов, А.В. Марлов, В. Ю. Храмов // Телекоммуникации. №6, 2004.- С.21-25.
87. Морозов, В.П. Визуальный контактор для моделирования искусственных нейронных сетей / В.П. Морозов, А. С. Ларин, А.И. Сырин // Математическое моделирование в технике и технологии: Всероссийская конференция с элементами научной школы. – Воронеж: ВИВТ-ВГТУ, 2011. – С. 247-249.
88. Морозов, В.П. Влияние внешней среды на функционирование предприятия / В.П. Морозов., А.И. Сырин // Управление большими системами:

матер. IX Всерос. школы-конфер. Молод. Ученых. Том I / Липецк. гос. тех. универ.-Тамбов-Липецк: Изд-во Першина Р.В., 2012. – С. 211-214.

89. Морозов, В.П. Влияние организационно-технических факторов внутренней среды на функционирование предприятия / В.П. Морозов., А.И. Сырин // Управление большими системами: матер. IX Всерос. школы-конфер. Молод. Ученых. Том I / Липецк. гос. тех. универ.-Тамбов-Липецк: Изд-во Першина Р.В., 2012. – С. 209-211.
90. Морозов, В.П. Возможности автоматизации управления информационной безопасностью организации на основе инноваций / В.П. Морозов, А.И. Сырин, Л. Г. Попова // Информация и безопасность, 2012. – №3. – С. 383 - 388.
91. Морозов, В.П. Выбор нейронных сетей для управления стабильным функционированием организации / Д.Г. Кобзарь., В.П. Морозов., А.И. Сырин // Современные сложные системы управления. HTCS – 2012: Международная науч.-тех. конф. - Ст. Оскол, 2012. – С.38-41.
92. Морозов, В.П. Высокие технологии как инструмент информатизации современной организации / С.А. Баркалов, В.П. Морозов, А.И. Сырин, Л.Г. Попова // Информатизация и информационная безопасность правоохранительных органов: матер. XXI Всерос. научн. конфер. -М.: Академия управления МВД России, 2012. – С. 21-26.
93. Морозов, В.П. Генетический алгоритм распределения финансовых инвестиций / В.П. Морозов // Развитие экономики, педагогики и информатики в социальных системах: Матер. ежегодной научно-практической конференции. – Воронеж: Воронежский госпедуниверситет, 2010. – С.53-55.
94. Морозов, В.П. Документальная информационно-поисковая система «Научные труды» / С.А. Баркалов, Н.П. Курочка, В.П. Морозов, А.И. Сырин // Государственный информационный фонд неопубликованных документов ФГАНУ «Центр информационных технологий и систем органов исполнительной власти». Рег.№50201450824 от 12.12.2014 г.
95. Морозов, В.П. Инструментарий интегрированного менеджмента ста-

бильного функционирования организации / В.П. Морозов // Инновации в сфере науки, образовании и высоких технологий. Малое инновационное предпринимательство (INNOV -2012). Электронный сборник тезисов научно-образовательного форума. Воронежский ГАСУ, Инновационный бизнес-инкубатор, Воронеж 2012, №531 (9 с).

96. Морозов, В.П. Интегрированный менеджмент – основа обеспечения стабильного функционирования организации / В.П. Морозов // Перспективы инновационного развития современного мирового сообщества: экономико-правовые и социальные аспекты: матер. Междунар. юбил. науч.-практ. конфер. ВЭПИ-ВГЛТА-2012. – С. 119-121.

97. Морозов, В.П. Интегрированный менеджмент XXI века: Парадигма безопасного и устойчивого (антикризисного) развития: монография / С.В. Барковская, Е.А. Жидко, В.П. Морозов, Л.Г. Попова. - Воронеж: ГОУ ВПО ВГАСУ, 2011. – 168с.

198. Морозов, В.П. Интегрированный менеджмент XXI века: Проектное управление устойчивостью развития: учебное пособие / С.В. Барковская, Е.А. Жидко, В.П. Морозов, Л.Г. Попова. - Воронеж: ФГБО ВПО ВГАСУ, 2011. – 167 с.

199. Морозов, В.П. Интегрированный менеджмент телекоммуникационной компании: система принципов / В.П.Морозов, Д.Е. Кобзарь, А.И. Сырин // Телекоммуникации. №6, 2013.- С.43-50.

100. Морозов, В.П. Интеллектуальная система интегрированного менеджмента XXI века / Е. А. Жидко, В.П. Морозов, Л.Г. Попова // Информация и безопасность, 2011. – №1 (18). – С. 59 - 64.

101. Морозов, В.П. Информационная система поддержки принятия инвестиционных решений в условиях неопределенности внешней среды: монография / Л.Е. Мистров, В.П. Морозов. - Воронеж: Воронежский ГАСУ. - 2016. – 245 с.

102. Морозов, В.П. Информационная система распределения финансовых инвестиций малого предприятия / Н.П. Курочка, Д.О. Косенко, В.П. Морозов, В.Г. Сырин, А.Н. Чекомазов // Государственный ин-

формационный фонд неопубликованных документов ФГАНУ «Центр информационных технологий и систем органов исполнительной власти». Рег.№50201450825 от 12.12.2014 г.

103. Морозов, В.П. Информационное обеспечение интегрированного менеджмента стабильного функционирования организации. Вербальная и математическая постановка проблемы / В.П. Морозов, А.И. Сырин // Научный вестник Воронежского государственного архитектурно-строительного университета. Серия: Управление строительством. Выпуск №1(4), 2013. – С. 244-251.
104. Морозов, В.П. Информационное обеспечение модернизации России в XXI веке: высокие технологии / С.В. Барковская, Е.А. Жидко, А. С. Ларин, В.П. Морозов, Л.Г. Попова // Вестник Воронежского государственного технического университета. 2011. Т.7. -№9. –С.33-37.
105. Морозов, В.П. Информационное обеспечение системы распределения финансовых инвестиций / В.П. Морозов, А. С. Ларин, А.М. Яцев // Инновационные механизмы экономики и педагогики в современной России: теория и практика: Материалы всероссийской научно-практической конференции. - Воронеж: ГООУ ВПО ВИИС, 2011.- С. 250 - 252.
106. Морозов, В.П. Информационные задачи обеспечения стабильного функционирования организации / В.П. Морозов, Д. Г. Кобзарь, А. В. Никитенко, А.И. Сырин // Материалы Междунар. молодежная конфер. «Математические проблемы современной теории управления системами и процессами». - Воронеж: ИПЦ «Научная книга», 2012. – С. 135 - 139.
107. Морозов, В.П. Информационные риски модернизации России: эвентологическое моделирование / С.В. Барковская, Е. А. Жидко, В.П. Морозов, Л.Г. Попова // Вестник Воронежского государственного технического университета. 2011. Т.7. -№8. –С.123-127.
108. Морозов, В.П. Информационные системы в производственном менеджменте: учебное пособие / Б.Г. Лютиков, В.П. Морозов, П.В. Селезнев. - Воронеж: АНО МОК ВЭПИ, 2006. – 132 с.
109. Морозов, В.П. Информационные системы: учебное пособие / В.П.

Морозов, А.И. Кустов, В.Ю. Храмов. - Воронеж: ВЭПИ, 2004. – 56 с.

110. Морозов, В.П. Информационный менеджмент: учебное пособие / В.П. Морозов, П.В. Селезнев, А.Ю. Скрипник. - Воронеж: АНО МОК ВЭПИ, 2006. – 69 с.
111. Морозов, В.П. Классификация искусственных нейронных сетей / Д.Г. Кобзарь, В.П. Морозов. А.И. Сырин // Инновационные механизмы экономики и педагогики в современной России: теория и практика (экономические, социальные, педагогические, правовые тенденции): матер. межвузовск. научн. – практ. конфер. – Воронеж: Изд-во «Наука - Юнипресс», 2012. – С. 80-85.
112. Морозов, В.П. Классификация коммуникаций в рамках системы управления интегрированного менеджмента организационной системы / В.П. Морозов, Д. Г. Кобзарь, А.И. Сырин // Материалы Междунар. молодежная конфер. «Математические проблемы современной теории управления системами и процессами». - Воронеж: ИПЦ «Научная книга», 2012. – С. 118 - 120.
113. Морозов, В.П. Классификация, цена и доходность акций / В.П. Морозов, Б.Г. Лютиков, С.В. Ковалевский // Развитие экономической науки России в XXI в.: Материалы междунар. нучно.-практич. конф. - Воронеж: АНО МОК «ВЭПИ», 2009.- С. 21-23.
114. Морозов, В.П. Классификация, цена и доходность облигаций / В.П. Морозов, Б.Г. Лютиков, А.В. Харламов // Развитие экономической науки России в XXI в.: Материалы междунар. нучно.-практич. конф. - Воронеж: АНО МОК «ВЭПИ», 2009.- С. 18-20.
115. Морозов, В.П. Концептуальные основы управления проектом разработки информационных систем интегрированного менеджмента социально-экономических организаций / В.П. Морозов // Интернет-журнал «Науковедение». 2013 №1 (14) (05ТВН113) [Электронный ресурс]. - М. 2013. – Режим доступа: http://naukovedenie.ru/index.php?p=issue-1-13, свободный – Загл. с экрана.
116. Морозов, В.П. Концепция и технология управления проектами со-

здания и развития информационных систем интегрированного менеджмента социально-экономических организаций / В.П. Морозов, А. В. Никитенко, А.И. Сырин // Современные сложные системы управления: сб. статей по материалам международной конференции. – Воронежский ГАСУ. – Воронеж, 2013. – С. 72-80.

117. Морозов, В.П. Логико-лингвистический подход к моделированию радиоэлектронной обстановки / В.П. Морозов, М.М. Ступницкий // Электросвязь. 1997. – №1. – С. 21 - 23.
118. Морозов, В.П. Математическая модель весовой распределенной информационной системы на тезаурусе / В.П. Морозов // Системы управления и информационные технологии. 2012. №3 (49). – С. 54 – 56.
119. Морозов, В.П. Математическая модель подсистемы весовой обработки информации на заданном терминологическом портрете / В.П. Морозов // Ин-тернет-журнал «Науковедение». 2013 №1 (14) (27TBH113) [Электронный ре-сурс]. - М. 2013. – Режим доступа: http://naukovedenie.ru/index.php?p=issue-1-13, свободный – Загл. с экрана.
120. Морозов, В.П. Метод повышения устойчивости функционирования информационной системы поддержки принятия инвестиционных решений на основе разрешения ресурсного конфликта элементов / Л.Е. Мистров, В.П. Морозов // Интернет-журнал «Науковедение». 2016, №5 (8) (35EVN516) [Электронный ресурс]. - М. 2016.
121. Морозов, В.П. Методологические основы интегрированного менеджмента стабильного функционирования строительной организации / Д. Г. Кобзарь, В.П. Морозов, А. В. Никитенко, А.И. Сырин // Экономика и менеджмент систем управления. 2012, №4.1 (6). – С. 138 - 147.
122. Морозов, В.П. Методология и технология интегрированного менеджмента строительной организации. - Воронеж: Воронежский ГАСУ. - 2012. – 150 с.
123. Морозов, В.П. Метод моделирования конфликтного взаимодействия социально-экономических организаций / Л.Е. Мистров, В.П. Морозов, А.В. Петриченко // Динамика сложных систем – XXI век. 2017,

№1. – С. 12 - 19.

124. Морозов, В.П. Методы моделирования сезонных волн с предварительным элиминированием тренда / В.П. Морозов, Б.Г. Лютиков, П.В. Селезнев // Актуальные проблемы науки в студенческих исследованиях: Труды VII межвузовской студенческой научной конференции. – ВЭПИ. Воронеж, 2006. – С. 6-11.
125. Морозов, В.П. Модели управления проектами создания и развития систем поддержки принятия инвестиционных решений / В.П. Морозов // Современные сложные системы управления: матер. XI междунар. конфер. – Воронеж: Воронежский ГАСУ. – 2014. Ч 1. – С. 121-130.
126. Морозов, В.П. Модели и алгоритмы проектирования и разработки систем поддержки принятия инвестиционных решений: монография / В.П. Морозов, С.А. Баркалов, А.И. Сырин. - Воронеж: Воронежский ГАСУ. - 2015. – 232 с.
127. Морозов, В.П. Моделирование влияния рельефа местности на распространение радиоволн при выборе позиций средств радиоэлектронной борьбы. Автореф. и дисс. к.т.н. Воронеж. 2002.
128. Морозов, В.П. Моделирование функционирования социально-экономических систем в конфликтах: постановка задачи / Л.Е. Мистров, В.П. Морозов // Комплексные проблемы техносферной безопасности: материалы междунар. науч. - практ. конф. Воронеж: ФГБОУ ВО «ВГТУ», 2016. Ч. II. – С. 243 – 247.
129. Морозов, В.П. Модель информационного снижения неопределенности внешней среды в задачах инвестиционного анализа / С.А. Баркалов, И.В. Горошко, В.П. Морозов // Моделирование и наукоемкие информационные технологии в технических и социально-экономических системах: материалы IV международной научно-практической конференции. - Новокузнецк: СибГИУ, 2016. – С. 22 - 26.
130. Морозов, В.П. Модель оценки влияния информационных воздействий на качество информации в автоматизированных системах управления / В.П. Морозов, А.В. Неволин, О.Г. Никитин, В.А.

