ISSN 1998-0663

№3(21)–2012
http://bijournal.hse.ru

БИЗНЕС-ИНФОРМАТИКА

МЕЖДИСЦИПЛИНАРНЫЙ НАУЧНО-ПРАКТИЧЕСКИЙ ЖУРНАЛ НИУ ВШЭ

BUSINESS INFORMATICS

Учредитель:
Национальный исследовательский университет «Высшая школа экономики»

Редакционная коллегия
Абдульраб А. (Франция)
Авдошин С.М.
Алескеров Ф.Т.
Бабкин Э.А.
Баранов А.П.
Беккер Й. (Германия)
Белов В.В.
Грибов А.Ю.
Громов А.И.
Зандкуль К. (Германия)
Ильин Н.И.
Калягин В.А.
Каменнова М.С.
Козырев О.Р.
Кузнецов С.О.
Мальцева С.В.
Миркин Б.Г. (Великобритания)
Моттль В.В.
Пальчунов Д.Е.
Пардалос П. (США)
Силантьев А.Ю.
Таратухин В.В.
Ульянов М.В.

В ЭТОМ НОМЕРЕ:

АЗЕРБАЙДЖАН: ИНДЕКСЫ «ОХВАЧЕННОСТИ» ИНТЕРНЕТ

ХАОС НА ФОНДОВЫХ РЫНКАХ

ТЕМПОРАЛЬНЫЕ ОРГРАФЫ

УПРАВЛЕНИЕ ЭФФЕКТИВНОСТЬЮ

В соответствии с решением президиума ВАК РФ журнал «Бизнес-информатика» с 19.02.2010 включён в Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученых степеней кандидата и доктора наук.

БИЗНЕС-
ИНФОРМАТИКА
№3(21)–2012

СОДЕРЖАНИЕ

Принятие решений и бизнес-интеллект

Р.А. Караев
Стратегия развития информационно-коммуникационных технологий в Республике Азербайджан. Форсайт-2020 3

В.В. Таратухин, Е.А. Баженова
Влияние информационно-коммуникационных технологий на управление бизнес-процессами малых и средних предприятий в развивающихся странах..... 9

Математические модели социальных и экономических систем

Е.С. Кузнецов
Особенности прогнозирования экономических процессов на основе определения оптимальных базовых параметров 17

А.В. Зиненко
R/S анализ на фондовом рынке..... 24

Математические методы и алгоритмы бизнес-информатики

Б.Г. Миркин, Е.Л. Черняк, О.Н. Чугунова
Метод аннотированного суффиксного дерева для оценки степени вхождения строк в текстовые документы 31

В.А. Кохов, В.В. Кохов
Методы анализа изменений глобальных и локальных свойств темпоральных орграфов 42

Анализ данных и интеллектуальные системы

А.Г. Исавнин, Д.Р. Галиев
Программный комплекс для управления структурой инвестиционного портфеля 52

Информационные технологии в бизнесе

Д.В. Исаев
Мониторинг и планирование развития систем информационной поддержки корпоративного управления и стратегического менеджмента..... 63

М.А. Аракелян, Е.В. Чепин, А.П. Шабанов
Об инновационном подходе к созданию и испытаниям контрольно-учётных модулей для корпоративных информационных систем 70

Annotations..... 78

БИЗНЕС-ИНФОРМАТИКА
№3(21)–2012

Междисциплинарный
научно-практический журнал
НИУ ВШЭ

Журнал рекомендован ВАК
для научных публикаций

Подписной индекс издания
в каталоге агентства
«Роспечать» –72315

Учредитель:
Национальный
исследовательский университет
«Высшая школа экономики».
Выходит 4 раза в год.

Главный редактор
Голосов А.О.

Заместители главного редактора
Горбунов А.Р., Исаев Д.В.

Научный редактор
Лычкина Н.Н.

Технический редактор
Осипов В.И.

Дизайн обложки
Борисова С.Н.

Компьютерная вёрстка
Богданович О.А.

Администратор веб-сайта
Проценко Д.С.

Адрес редакции:
105187, г. Москва,
ул. Кирпичная, д. 33/5.
Тел.+7 (495) 771-32-38,
e-mail: bijournal@hse.ru

За точность приведённых сведений
и содержание данных,
не подлежащих открытой публикации,
несут ответственность авторы

При перепечатке ссылка на журнал
«Бизнес-информатика» обязательна

Тираж 500 экз.

Отпечатано в типографии НИУ ВШЭ
г. Москва, Кочновский проезд, 3.

© Национальный
исследовательский университет
«Высшая школа экономики»

СТРАТЕГИЯ РАЗВИТИЯ ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ В РЕСПУБЛИКЕ АЗЕРБАЙДЖАН ФОРСАЙТ-2020

Р.А. Караев,
профессор, руководитель лаборатории Института кибернетики НАН Азербайджана

Адрес: Азербайджан, г. Баку, ул. Б. Вагабзаде, д. 9
E-mail: karayevr@rambler.ru

Дается анализ Национальной стратегии информатизации на период до 2015 г. Отмечаются ограниченные возможности рейтинговой концепции ITU при решении вопроса дальнейшего развития ИКТ, связанного с Концепцией развития страны до 2020 г. Поднимается вопрос разработки рыночной стратегии, отражающей миссию ИКТ на этапе перехода Республики от ресурсно-экспортной к ресурсно-инновационной («ненефтяной») экономике.

Ключевые слова: информационно-телекоммуникационные технологии, стратегия развития, «ненефтяная» экономика, миссия ИКТ, Азербайджан

1. Введение

В последние годы Азербайджан столкнулся с системными вызовами, отражающими мировые тенденции и внутренние особенности социально-экономического развития. К главным из них относится – укрепление роли информационно-коммуникационных технологий (ИКТ) на этапе перехода от ресурсно-экспортной к ресурсно-инновационной («ненефтяной») экономике.

Глубокое понимание и адекватная оценка роли ИКТ нашли отражение в двух основополагающих документах, принятых высшим руководством страны:

1. «Национальной стратегии (на 2003-2012 годы) по информационным и коммуникационным технологиям для развития Азербайджанской Республики» (далее – Национальная стратегия), утвержденной 17.02.2003 г. общенациональным лидером Президентом Г.Алиевым;

2. «Государственной Программе на 2005-2008 годы по развитию связи и информационных технологий Азербайджанской Республики (Электронный Азербайджан), утвержденной 21.10.2005 г. нынешним Президентом И.Алиевым.

Достижение приоритетов, определенных Национальной Стратегией, и целей, определенных Государственной Программой, составили основу всего последующего развития отрасли.

Новый этап в развития ИКТ знаменовал Указ главы государства «О некоторых мерах в области организации оказания государственными органами электронных услуг» от 23.05.2011 г., открывший широкие возможности для применения электрон-

ных услуг в стране. В рамках выполнения указа создан Национальный центр сертификации, введена электронная подпись, с целью организации и использования электронных услуг госорганов на основе принципа «единого окна» создан Портал электронного правительства, ведутся работы по размещению здесь услуг госорганов.

В результате развития национального сегмента Интернета, неоднократного понижения цен, пропаганды современных средств подключения к Интернету, в том числе широкополосных, беспроводных и мобильных средств, создания благоприятных условий для деятельности провайдеров и интернет-клубов параметры охвата Азербайджана доступом к сети Интернет существенно возросли.

Если в 2010 г. на каждые 100 человек приходилось 50 интернет-пользователей, а в 2011 г. их число достигло 60 процентов всего населения страны. Приблизительно половина интернет-пользователей обзавелась широкополосным доступом. Это больше общемирового среднего показателя приблизительно в 2,5 раза

По мнению министра связи и информационных технологий Азербайджана акад. А. Аббасова (http://www.e-azerbaijan.info/ru/standpoint/read/26), в целом развитие отрасли соответствует Национальной стратегии. Перед отраслью была поставлена цель: до 2009 г. достичь среднемировых показателей в отрасли. Текущее состояние отрасли свидетельствует о том, что это получилось. Цель второго этапа Национальной стратегии состояла в том, чтобы к 2013 г. приблизиться к показателям развитых стран.

Основная задача проведенного нами исследования состояла в том, чтобы оценить возможность достижения этой глобальной цели с учетом сложившихся трендов и определить факторы дальнейшего развития отрасли на период до 2020 г., определенного Концепцией развития страны «Азербайджан-2020: взгляд в будущее».

2. Трудности оценки стратегии

Работы по созданию инструментов измерения прогресса в построении информационного общества и использовании ИКТ ведутся с 2001 г. В настоящее время разработано несколько систем мониторинга, которые используются для межстранового рейтингового анализа. Практика применения этих систем повсеместно демонстрирует многочисленные трудности, с которыми сталкиваются и непосредственные пользователи систем, и международные статистические агентства при сборе данных. В частности:

- ✦ системы мониторинга имеют разную целевую ориентацию (например, «Индекс сетевой готовности», «Индекс развития ИКТ», «Индекс экономики знаний» [1, 2] и др.) и соответственно используют разные наборы показателей, разные шкалы и разные модели агрегирования;
- ✦ наборы показателей и методы агрегирования непрерывно корректируются по мере прогресса в методологиях оценки;
- ✦ имеют место существенные задержки с публикацией результатов мониторинга;
- ✦ подробные сведения о методах сбора данных, экспертах и методологиях оценки не всегда находятся в открытом доступе;
- ✦ большинство исследований дают мало информации о том, каким образом их показатели были построены и почему, или как они могут быть скорректированы;
- ✦ отсутствуют данные по отдельным пунктам опросных листов;
- ✦ некоторые страны не представляют статистические отчеты из-за отсутствия необходимых ресурсов для сбора данных;
- ✦ имеет место излишнее доверие к данным из административных источников и отсутствует возможность проверки их достоверности.

Указанные трудности присущи и местной практике и, несомненно, говорят об условности выбора некоторого общего индикатора, позволяющего оценить текущее состояние ИКТ в Республике и прогнозировать достижимость показателей развитых стран, провозглашенную Национальной стратегией.

В настоящей статье нами в качестве такого индикатора выбран «Индекс развития ИКТ (IDI)», предложенный International Telecommunication Union (ITU) [1]. Индекс широко используется в настоящее время в международной рейтинговой практике.