Скляров // Телекоммуникации. №10, 2004.- С.35-41.

131. Морозов, В.П. Модель среднесрочного прогноза рентабельности авиационных рейсов для воздушно-транспортной системы региона / В.П. Морозов, Б.Г. Лютиков, П.В. Селезнев // Актуальные проблемы науки в студенческих исследованиях: Труды VI межвузовской студенческой научной конференции. – ВЭПИ. Воронеж, 2005. – С. 21-26.
132. Морозов, В.П. Модификация модели Шарпа / В.П. Морозов, А. С. Ларин, А.И. Сырин // Математическое моделирование в технике и технологии: Всероссийская конференция с элементами научной школы. – Воронеж: ВИВТ-ВГТУ, 2011. – С. 244-246.
133. Морозов, В.П. Модификация модели Шарпа на основе комитета нейронных сетей / В.П. Морозов // Успехи современной науки. 2017, №2, Т.4. – С. 73 - 75.
134. Морозов, В.П. Модификация сети Кохонена для кластеризации многомерных данных. Постановка задачи / В.П. Морозов, А. С. Ларин, О. А. Седунова // Инновационные механизмы экономики и педагогики в современной России: теория и практика: Материалы всероссийской научно-практической конференции. - Воронеж: ГООУ ВПО ВИИС, 2011.- С. 238-240.
135. Морозов, В.П. Модифицированный генетический алгоритм выбора конфигурации оборудования проектируемой телекоммуникационной сети / Б.Г. Лютиков, В.П. Морозов, Д.В. Морозов // Телекоммуникации. №5, 2008 г.-С.10-12.
136. Морозов, В.П. Модуль оценки облигаций как инструмент информационной системы поддержки портфельной оптимизации компании / В.П. Морозов, В.Л. Порядина, А.И. Сырин, О.И. Торубарова // Современные тенденции развития науки и технологий: материалы XXII международной заочной научно-практической конференции. – Белгород: АПНИ, 2017. – С. 98 - 102.
137. Морозов, В.П. Нечеткий поиск дубликатов текстов на основе алгоритма шинглов / В.П. Морозов, А. С. Ларин, В.Ю. Наумов // Иннова-

ционные механизмы экономики и педагогики в современной России: теория и практика: Материалы всероссийской научно-практической конференции. - Воронеж: ГООУ ВПО ВИИС, 2011.- С.234,235.

138. Морозов, В.П. Обеспечение информационной безопасности модернизации России в условиях состязательности: эвентология / Е. А. Жидко, В.П. Морозов, Л.Г. Попова // Информация и безопасность, 2011. – №1 (18). – С. 51 - 58.

139. Морозов, В.П. Обеспечение стабильного функционирования телекоммуникационной компании в современных условиях / В.П. Морозов // Телекоммуникации. №1, 2013 г.-С.42-48.

140. Морозов, В.П. Облик системы управления интегрированного менеджмента организации / В.П. Морозов, Д. Г. Кобзарь, А. В. Никитенко, А.И. Сырин // Материалы Междунар. молодежная конфер. «Математические проблемы современной теории управления системами и процессами». - Воронеж: ИПЦ «Научная книга», 2012. – С. 157 - 159.

141. Морозов, В.П. Обоснование инвестиционных решений как элемент инновационного механизма в организационной системе / В.П. Морозов, Д. Г. Кобзарь, А. В. Никитенко, А.И. Сырин // Междунар. науч. – практ. конфер. «Перспективы и проблемы инновационного развития социально-экономических систем. – Воронеж: ВГУИТ, 2012. – С. 40-43.

142. Морозов, В.П. Обоснование математического обеспечения подсистемы поиска экономических кризисных тенденций внешней среды / В.П. Морозов // Междунар. науч. – практ. конфер. «Перспективы и проблемы инновационного развития социально-экономических систем. – Воронеж: ВГУИТ, 2012. – С. 38-40.

143. Морозов, В.П. Обоснование облика интеллектуальной информационной системы интегрированного менеджмента организации XXI века / Е.А. Жидко, А. С. Ларин, В.П. Морозов, Л.Г. Попова // Информация и безопасность, 2011. – №1 (18). – С. 65 - 70.

144. Морозов, В.П. Обоснование состава информационной системы поддержки принятия инвестиционных решений / Е.В. Баутина, Л.Е.

Мистров, В.П. Морозов, А.И. Сырин // Экономика и менеджмент систем управления. 2016, №2.1 (20). – С. 196 - 200.

145. Морозов, В.П. Обоснование специального математического и программного обеспечения системы распределения финансовых инвестиций / В.П. Морозов, А. С. Ларин // Инновационные механизмы экономики и педагогики в современной России: теория и практика: Материалы всероссийской научно-практической конференции. - Воронеж: ГООУ ВПО ВИИС, 2011.- С. 232-234.
146. Морозов, В.П. Обработка временных рядов на основе модифицированной сети Кохонена / В.П. Морозов, А. В. Никитенко // Материалы Междунар. молодежная конфер. «Математические проблемы современной теории управления системами и процессами». - Воронеж: ИПЦ «Научная книга», 2012. – С. 143 - 145.
147. Морозов, В.П. Определение высоких информационной и интеллектуальной технологий с позиций системного подхода / В.П. Морозов, А. В. Никитенко // Научный вестник Воронежского государственного архитектурно-строительного университета. Серия: Управление строительством. Выпуск №1(4), 2013. – С. 252-258.
148. Морозов, В.П. Оптимизация коммуникаций в строительной организации / С.А. Баркалов, В.П. Морозов, Д. Г. Кобзарь, А.И. Сырин // Междунар. науч. – практ. конфер. «Перспективы и проблемы инновационного развития социально-экономических систем. – Воронеж: ВГУИТ, 2012. – С. 140-143.
149. Морозов, В.П. Оптимизация функциональной структуры автоматизированной информационной системы поддержки принятия инвестиционных решений / В.П. Морозов // Успехи современной науки. 2016, №11, Т.4. – С. 74 - 76.
150. Морозов, В.П. Оценка эффективности функционирования информационной системы поддержки портфельной оптимизации в составе социально-экономической организации / В.П. Морозов // Интернет-журнал «НАУКОВЕДЕНИЕ» Том 9, №2 (2017)

http://naukovedenie.ru/PDF/117TVN217.pdf. Загл. с экрана. Яз. рус., англ.

151. Морозов, В.П. Поддержка мотивационных решений в системе интегрированного менеджмента организации строительной индустрии / В.П. Морозов, Д. Г. Кобзарь, А.И. Сырин // Материалы Междунар. молодежная конфер. «Математические проблемы современной теории управления системами и процессами». - Воронеж: ИПЦ «Научная книга», 2012. – С. 151 - 154.
152. Морозов, В.П. Подходы интегрированного менеджмента современной телекоммуникационной компании / В.П.Морозов, А.И. Сырин // Телекоммуникации. №9, 2013.- С.36-42.
153. Морозов, В.П. Постановка задачи стохастического учета влияния неопределенности внешней среды при управлении проектами создания и развития систем интегрированного менеджмента / В.П. Морозов // Интернет-журнал «Науковедение». 2013 №1 (14) (28TBH113) [Электронный ресурс]. - М. 2013. – Режим доступа: http://naukovedenie.ru/index.php?p=issue-1-13, свободный – Загл. с экрана.
154. Морозов, В.П. Построение дерева целей для автоматизированной информационной системы поддержки принятия инвестиционных решений / В.П. Морозов // Успехи современной науки. 2016, №11, Т.4. – С. 54 - 57.
155. Морозов, В.П. Применение моделей фундаментального анализа при управлении финансовыми инвестициями малого предприятия / С.А. Баркалов, В.П. Морозов, А. В. Никитенко, А.И. Сырин // Моделирование и наукоемкие информационные технологии в технических и социально-экономических системах: материалы IV международной научно-практической конференции. - Новокузнецк: СибГИУ, 2016. – С. 16 - 22.
156. Морозов, В.П. Применение нейро - нечетких сетей в системах поддержки принятия решений / В.П. Морозов, А. С. Ларин, М. А. Маслова // Инновационные механизмы экономики и педагогики в современной России: теория и практика: Материалы всероссийской научно-практической конференции. - Воронеж: ГООУ ВПО ВИИС, 2011.- С. 240-242.

157. Морозов, В.П. Применение островной модели параллельных вычислений для принятия инвестиционных решений / В.П. Морозов, А. В. Никитенко // Материалы Международной молодежной научной школы «Летняя Суперкомпьютерная Академия» . - Воронеж: ИПЦ «Научная книга», 2012. – С.64 - 68.
158. Морозов, В.П. Приоритеты сравниваемых элементов квалиметрическим методом / М.И. Ломакин, Л.Е. Мистров, В.П. Морозов // Стандарты и качество. 2017, №1 (955). – С. 76 - 79.
159. Морозов, В.П. Программная реализация автоматизированного терминологического поиска в сети интернет / В.П. Морозов, А. С. Ларин, В.Ю. Наумов // Инновационные механизмы экономики и педагогики в современной России: теория и практика: Материалы всероссийской научно-практической конференции. - Воронеж: ГООУ ВПО ВИИС, 2011.- С. 243-245.
160. Морозов, В.П. Программный комплекс нейросетевого прогнозирования временных рядов интегрированного менеджмента / Д.Г. Кобзарь, В.П. Морозов, А.В. Никитенко, А.И. Сырин // Гос. инф. фонд неопубликованных документов ФГАНУ «Центр информационных технологий и систем органов исполнительной власти». Рег.№ №50201350236 от 13.03.2013 г.
161. Морозов, В.П. Программный комплекс формирования портфеля инвестора / В.П. Морозов // Материалы научной сессии преподавателей ВЭПИ-2010. - Воронеж: АНО МОК «ВЭПИ», 2010.- С. 83-86.
162. Морозов, В.П. Проектное управление разработкой информационных систем интегрированного менеджмента / В.П. Морозов, А. В. Никитенко, А.И. Сырин // Научный вестник Воронежского государственного архитектурно-строительного университета. Серия: Управление строительством. Выпуск №2(5), 2013. – С. 66-73.
163. Морозов, В.П. Разработка информационно - справочной системы для обоснования мотивационных решений / С.А. Баркалов, В.П. Морозов, Д. Г. Кобзарь, А. В. Никитенко // Междунар. науч. – практ. конфер. «Перспективы и проблемы инновационного развития социально-

экономических систем. – Воронеж: ВГУИТ, 2012. – С. 143-145.

164. Морозов, В.П. Реализация генетического алгоритма распределения финансовых инвестиций на основе островной модели параллельных вычислений / В.П. Морозов, Б.Г. Лютиков // Проблемы развития экономической системы России в XXI в.: Материалы междунар. нучно.-практич. конф. - Воронеж: АНО МОК «ВЭПИ», 2008.- С. 6-10.
165. Морозов, В.П. Реализация модели Марковица в системе поддержки принятия инвестиционных решений / В.П. Морозов, А. С. Ларин, Е. В. Ватутина // Инновационные механизмы экономики и педагогики в современной России: теория и практика: Материалы всероссийской научно-практической конференции. - Воронеж: ГООУ ВПО ВИИС, 2011.- С.246,247.
166. Морозов, В.П. Реализация модели Шарпа в системе поддержки принятия инвестиционных решений / В.П. Морозов, А. С. Ларин, О.В. Корчагина // Инновационные механизмы экономики и педагогики в современной России: теория и практика: Материалы всероссийской научно-практической конференции. - Воронеж: ГООУ ВПО ВИИС, 2011.- С. 248, 249.
167. Морозов, В.П. Рекуррентная сеть Эльмана для моделирования временных рядов / В.П. Морозов, А. С. Ларин, М.В. Самарина // Инновационные механизмы экономики и педагогики в современной России: теория и практика: Материалы всероссийской научно-практической конференции. - Воронеж: ГООУ ВПО ВИИС, 2011.- С. 236-238.
168. Морозов, В.П. Рыночное моделирование в интересах принятия инвестиционных решений / В.П. Морозов, А. В. Никитенко // Материалы Междунар. молодежная конфер. «Математические проблемы современной теории управления системами и процессами». - Воронеж: ИПЦ «Научная книга», 2012. – С. 293 - 295.
169. Морозов, В.П. Сетевая экономика. Теория и практика: учебное пособие / В.П. Морозов, П.В. Селезнев, А.Ю. Скрипник. - Воронеж: АНО МОК ВЭПИ, 2006. – 190 с.
170. Морозов, В.П. Система виртуального распределенного терминологического поиска / Н.П. Курочка, В.П. Морозов, В.Г. Сидорова, А.И.

Сырин, А.Л. Ухин // Государственный информационный фонд неопубликованных документов ФГАНУ «Центр информационных технологий и систем органов исполнительной власти». Рег.№50201450823 от 12.12.2014 г.

171. Морозов, В.П. Система информационной поддержки оценки портфеля ценных бумаг / В.П. Морозов, А.В. Никитенко // Интернет-журнал «Науковедение». 2013 №2 (15) (15ТВН213) [Электронный ресурс]. - М. 2013. – Режим доступа: http://naukovedenie.ru/index.php?p=issue-1-13, свободный – Загл. с экрана.
172. Морозов, В.П. Система поддержки принятия инвестиционных решений интегрированного менеджмента / Д.Г. Кобзарь, В.П. Морозов, А.В. Никитенко, А.И. Сырин // Гос. инф. фонд неопубликованных документов ФГАНУ «Центр информационных технологий и систем органов исполнительной власти». Рег.№50201350235 от 13.03.2013 г.
173. Морозов, В.П. Система поддержки принятия инвестиционных решений малого предприятия / С.А. Баркалов, В.П. Морозов, А.В. Никитенко, А.И. Сырин. - Воронеж: Воронежский ГАСУ. - 2014. – 162 с.
174. Морозов, В.П. Система поддержки принятия инвестиционных решений. Постановка задачи / В.П. Морозов, А. С. Ларин, А.И. Сырин // Математическое моделирование в технике и технологии: Всероссийская конференция с элементами научной школы. – Воронеж: ВИВТ-ВГТУ, 2011. – С. 225-227.
175. Морозов, В.П. Система управления и методы интегрированного менеджмента стабильного функционирования строительной организации / С.А. Баркалов, В.П. Морозов, А. В. Никитенко, А.И. Сырин // Экономика и менеджмент систем управления. 2012, Т.6, №4 (6). – С. 10 - 18.
176. Морозов, В.П. Снижение неопределенности внешней среды на основе информационной системы поддержки принятия инвестиционных решений / С.А. Баркалов, Л.Е. Мистров, В.П. Морозов, А.И. Сырин // Экономика и менеджмент систем управления. 2016. №2.1 (20). – С. 186 - 191.

177. Морозов, В.П. Снижение неопределенности внешней среды на основе системы поддержки принятия инвестиционных решений: постановка задачи / В.П. Морозов, А.И. Сырин // Теория активных систем ТАС-2014: материалы международной научно-практической конференции 17-19 ноября 2014 г., Москва 2014 г. - С. 318-320.
178. Морозов, В.П. Совершенствование визуальной технологии моделирования искусственных нейронных сетей / В.П. Морозов, В.В. Ряполов // Развитие инноваций в России: Материалы между-нар. нучно.-практич. конф. - Воронеж: АНО МОК «ВЭПИ», 2010.- С. 37-41.
179. Морозов, В.П. Содержательные аспекты мотивационных моделей в системе интегрированного менеджмента организаций строительной индустрии / В.П. Морозов, Д. Г. Кобзарь, А.И. Сырин // Материалы Междунар. молодежная конфер. «Математические проблемы современной теории управления системами и процессами». - Воронеж: ИПЦ «Научная книга», 2012. – С. 333 - 335.
180. Морозов, В.П. Способ распознавания функциональных подсистем управления в интересах оценки радиоэлектронной обстановки / В.П. Морозов, А. Ю. Ларшин // Военная радиоэлектроника: опыт использования и проблемы подготовки специалистов: Труды VII межвузовской научно-технической конференции. – Воронеж: Издательство ВИРЭ, 2001. – С. 73 - 77.
181. Морозов, В.П. Текст как объект моделирования для информационно-аналитических систем / В.П. Морозов, Б.Г. Лютиков, А.Ю. Скрипник // Актуальные проблемы науки в студенческих исследованиях: Труды VI межвузовской студенческой научной конференции. – ВЭПИ. Воронеж, 2005. – С. 18-21.
182. Морозов, В.П. Терминологический поиск в метрологических информационных системах / Л.Е. Мистров, В.П. Морозов // Метрология. 2017, №1. – С. 33 - 41.
183. Морозов, В.П. Технология обеспечения стабильного функционирования организации / В.П. Морозов // Территория науки, 2012. – №4 . – С. 52 - 56.

184. Морозов, В.П. Технология управления проектом создания и развития информационных систем интегрированного менеджмента социально-экономических организаций / В.П. Морозов // Интернет-журнал «Науковедение». 2013 №1 (14) (06ТВН113) [Электронный ресурс]. - М. 2013. – Режим доступа: http://naukovedenie.ru/index.php?p=issue-1-13, свободный – Загл. с экрана.
185. Назаров, А.В. Нейросетевые алгоритмы прогнозирования и оптимизации систем / А.И. Лоскутов, А.В. Назаров. - СПб.: Наука и техника, 2003.-384 с.
186. Найк, Дайлип. Системы хранения данных в Windows.:Пер. с англ / Дайлип Найк. - М.: Издательский дом «Вильямс», 2005. – 432 с.81
187. Науменко, В.В. Реструктуризация крупных портфелей ценных бумаг в условиях низкой ликвидности рынка. Автореф. и дисс. к.э.н. Москва. 2012.
188. Новиков, Д.А. Математические модели организаций: Учебное пособие / А.А. Воронин, М.В. Губко, С.П. Мишин, Д.А. Новиков. - М.: ЛЕНАНД, 2008. - 360 с.
189. Новиков, А.М. Методология / А.М. Новиков, Д.А. Новиков. – М.: Синтег. 2007. – 668 с.
190. Новиков, Д.А. Теория управления организационными системами / Д.А. Новиков. М.: МПСИ, 2005. – 584 с.
191. Новосельцев, В.И. Теоретические основы системного анализа / под ред. В.И. Новосельцева. – М.: Майор, 2006. – 592 с.
192. Новосельцев, В.И. Системный анализ: современные концепции. Изд. 2-е испр. и дополн. / В.И. Новосельцев. – Воронеж: Кварта, 2003.- 280 с.
193. Ньютон, И. Об универсальном языке / И. Ньютон // Семиотика и информатика. Выпуск 35. М.: Майор, 1997. - 90 с.
194. Орлова, Е.И. Инвестиции / Е.И. Орлова. – СПб.: «Омега-Л». 2003. – 192 с.
195. Орлянская, И.В. Современные подходы к построению методов глобальной оптимизации // Электронный журнал «ИССЛЕДОВАНО В РОССИИ».– (http://zhurnal.ape.relarn.ru/articles/2002/189.pdf)
196. Оссовский, С. Нейронные сети для обработки информации / С. Ос-

совский. – М.: Финансы и статистика. 2002.– 344 с.

197. Перегудов, Ф. И. Введение в системный анализ / Ф.И. Перегудов, Ф.П. Тарасенко. – М.: Высшая школа, 1998. – 288 с.
198. Першиков, В.И. Толковый словарь по информатике / В.И. Першиков, В.М. Савинков. – М.: Финансы и статистика, 1991. – 543 с.
199. Петровский, А.Д. Теория принятия решений / А.Д. Петровский . - М.: Издательский центр «Академия». 2009.– 400 с.
200. Погодин, С.К. Методы оценки портфелей инвестиций, включающих ценные бумаги и недвижимость. Автореф. и дисс. к.э.н. Москва. 2006.
201. Попов, В. Ю. Инвестиции: математические методы: учеб. пособие / В.Ю. Попов, А.Б. Шаповал. - 2-е изд. - М.: ФОРУМ. 2008. – 144 с.
202. Прангишвили, И.В. Системный подход и общесистемные закономерности. Серия «Системы и проблемы управления» / И.В. Прангишвили. - М.: СИНТЕГ, 2000, 528 с.
203. Реализация баз данных Microsoft SQL Server 7.0: учеб.курс:Офиц.пособие Microsoft для самостоят.подготовки:Пер.с англ. -М.: Рус.ред., 2000. -483 с.
204. Редкозубов, С.А. Управление динамикой рынка: системный подход / Н.В. Аржакова, В.И. Новосельцев, С.А. Редкозубов. - Воронеж: ВГУ, 2004. – 148 с.
205. Резников, Б.А. Системный анализ и методы системотехники. Часть 1. Методология системных исследований. Моделирование сложных систем / Б.А. Резников. – М.: Воениздат, 1990. – 522 с.
206. Рекундаль, О.И. Формирование и управление инвестиционным портфелем пенсионных накоплений в деятельности негосударственных пенсионных фондов. Дисс. к.т.н. Томск. 2016.
207. Рогов, С.В. Формирование информационного обеспечения распределенных систем поддержки принятия решений. Автореф. и дисс. к.т.н. Красноярск. 2008.
208. Романов, А.Н. Советующие информационные системы в экономике: Учеб. Пособие для вузов / А.Н. Романов, Б.Е. Одинцов. – М.: Юнити-ДАНА, 2000. – 487 с.
209. Рутковская, Д. Нейронные сети, генетические алгоритмы и нечеткие

системы / М. Пилиньский, Д. Рутковская, Л. Рутковский Л. – М.: Горячая линия – Телеком, 2004. – 230 с.

210. Рыков, А.С. Системный анализ: модели и методы принятия решений и поисковой оптимизации / А.С. Рыков. – М.: Издательский Дом МИСиС, 2009. – 608 с.
211. Саркисян, А.К. Развитие механизма управления портфелем акций на российском рынке ценных бумаг. Дисс. к.э.н. Ростов-на Дону. 2014.
212. Селезнев, П.В. Разработка моделей прогнозирования для информационно-аналитической системы оценки рентабельности региональной транспортной компании. Автореф. и дисс. к.т.н. Воронеж. 2007.
213. Смирнова, Г.Н. Проектирование экономических информационных систем: учебник/ Г.Н. Смирнова, А.А. Сорокин, Ю.Ф. Тельнов; под ред. Ю.Ф. Тельнова. – М.: Финансы и статистика, 2001. – 512 с.
214. Смолин, Д. В. Введение в искусственный интеллект: конспект лекций / Д.В. Смолин. - М.: ФИЗМАТЛИТ, 2004. - 208 с.
215. Снежко, Ю.С. Технический анализ как инструмент выбора торговой стратегии на российском рынке акций. Дисс. к.э.н. Москва. 2016.
216. Советсткий энциклопедический словарь / под ред. А.М. Прохорова. - М.: «Советская энциклопедия», 1987. - 1600 с.
217. Субботин, А.В. Управление инвестиционным портфелем на основе индикаторов рыночной волатильности. Автореф. и дисс. к.э.н. Москва. 2008.
218. Сырин, А.И. Алгоритм ранговой оптимизации доступа к данным в информационной системе / А.И. Сырин // Интернет-журнал «Науковедение». 2016 №4 (8) (49TVN416) [Электронный ресурс]. - М. 2016. – Режим доступа: http://naukovedenie.ru/ index.php?p=issue-1-13, свободный – Загл. с экрана.
219. Титоренко, Г.А. Информационные системы в экономике: учебник для студентов вузов, обучающихся по специальностям «Финансы и кредит», «Бухгалтерский учет, анализ и аудит» и специальностям экономики и управления / Г.А. Титоренко. – М.: ЮНИТИ-ДАНА, 2008. – 463 с.