3. Индекс развития ИКТ (IDI)

Индекс развития ИКТ (ICT Development Index, IDI) является индикатором общего прогресса отдельных стран в направлении построения информационного общества. Индекс IDI разработан ITU в 2007 г., и служит инструментом для измерения и сравнительного анализа состояния ИКТ на глобальном, региональном и национальном уровнях.

Место Азербайджана в рейтинге стран по этому индексу входит в число контрольных показателей Национальной стратегии (2003-2012 гг.) и Государственной программы (2005-2008 годы).

Название объединенного индекса «ICT Development Index» отражает его основную цель – отслеживать прогресс в развитии ИКТ в разных странах, а также показывать глобальное цифровое неравенство, т.е. различия между странами, имеющими разные уровни развития ИКТ.

3.1. Структура индекса IDI

Основу индекса IDI составила трехуровневая модель движения к информационному обществу (*рис. 1*), определившая три субиндекса, образующих IDI:

✧ Субиндекс 1 «Инфраструктура и доступность ИКТ» (ICT Readiness (Infrastructure and Access));

✧ Субиндекс 2 «Использование ИКТ» (ICT Use (Intensity));

✧ Субиндекс 3 «Знания и навыки в сфере ИКТ» (ICT Capability (Skills)).

В процессе разработки IDI для каждого субиндекса были предложены наборы потенциальных показателей, из которых затем было выбрано 11 наиболее подходящих.

Субиндекс 1 «Инфраструктура и доступность ИКТ» включает показатели: «количество фиксированных телефонных линий на 100 жителей», «количество пользователей мобильной телефонной связи на 100 жителей», «пропускная способность международного Интернет-канала (бит/с) на одного пользователя», «доля домохозяйств, имеющих компьютер», «доля домохозяйств, имеющих доступ к Интернету дома.

Субиндекс 2 «Использование ИКТ» основывается на результатах показателей субиндекса 1, т.е. показывает фактическое использование, а также уровень интенсивности использования ИКТ, обозначенных в субиндексе 1. К сожалению, данные по использованию ИКТ по многим показателям, которые предполагалось включить в субиндекс 2, недоступны для ряда развивающихся стран, поэтому в дальнейшем, по мере появления соответствующей статистики, количество показателей данного субиндекса предполагается увеличить. Субиндекс 2 включает показатели: «количество Интернет-пользователей на 100 жителей», «количество абонентов фиксированного Интернет-соединения», «количество абонентов мобильного Интернет-соединения».

Субиндекс 3 «Знания и навыки в сфере ИКТ». При разработке IDI предполагалось, что в этот субиндекс войдут показатели, отражающие уровень знаний и навыков населения страны в сфере ИКТ. Однако по большинству развивающихся стран подобные данные недоступны, поэтому в качестве замены были выбраны показатели уровня образования. В

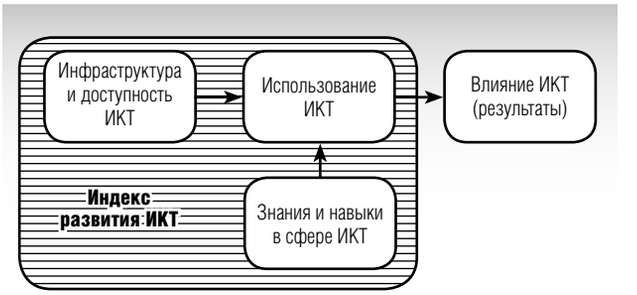


Рис. 1. Стадии развития стран на пути к информационному обществу, Источник: [1, P. 8].

дальнейшем, по мере появления соответствующей статистики, существующие показатели будут заменены на показатели, отражающие уровень знаний и навыков в сфере ИКТ, а не уровень образования в целом. Вся информация по показателям субиндекса 3 предоставляется Институтом статистики ЮНЕСКО. Показатели субиндекса: «уровень образования взрослого населения», «валовой охват средним образованием», «валовой охват высшим образованием».

3.2. Модель агрегирования индекса IDI

Для расчета сводного индеса IDI используется модель «взвешенной суммы». При этом все показатели в рамках субиндексов имеют равные веса. Как объясняют разработчики индекса IDI, это сделано для того, чтобы применяемая методология расчета была максимально простой.

В отличие от показателей субиндексы имеют разные веса в конечном индексе. Субиндекс «Инфраструктура и доступность ИКТ» и субиндекс «Использование ИКТ» имеют вес 40%. Субиндекс «Знания и навыки в сфере ИКТ» имеет вес 20%, так как на данный момент он представлен показателями-заменителями, отражающими общий уровень образования, а не показателями уровня знаний и навыков в сфере ИКТ.

4. Текущее состояние ИКТ в Республике

Индекс IDI был разработан для оценки текущего состояния и динамики изменений в сфере ИКТ как в развитых, так и в развивающихся странах. За последние годы целый ряд стран, в том числе и некоторые развивающиеся страны, добились весьма существенных успехов по индексу IDI, как в абсолютной, так и в относительной оценках. Азербайджан входит в число таких стран с «динамично развивающимися» ИКТ. В стране высокие и выше средних темпы роста наблюдались как по субиндексу доступа, так и по субиндексу использования, что указывает на развитие в Азербайджане информационного общества.

С точки зрения развития ИКТ между странами наблюдаются значительные различия. По данным [1] уровень индекса IDI колеблется от 0,8 для страны, находящейся в конце рейтинга, до 8,4 для страны, занимающей в этом рейтинге первое место (по шкале от 1 до 10). В динамике, увеличилась разница между странами в начале и в конце рейтинга, при этом более существенный прогресс наблюдается у лидеров рейтинга, и практически отсутствует у стран, занимающих самые последние его места.

В целом в 2011 году уровни индекса IDI в развивающихся странах составили по своей величине примерно половину от уровней индекса в развитых странах, при этом относительные темпы роста в развивающихся странах являются более высокими.

Азербайджан вошел в число стран с высоким показателем развития по индексу ИКТ. В период с 2002 по 2010 г. Азербайджан в мировом рейтинге поднялся на восемь позиций – с 100-го на 74-е место, при этом абсолютное значение индекса увеличилось на 2,07 – с 1,71 до 3,78, что выше среднего прироста индекса для всей группы анализируемых стран.

5. Оценка перспектив развития ИКТ и достижимости цели, установленной Национальной стратегией

Анализ многолетней динамики индекса IDI Азербайджана за период 2002 – 2010 гг. показывает следующее:

1. В группе стран СНГ индекс развития ИКТ Азербайджана (74) относительно диапазона ведущих странах СНГ (Россия – 47, Казахстан – 68) все еще существенно отстает.

2. Сопоставление Азербайджана с ведущими 20 странами – мировыми лидерами рейтинга IDI свидетельствует о значительном отставании Азербайджана по интегральному индексу развития ИКТ.

В контексте Национальной стратегии интерес представляет исследование возможности преодоления этого отставания и достижения показателей ведущих стран мира. Такое исследование может быть проведено на основе наблюдаемой многолетней динамики индекса IDI и прогнозирования трендов изменения индекса IDI в 20 ведущих странах мира и в Азербайджане.

Горизонт прогноза был принят равным 2015 г., т.е. он перекрывает сроки достижения глобальной цели Национальной стратегии (2013 г.) и дает возможность проследить за темпами приближения страны к показателям IDI развитых стран в среднесрочной перспективе. Прогнозирование трендов было осуществлено на основе методов регрессионного анализа и выполнено с помощью пакета Microsoft Office Excel 2003. Пакет имеет специальный аппарат для графического анализа временных рядов данных и построения линий тренда, которые могут использоваться для целей прогнозирования. В использованной версии Excel существует 10 различных видов линий тренда. Из возможных видов линий тренда была выбрана «экспоненциальная» функция, которая используется в случаях, когда скорость изменения данных непрерывно возрастает. Это соответствует характеру изменения индекса IDI во всех рассматриваемых странах.

Позиционирование Азербайджана по отношению к 20 ведущим странам осуществлялось в рамках диапазона индексов IDI, демонстрируемых страной, открывающей (Республика Корея) и страной замыкающей (Израиль) эту двадцатку на период отчета ITU [1]. Рассматривались также динамика и ожидаемые тренды индекса IDI России, которая на текущий момент имеет самый высокий рейтинг IDI среди стран СНГ. Параметр R, приводимый для трендов на диаграммах, показывает степень достоверности аппроксимации. В настоящем исследовании в качестве принятых трендов, за исключением случаев, противоречащих здравому смыслу, выбирались аппроксимации с максимальными значениями R. Все исходные данные получены из докладов ITU [1] и WEF [2]. Эти данные сформатированы и масштабированы для построения соответствующих графиков и представлены в *табл. 1.*

Таблица 1.

Значения индексов IDI (2002 – 2010 гг.)

Год	Республика Корея	Израиль	Россия	Азербайджан
2002	5,83	4,24	3,93	1,71
2003	6,41	4,81	3,88	1,99
2004	6,56	5,01	3,74	2,10
2005	6,85	5,15	3,98	2,23
2006	7,05	5,53	4,30	2,55
2007	7,26	5,6	4,27	2,71
2008	7,85	6,20	4,42	3,29
2009	7,91	6,39	4,47	3,31
2010	8,4	6,87	5,38	3,78

На рис. 2 приведены графики изменения индекса IDI Азербайджана и диапазон 20 ведущих стран мира, а также их тренды до 2015 года.

По результатам рейтинга ITU 2011 г. [1] в качестве стран открывающих и закрывающих диапазон 20 ведущих стран мира были выбраны соответственно Республика Корея и Израиль.

Приведенные графики показывают, что хотя отставание Азербайджана (3,78) от Израиля (6,87), имевшее место в 2010 г., к 2015 г. несколько сократится, и Азербайджан в 2014 г. может догнать Россию, тем не менее, индекс развития ИКТ (IDI) Азербайджана будет примерно в полтора раза меньше значения нижней границы диапазона 20 ведущих стран.

При этом, согласно приведенным графикам, значение индекса IDI Азербайджана в 2015 году приблизится к индексу IDI Израиля, представляющего нижнюю границу диапазона 20 ведущих стран, которую он имел в 2008 году. Это означает, что отставание Азербайджана от ведущих стран составляет 7 лет.

6. ИКТ Азербайджана – 2020: взгляд в будущее

Таким образом, если развитие ИКТ в Азербайджане в течение прогнозируемого периода будет происходить в условиях действия прежней совокупности факторов внутренней и внешней среды, то заданная Национальной стратегией цель достигнута не будут.