220. Титоренко, Г.А. Информационные технологии управления: учеб. пособие для вузов; под ред. проф. Г.А. Титоренко. – 2-е изд., доп. – М.: ЮНИТИ-ДАНА, 2005. – 439 с.
221. Тихненко, А.Н. Разработка модели анализа активов в рамках гипотезы когерентных фондовых рынков. Дисс. к.э.н. Москва. 2015.
222. Томас Коннолли. Базы данных. Проектирование, реализация и сопровождение. Теория и практика: пер. с англ. / Томас. Коннолли, Каролин, Бегг. -3-е издание. - М. : Издательский дом "Вильяме", 2003. — 1440 с.
223. Трофимов, В.В. Информационные системы и технологии в экономике и управлении.: учебник под ред. В.В. Трофимова – 3-е изд., перераб. и доп. – М.: Издательство Юрайт, 2011. – 521 с.
224. Уилкс, С. Математическая статистика / С. Уилкс. М.: Наука. Гл. ред. Физ.-мат. лит., 1967. -632 с.
225. Уинстон, П. Искусственный интеллект / Пер. с англ. / П. Уинстон. - М.: Мир, 1980.-520 с.
226. Управление инвестициями / под ред. В.В. Шеремета. - М.: «Высшая школа», 1998. Т 1, 2. – 416; 512 с.
227. Урман, С. Oracle9i. Программирование на языке PL/SQL / С. Урман. – М.: Лори, 2005. – 528 с.
228. Устинова, Г.М. Информационные системы менеджмента: Основные аналитические технологии в поддержке принятия решений / Г.М. Устинова. – СПб.: Издательство «ДиаСофтЮП», 2000. – 368 с.
229. Флейшман, Б.С. Основы системологии / Б.С. Флейшман. – М., 1982. – 214 с.
230. Хансен, Г. Базы данных: разработка и управление: пер с англ. С. Каратыгина /Г. Хансен, Д.Хансен. - М.: ЗАО «Издательство БИНОМ», 2000. - 704 с.
231. Хомяков, Д.М. Основы системного анализа / Д.М. Хомяков, П.М. Хомяков, [предисл. М.Я. Лемешева] – М.: Издательство механико-математического факультета МГУ им. М.В. Ломоносова, 1996. – 108 с.
232. Цибульникова, В.Ю. Оценка финансовых рисков в процессе взаимодействия брокера и инвестора. Дисс. к.э.н. Томск. 2016.
233. Шапошникова, А.Г. Формирование портфеля ценных бумаг на основе

комплексных индексных мер риска. Автореф. дисс. к.э.н. Уфа. 2011.

234. Шарп, У.Ф. Инвестиции. ч. 1 / У.Ф. Шарп, Г.Дж. Александер, Дж. В. Бэйли. – М.: Инфра-М, 2001. – 1028 с.
235. Шорин, В.Г. Системный анализ и структуры управления / Л.П. Стрельникова, Н.Л. Федоренко. - М.: Знание, 1975. - 304 с.
236. Янукян, М.Г. Практикум по рынку ценных бумаг / М.Г. Янукян. - СПб.: Питер, 2006. – 192 с.
237. Bellman, R.E. Adaptive Control Processes. Princeton University Press, Princeton, 1961, NJ.
238. Bertalanffy L. von. General systems theory. Foundation, development, applications, 2 ed. / L. Bertalanffy - N.Y., 1969. – 180 p.
239. Burbea J., Rao C.R. On the convexity of some divergence measures based on entropy functions, IEEE Trans. Information Theory, Vol. 28, 1982.
240. Cohen, W.W. Context-Sensitive learninq methods for text cateqorization / W.W. Cohen, V. Sinqer // ACM Transaction jn Information Systems 17, 2, 1999.-pp. 141-173.
241. Couturier A., Fioleau B. Debt Level and Company Efficiency: Independence or Implication? An Evaluation of Fuzzy Implication // European Journal of Economic and Social Systems, 14, 1 (2002). 119.
242. Dourra H., Siy P. Investment Using Technical Analysis and Fuzzy Logic // Fuzzy Sets and Systems 127 (2002). 124.
243. Edwards, J.S. Expert Systems in Management and Administration — Are they really different from Decision Support Systems? // European Journal of Operational Research, 1992. — Vol. 61. — pp. 114—121.
244. Eiben, A. E. Genetic Algorithms with multiparent re-combination / A. E. Eiben, P. E. Raue, Zs. Ruttkay // Parallel Problem Solving from Nature III. Berlin: Springer Verlag, (LNCS), v866 (1994), pp 78-87.
245. Goldberg, D.E. Genetic Algorithms in Search, Optimization, and Machine Learning / D.E. Goldberg. Addison-Wesley, 1989.
246. Herrera, F. Tackling real-coded genetic algorithms: operators and tools for the behaviour analysis / F. Herrera, M. Lozano, J.L. Verdegay // Arti-

ficial Intelligence Review. – 1998 – V. 12, № 4. – 319 p.

247. Hirschey, M., Nofsinger, M. (2008), Investments: Analysis and Behavior, Irwin/McGraw-Hill: McGraw-Hill Companies, Inc., NY.
248. http://ru.wikipedia.org/
249. Jaeger R.A. All about Hedge Funds: The Easy Way to Get Started. McGraw-Hill, 2003. P. 298.
250. Kumar, Deb. A Realcoded genetic algorithms with simulated binary crossover: Studies on multimodal and multiobjective problems. Complex Systems, 1995. – V. 9 №6. – 454 p.
251. Little I.D.C. Models and Managers: The Concept of a Decision Calculus // Management Science, 1970. — v.16. — N 8.
252. Morozov, V.P. Lodical_Linquistic Approach to Elaboration the Models of Radioelectronic Environment / В.П. Морозов, М.М. Ступницкий // Моделирование и ЭМС. 12 международный симпозиум. – Вроцлав, 1994. – С. 18-21.
253. Morozov, V.P. The concept of project management for creation and development of integrated management of the social and economic organizations system / V.P. Morozov, A. V. Nikitenko // Modern informatization problems in economics and safety: Proceedings of the XVIII-th International Open Science Conference (Lorman, MS, USA, January 2013)/ MS, USA: Science Book Publishing House, 2013. – P. 104 -108.
254. Mutual Fund Fact Book // Investment Company Institute, 42nd Edition, 2002. – P. 61. 134.
255. Power D. J. Web-based and model-driven decision support systems: concepts and issues. Americas Conference on Information Systems, Long Beach, California, 2000.
256. Quillan, M. R. Semantic memory / M. R. Quillan // In Semantic Information Processing. — M.I.T. Press, Cambridge, Mass., 1968. — P. 227-268.
257. Rosenblatt F. "On the convergence of reinforcement procedures in simple perceptrons", Cornell Aeronautical Laboratory Report, VG-1196-G-4, Buffalo, NY,1960. – 305 pp.

258. Rosenblatt F. "Perception simulation experiments", Proceedings of the Institute of Radio Engineers, 1960, vol. 48, p. 301-309.
259. Rosenblatt F. "The Perceptron: A probabilistic model for information storage and organization in the brain", Psychological Review, 1958, vol. 65, p. 386-408.
260. Rosenblatt F. Principles of Neurodynamics, Washington, DC: Spartan Books, 1962. - 203 p.
261. Steinhaus H. (1956). Sur la division des corps materiels en parties. Bull. Acad. Polon. Sci., C1. III vol IV: 801—804, 1956.
262. Timmer, J. H.. Understanding the Fed Model, Capital Structure, and Then Some. Capital Structure, and Then Some (March 4, 2012).
263. Wright, A. Genetic algorithms for real parameter optimization // Foundations of Genetic Algorithms. – 1991. – V. 1. – 218 p.

Приложение А. Копии актов внедрения, регистрации программ и сертификатов

«УТВЕРЖДАЮ»
Генеральный директор
АО НВП «ПРОТЕК»
кандидат технических наук, доцент
В. Шуваев
«16» 11 2017 г.

АКТ

об использовании в АО НВП «ПРОТЕК»положений и результатов диссертационной работы на соискание ученой степени доктора технических наук Морозова Владимира Петровича «Методы, модели и алгоритмы синтеза информационных систем поддержки портфельной инвестиционной деятельности социально-экономических организаций» при выполнении работ по теме «Создание федеральной государственной информационно-аналитической системы по управлению использованием радиочастотного спектра»

Научно-техническая комиссия в составе:председателя – главного конструктора – заместителя генерального директора кандидата технических наук профессора Исаева В.В. и членов комиссии – начальника научно-технического управления кандидата технических наук Кузьменко Ю.В., начальника отдела кандидата технических наук Кожухова Е.А.установила, что математические модели весовых распределенных информационных систем на тезаурусе и терминологическом портрете, а также алгоритм обмена данными на основе эмпирической оценки их ранговой популярности, разработанные в ходе проведения диссертационных исследований Морозовым В.П., нашли применение в АО «НВП «ПРОТЕК» при выполнении работ по теме «Создание федеральной государственной информационно-аналитической системы по управлению использованием радиочастотного спектра» в рамках государственного контракта от 21 февраля 2017 г. № 0410/37.

Председатель комиссии: В.В. Исаев

Члены комиссии: Ю.В. Кузьменко

Е.А. Кожухов

«16» 11 2017 г.

УТВЕРЖДАЮ
Генеральный директор
ООО управляющая компания «Жилпроект»
_____ к.т.н. П.В. Михин
_____21 июня_____2017 г.

А К Т

«21.06. 2017» г. Воронеж

О внедрении результатов докторской диссертационной работы Морозова Владимира Петровича «Методы, модели и алгоритмы синтеза информационных систем поддержки портфельной инвестиционной деятельности социально-экономических организаций»

В период с 12 мая 2016 г. по настоящее время в ООО УК «Жилпроект» в интересах проектной деятельности и решения инвестиционных задач организации в опытной эксплуатации находится прототип информационной системы поддержки портфельной инвестиционной деятельности, разработанный Морозовым В.П. в ходе проведения диссертационных исследований. Результаты использования показали, что, несмотря на определенную функциональную избыточность информационной системы (например, наличие программного комплекса исследования нейронных сетей), польза отдельных подсистем очевидна. Отмечается качественная работа подсистем формирования портфеля инвестора и поиска терминологической информации.

Положительным моментом является возможность автономного использования отдельных программных модулей, реализующих модели и алгоритмы наиболее востребованные для решения текущих задач.

В области инвестиционной деятельности к числу таких моделей и алгоритмов следует отнести: одноиндексную модель Шарпа; нейромодифицированную модель Шарпа; нейрокомитетную модель Шарпа; модифицированный генетический алгоритм.

В интересах поиска новой полезной информации применительно к проектной деятельности организации активно использовалась подсистема терминологического поиска. В ходе эксплуатации выявлено ее дополнительное положительное свойство, как эффективного средства обучения молодых специалистов организации.

Положительные результаты опытной эксплуатации системы отражены в отзывах членов технического совета организации.

Экономический эффект от внедрения инвестиционных моделей и алгоритмов за истекший период составил более 1,1 млн. рублей.

Начальник управления кадрами

Лобач / Лобанская С.А /

ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ
ВОРОНЕЖСКИЙ ФОНДОВЫЙ ИНТЕРНЕТ ЦЕНТР

Россия, 394030, г. Воронеж, ул. Свободы, д. 69А, офис 102
тел/факс (473) 220-57-18;, e-mail: vrnfic@alor.ru; сайт – vrn.alor.ru

УТВЕРЖДАЮ
Генеральный директор
ООО «Воронежский фондовый интернет центр»
_____ С.С. Овечкин
13 ноября 2017 г.

АКТ

«13.11. 2017 г» г. Воронеж

о внедрении научных и практических результатов диссертационной работы Морозова Владимира Петровича «Методы, модели и алгоритмы синтеза информационных систем поддержки портфельной инвестиционной деятельности социально-экономических организаций»

Настоящий акт составлен в том, что ряд научных и практических результатов, полученных в диссертационной работе Морозова Владимира Петровича, используется в работе организации, в частности:

- для повышения эффективности функционирования корпоративной инвестиционной системы в ее состав включены программные реализации моделей Марковица, Шарпа и их модификации (нейромодифицированная и нейрокомитетная). Точность расчета доходности портфеля с использованием данных моделей возросла более чем на 12%, по сравнению с используемыми;
- в состав средств фундаментального анализа ценных бумаг включен терминологический поиск, оказавшийся своевременным и полезным, по сравнению с существовавшим предоствлением новостей;
- положительный эффект показал алгоритм обмена данными на основе их ранговой популярности, внедренный в состав корпоративной базы данных.

Общий годовой эффект от проведенной модернизации составил около 1,5 млн. рублей.

ЗАО “ВОРОНЕЖ-ДОМ”

394030, г. Воронеж, ул. Плехановская, 10 а, Ген.директор 22-74-54, гл.инженер 22-74-55, отдел недвижимости 20-77-44 , факс 22-74-54 , ИНН 3662028490, БИК 042007835 , Филиал ОАО Банк ВТБ в г. Воронеже р/с 40702810525000002134, к/с 30101810100000000835.

УТВЕРЖДАЮ
Генеральный директор
ЗАО «Воронеж-Дом»
_____ Семенова О.Ю
15 ноября 2017 г.

АКТ

«15.11. 2017» г. Воронеж

реализации результатов и выводов диссертационной работы Морозова Владимира Петровича «Методы, модели и алгоритмы синтеза информационных систем поддержки портфельной инвестиционной деятельности социально-экономических организаций» представленной на соискание ученой степени доктора технических наук в ЗАО «Воронеж-Дом»

Настоящим актом констатируется тот факт, что некоторые результаты и выводы диссертационной работы Морозова В.П. внедрены в деятельность организации.

1. Программный комплекс нейросетевого прогнозирования временных рядов и, связанные с ним модели (нейромодифицированная и нейрокомитетная модели Шарпа), встроены в информационную систему финансовых инвестиций. Модульный принцип построения, обеспечил их внедрение в короткие сроки и обеспечил формирование более эффективных портфелей (показавших более высокую прибыль) в том числе с рисковыми ценными бумагами (чего ранее не проводилось), тем самым расширив инвестиционные возможности организации.

2. Идеи и механизмы организации взаимодействия хранилища и витрины данных использованы для реконфигурации корпоративного хранилища данных, что позволило ускорить обработку данных (загрузку и выгрузку) почти в 1,5 раза.

3. Инструментарий терминологического информационного поиска позволил найти множество полезной информации методического характера и обосновать ряд новых инвестиционных стратегий.

ООО "Декарт"
394016, г. Воронеж, ул.45-Стрелковой дивизии, д.223,
ИНН 3666131467, КПП 366201001
р/с 40702810905390000073 в ПАО «МИНБАНК» Г. МОСКВА
БИК 044525600, к/с 30101810300000000600
тел. (473) 227-19-19 (многоканальный)
http://www.dekart-vn.ru , e-mail: dekart-vrn@mail.ru;

А К Т

«15.11. 2017» г. Воронеж

внедрения практических результатов и теоретических выводов диссертационной работы Морозова Владимира Петровича «Методы, модели и алгоритмы синтеза информационных систем поддержки портфельной инвестиционной деятельности социально-экономических организаций»

В рамках инвестиционной деятельности в ООО «Декарт» нашли применение следующие практические результаты и теоретические выводы диссертационной работы Морозова В. П.:

- специальное математическое и программное обеспечение терминологического информационного поиска в интересах оценки существующих и определения перспективных активов для инвестиционного портфеля;
- технологические решения организации виртуального вычислительного пространства для решения исследовательских инвестиционных задач;
- оригинальный метод индексирования базы данных, позволивший увеличить скорость обработки запросов почти в 2 раза;
- программный комплекс прогнозирования временных рядов, как исследовательская база нейромодифицированных моделей Шарпа, обеспечивающих обоснование оправданного использования высокодоходных активов.

В результате внедрения данных решений, прирост доходности инвестиционного портфеля увеличился на 28%.

Генеральный директор
ООО «Декарт» _____ Е.В. Михальский

"УТВЕРЖДАЮ"
Заместитель директора по учебной и воспитательной работе
кандидат военных наук, доцент
В.Д. Еременко
«13» ноября 2017 г.

АКТ

реализации научных положений и выводов докторской диссертации Морозова Владимира Петровича «Методы, модели и алгоритмы синтеза информационных систем поддержки портфельной инвестиционной деятельности социально-экономических организаций»

Комиссия в составе:

председателя: заведующего кафедрой правовой информатики, информационного права и естественнонаучных дисциплин, кандидата технических наук, доцента Мишина А.В.;

членов комиссии: заведующая кафедрой гуманитарных и социально-экономических дисциплин кандидата экономических наук, доцента Кочетовой Л.В.;
доцента кафедры правовой информатики, информационного права и естественно-научных дисциплин, кандидата технических наук, доцента Голикова В.К.

установила, что научные положения и вывода в части содержания методологии внешнесистемного и внутрисистемного облика информационной системы поддержки портфельной инвестиционной деятельности социально-экономических организаций и организации проектного управления использованы в Центральном филиале Российского государственного университета правосудия при разработке учебно-методических материалов для студентов по дисциплинам «Информационные технологии в юридической деятельности», «Компьютерные технологии в экспертной деятельности» и «Экономика».

Председатель комиссии:
к.т.н., доцент А.В. Мишин

Члены комиссии:
к.э.н., доцент Л.В. Кочетова
к.т.н., доцент В.К. Голиков

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования

"ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ АРХИТЕКТУРНО - СТРОИТЕЛЬНЫЙ УНИВЕРСИТЕТ"

394006 Россия, г. Воронеж, ул. XX-летия Октября, 84 Тел. (473) 271-52-68 Факс (473) 271-59-05 E-mail: rector@vgasu.vrn.ru
http://www.vgasu.vrn.ru

"18" 05 2016 г. №002-14

"УТВЕРЖДАЮ"
Проректор по научной работе
Федерального государственного бюджетного
образовательного учреждения
высшего профессионального образования
«Воронежский государственный
архитектурно-строительный университет»

Рудаков О. Б.

"18" 05 2016 г.

АКТ
внедрения документальной информационно-поисковой системы
«Научные труды» в деятельность кафедры «Управление строительством»

Комиссия в составе:
председателя: заведующего кафедрой «Управление строительством»,
д.т.н., проф. Баркалова Сергея Алексеевича

членов комиссии: профессора кафедры «Управление строительством»,
д.т.н., проф. Курочки Павла Николаевича;

профессора кафедры «Управление строительством»,
д.т.н., доц. Половинкиной Аллы Ивановны

установила, что документальная информационно-поисковая система «Научные труды», разработанная профессорами, д.т.н. Баркаловым Сергеем Алексеевичем и к.т.н. Морозовым Владимиром Петровичем, а также аспирантами Курочкой Николаем Павловичем и Сыриным Александром Ивановичем, используется в интересах ведения базы данных научных трудов профессорско-преподавательского состава кафедры и формирования отчетных материалов об их публикациях.

Реализованный в данной системе терминологический поиск показал высокую достоверность и оперативность выборки требуемой информации в соответствии с введенным термином, как на локальном компьютере, так и в глобальной сети Интернет.

Председатель комиссии: С.А. Баркалов

Члены комиссии: П.Н. Курочка
А.И. Половинкина

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования

“ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ АРХИТЕКТУРНО - СТРОИТЕЛЬНЫЙ УНИВЕРСИТЕТ”

394006 Россия, г. Воронеж, ул. XX-летия Октября, 84 Тел. (473) 271-52-68 Факс (473) 271-59-05 E-mail: rector@vgasu.vrn.ru
http://www.vgasu.vrn.ru

“18” 05 2016 г. № 003-14

"УТВЕРЖДАЮ"
Проректор по научной работе
Федерального государственного бюджетного
образовательного учреждения
высшего профессионального образования
«Воронежский государственный
архитектурно-строительный университет»

Рудаков О. Б.

"18" 05 2016 г.

АКТ

внедрения программного комплекса «Информационная система распределения финансовых инвестиций малого предприятия» в учебный процесс
кафедры «Управление строительством»

Комиссия в составе:
председателя: заведующего кафедрой «Управление строительством»,
д.т.н., проф. Баркалова Сергея Алексеевича

членов комиссии: профессора кафедры «Управление строительством»,
д.т.н., проф. Курочки Павла Николаевича;

профессора кафедры «Управление строительством»,
д.т.н., доц. Половинкиной Аллы Ивановны

установила, что программный комплекс «Информационная система распределения финансовых инвестиций малого предприятия», разработанная профессором, к.т.н., доцентом Морозовым Владимиром Петровичем и аспирантами Косенко Дмитрием Олеговичем, Курочкой Николаем Павловичем, Сыриным Александром Ивановичем и Чекомазовым Александром Николаевичем, используется при проведении практических занятий в рамках дисциплины «Инвестиционный менеджмент».

Разработанный программный комплекс обеспечивает автономный ввод исходных данных по используемым финансовым инструментам, комплексную настройку генетического алгоритма оценки и островной модели параллельных вычислений и позволяет находить оптимальное распределение заданного объема финансовых инвестиций по исследуемым финансовым инструментам с погрешностью, не превышающей 1,5%.

Председатель комиссии: С.А. Баркалов

Члены комиссии: П.Н. Курочка
А.И. Половинкина

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования

“ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ АРХИТЕКТУРНО - СТРОИТЕЛЬНЫЙ УНИВЕРСИТЕТ”

394006 Россия, г. Воронеж, ул. XX-летия Октября, 84 Тел. (473) 271-52-68 Факс (473) 271-59-05 E-mail: rector@vgasu.vrn.ru
http://www.vgasu.vrn.ru

“ 18 ” 05 2016 г. № 001-14

"УТВЕРЖДАЮ"
Проректор по научной работе
Федерального государственного бюджетного
образовательного учреждения
высшего профессионального образования
«Воронежский государственный
архитектурно-строительный университет»

Рудаков О. Б.

"18" 05 2016 г.

АКТ

внедрения системы виртуального распределенного терминологического поиска в учебный процесс и деятельность кафедры «Управление строительством»

Комиссия в составе:
председателя: заведующего кафедрой «Управление строительством», д.т.н., проф. Баркалова Сергея Алексеевича

членов комиссии: профессора кафедры «Управление строительством», д.т.н., проф. Курочки Павла Николаевича;

профессора кафедры «Управление строительством», д.т.н., доц. Половинкиной Аллы Ивановны

установила, что система виртуального распределенного терминологического поиска, разработанная профессором, к.т.н., доцентом Морозовым Владимиром Петровичем и аспирантами Курочкой Николаем Павловичем, Сыриным Александром Ивановичем, Сидоровой Викторией Григорьевной, Ухиным Александром Леонидовичем используется при проведении практических занятий в рамках дисциплины «Интеллектуальные технологии и представление знаний» и в документальной информационно-поисковой системе «Научные труды».

Разработанная система позволяет осуществлять распределенный поиск информации о терминах, вводимых пользователем и извлечение, связанных с ними, новых дополнительных сведений как на локальной ЭВМ и в локальной вычислительной сети, так и в виртуальном пространстве. Экспериментальная апробация фрейм - семантической модели обработки текста и системы в целом, показала высокие релевантность найденных дополнительных сведений, связанных с вводимыми терминами и оперативность их получения.

Председатель комиссии: С.А. Баркалов

Члены комиссии: П.Н. Курочка
А.И. Половинкина

Федеральное агенство по образованию
Государственное образовательное учреждение
высшего профессионального образования
«Воронежский государственный технический
университет»
Воронежский областной центр
новых информационных технологий
CNIT
394026, Воронеж,
Московский проспект, 14,
ОЦ НИТ
тел. (4732) 71-85-49
факс (4732) 77-92-12
E-mail – yurasov@vorstu.ru

31.01.07 № 0.114.1р-009

На № _____

Настоящим удостоверяется, что программное средство «Автоматизированное прогнозирование объемов авиаперевозок воздушно-транспортной системы региона с учетом сезонных волн» (**авторы** – Лютиков Б.Г., Морозов В.П., Морозов Д.В., Селезнев П.В.), разработанное в Воронежском отделении АСТ, зарегистрировано в Государственном фонде алгоритмов и программ (г. Москва), регистрационный номер – **50200700047 от 12.01.07.**

Директор
Воронежского областного
центра новых информационных
технологий при ГОУВПО «ВГТУ» _____ В.Г. Юрасов

Министерство образования и науки РФ
Федеральное государственное бюджетное
образовательное учреждение
высшего профессионального образования
«Воронежский государственный технический
университет»
Воронежский областной центр
новых информационных технологий
CNIT
394026, Воронеж,
Московский проспект, 14,
ОЦ НИТ
тел. (473) 271-85-49
факс (473) 277-92-12
E-mail – yurasov@vorstu.ru
27.03.2013 № 0114/0-017
На № _____

Настоящим удостоверяется, что **«Система поддержки принятия инвестиционных решений интегрированного менеджмента»**, разработанная в ФГБОУ ВПО «Воронежский государственный архитектурно-строительный университет» (**авторы** – Морозов В.П., Кобзарь Д.Г., Никитенко А.В., Сырин А.И., **научный руководитель** – профессор Морозов В.П.), зарегистрирована в государственном информационном фонде неопубликованных документов ФГАНУ «Центр информационных технологий и систем органов исполнительной власти» (**№ 50201350235 от 13.03.2013**).