Здесь уместен вопрос: а отражает ли в полной мере индекс IDI положение дел в отрасли и достаточен ли он для дальнейшего управления развитием ИКТ? Очевидно, что нет.

Сегодня, на этапе перехода Республики к «ненефтяной» экономике, ИКТ становятся важным стратегическим ресурсом. Поэтому необходим более действенный адекватный подход к эффективному использованию этого ресурса, позволяющий перейти от узкоотраслевой рейтинговой доктрины ITU («догнать и перегнать»), к разумной и сбалансированной стратегии информатизации, отражающей социально-экономическую миссию ИКТ и стратегические цели информатизации на данном этапе.

Вопрос перехода к такой стратегии актуален и в контексте позиции ОЭСР, отмечающей, что сегодня в условиях информационного бума регионы могут инвестировать чрезмерные средства в ИКТ «либо в стремлении компенсировать недостаток квалификации, либо из-за отсутствия четкой рыночной стратегии информатизации» [5, с. 37].

Разработка четкой рыночной стратегии информатизации жизненно важна и для инвестиционной политики отрасли, и для экономики в целом. В настоящее время в Республике намечаются и осуществляются многомиллиардные (в долл. США) амбициозные проекты [3, 4]:

✦ создание Общенациональной Инновационной Зоны в сфере ИКТ;

✦ запуск двух геостационарных телекоммуникационных спутников, ориентированных на предоставление теле-радио услуг населению страны и арендное обслуживание соседних стран региона;

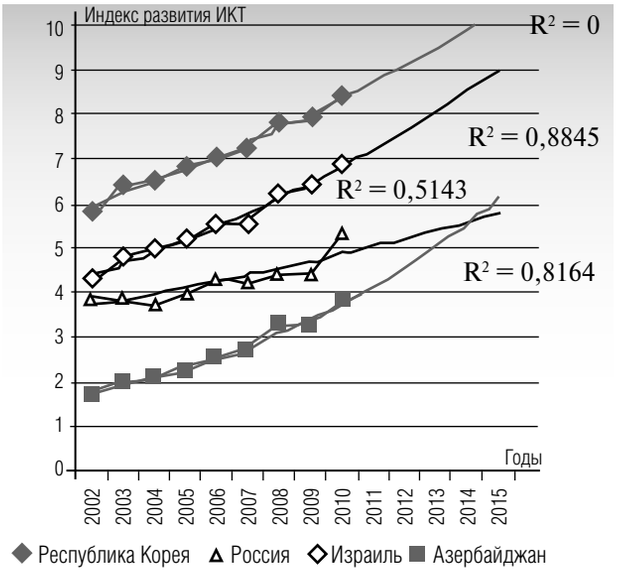


Рис. 2. Графики изменения индексов IDI и их тренды до 2015 года. Диапазон 20 ведущих стран мира

✦ запуск спутника для дистанционного зондирования Земли в народнохозяйственных целях;

✦ участие в строительстве «ТрансЕвразийской Информационной СуперМагистрали», соединяющей Европу с Китаем;

✦ разработка и реализация крупномасштабной (2 млрд. долл. США) Государственной программы по обеспечению кибербезопасности и информационной безопасности;

✦ развитие технологий «облачных вычислений» и др.

Эффективность этих проектов уже не может оцениваться традиционными методами ТЭО. Расчеты, выполненные западными консалтинговыми компаниями по отдельным проектам, не учитывают макроэкономическую ситуацию в стране и в значительной степени отвечают интересам «поставщиков» ИКТ, а не «заказчика». Сегодня эффективность проектов должна оцениваться в рамках единой рыночной стратегии, учитывающей:

1) ожидаемое истощение запасов углеводородных ресурсов,

2) растущую турбулентность внешней среды,

3) цели и приоритеты Концепции развития страны на период до 2020 г. [3],

4) серьезные внутренние проблемы (Р) национальной экономики (*табл. 2*), в решении которых государство хотя и отводит приоритетную роль отрасли ИКТ (показатель «Важность ИКТ для государства» – 14 (142) свидетельствует об исключительной поддержке государством отрасли ИКТ, опережающей

многие развитые страны), но сталкивается с серьезными трудностями, преодоление которых потребует определенного времени.

Таблица 2.

Р-профиль экономики Азербайджана [1, 2, 6]

Доля нефтегазового сектора в экспорте страны (2006-2012)	75-95%
Важность ИКТ для государства	14(142)
Индекс конкурентоспособности	57(139)
Индекс экономической свободы	91(179)
Качество бизнес-среды	80(142)
Уровень коррупции	134(181)
Знания и навыки в сфере ИКТ-менеджмента	100(142)
Доступ к венчурному капиталу	54(142)
Степень использования ИКТ бизнесом	108(142)

Хорошие возможности для разработки долгосрочной рыночной стратегии развития ИКТ, отвечающей Концепции развития страны до 2020 г. в сегодняшних сложных условиях, открывает парадигма форсайт-прогноза («коллективного архетипа» будущего) [7] и такие эффективные инструменты поддержки, как динамический PEST&SWOT-анализ, методы стратегического и сценарного планирования, технологии когнитивного моделирования [8].

Ключевым в решении вопроса является переход от волюнтаристского метода формирования стратегии, лежащего в основе сегодняшней политики информатизации, к научному, рассматривающему стратегию развития ИКТ не только как объект измерения, но и как объект управления [8]. Это – принципиальный момент, создающий предпосылки для синтеза эффективных стратегий.

Особую роль здесь призваны сыграть когнитивные технологии. В 2006 г. Национальный научный фонд США, под эгидой которого ведется львиная доля научных исследований в Америке, выпустил отчет, прогнозирующий развитие науки в ближайшие 50 лет [9]. Отчет был назван «NBIC-конвергенция» (N - это нанотехнологии, B - биотехнологии, I - информационные технологии, C - когнитивные технологии) и определил главные тренды в мировой науке ближайших десятилетий. Революция информационных технологий началась еще в 60-е, бурный прогресс биотехнологий развернулся в 90-е годы, нанотехнологий – в начале этого века. Сегодня начинают стремительно развиваться когнитивные технологии, в том числе, и когнитивные технологии управления развитием социально-экономических объектов.

Когнитивные технологии открывают новую (не исключено, уникальную) возможность осуществлять афферентный синтез стратегии развития ИКТ (в смысле Г. Минтцберга [10, с. 127]) с учетом:

- ✧ всех перечисленных *в табл. 2* проблем национальной экономики;
- ✧ противоречивости интересов стейкхолдеров (государство, бизнес, наука, население);
- ✧ неполноты и неточности статистики;
- ✧ многовариантности будущего (оптимистический, инерционный, пессимистический сценарии). ■

Литература

1. Measuring the Information Society 2011 / International Telecommunication Union (http://www.itu.int/ITU-D/ict/publications/idi/2011/index.html).
2. The Global Information Technology Report 2012 / World Economic Forum, INSEAD http://www.weforum.org/en/initiatives/gcp/Global%20Information%20Technology%20Report).
3. Аббасов А.М. Азербайджан-2020: Взгляд в будущее. Реалии и перспективы ИКТ. Газета «Азербайджан», 3 июля 2012 г. (www.azerbaijan.news.az).
4. Aliguliyev R., Gurbanov G. Big Ambitions in Rapidly Changing World: Azerbaijan // Measuring the Information Society, International Telecommunication Union. 2011. (http://www.itu.int/ITU-D/ict/publications/idi/2011/index.html).
5. Организация Экономического Сотрудничества и Развития (ОЭСР). Наука, Технологии и Промышленность: Перспективы 2008. (http://www.oecd.org/html).
6. Индекс восприятия коррупции (http:// ru.wikipedia.org/wiki/).
7. Hines A., Bishop P. Thinking about the future: Guidelines for Strategic Foresight. – N.Y.: Fre Press. 2007. 357 p.
8. Караев Р.А., Гюльмамедов Р.Г. Проблема выбора стратегии реализации ИКТ– потенциала инновационных регионов: когнитивный анализ // Экономика и управление. – 2010. – № 7 (57). – С. 34-39.
9. Managing Nano-Bio-Info-Cogno Innovations: Converging Technologies in Society. – National Science Foundation.USA. 2006. (http://www.wtec.org/ConvergingTechnologies/3/NBIC3_report.pdf)
10. Mintzberg H. The Rise and Fall of Strategic Planning. – N.Y.: Free Press, 1994.

ВЛИЯНИЕ ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ НА УПРАВЛЕНИЕ БИЗНЕС-ПРОЦЕССАМИ МАЛЫХ И СРЕДНИХ ПРЕДПРИЯТИЙ В РАЗВИВАЮЩИХСЯ СТРАНАХ

В.В. Таратухин,
доктор философии, кандидат технических наук, заведующий базовой кафедрой SAP Национального исследовательского университета «Высшая школа экономики»
Е.А. Баженова,
аспирант кафедры инноваций и бизнеса в сфере информационных технологий Национального исследовательского университета «Высшая школа экономики»

E-mail: vtaratoukhine@hse.ru, kbazhenova@hse.ru
Адрес: г. Москва, ул. Кирпичная, д. 33/5

В работе осуществлен анализ роли информационно-коммуникационных технологий при ведении малого и среднего бизнеса в развивающихся странах. Предлагается подход для разработки концепции управления бизнес-процессами на малых и средних предприятиях.

Ключевые слова: информационно-коммуникационные технологии, мобильные технологии, бизнес-процессы, управление предприятием, малый бизнес.

1. Введение

К среднему и малому бизнесу, согласно определению, сформулированному Европейской Комиссией [9], относятся предприятия с численностью сотрудников до 250 человек и максимальным годовым оборотом в 50 миллионов евро. Важность малых и средних предприятий (МСП) на сегодняшний день неоспорима как для развитых, так для развивающихся стран [1, 2, 3]. Предоставляя миллионы рабочих мест, институт малых и средних предприятий является основным средством устойчивой индустриальной и социальной диверсификации общества, таким образом выступая одним из основных драйверов экономического развития большинства стран [3, 4]. Однако, такие явления как глобализация, интернационализация национальных рынков, мировой

экономический кризис, волатильность финансовых рынков, сокращение инвестиций, быстро меняющийся потребительский спрос, — оказывают возрастающее давление на МСП, стимулируя их к поиску путей для выживания и развития в современной бизнес-среде. Каким же образом МСП находят эти пути, преодолевают вышеперечисленные трудности?