Директор ОЦ НИТ ФГБОУ ВПО «ВГТУ» В.Г. Юрасов

Министерство образования и науки РФ
Федеральное государственное бюджетное
образовательное учреждение
высшего профессионального образования
«Воронежский государственный технический
университет»
Воронежский областной центр
новых информационных технологий
CNIT
394026, Воронеж,
Московский проспект, 14,
ОЦ НИТ
тел. (473) 271-85-49
факс (473) 277-92-12
E-mail – yurasov@vorstu.ru
27.03.2013 № 0114/0-014
На № _____

Настоящим удостоверяется, что **«Программный комплекс нейросетевого прогнозирования временных рядов интегрированного менеджмента»**, разработанный в ФГБОУ ВПО «Воронежский государственный архитектурно-строительный университет» (**авторы** – Морозов В.П., Кобзарь Д.Г., Никитенко А.В., Сырин А.И., **научный руководитель** – профессор Морозов В.П.), зарегистрирован в государственном информационном фонде неопубликованных документов ФГАНУ «Центр информационных технологий и систем органов исполнительной власти» (**№ 50201350236 от 13.03.2013**).

Директор ОЦ НИТ ФГБОУ ВПО «ВГТУ» В.Г. Юрасов

Аккредитованная Испытательная лаборатория программных средств ВГТУ * Воронежский ОЦ НИТ

Министерство образования и науки РФ
Федеральное государственное бюджетное
образовательное учреждение
высшего профессионального образования
«Воронежский государственный технический
университет»
Воронежский областной центр
новых информационных технологий
CNIT
394026, Воронеж,
Московский проспект, 14,
ОЦ НИТ
тел. (4732) 271-85-49
факс (4732) 277-92-12
E-mail – yurasov@vorstu.ru

18.12.2014 № 0114/0-465

На № _____

Настоящим удостоверяется, что программное средство **«Информационная система распределения финансовых инвестиций малого предприятия»** (**авторы** – Курочка Н.П., Косенко Д.О., Морозов В.П., Сырин А.И., Чекомазов А.Н., **научный руководитель** – профессор Морозов В.П.), разработанное в ФГБОУ ВПО «Воронежский государственный архитектурно-строительный университет», зарегистрировано в государственном информационном фонде неопубликованных документов ФГАНУ «Центр информационных технологий и систем органов исполнительной власти» (**№ 50201450825 от 12.12.2014**).

Директор
Воронежского областного
центра новых информационных
технологий ФГБОУ ВПО «ВГТУ» В.Г. Юрасов

Министерство образования и науки РФ
Федеральное государственное бюджетное
образовательное учреждение
высшего профессионального образования
«Воронежский государственный технический
университет»
Воронежский областной центр
новых информационных технологий
CNIT
394026, Воронеж,
Московский проспект, 14,
ОЦ НИТ
тел. (4732) 271-85-49
факс (4732) 277-92-12
E-mail – yurasov@vorstu.ru
18.12.2014 № 0119/0-456
На № _____

Настоящим удостоверяется, что программное средство **«Система виртуального распределенного терминологического поиска»** (**авторы** – Курочка Н.П., Морозов В.П., Сидорова В.Г., Сырин А.И., Ухин А.Л., **научный руководитель** – профессор Морозов В.П.), разработанное в ФГБОУ ВПО «Воронежский государственный архитектурно-строительный университет», зарегистрировано в государственном информационном фонде неопубликованных документов ФГАНУ «Центр информационных технологий и систем органов исполнительной власти» (**№ 50201450823 от 12.12.2014**).

Директор
Воронежского областного
центра новых информационных
технологий ФГБОУ ВПО «ВГТУ» В.Г. Юрасов

Аккредитованная испытательная лаборатория программных средств ВГТУ * Воронежский ОЦ НИТ

Министерство образования и науки РФ
Федеральное государственное бюджетное
образовательное учреждение
высшего профессионального образования
«Воронежский государственный технический
университет»
Воронежский областной центр
новых информационных технологий
CNIT
394026, Воронеж,
Московский проспект, 14,
ОЦ НИТ
тел. (4732) 271-85-49
факс (4732) 277-92-12
E-mail – yurasov@vorstu.ru

18.12.2014 № 0114/0-460

На № _____

Настоящим удостоверяется, что программное средство **«Документальная информационно-поисковая система «Научные труды»»** (**авторы** – Баркалов С.А., Курочка Н.П., Морозов В.П., Сырин А.И., **научный руководитель** – профессор Морозов В.П.), разработанное в ФГБОУ ВПО «Воронежский государственный архитектурно-строительный университет», зарегистрировано в государственном информационном фонде неопубликованных документов ФГАНУ «Центр информационных технологий и систем органов исполнительной власти» (**№ 50201450824 от 12.12.2014**).

Директор
Воронежского областного
центра новых информационных
технологий ФГБОУ ВПО «ВГТУ» _____ В.Г. Юрасов

СИБИРСКИЙ ГОСУДАРСТВЕННЫЙ
ИНДУСТРИАЛЬНЫЙ
УНИВЕРСИТЕТ

СЕРТИФИКАТ

выдан

Морозову В.П., Никитенко А.В.

Воронежский государственный архитектурно-строительный университет, г. Воронеж

за участие в IV Всероссийской научно-практической конференции (с международным участием) «Моделирование и наукоемкие информационные технологии технических и социально-экономических системах, 2016»

с докладом

«Применение моделей фундаментального анализа при управлении финансовыми инвестициями малого предприятия»»

12-15 апреля 2016г.

Ректор, профессор **Е.В. Протопопов**

выдан

АПНИ
АГЕНТСТВО ПЕРСПЕКТИВНЫХ НАУЧНЫХ ИССЛЕДОВАНИЙ
СЕРТИФИКАТ
участника XXII Международной
научно-практической конференции
«СОВРЕМЕННЫЕ ТЕНДЕНЦИИ
РАЗВИТИЯ НАУКИ И ТЕХНОЛОГИЙ»
31 января 2017 года, г. Белгород
настоящий сертификат подтверждает участие в конференции
Морозова Владимира Петровича
с докладом
МОДУЛЬ ОЦЕНКИ ОБЛИГАЦИЙ КАК
ИНСТРУМЕНТ ИНФОРМАЦИОННОЙ
СИСТЕМЫ ПОДДЕРЖКИ ПОРТФЕЛЬНОЙ
ОПТИМИЗАЦИИ КОМПАНИИ
Руководитель Агентства
перспективных научных
исследований (АПНИ)
Ткачева Е.П.

Приложение Б. Результаты сравнительной оценки основных моделей портфельного инвестиционного анализа

Название	Краткое содержание	Достоинства	Недостатки
Модель Г. Марковица [14, 234]	Доходность используемого финансового инструмента в инвестиционном портфеле, а, следовательно, его общая доходность является случайной величиной, которая изменяется в некотором диапазоне. Определение вероятности доходности каждого актива, позволит получить распределение вероятностей доходности альтернативных средств вложения. В математическом плане, данная модель относится к классу задач квадратической оптимизации при линейных ограничениях.	1. Предложен вероятностный подход к формализации понятий доходности и риска. 2. Предложена математическая основа проведения рыночных операций.	1. Сложность реализации. 2. Формирует множество эффективных портфелей. 3. Требует значительного объема информации. 4. Приемлема для стабильного состояния фондового рынка
Модель У.Шарпа [21, 234]	Описывает взаимосвязь между доходностью выбранного актива и ожидаемой доходностью рынка в виде линейной зависимости, имеющей вид: $E(r_i) = y_i + \beta_i E(r_m) + e_i$ где $E(r_i)$ - прогнозируемая доходность актива; y_i - доходность актива без воздействия рынка; β_i - коэффициент β актива; $E(r_m)$ - прогнозируемый доход финансового рыночного портфеля; e_i - случайная ошибка. Для сравнения эффективности различных финансовых инструментов выбирается один определенный рыночный индекс. Такая модель является одноиндексной. Параметр β описывает ковариацию доходности актива с доходностью рыночного индекса. На основе однофакторной модели разработан эмпирический метод выбора оптимального портфеля. Он сводит задачу оптимизации второго порядка к линейной.	1. Простота. 2. Делает возможным разделение риска актива на диверсифицируемую и недиверсифицируемую части. 3. Не требует значительных объемов исходных данных для проведения анализа.	1. Выдвигается условие существования эффективного рынка.
Модель Д. Тобина [26]	В основу модели положен макроэкономический подход. В состав портфеля включаются безрисковые активы (например, государственные облигации). Совокупный капитал разделен на две части наличную (денежную) и неналичную (например, ценные бумаги). Суть проводимого анализа заключается в оценке факторов, влияющих на предпочтение инвесторов иметь активы, а не содер-	1. Гибкость. 2. Практическая направленность. 3. Дан анализ адекватности количественных характеристик активов и портфелей.	1. Сложность. 2. Высокие вычислительные ресурсы.

	жать свой капитал в единственной форме (например, налично-денежной).		
CAPM (модель оценки капитальных активов) [30, 234]	Определяет взаимосвязь между портфельным риском или риском единичного актива и ожидаемой доходностью равновесного РЦБ. Полный риск портфеля (актива) разделяется на рыночный, и не рыночный. Важное значение при оптимизации ИП имеет значение недиверсифицируемого риска. Данный вид риска представляет собой зависимость доходности актива от доходности РЦБ. В процессе инвестиционной деятельности целесообразно учитывать недиверсифицируемый (систематический) риск. Он описывается параметрами «альфа» и «бета». Параметр «бета» соответствует своему аналогу β в однофакторной модели Шарпа. Данный параметр устанавливает взаимосвязь между доходностями портфеля (актива) и РЦБ. Параметр «альфа» позволяет оценить величину недооцененности актива рынком, в сравнении с уровнем его доходности когда он находится в состоянии равновесия. Если его величина положительна, то имеет место недооценка, если отрицательна, он переоценен.	1. Реализовано точное аналитическое представление правила "большая доходность - большой риск". 2. Уточняет учитываемый риск.	1. Предполагает наличие эффективного рынка. 2. Требует значительный объем информации. 3. Не допускает эмпирической проверки.
Модель Блека-Шоулса [38, 234]	Оценивает возможность проведения безрисковой сделки в случае совместного применения акции и оформленного на нее опциона. Цена подобной сделки должна соответствовать стоимости безрисковых активов на рынке. Если учесть, что цена акции имеет временное изменение, то и стоимость оформленного опциона, который обеспечивает сделку с нулевым риском, также должна изменяться аналогично. На основе вышеизложенных положений, может быть сформирована стохастическая оценка стоимости опциона.	1. Первая динамическая вероятностная модель, основанная на теории случайных процессов. 2. Практическая реализуемость.	1.Сложность. 2. Низкое быстродействие.

Приложение В. Характеристика современного методического обеспечения портфельной инвестиционной деятельности социально-экономической организации

Название	Содержание	Достоинства	Недостатки
Инвестиционные стратегии управления ИП на основе модифицированного индикатора волатильности РЦБ [217]	Предложена новая шкала волатильности РЦБ, на основе применения вейвлетной трансформации. Введены новые способы определения волатильности и корреляции активов РЦБ на различных временных горизонтах. Предложен класс инвестиционных стратегий управления ИП, учитывающий ограничения волатильности. Предложен алгоритм расчета индикаторов в соответствии со шкалой волатильности РЦБ.	1. Возможность учета свойств доходностей, проявляющихся при их временном агрегировании (например, изменение форм распределений вероятности активов на различных временных интервалах, в пределах которых определяются доходности активов, и взаимосвязанные с ними мультифрактальные характеритсики, соответствующих финансовых временных рядов). 2. Возможность получения актуальной информация о параметрах активов, на разных временных горизонтах, полезной для их управления.	1. Сложность реализации. 2. Вероятностный характер получаемых результатов. 3. Требуется значительный объем информации. 4. Решают частные задачи повышения эффективности диверсификации ИП на различных временных интервалах.
Методика формирования ИП с учетом	Разработана оптимизационная модель формирования ИП, учитывающая особенности РЦБ РФ. Предложена уникаль-	1. Простота. 2. Реализует этапы	1. Предполагает наличие эффективного рынка.

375

особенностей РЦБ РФ [51]	ная целевая функция, которая учитывает волатильность активов РЦБ и корреляцию их доходности. .	исследования, моделирования и прогнозирования в цикле формирования ИП.	2. Отсутствие реализации этапа корректировки в цикле формирования ИП.
Алгоритм формирования оптимального ИП на основе комплексных мер риска [66]	Предложен алгоритм оптимизации структуры ИП ЦБ, основанный на методе стохастического моделирования и модификации метода Хука—Дживса, учитывающий квантильные меры и меры рассеяния при оценке риска, позволяющий учитывать различные параметры распределения доходности ЦБ.	1. Учет различных мер риска. 2. Учет различных параметров распределения доходности ЦБ.	1. Сложность реализации. 2. Вероятностный характер получаемых результатов. 3. Значительный объем требуемой информации.
Динамическая модель управления ИП [29]	Разработана динамическая модель управления ИП, обеспечивающая синтез оптимальных управляющих стратегий, реализующих максимальную гладкость кривой роста капитала, заданной ЛПР на всех горизонтах инвестирования при наличии ограничений на инвестируемый капитал, с учетом различных моделей динамики цен активов (Блэка-Шоулса, стохастической волатильности, финансового рынка с варьирующимися режимами).	1. Впервые предложены сетевые динамические модели управления ИП большой размерности. 2. Учет скачкообразных изменений РЦБ.	1. Сложность реализации. 2. Значительный объем требуемой информации. 3. Большая погрешность получаемых результатов.
Методика управления реструктуризацией ИП [187]	Предложена методика реструктуризации ИП, основанная на его потенциальной ликвидационной стоимости, используемой в качестве меры эффективности. Разработаны стратегии ликвидации ЦБ для различных временных интервалов, основанные на их длинах.	1. Простота. 2. Реализация этапа пересмотра (корректировки) ИП в общем цикле управления.	1. Невысокая точность получаемых результатов. 2. Сложность получения информации о ликвидационной стоимости ЦБ. 3. Реализация частного этапа в общем цикле управления ИП.
Методический подход инвестирования на базе алгоритмической торговли [64]	Предложен методический подход инвестирования, основанный на алгоритмической торговле, содержащий: а) прогнозные математические модели динамики актива EUR/USD для различных тайм-фреймов; б) стратегии инвестирования для механических торговых систем; в) метод ситуационного риск-контроля, основанный на показателе –	1. Высокая оперативность получения результата. 2. Исключение человеческого фактора (эмоционально-	1. Сложность реализации. 2. Проблемы с настройкой. 2. Постоянный мониторинг функционирования и настройки.

	«коэффициенте покрытия убытка текущей операции».	го) в процессе инвестиционного процесса.	
Модель прогнозирования распределений доходностей ЦБ и формирования ИП на основе многомерного *t*-распределения с вектором степеней свободы [6]	Предложено при формировании ИП использовать многомерное *t*-распределение с вектором степеней свободы. Выведены формула смешанного момента общего вида с условиями существования, формулы одномерных маргинальных функций плотности и характеристических функций, получена методика построения произвольной многомерной GARCH модели, предложен алгоритм симулирования, асимметричной модификации и построена копула.	1. Реализовано прогнозирование совместного распределения доходностей активов. 2. Предложен новый класс функций: многомерное *t*-распределение с вектором степеней свободы.	1. Сложность реализации. 2. Значительный объем требуемой информации. 3. Невысокая точность получаемых результатов.
Экономико-математическая методика формирования ИП на основе комплексных индексных мер риска [233]	Предложены комплексные индексные меры финансового риска, основанные на высших моментах распределения доходности ЦБ. Разработана методика формирования ИП на основе учета комплексных индексных мер финансового риска.	1. Уточнена мера риска.	1. Сложность реализации. 2. Значительный объем требуемой информации. 3. Невысокая точность получаемых результатов.
Модель формирования структуры ИП с нечетко-множественными входными параметрами, учитывающая различные меры риска. [28]	Разработана структурная модель формирования ИП, в качестве меры риска в которой, предложено использовать коэффициент эксцесса доходности ЦБ. В качестве способа оценки доходности ЦБ предложены методы распознавания паттернов. Для решения оптимизационных задач формирования ИП предложен аппарат генетических алгоритмов.	1. Расширены возможности оптимизации ИП. 2. Разработаны альтернативные меры риска.	1. Сложность реализации. 2. Значительный объем требуемой информации. 3. Наличие экспертных оценок.
Подход к формированию агрегированных ИП [200]	Предложен подход к формированию агрегированных ИП, содержащих наряду с ЦБ, стоимостные значения объектов недвижимости. Разработана стоимостная модель оценки капитала при инвестировании в объекты недвижимости на рынке РФ.	1. Развитие инструментов хэджирования ИП.	1. Реализация частного этапа в общем цикле управления ИП.
Долгосрочная страте-	Разработана долгосрочная стратегия развития РЦБ РФ, ос-	1. Впервые система-	

377

гия развития РЦБ РФ [81]	нованная на теоретическом исследовании фундаментальных факторов, воздействующих на конкурентоспособность фондовых рынков, объективном анализе результатов формирования отрасли ЦБ РФ в 90-е годы, оценке мирового опыта развития РЦБ и исследовании государственных макроэкономических стратегий в данной области.	тизированы результаты исследования и прогнозирования РЦБ РФ.	
Алгоритм заключения опционного контракта для ИП [37]	Предложено аналитическое решение задачи определения стоимостей стандартных и специфических Европейских опционов для ИП, обеспечивающей хеджи капиталов с выплатой процентов по активу с высоким риском при квантильном хеджировании.	1. Расширение возможностей формирования ИП. 2. Возможность получения высоких дивидендов.	1. Сложность реализации. 2. Значительные уровни риска.
Методики и модели формирования и управления ИП пенсионных накоплений [206]	Разработаны: - методика оценки инвестиционной привлекательности облигаций муниципальных и субфедеральных заемщиков, основанная на анализе основных факторов риска для облигаций; - интегрированная модель формирования ИП пенсионных накоплений, основанная на портфельной теории Марковица и безрисковых ЦБ; - динамическая модель управления ИП пенсионных накоплений, учитывающая рисковые и безрисковые активы.	1. Учет рисковых и безрисковых активов. 2. Практическая направленность работы.	1. Недостатки присущие модели Марковица. 2. Низкая точность получаемых результатов.
Теоретический подход аналитической оценки справедливой стоимости сложных финансовых инструментов [45]	Предложен теоретический подход аналитической оценки справедливой стоимости сложных финансовых инструментов, базирующийся на двумерных негауссовских асимметричных совместных распределениях вероятностей цен базовых активов, учитывающий негауссовский характер ценовых процессов и не предполагающий дробление структурированного деривативa на составные части.	1. Расширение возможностей формирования ИП. 2. Возможность получения высоких дивидендов.	1. Сложность реализации. 2. Значительный объем требуемой информации. 3. Вероятностный характер получаемых результатов 4. Невысокая точность получаемых результатов
Новые подходы к принятию инвестиционных решений [2]	Обоснованы концептуальные подходы, обеспечивающие принятие инвестиционных решений, более адекватных государственным особенностям движения финансовых потоков и динамики цен активов на финансовых рынках, осно-	1. Совершенствование методологических основ формирования ИП.	1. Теоретический характер работы.

378

	ванные на представлении ПИР как системного процесса с элементами фундаментального и технического анализа, поведенческих финансов и факторов ликвидности.		
Моментум стратегия управления ИП [81]	Обоснована моментум стратегия управления ИП, основанная на портфельном эффекте, проявляемом на моментуме и развороте, представляющим собой такую ценовую аномалию на активах одного класса (например, на акциях), когда ИП, реализованные с использованием прошлых результатов инвестирования на определенном временном интервале (краткосрочном, среднесрочном и долгосрочном) демонстрируют статистически значимую прибыль.	1. Совершенствование методического обеспечения формирования ИП.	1. Сложность реализации. 2. Вероятностный характер получаемых результатов.
Аналитический способ оценки финансовых рисков брокера и инвестора на РЦБ [232]	Предложен аналитический способ оценки финансовых рисков брокера и инвестора, основанный на нечетких множествах, позволяющий оценить влияние множества основных рисков на результат деятельности брокера, отличающийся комплектом учитываемых рискообразующих факторов, критериями их оценки на основе логических правил.	1. Совершенствование методического обеспечения формирования ИП.	1. Сложность реализации. 2. Вероятностный характер получаемых результатов. 3. Невысокая точность получаемых результатов.
Модели прогнозирования доходности ЦБ на финансовых рынках на основе ИНС [32]	Разработаны и построены оригинальные модели прогнозирования доходности ЦБ на финансовых рынках на основе: а) самоорганизующейся ИНС Кохонена – Хакена, содержащей конкурентные механизмы обработки входных данных, использующие методы независимых и главных компонент в совокупности с применением относительной энтропии Кульбака – Ляйблера; б) роевой модели, реализующей предварительную обработку эмпирических данных на основе кластеризации с использованием расстояния Дженсона-Шеннона и обучение на основе комитета ИНС.	1. Совершенствование методического обеспечения прогнозирования характеристик ЦБ. 2. Применение ИНС в ПИД.	1. Сложность реализации. 2. Не ясный механизм функционирования ИНС.
Методика управления ИП акций, сочетающая методы ФА и ТА [211]	Методика управления ИП акций, сочетающая методы ФА и ТА, обеспечивающая: а) обоснование состава ИП, учитывающего инвестиционную полезность акций; б) формирование структуры ИП, учитывающей уровень ликвидности акций; в) управление ИП, использующее сигналы индикатора MACD.	1. Совершенствование методического обеспечения формирования и управления ИП. 2. Рациональное комбинирование	1. Трудности получения экспертных оценок. 2. Реализация нескольких частных этапов в общем цикле управления ИП.

		методов ФА и ТА при формировании и управлении ИП.	
Стратегия формирования ИП долгосрочных ЦБ [215]	Обоснована торговая стратегия формирования ИП долгосрочных ЦБ, основанная на индикаторах ТА («баланс оборота», «относительная сила», «полоса Боллинджера», «стохастический осциллятор»).	1. Совершенствование методического обеспечения формирования ИП. 2. Доказательство успешного использования алгоритмических торговых систем при принятии инвестиционных решений.	1. Сложность реализации. 2. Реализация частного этапа в общем цикле управления ИП.
Метод сверхкраткосрочного прогнозирования цен рыночных активов [25]	Разработан метод сверхкраткосрочного прогнозирования цен активов на РЦБ, обеспечивающего доходность ИП на основе реализации высокочастотных операций.	1. Совершенствование методического обеспечения формирования ИП. 2. Доказательство успешного использования высокочастотных операций при принятии инвестиционных решений.	1. Сложность реализации. 2. Реализация частного этапа в общем цикле управления ИП.
Синтетическая модель анализа финансовых активов [221]	Разработана синтетическая модель анализа финансовых активов, фиксирующая временные периоды эффективной работы функционирующей торговой системы на основе синтетического индикатора Хольта и нелинейной статистической модели Веге.	1. Совершенствование математического и программного обеспечения ПИД СЭО. 2. Исключение из анализа экспертного подхода.	1. Сложность реализации. 2. Реализация частного этапа в общем цикле управления ИП. 3. Допущение когерентности фондовых рынков.

Приложение Г. Характеристика частных информационных систем, программных решений и систем поддержки принятия решений в портфельной инвестиционной деятельности социально-экономической организации

Название	Содержание	Достоинства	Недостатки
Система поддержки принятия решений при управлении ИП [66]	Разработана система поддержки принятия решений при управлении ИП, рассчитывающая различные меры риска, обеспечивающая оптимизацию ИП с учетом полученных значений мер.	1. Возможность формирования и управления ИП. 2. Возможность расчета различных мер риска.	1. Расчет ограниченного числа мер риска. 2. Реализация нескольких частных этапов в общем цикле формирования ИП. 3. Закрытость системы.
Программное обеспечение поддержки принятия инвестиционных решений при формировании ИП ЦБ [233]	Разработано программное обеспечение поддержки принятия инвестиционных решений при формировании ИП ЦБ с учетом обоснованных в работе мер риска.	1. Расширение диапазона определяемых мер риска. 2. Возможность формирования ИП.	1. Расчет ограниченного числа мер риска. 2. Реализация одного частного этапа в общем цикле формирования ИП. 3. Закрытость системы.
Система поддержки принятия решений формирования эффективной структуры ИП [28]	Разработана система поддержки принятия решений по формированию эффективной структуры ИП на основе применения различных подходов к измерению риска и доходности, а также с нечетко-множественными входными параметрами.	1. Возможность формирования эффективной структуры ИП. 2. Учет нечетко-множественных входных параметров активов.	1. Реализация одного частного этапа в общем цикле формирования ИП. 2. Закрытость системы.
Система поддержки принятия решения инвестора при управлении ИП пенсионных накоплений [206]	Предложена система поддержки принятия решения инвестора при управлении ИП пенсионных накоплений, обеспечивающая оптимальные оперативные воздействия на сформированный ИП.	1.Возможность формирования и управления ИП.	1. Реализация нескольких частных этапов в общем цикле формирования ИП. 2. Закрытость системы. 3. Большая погрешность, обусловленная несовершенством используемой модели Марковица.
Система поддержки принятия решений в процессе оценки и	Разработана СППР управления и оценки рисками в финансовой деятельности брокерской компании, обеспечивающая формирование оптимального хеджирующего ИП на	1. Расширение возможностей формирования и управле-	1. Реализация нескольких частных этапов в общем цикле формирования ИП.