Безусловным фактором повышения конкурентоспособности большинства предприятий является использование ими результатов научно-технического прогресса, а именно, информационно-коммуникационных технологий (ИКТ). Доступ МСП к ИКТ повышает эффективность ведения бизнеса и мировой экономики вообще, являясь фактором, входящим в цепочку создания добавленной стоимости продукта. Использование ИКТ на предприятии определяется различными подходами к их управлению, методологией их внедрения и использования. Именно поэтому необходимо рассматривать ИКТ в неразрывной связи со всеми реализуемыми бизнес-процессами на предприятии.

Мировой рынок ИКТ, поддерживающих управление бизнес-процессами (Business Process Management, BPM) и включающих в себя всевозможные коммуникации, программное обеспечение, веб-приложения и сервисы, ERP-системы и т.д. был оценен в 2008 году в 1,8 миллиард долларов и прогнозируется его увеличение до 6,2 миллиарда долларов в 2015 году [5]. В то же время, на сегодняшний день использование ИКТ в управлении бизнес-процессами малых и средних предприятий в развивающихся странах весьма умеренно [6, 7]. Следовательно, существует необходимость исследования использования и влияния ИКТ на таких предприятиях и разработки концепции адаптированного управления бизнес-процессами, учитывающей специфику ведения малого и среднего бизнеса в развивающихся странах. При этом среди развивающихся стран, обладающих наибольшим потенциалом для развития ИКТ, следует выделить группу BRICS: Бразилия, Россия, Индия, Китай и Южная Африка [8]. Развитие среднего и малого бизнеса является одной из главных задач государства как этих, так и большинства стран в мире.

Вышеперечисленные аргументы и факты стали предпосылками к проведению анализа существующих практик использования ИКТ и их влияния на управление бизнес-процессами на МСП.

2. Управление бизнес-процессами на малых и средних предприятиях

Согласно [10], BPM (управление бизнес-процессами) — концепция процессного управления организацией, согласно которой бизнес-процессы рассматриваются как особые ресурсы предприятия, непрерывно адаптируемые к постоянным изменениям. При этом основными принципами управления предприятием являются понятность и видимость бизнес-процессов за счёт их моделирования с использованием соответствующих формальных нотаций, программного обеспечения, симуляции, мониторинга и анализа, а также возможность динамического перестроения моделей бизнес-процессов силами участников и средствами программных систем.

Ключевая причина ограниченного применения большинства концепций BPM малым и средним бизнесом заключается в неуверенности предпринимателей в получении выгоды от использования BPM-решений, их приверженности устаревшим принципам управления, что в совокупности с неполнотой информации из-за ограниченности ресурсов и большими рисками при внедрении программного обеспечения, поддерживающего управление бизнес-процессами, препятствует развитию бизнеса [6].

Для понимания направлений развития BPM на малых и средних предприятиях, необходимо выделить отличительные свойства МСП, связанные с их оргструктурой, общими стратегиями, а также рассмотреть особенности социологического контекста. Согласно [11], МСП свойственны следующие характеристики, отличающие их от крупных предприятий:

1. Специализация и индивидуальность. МСП, работающие на бизнес-рынках, не покрытых крупными предприятиями, отличает глубокая специализация на определенном сегменте.

2. Доступность рынкам. В отличие от крупных предприятий, фокусирующихся на взаимозаменяемых товарах и услугах для анонимных рынков, МСП сильно концентрируются непосредственно на своих конечных потребителях

3. Быстрота реакции и бизнес-переориентации. Гибкость в принятии решений и внедрении организационных изменений достигается путем предпочтения простоты и трансформируемости процессов и организационных структур МСП.

4. Ограниченность ресурсов. МСП сильно ограничены ресурсами, особенно финансовыми и ИТ. Отсутствие «ноу-хау» компенсируется базовыми знаниями и навыками во многих областях разносторонне развитого персонала, выполняющего одновременно много задач.

5. Неоднородность технических ресурсов. МСП зачастую используют ИКТ не согласованно, без общей стратегии или при нехватке соответствующих компетенций. ИТ-ландашфт МСП зачастую представляет собой разнородные системы, начиная от ERP-систем до изолированных приложений, генерирующих расчетные документы для осуществления ежедневных бизнес-транзакций.

6. Противостояние глобализации. Растущая интернационализация рынков, с одной стороны, уменьшает барьеры входа на рынки для МСП. С другой стороны, возможности приобретения потенциальных бизнес-партнеров влекут за собой огромное адаптационное давление на МСП, которые имеют меньше опыта участия в глобальном электронном бизнесе, чем большие мероприятия.

На сегодняшний день в литературе описано определенное количество практик и подходов к управлению бизнес-процессами на предприятиях. Разработанная в [12] схема BPM является универсальной и может быть использована в качестве базы в том числе и для МСП. Неотъемлемой предпосылкой реализации проектов по внедрению или усовершенствованию управления бизнес-процессами на любом предприятии является моделирование процессов. По [12], порядок проведения проекта, ориентированного на реорганизацию процессов, представлен на *рис. 1*:

Таким образом, для разработки адаптированной модели управления бизнес-процессами на МСП можно разрабатывать данную схему, учитывая особенности малого и среднего бизнеса, описанные выше. Ясно, что для различных компаний данная схема будет существенно различаться по контексту в зависимости от условий на том или ином предприятии. Для внедрения BPM-решений необходимо для каждого МСП проводить анализ причинно-следственных связей для факторов, влияющих на принятие ИКТ, однако, необходимо подобрать правильную методологию такого анализа.

В литературных источниках, посвященных моделированию бизнес-процессов на малых и средних предприятиях, можно встретить такие методы как:

✦ проведение опросов представителей МСП о бизнес-структуре и процессах их предприятий с последующим их моделированием посредством вспомогательного программного обеспечения, такого как GRAI, ARIS, HDP и др. [13];

✦ построение бизнес-платформы МСП с помощью сбалансированной системы показателей (концепции переноса и декомпозиции стратегических целей для планирования операционной деятельности и контроля) [11];

✦ метод моделирования структурными уравнениями (выделение групп факторов, влияющих на принятие ИКТ в МСП и выведение причинно-следственной связи между ними [14];

✦ анализ информационных потоков данных предприятия [15]; и др.

Однако, в литературе имеется существенный пробел исследований, а именно, — отсутствие анализа использования ИКТ в управлении бизнес-процессами на предприятиях вообще, а тем более с учетом специфики МСП в развивающихся странах.



Рис. 1. Порядок проведения проекта, направленного на реорганизацию процессов

3. Использование ИКТ малыми и сверхмалыми предприятиями в странах с развивающейся экономикой

Эволюция технологий сильно влияет на ведение бизнеса, меняя инфраструктуру промышленности и бизнес-операций, а также создавая предпосылки для возникновения конкурентных преимуществ у тех организаций, который внедряют ИКТ в свои бизнес-процессы. Принятие ИКТ на МСП предоставляет возможность быстрого доступа, оценки, обработки и распространения больших объемов данных и информации об инфраструктуре бизнеса. Следовательно, только МСП, использующие соответствующие последнему слову науки технологии, имеют возможность выхода на международный рынок и оставаться конкурентоспособными несмотря на вызовы глобализации, либерализации и НТП [3].

Решения ИКТ помогают МСП увеличить их продуктивность, достичь более высоких бизнес-показателей. Однако, процент принятия ИКТ малыми и средними предприятиями в странах BRICS очень низок [16]. Таким образом, в настоящий момент можно наблюдать более чем умеренное использование ИКТ на МСП в странах с развивающейся экономикой. Это объясняется достаточно просто, ведь в развивающихся странах МСП сталкиваются с дополнительными трудностями – высокая стоимость телекоммуникаций, отсутствие законодательной поддержки от государства, использование устарелых технологий, общая технологическая неграмотность, отсутствие квалифицированного персонала, бедная коммуникационная инфрастуктура [3].

Препятствием к использованию ИКТ развивающимися странами является также неприменимость для них стандартных программных продуктов, используемых в развитых странах, в силу их высокой для развивающихся стран стоимости, отсутствия развитой поддерживающей инфраструктуры и ряда других причин. Поэтому для разработки решений, учитывающих особенности развивающихся стран, необходимо провести дополнительные исследования, изучить существующие проблемы, а также особенности микро и макро среды. Другими стратегиями преодоления вышеуказанных барьеров являются создание соответствующей инфраструктуры, создание финансовых и правовых фундаментов.

4. 1. Использование ERP-систем малыми и средними предприятиями в странах с развивающейся экономикой

Использование систем планирования ресурсов предприятия (Enterprise resource planning, ERP) считаются одним из путей выживания МСП на сегодняшний день [17]. Под ERP-системой подразумевается единая стандартная информационная система, которая оказывает комплексную поддержку всех БП компании [18]. Благодаря интеграции бизнес-процессов, которую реализуют ERP, становится возможным устранение функциональных границ между подразделениями, повышается доступность информации и ее беспрепятственное движение между различными отделами. Среди преимуществ, которые приносит внедрение ERP, также можно отметить такие улучшения в операционной деятельности предприятия, как снижение времени разработки продукта и его вывода на рынок, а так-

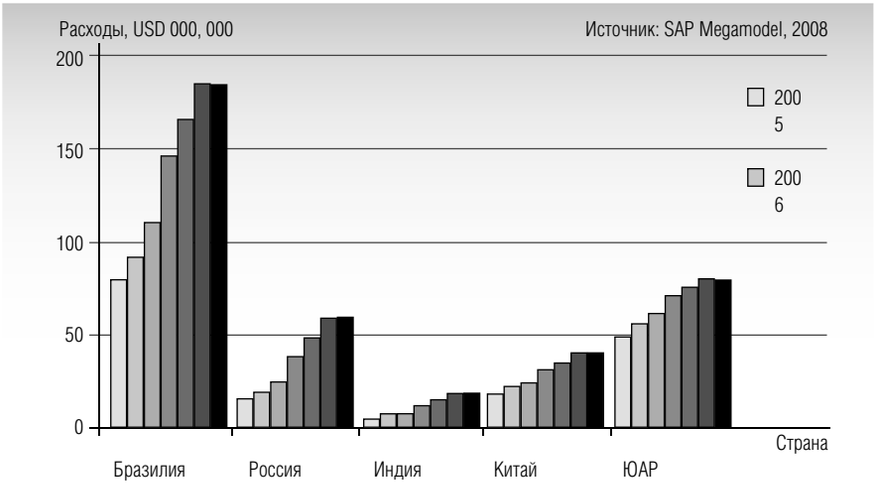


Рис. 2. Динамика расходов на ERP-системы малыми и средними предприятиями в странах BRICS

же появления возможности быстро реагировать на давление со стороны конкурентов и изменяющиеся условия на рынке [12].

До недавнего времени ERP системы создавались исключительно для крупных корпораций, обладающих значительными бюджетами на ИКТ и нуждающимися в стандартизации своих бизнес-процессов в соответствии с «лучшими мировыми практиками». Однако, в настоящий момент поставщики ERP ставят своей новой целью завоевание рынка МСП [19] и на рынке ERP-систем наблюдается устойчивая тенденция разработки решений для среднего и малого бизнеса. Согласно исследованиям и прогнозам компании SAP, динамика расходов на ERP-системы для МСП в странах BRICS отражает рост их использования, см. *рис. 2.*

С помощью ERP-системы и процессно-ориентированного подхода МСП развивающихся стран могут улучшить конкурентное положение, снизить издержки, а также повысить эффективность работы и контроля деятельности [12]. Однако, внедрение ERP-систем на МСП далеко не всегда дает позитивные результаты. Монолитная структура ERP со сложным процессом настройки параметров не всегда приемлема для МСП, которым свойственна ограниченность ресурсов, таких как деньги, персонал и навыки в области информационных технологий. Таким образом возникает потребность в создании специализированного решения для МСП, отличного от применяемых в крупных компаниях.

4.2. Использование принципа «программное обеспечение как сервис» малыми и средними предприятиями в странах с развивающейся экономикой

Для МСП, ИКТ-бюджеты которых обычно крайне ограничены, все более актуальным становится получение ПО через Интернет. Такое решение также оправдано с связи с ограниченным количеством специалистов в области ИКТ, работающих на предприятии. Малой компании часто гораздо выгоднее вообще не иметь собственный ИКТ-отдел, а нанимать сторонних специалистов, когда это необходимо [12]. Основной принцип SaaS (Software as a Service, «программное обеспечение как сервис») заключается в том, что та или иная программа не устанавливается на компьютер пользователя, а функционирует на мощностях сервисной компании. На *рис. 3* представлена реализация принципа SaaS компанией SAP во взаимодействии малых, средних и крупных предприятий. Клиенты при этом работают с программой через веб-интерфейс. Главное преимущество SaaS состоит в прямой экономии денег, – необходимы лишь компьютер и подключение к Интернету.

Благодаря развитию открытых технологий, разработчики ПО для малых и нишевых компаний получили возможность конкурировать с крупными производителями ERP, предоставляя свои решения МП по более низким ценам или вообще бесплатно.

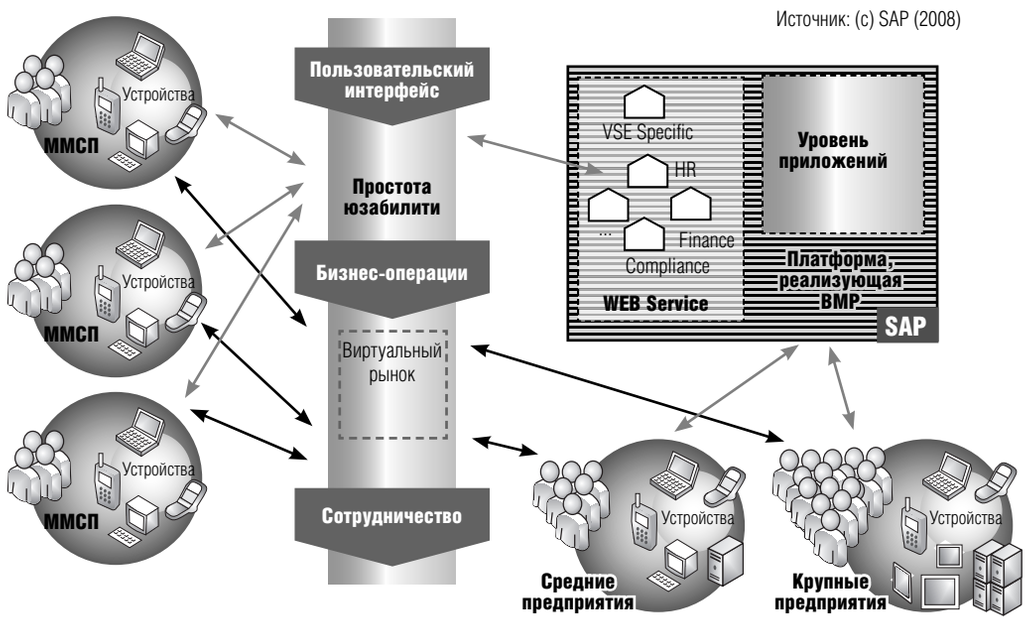


Рис. 3. Реализация принципа SaaS компанией SAP

4.3. Использование мобильных технологий малыми и средними предприятиями в странах с развивающейся экономикой

Становясь более мобильными, МСП заинтересованы в интеграции новых технологий с существующими бизнес-процессами и ждут соответствующих предложений от поставщиков. Согласно анализу [20], почти 80% опрошенных малых и средних предприятий BRICS имеют доступ к мобильным телефонам, хотя только 48% в настоящее время используют его для деловых целей. Согласно прогнозу eMarketer «Mobile BRIC: Extreme Growth Ahead», в странах BRICS будет более 1,7 млрд мобильных абонентов среди МСП к 2012 году, с более чем 680 миллионов пользователей, имеющих доступ к Интернету.

При разработке и внедрении новых инновационных технологий следует учитывать компетенции работников организаций. Так, например, авторами [20] проводится опрос менеджеров МСП стран BRICS, в котором респондентам было предложено самостоятельно оценить свой уровень владения различными вспомогательными устройствами для ведения бизнеса по шкале от 1 до 4 (1 – респондент никогда не использует устройство, 2 – респондент хочет научиться пользоваться устройством, 3 – у респондента имеется опыт использования устройства, 4 – респондент является продвинутым пользователем). Результаты опроса отражены *на рис. 4:*

Наибольший опыт менеджеры МСП стран BRICS имеют во владении компьютерами и мобильными телефонами. Это означает, что разрабатываемые

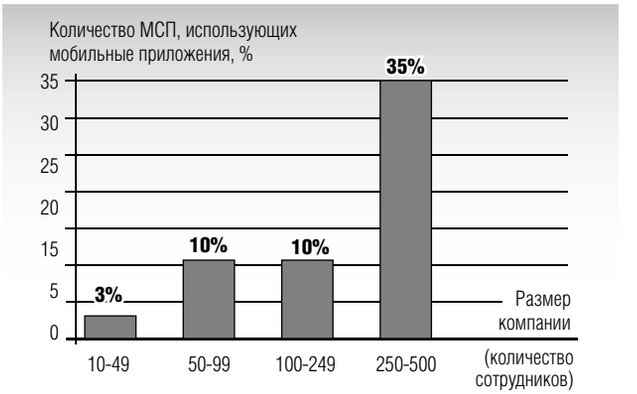


Рис. 5. Процент МСП из опроса [7], использующих мобильные интернет-приложения B2E

для малых и средних предприятий решения должны как минимум подходить именно к этим устройствам.

Таким образом, все больше представителей малых и сверхмалых предприятий в развивающихся странах изъявляют готовность инвестировать в интегрированное мобильное ПО бизнес-приложений и услуг для роста и поддержания своего бизнеса. Однако, они сталкиваются с проблемами, в силу которых предлагаемые а рынке текущие продукты и услуги не являются приемлемыми. Наглядно демонстрирует данный тезис исследование ученой группы из Италии, занимающейся анализом использования мобильных интернет-приложений «от бизнеса к сотрудникам» (business-to-employee, B2E) на итальянских МСП [7]. Под B2E подразумеваются мобильные интернет-приложения, которые могут принести существенную пользу реализации бизнес-процессов предприятия без их существенных модификаций и крупных инвестиций. По

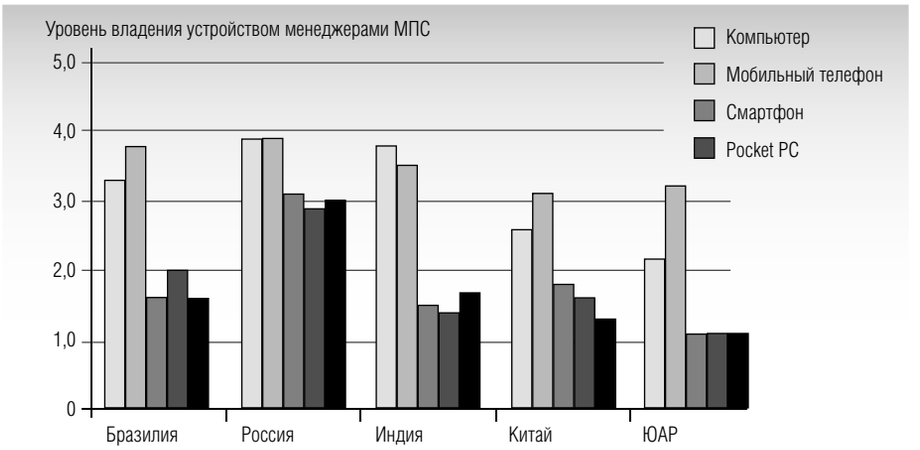


Рис. 4. Уровень владения менеджерами малых и средних предприятий различных устройств

этим причинам приложения B2E особенно подходят МСП, обычно не имеющим больших возможностей и ресурсов для использования ИКТ. По результатам опроса более 600 производственных МСП Италии и посредством использования метода районированной выборки (с учетом размера предприятия), был определен коэффициент диффузии мобильных приложений B2E в малом и среднем бизнесе (*рис. 5*).

Из рисунка видно, что процент использования МСП B2E приложений очень мал и наблюдается явная зависимость роста их использования в зависимости от размера компаний. Среди причин непопулярности использования таких мобильных технологий в [7] указываются недостаток знаний руководителей МСП о них и трудности в затрудненной оценке эффективности и выгод использования данных технологий до их внедрения.

Таким образом, роль инновационных мобильных технологий, таких как мобильные решения ERP-систем, в ведении бизнеса на МСП в условиях развивающихся рынков является весомым фактором развития малого и среднего бизнеса. Однако, существует потребность адаптации существующих решений к реальным нуждам и возможностям малого и среднего бизнеса.