381

управления финансовыми рисками брокерской компании [232]	основе данных о торговых сделках, позволяющая оценить структуру ИП и определить оптимальную стратегию инвестиционного управления.	ния ИП. 2. Возможность учета дополнительных рисков брокерской деятельности.	2. Закрытость системы.
Программный комплекс обеспечения проведения торгов на фондовом рынке [4]	Создан программный комплекс, совершенствования системы поддержки принятия решений в рамках регулирования процессов проведения торгов на фондовом рынке, основанный на механизме двойного непрерывного аукциона, позволяющий оценивать влияние регулирования на взаимодействие участников в процессе ценообразования и реализовать процедуру валидации построенных моделей	1. Расширение возможностей СППР в ПИД СЭО. 2. Реализация валидации разработанных моделей.	1. Реализация одного частного этапа в общем цикле формирования ИП. 2. Закрытость системы.
Система поддержки принятия инвестором управленческих решений при работе на фондовом рынке[221]	Разработана система поддержки принятия инвестором управленческих решений при работе на фондовом рынке, реализующая формализованный механизм синхронной коррекции возникающих торговых и модельных сигналов, позволяющей снизить риск инвестора.	1. Расширение функциональных возможностей СППР.	1. Сложность системы. 2. Проблематичность совмещения СППР и торговой системы. 3. Закрытость системы.

Приложение Д. Характеристика зарубежных и отечественных информационных систем и программ для работы на рынке ценных бумаг

Наименование	Назначение и возможности	Достоинства	Недостатки
OmegaResearchProSuite 2000i	Назначение: инструментальное средство ТА для организации торговых систем. Разработчик: компания OmegaResearch (США). Состав (набор программных модулей): TradeStation 2000i, OptionStation 2000i, RadarScreen 2000i, GlobalServer, PowerEditor. Возможности: 1) обработка ценовых графиков и индикаторов (TradeStation 2000i); 2) работа с опционами (OptionStation 2000i); 3) сортировка инструментов в реальном масштабе времени в соответствии с заданными критериями (RadarScreen 2000i); 4) прием, хранение и выдача данных о ЦБ (GlobalServer); 5) создание уникальных функций, сигналов и индикаторов на встроенном языке (PowerEditor).	1. Возможность формирования уникальных функций, сигналов и индикаторов. 2. Наличие множества встроенных функций, формул, примеров и индикаторов (более 300). 3. Реализация различных стратегий управления капиталом. 4. Ведение статистики проведения торгов. 5. Вывод множества окон с информацией.	1. Сложность полной настройки. 2. Низкое качество отображения графиков индикаторов и цен. 3. Отсутствие русификации. 4. Возможность пропажи данных. 5. Ограниченный размер текста. 6. Закрытость системы. 7. Высокая стоимость (сотни тысяч долларов).
MetaStock	Назначение: инструмент ТА различных финансовых рынков для трейдеров и профессиональных инвесторов. Включает множество торговых систем, инструментов и экспертов. Разработчик: компания Eguis. Возможности: формализация торговых стратегий; конструирование индикаторов и новых торговых систем; тестирование торговых систем на основе исторических данных; математическая формализация и включение дополнительных компонент (комиссия, StopLoss, TakeProfit); проведение анализа разномасштабных графиков (дневных,	1. Возможность формирования уникальных функций, сигналов и индикаторов. 2. Наличие библиотеки экспертов. 3. Наличие множества готовых индикаторов. 4. Возможность	1. Неудобство редактирования формул (проблемы со шрифтами и др.). 2. Проблемы написания сложных формул. 3. Невозможность установки стоп на закрытие или открытие позиции по рынку или по цене сигнала. 4. Ограниченные методы

	недельных, годовых). Стоимость: более 100000 руб.	экспорта и импорта данных.	управления капиталом. 5. Некорректный расчет ряда параметров. 6. Большие величины значений ошибок. 7. Возможность пропажи данных. 8. Закрытость системы.
Neuroshell DayTrader	Назначение: анализ рынка с использованием ИНС. Возможности: конструирование индикаторов и новых торговых систем; нейросетевые прогнозы; выбор целевой функции для нейросетевой модели; тестирование торговых систем на основе исторических данных; автоматическая загрузка данных; экспорт и импорт данных; генетическая оптимизация набора и параметров торговых правил; возможность внутридневной работы; тестирование стратегий. Стоимость: более 50000 руб.	1. Реализация внутридневной торговли. 2. Наличие оптимальной тренировки ИНС.	1. Работа с короткими временными периодами. 2. Проблемы написания сложных формул. 3. Возможность пропажи данных. 4. Закрытость системы.
AmiBroker	Назначение: для ТА и разработки торговых систем. Возможности: предоставление широкого набора индикаторов; создание уникальных индикаторов; автоматическая загрузка данных из различных источников; настройка интерфейса; отрисовка примитив; поддержка различных внутридневных интервалов; управление позициями. Стоимость: более 25000 руб.	1. Высокая оперативность функционирования. 2. Широкое распространение. 3. Открытая архитектура для программ подобного класса. 4. Сбор статистических данных.	1. Отсутствие возможности работы со средними позициями. 2. Несовместимость с современными операционными системами. 3. Возможность пропажи данных. 4. Трудности сопряжения с другими программами.
Multicharts	Назначение: ведение торгов на валютном и фондовом рынках. Возможности: создание уникальных индикаторов и торговых систем; отрисовка примитив; тестирование; предоставление встроенных функций; настройка временных и	1. Возможность импорта индикаторов и систем. 2. Гибкая настройка интерфейса.	1. Ограниченное число платформ для импорта индикаторов и систем. 2. Возможность пропажи данных.

	тиковых фреймов; настройка уникальных стратегий; автоматическая загрузка данных из различных источников; трехмерное представление результатов; формирование и тестирование торгового портфеля на исторических данных. Стоимость: более 100000 руб.	3. Трехмерная визуализация. 4. Сопряжение с другими системами.	3. Ограниченные методы управления капиталом.
ELWAVE	Назначение: проведение технического анализа на основе волновой теории Элиота. Возможности: предоставление набора индикаторов; прогнозирование состояния рынка и цен; автоматическое определение возможности заключения потенциально прибыльной сделки и моментов закрытия позиций; визуализация полученных результатов; определение «точек разворота» тренда; обучение персонала.	1. Возможность анализа и обучения торговле. 2. Гибкая настройка интерфейса. 3. Возможность прогнозирования цен активов.	1. Отсутствие возможности работы в режиме реального времени. 2. Необходимость корректировки построенных графиков. 3. Закрытость системы.
FXAnalytics	Назначение: проведение фундаментального анализа рынка. Возможности: получение новостей в реальном режиме времени; котировки активов с различных рынков; календарь экономических событий; торговые рекомендации; скачивание торговых систем и др.	1. Получение новостей в реальном режиме времени.	1. Отсутствие аналитических средств проведения ФА. 2. Предназначена для Forex/ 3. Закрытость системы.
София	Назначение: мониторинг портфеля ЦБ. Разработчик: фирма «Инфострой» (РФ). Функционирует в Санкт-Петербургском Промстройбанке. Возможности: ведение ИП инвесторов; характеристика состава, структуры и состояния ИП; расчет характеристик ИП (число ЦБ, средняя цена покупки, балансовая стоимость ЦБ, сумма выплаченных налогов, средний инвестированный капитал за период и др.); проведение операций с ЦБ (покупка, продажа, закрытие, оприходование, списание); предоставление графиков изменения рыночных параметров; сведения о ЦБ (котировки, объявляемые дивиденды и др.); отображение курсов валют; отображение экономических и биржевых индексов; генерация отчетов и предоставление графиков..	1. Широкий спектр предоставляемой информации. 2. Решение разнородных задач. 3. Значительные функциональные возможности.	1. Отсутствие моделей ИП. 2. Закрытость системы. 3. Высокая стоимость (более 300000 руб.).

GAMA	Назначение: поддержка принятия решения менеджером ИП и обеспечение единого информационного пространства в цикле управления ИП. Разработчик: фирма «Инфострой» (РФ), ·ская компания «Insight System ApS», итальянская фирма APL Italiana. Возможности: мониторинг состояния ИП, банковских счетов и ситуации на рынке; ввод, контроль и анализ транзакций; анализ доходности ИП и активов; имитационное моделирование ИП; моделирование по Блэку-Шоулсу; оценка рисков и др.	1. Полноценная СППИР. 2. Реализация нескольких этапов в цикле управления ИП. 3. Широкий спектр предоставляемых возможностей. 4. Возможность моделирования ИП.	1. Ограниченное число используемых моделей. 2. Закрытость системы. 3. Отсутствие возможности работы со стохастическими данными. 4. Высокая стоимость (более 700000 руб.).

Приложение Е. Результаты сравнительного анализа основных моделей формализации текстовой информации

Название	Краткое содержание	Достоинства	Недостатки
Модель И. Ньютона [196]	Общность языков обусловлена их предназначением для обозначения субстанций (духов или тел), которые у всех народов одинаковы. Носитель слов – предикат, обозначающий отношения. Корень универсального языка относится к какой-то одной субстанции (например, необходима специальная буква для обозначения видов людей, ангелов, домов и т.д.). Предпринята попытка первичной лексической категоризации	1. Впервые предложена оценочная шкала использования грамматических префиксов и аффиксов (плохой-хороший, большой-маленький и др.) 2. Проработаны глубинные падежи 3. Разработана шкала времени	1. Малое число глубинных падежей и их примеров 2. Временная шкала не походит в качестве формальной модели времени в силу значительной идеализации
Модель Г.Лейбница [243]	Все имена собственные произошли от нарицательных (общих). Суть механизма метонимического переноса локативных предлогов заключается в их переходе от "чувствительного" пространственного значения к "нечувственному" ментальному осмыслению. Реальное определение – это определение, которое задает объект, существование которого подтверждено нашим опытом или конструктивным доказательством	1. Категоризация лексики 2. Введение первого шага семантического анализа на основе номинальных и реальных определений	Недостаточная категоризация и примитивный семантический анализ
Лямбда-исчисление	Разработка формальной семантики на основе использования λ-оператора и двух (α и β) операций конверсии. λ-оператор является синтаксическим аналогом кванторов всеобщности и существования. Он ставится перед переменной, после чего эта переменная α считается связанной и ждет заполнения формулой. Операция β-конверсия реализует подобные заполнения	Применение λ-исчисления для представления смысла высказываний и построения теоретико-множественных интерпретаций	1. Сложность получаемых формул 2. Низкая эффективность при работе с насыщенными фрагментами языка
Семантическая сеть	Представляет собой множество взаимосвязанных элементов - слов и словосочетаний. Они несут основную смысловую нагрузку и наиболее часто встречаются в тексте. Статистическая обработка элементов (оценка частоты встре-	1. Приемлемая точность тематического поиска, реферирования и др.	1.Высокие требования к вычислительным ресурсам 2. Низкое быстродействие

	чаемости) позволяет выявить их вклад в общее содержание документа, а последующая весовая обработка позволяет определить плотность их распределения в документе	2. Простота и наглядность	3.Необходимость привлечения экспертов для настройки сети
Искусственная нейронная сеть	Для смысловой обработки текста используется нейронная функциональная алгебра. Нейроны выполняют операции дизъюнкции, конъюнкции и отрицания. Структура ИНС определяет порядок применения базовых операций к тексту Отдельный нейрон из состава сети, извлекающей смысл из текста на естественном языке, соответствует элементарному понятию (слово, термин, абзац и др.) анализируемого языка	1. Высокая оперативность 2. Приемлемое качество обработки текстов	1. Сложность 2. Проблемы обучения