5. Заключение

Представленный анализ существующих практик использования ИКТ в управлении бизнес-процессами показал, что существуют различные подходы к осуществлению BPM на предприятии, но, в основном, для больших предприятий. Для МСП в литературе встречается идентификация определенных свойств, которые необходимо учитывать при разработке BPM, однако конкретные разработки таких моделей встречаются только в единичных случаях в пилотных проектах. Таким образом, существует необходимость в разработки обобщенной модели управления BPM для малого и среднего бизнеса, которая могла быть использована различными МСП и производителями соответствующего программного обеспечения.

Также в статье был представлен такой анализ воздействия ИКТ на ведения бизнеса на МСП, — были рассмотрены ERP-системы, программное обеспечение типа SAAS и мобильные технологии. Исследования указывают на сильную положительную корреляцию между ИКТ и экономической эффективностью и конкурентоспособностью МСП. Однако, для развивающихся стран BRICS использование ИКТ малым и средним бизнесом затруднено в силу ряда таких особенностей, как недостаточное финансирование, отсутствие ресурсов и квалифицированных кадров, что определяет неприменимость к ним стандартных ERP-решений для развитых стран. При этом перспективными технологиями, применяемыми в управлении на МСП в развивающихся странах, представляются такие инновационные технологии, как мобильные решения ERP-систем.

Развитием данной работы предполагается разработка адаптированной концепции процессного управления МСП, работающем в развивающейся стране, учитывающей особенности управления бизнес-процессами в малом и среднем бизнесе. Кроме того, так как МСП существуют не в бизнес-вакууме, а являются элементами глобальной коммерческой экосистемы, представляется целесообразным рассмотреть малые и средние предприятия как клиентов и поставщиков крупных компаний. В ракурсе разработки методологии управления бизнес-процессами на предприятии такой подход подразумевает включение в рассмотрение не только внутренних, но и внешних бизнес-процессов компании.

Вышеперечисленные факторы являются предпосылками к проведению дальнейшей научной работы, направленной на разработку расширенной модели управления бизнес-процессами на малых и сверхмалых предприятиях с учетом влияния ИКТ в условиях развивающегося рынка и исследование возможности и выгод ее применения для малого бизнеса и общества в целом. ■

Литература

1. Julien P.-A. Introduction. The State of the Art in Small Business and Entrepreneurship // Ashgate, Aldershote. – 1998. – P. 1-3.
2. Placide P.-N., Raymond L., Fabi B. Adoption and risk of ERP systems in manufacturing SMEs: a positivist case study // Business Process Management Journal. – 2008. – Vol. 14, № 4. – P. 530-545.
3. Ongori. H. Information and communication tecnologies adoption in SMEs: literature review // Journal of Chinese Entrepreneurship. – 2010. Vol. 2, № 1. – P. 93-101.

4. 4. Apulu I., Latham A., Moreton R. Factors affecting the effective utilisation and adoption of sophisticated ICT solutions. Case studies of SMEs in Lagos, Nigeria // Journal of Systems and Information Technology. – 2011. – Vol. 13, № 2. – P. 98-100.
5. WinterGreen Research. Worldwide Business Process Management (BPM) Market Opportunities, Strategies and Forecasts, 2009 to 2015 // WinterGreen Research Inc. – 2006. – Report № SH24101315.
6. Chong S. Business process management for SMEs: an exploratory study of implementation factors for the Australian wine industry // Journal of Information Systems and Small Business. – 2007. – Vol. 1. – P. 1-2.
7. Balocco R., Mogre R., Toletti G. Mobile internet and SMEs: a focus on the adoption // Emerald Gorup Publishing Limited. – 2009. – № 109. – P. 2.
8. Баженова Е.А., Горбачева Е.В., Таратухин В.В. Роль инновационных и мобильных технологий в развитии малых и сверхмалых предприятий на развивающихся рынках // Труды Международного конгресса по интеллектуальным системам и информационным технологиям IS&IT'2011. – Физматлит, 2011. – С. 35-36.
9. European Commision. Recommendation 2003/361/EC: SME definition [Электронный ресурс]. – Режим доступа: http://europa.eu.int/comm/enterprise/enterprise_policy/sme_definition/index_en.htm (дата обращения 01.12.2011).
10. Gartner. BPM (business process management). [Электронный ресурс]. – Режим доступа: http://www.gartner.com/technology/it-glossary/#bpm (дата обращения 01.09.2011).
11. Hoyer V. Modeling Collaborative e-Business Processes in SME enviroments // Journal of Information Scence and Technology. – 2008. – Vol. 5, № 2. – P. 46-50.
12. Менеджмент процессов / Й.Беккер, Л.Вилков, В.Таратухин, М.Кугелер, М.Роземанн / Пер. с нем. – М.: Эксмо, 2007.
13. Santos J., Sarriegi J.M., Serrano N. A support methodology for EAI and BPM projects in SMEs // Taylor & Francis. – 2008. – Vol. 2, № 3. – P. 63-70.
14. An empirical test of causal relationships of factors affecting ICT adoption for building project management : an Indian SME case study / V.Ahuja, J.Yang, M.Skitmore, R.Shankar // Construction Innovation: Information, Process, Management. – 2010. – Vol. 10, № 2. – P.1-5.
15. Tam A.S.M., Chu L.K., Sculli D. Business process modelling in small- to meidum-sized enterprises // MCB University Press. – 2001. – Vol. 101, № 4. – P. 56-60.
16. Shantanu B., Soumya R. A Shared ICT Infrastructure for Indian SME Clusters / International Conference on Information and Communication Technologies and Development, 2007. – ICTD, 2007. – P. 10-15.
17. Rao S.S. Enterprise resource planning: business needs and technologies // Industrial Management & Data Systems. – 2000. – Vol.100, № 2. – P. 35-38.
18. Blackwell P., Shebab E.M., Kay J.M. An effective decision-support framework for implementing enterprise information systems within SMEs // International Journal of Production Research. – 2006. – Vol. 44, № 17. – P. 43-47.
19. 19. Kale P.T., Banwait S.S., Laroiya S.C. Performance evaluation of ERP implementation in Indian SMEs // Journal of Manufacturing Technology Management. – 2010. – Vol. 21, № 6. – P. 20-24.
20. The extent of ICT usage by SMEs in emerging economies — result from a pilot study of SMEs in Brazil, Russia, India, China and South Africa / R.Adam, C.Chuang, M.Herrington, F.Modiba, M.Zielinski // Ref. Libr. – 2009. – P. 1-5.
21. Шеер А.-В. Бизнес-процессы. Основные понятия. Теория. Методы / Пер. с англ. – М.: Весть-МетаТехнология, 2000.

ОСОБЕННОСТИ ПРОГНОЗИРОВАНИЯ ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ НА ОСНОВЕ ОПРЕДЕЛЕНИЯ ИХ ОПТИМАЛЬНЫХ БАЗОВЫХ ПАРАМЕТРОВ

Е.С. Кузнецов,
аспирант кафедры компьютерных технологий в проектировании и производстве Нижегородского государственного технического университета им. Р.Е. Алексеева

E-mail: yegor_s@rambler.ru
Адрес: г. Нижний Новгород, ул. Минина, д. 24

Рассмотрены особенности нового метода и алгоритмов прогнозирования, основанных на определении оптимальных базовых параметров процессов, оптимально дискретизированных по уровню и времени во временные ряды. Показано, что информационные системы прогнозирования, разработанные на основе этого метода, позволяют эффективно прогнозировать экономические процессы.

Ключевые слова: модель исходных данных, оптимальная дискретизация данных, оптимальные базовые параметры данных, прогнозирующие операторы, прогнозирование прогнозируемость.

1. Введение

Развитие прогностики как науки в последние десятилетия привело к созданию множества методов, процедур, приемов прогнозирования. По оценкам зарубежных и отечественных систематиков прогностики [1] насчитывается свыше ста методов прогнозирования. В связи с этим перед экономистами и специалистами других специальностей возникает задача выбора метода, который давал бы адекватные прогнозы для изучаемых систем и связанных с ними процессов.

Опыт показывает, что собственную оценку сложности реализации конкретного метода можно выполнить, если иметь чёткое математическое описание – математическую модель (ММ) конкретного метода, выраженную, например, в лингвистической форме в виде прогнозирующего оператора (ПО):

$$f(\{y_i\};\{p_j\}) \to \{y_{sf}\}, \qquad (1)$$

где $\{y_i\}, i \in [1, M]$ –исходный ряд данных длины M; $\{p_j\}, p_j \in [p_1, p_2, ..., p_n]$ – подбираемые параметры или оптимизируемые по выбранному критерию оптимальности; $\{y_{sf}\}, sf \in [1, 2, ..., L]$ – прогнозируемые выборки ряда; f – прогнозирующий оператор, характеризующий каждый из известных методов.

В практических исследованиях в качестве модели ПО, в основном, используются следующие функции f: линейная (ARMA, ARIMA) [2], квадратичная, степенная, показательная, экспоненциальная (экспоненциального сглаживания), логистическая. При этом не все процессы удаётся прогнозировать такими моделями, хотя в ряде случаев их удаётся заменять линейной комбинацией гармонических или иных функций.

В последнее время набирают популярность методы прогнозирования, ориентированные на обучение по прецедентам (относящиеся к разделу машинного обучения) или индуктивное обучение, основанные на выявлении общих закономерностей по частным эмпирическим данным. В этом случае, если линейная модель регрессии представляется необоснованной и предложить адекватную нелинейную модель $f(\{y_i\};\{p_j\}) \to \{y_{sf}\}$ также не удаётся, в качестве компромисса строится модель вида:

$$f(\{y_i\};\{p_j\}) = \sum_{j=1}^{n} \varphi_j f_j(\{y_i\}), \qquad (2)$$

где φ_j – некоторые преобразования исходных признаков, в общем случае нелинейные. Задача состоит в том, чтобы подобрать неизвестные одномерные преобразования φ_j, при которых достигается минимум квадратичного функционала ошибок [3]. Поэтому чаще используются такие модели ПО, сложность идентификации параметров которых не сильно зависит от вида их нелинейностей.

Одной из главных задач в моделях прогнозирования является нахождение порядка n ПО, который определяется преимущественно числом её параметров [2] $\{p_j\}, p_j \in [p_1, p_2, ..., p_n]$ и определяет точность прогноза.

Заметим, что на практике иногда требуется прогнозировать непрерывные (аналоговые) процессы конечной длительности T. Во многих таких случаях частота дискретизации исходного сигнала f_d при дальнейших расчётах не меняется, что не соответствует строго теореме В.А. Котельникова [4-5]. Чаще она выбирается из соображений практики кратной секунде, минуте, часу и т.д., что свидетельствует о возможной потере необходимой информации в исходных данных уже на стадии дискретизации сигнала. Выбор f_d по Котельникову осложняется еще и тем, что не всегда просто указать верхнюю частоту f_d. В нашем случае, как увидим далее, модель ПО настраивается на оптимальную $f_d = 1/\Delta t$, учитывающую минимум потерь исходной информации.

2. Метод прогнозирования на основе оптимальных базовых параметров

В известных наиболее популярных методах определение (подбор) порядка модели и значений других параметров ПО, влияющих на точность прогнозирования модели, осуществляется независимыми между собой способами (только для n или только для Δt).

Поэтому нами выбрана такая модель ПО, при которой её параметры были бы согласованы между собой и находились (идентифицировались) по единому критерию специально введенных оптимальных базовых параметров (ОБП) [6-8].

Модель основана на предварительной дискретизации исходных векторных процессов продолжительности T одновременно по времени, с периодом $\Delta t = T/M$, и по значению в q-уровневые временные ряды исходной длины M

$$y = \left\{ \begin{matrix} y_0^1, ..., y_{n-1}^1, ..., y_{l-n+1}^1, ..., y_l^1, ..., y_{k-n+1}^1, ..., y_k^1, ..., y_{M-1}^1 \\ ..... \\ y_0^r, ..., y_{n-1}^r, ..., y_{l-n+1}^r, ..., y_l^r, ..., y_{k-n+1}^r, ..., y_k^r, ..., y_{M-1}^r \end{matrix} \right\}, \qquad (3)$$

имеющие ограничения по диапазону изменения параметров $MM - y_i$ и продолжительности:

$$-\infty < y_{min}^v \le y_i^v(k) \le y_{max}^{vj} < \infty, k \in [0, M-1], v \in [1, r], \qquad (4)$$

где r-количество компонент векторного процесса.

Определение значений ОБП заключается в нахождении такой «тройки» $\{\Delta t_{opt}, q_{opt}, n_{opt}\}$ или «пары» БП $\{q_{opt}, n_{opt}\}$ (если есть доверие к дискретизации исходных данных и $\Delta t_{opt} = \Delta t$), при которой энтропия (3) по БП временного ряда $\{y^v_k\}$ будет минимальна:

$$E_{y;min} = min(log_2 N(\Delta t, q, n; y_k^v)) = \\ = log_2 N_{y,min} = n_{y;opt} \cdot log_2 q_{y;opt}, \qquad (5)$$

где $q \in [q_{min}, q_{max}], n \in [n_{min}, n_{max}] N_{y_i;min} = q_{y;opt}^{n_{y;opt}}$ (6)

При этом ПО на основе локализованных ОБП $\{\Delta t_{opt}, q_{opt}, n_{opt}\}$ представим в виде:

$$f(\{y_i^r\}, \{\Delta t_{opt}, q_{opt}, n_{opt}\}) \longrightarrow \{y_{sf}^r\} \quad (7)$$

Если изначально дан дискретный процесс с фиксированным шагом Δt, то для прогнозирования определяются только q_{opt} и n_{opt}. Если же изначально имеется выборка из непрерывного процесса с неоптимальным и требующим уточнения шагом Δt, то по имеющимся M отсчетам сигнал методом сплайнов восстанавливается в «непрерывный». Затем образуются новые наборы выборок процессов с разными шагами дискретизации в интервале

$$T/M_{max} \le \Delta t \le T/M_{min},$$

где T – длительность исходного процесса.

Тот набор из упомянутых наборов БП будет иметь число компонент $M_{opt} = T/\Delta t_{opt}$, на котором параметры q, n дают среди всех наборов выборок процессов наименьшее значение энтропии (3). Так определяется вся тройка ОБП $\{\Delta t_{opt}, q_{opt}, n_{opt}\}$.

3. Построение прогнозирующего оператора при $k < M$

Предлагаемый метод прогнозирования, используя только динамические параметры исходных данных, позволяет построить по ряду (2) ПО для любого $k = n, n+1, ..., M-1$ в виде q-значной логической функции с ОБП – $\{\Delta t, q, n\}$.

$$(y^1_{k+1},..., y^r_{k+1})^T = \quad (8)$$
$$f((y^1_{k-n+1},...,y^r_{k-n+1})^T, (y^1_{k-n+2},...,y^r_{k-n+2})^T, ..., (y^l_k,...,y^r_k)^T) \equiv f_k,$$

или эквивалентной таблице истинности (ТИ) [8].

Строки ТИ ПО строятся по всем идущим подряд n членам ряда отсчетов и следующего за ними отсчета, в качестве прогнозируемого ими значения. Величина n, является порядком математической модели ПО данных (2).

Порядок прогнозирующего оператора может определяться как минимальное n, при котором по одной и той же n-последовательности отсчетов прогнозируются одинаковые значения.

Когда числа n и q определяют ОБП, то ТИ оказывается непротиворечивой и ПО может допускать минимизацию в базисах функций q-значной алгебры логики.

4. Модификация прогнозирующего оператора прогнозирование при $k \ge M$

Для прогнозирования неизвестных выборок вне заданного ряда (3) при $k \ge M$ требуется модификации алгоритма (8).

Прогнозирование при $k \ge M$ заключается в пошаговом построении продолжения ТИ с $M-n+1$-ой по $M+sf$-ю строку, где $M+sf = 1,2,...,L$, а L – номер максимального шага прогнозирования или т.н. «прогнозного горизонта» для пополнения выборок данных (8), имеющихся в исходной ТИ.

Для определения y_{M+sf} используется последовательное сравнение $M-n+sf$-ой n-последовательности со всеми n-последовательностями, уже имеющимися в исходной таблице, рассматриваемыми как опорные («эталонные») по критерию «минимума расстояния» между ними [8].

$$y_{M+sf} = arg \underset{y_k \in [y_{n+1}, \overset{min}{y_{n+2}}, ... y_{M+sf-1}]}{} d_k^{w_j^i}(s_f) \quad (9)$$

где

$$d_k^i(s_f) = \sum_{j\in[1,n]} \sum_{j\in[1,r]} w_j^i \mid y^l_{k+1-j} - y^l_{M+s_f-j} \mid^p, \quad (10)$$
$$k = n+1, n+2, ...M+sf-1$$

Таблица 1.

Таблица истинности

Аргументы прогнозирующего оператора				Прогноз
$(y^1_0,..., y^r_0)^T$	$(y^1_1,..., y^r_1)^T$		$(y^1_{n-1},..., y^r_{n-1})^T$	$(y^1_n,..., y^r_n)^T$
$(y^1_1,..., y^r_1)^T$	$(y^1_2,..., y^r_2)^T$		$(y^1_n,..., y^r_n)^T$	$(y^1_{n+1},..., y^r_{n+1})^T$
…				…
$(y^1_{k-n+1},..., y^r_{k-n+1})^T$	$(y^1_{k-n+2},..., y^r_{k-n+2})^T$		$(y^1_k,..., y^r_k)^T$	$(y^1_{k+1},..., y^r_{k+1})^T$
…				…
$(y^1_{M-n-1},..., y^r_{M-n-1})^T$	$(y^1_{M-n},..., y^r_{M-n})^T$	...	$(y^1_{M-2},..., y^r_{M-2})^T$	$(y^1_{M-1},..., y^r_{M-1})^T$
$(y^1_{M-n},..., y^r_{M-n})^T$	$(y^1_{M-n+1},..., y^r_{M-n+1})^T$	...	$(y^1_{M-1},..., y^r_{M-1})^T$	-

В критерии близости (9) используются весовые функции индекса $j = 1, 2, ..., n$ с типом веса $i = \{c, l, e, h\}$:

$$w^{(c)}_j = 1,\ w^{(l)}_j = 1 + (1 - j)/n,\ w^{(e)}_j = e^{-j},\ w^{(h)}_j = j^{-1} \quad (11)$$

Возможна модификация формулы (7) с учетом классов эквивалентности, каждому из которых соответствует одинаковое прогнозируемое значение исходного ряда [9].

5. Связь оптимальных базовых параметров с предсказуемостью и восстанавливаемостью

Из введенных обозначений и алгоритма прогнозирования (8) для векторного временного ряда (3), удовлетворяющего условиям (4), существует ММ ПО, который по n_{opt} начальным, следующим подряд с шагом $\Delta t_{opt} \equiv T / M_{opt}$ q_{opt}-значным выборкам исходного ряда (3) позволяет вычислить все оставшиеся $M - n_{opt}$ выборок.

Исходя из этого, характеризуем «голографическое» свойство векторного *M, q, n*-процесса – «восстанавливаемость» (с точностью $1/q_{opt}$) с помощью прогнозирующего оператора (7) по n_{opt} известным, следующим подряд векторам-столбцам, последующих вектор-столбцов с любыми изменениями в них символов текста. И это позволяет естественным образом ввести понятие «предсказуемости (прогнозируемости)» (*Pr*) поведения векторного ПО, как отношения энтропии предсказываемой части векторных процессов к энтропии базовой (начальной части)

$$P_r = \left(\frac{log_2 q_{opt}^{M - n_{opt}}}{log_2 q_{opt}^{n_{opt}}} \right) + 1 = \frac{(M - n_{opt})}{n_{opt} + 1} = \frac{M}{n_{opt}} \le M, \quad (12)$$

из которого следуют следующие выводы:

Предсказуемость (12) векторного ряда (3) обладает следующими свойствами:

а) зависит явно только от длины M и порядка n_{opt} ПО (8);

б) не может превышать длины M векторного ряда;

в) возрастает с ростом M, если текст *M, q, n* при своем продолжении сохраняет ОБП, то есть при $\Delta M > 0$ остается стационарным «$M + \Delta M$, q_{opt}, n_{opt} - рядом»;

г) при $n_{opt} = 1$ предсказуемость *Pr* текста максимальна и равна M

Доказательство перечисленных свойств следует из возможности представления функции *f* ПО (5) в форме ТИ по тексту полной длины M [6].

6. Отличительные особенности предложенного метода

1. Предложен и проверен экспериментально на информационной системе прогнозирования метод, основанный на предварительной оптимальной дискретизизации («оптимальном загрублении») исходных данных во временные ряды.

2. Используемые ОБП $\{\Delta t, q, n\}$ находятся одновременно по единому энтропийному критерию, а не по различным известным ранее критериям оптимальности каждого из параметров, например, H.Akaike (только для n)[10] и В.А.Котельникова (только для $\{\Delta t\}$) [4].

3. Применяемый ПО един как для скалярных, так и для векторных процессов.

4. Метод позволяет уточнить частоту дискретизации исходных данных, если она была выбрана не оптимально.

5. Используемый ПО определяется в общем случае в виде нелинейной *q*-значной логической функции для любого прогнозируемого процесса.

6. Исходя из п.5 данный метод может быть применим в случае неизвестной функциональной зависимости прогнозируемого значения от предыдущих.

7. Прогнозирование экономических процессов

Экономические ряды имеют четкую восходящую или нисходящую тенденцию (тренд). Очевидно, что при этом количество уровней квантования будет не постоянным (не выполняется условие (4)). В данном случае целесообразно прогнозировать не сами значения ряда, а отклонения относительно предыдущего значения, т.е. разности 1-го порядка:

$$\overline{y^v_k} = y^v_{k+1} - y^v_k, \overline{k} \in [1, M\text{-}1], v \in [1, r] \quad (13)$$

Используя предположение, что количество уровней квантования для ряда (13) будет постоянным, после оптимальной дискретизации используя формулы (4)-(6) можно спрогнозировать ряд разностей 1-го порядка. Поскольку в конечном итоге важны не относительные изменения, а абсолютные значения, то используется обратное преобразование:

$$y^v_k = y^v_0 + \sum_{i=1}^{j-1} \overline{y^v_i},\ j \in [1, M\text{-}1\text{+}L], v \in [1, r] \quad (14)$$

При прогнозировании рыночной стоимости акции также необходимо учесть объем сделок, который был совершен в этот же временной про-

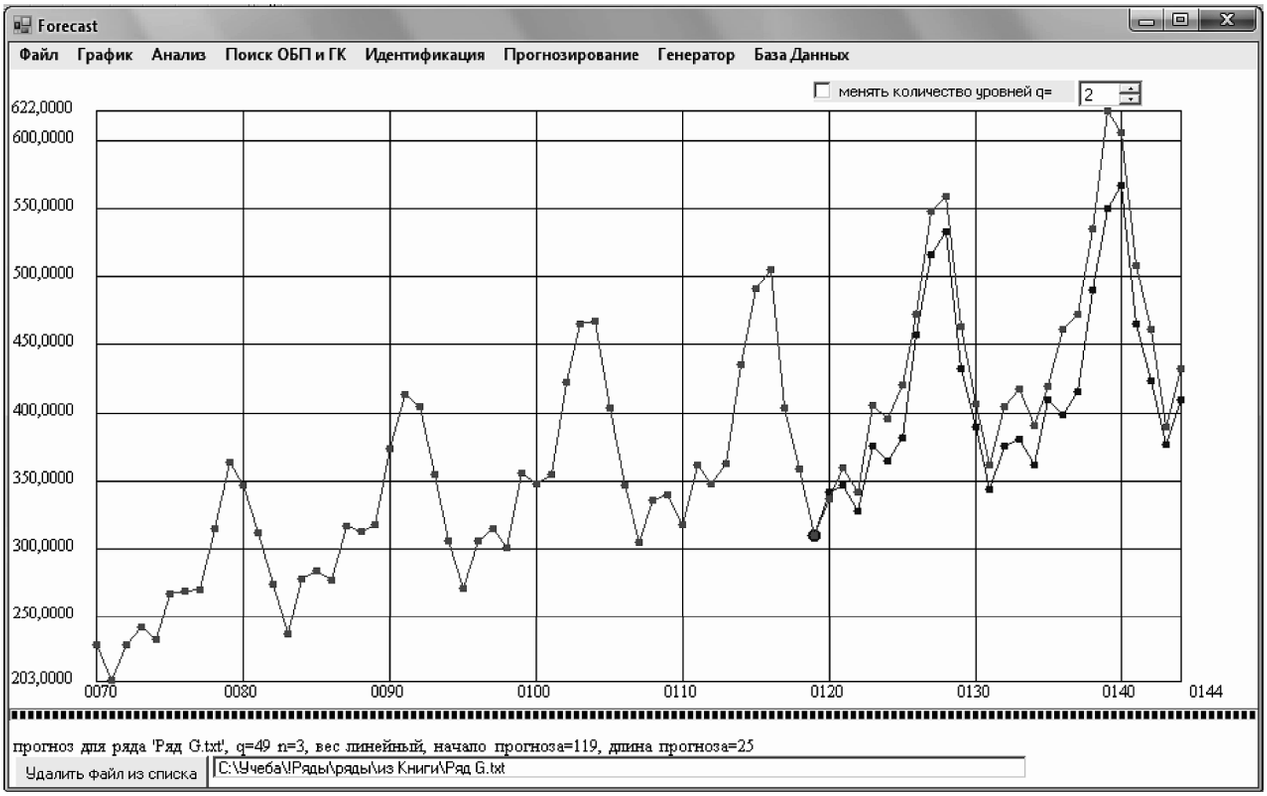


Рис. 1. Прогноз международных перевозок

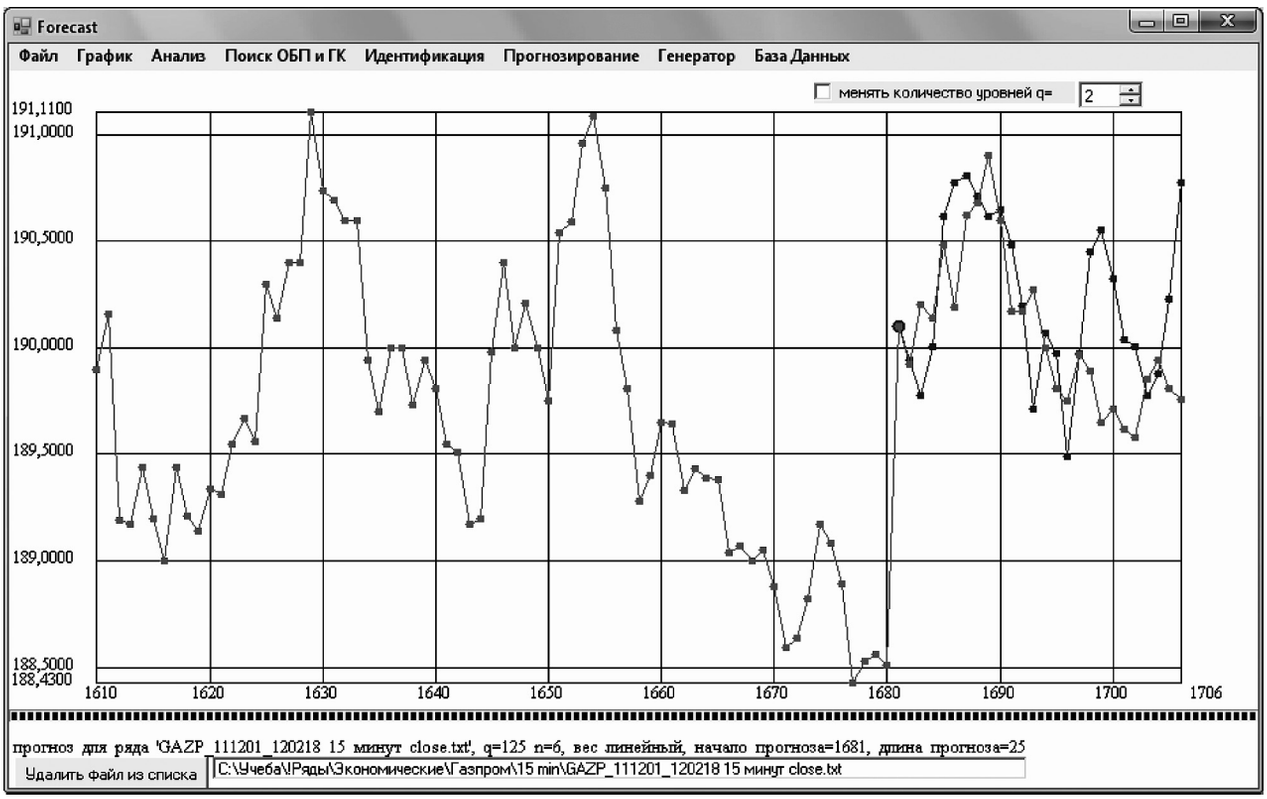


Рис. 2. Прогноз цены акции ОАО «Газпром»

межуток. Для этого прогнозный ряд можно представить в виде векторного процесса, состоящего из двух компонент: цена при окончании временного промежутка (Δt) и количество сделок, совершенных за данный временной промежуток времени.

8. Методика экспериментов. Оценка точности прогнозирования

Механизм верификации [1] прогноза позволяет выполнить оценку достоверности и точности прогноза по участку эталонного временного ряда. В качестве оценки точности прогноза используется формула

$$Er(L) = \sqrt{\frac{1}{L\text{-}1} \cdot \sum_{sf=1}^{L} \left| y_{M+sf\text{-}1} - y'_{sf} \right|^2}, \qquad (15)$$

где y' – исходный (реальный) процесс; s – шаг прогноза; L – длина прогноза. Сравниваются исходный процесс $(y'(s), s \in [M\text{-}L, M\text{-}1])$ и спрогнозированный процесс $(y(s), s \in [M\text{-}L, M\text{-}1])$ на основе исходного процесса с отброшенным концом $(y'(s), s \in [0, M\text{-}L\text{-}1])$. Поиск оптимальных базовых параметров и построение прогнозирующего оператора (12) осуществлялись на участке $k \in [0, M\text{-}L\text{-}1]$.

В качестве экспериментальных данных были выбраны 2 временных ряда: выборка цен акции ОАО «Газпром» за период *01.12.2011-18.02.2012* и «классический» ряд *Ряд G – Series G* [2], представляющий месячные международные авиаперевозки (в тысячах) в течение 12 лет с 1949 по 1960. Отличительной особенностью международных перевозок является наличие ярко выраженной сезонной составляющей. Ниже приведена таблица сравнения точности прогноза предложенного метода с двумя наиболее популярными методами – экспоненциального сглаживания и ARIMA.

Таблица 2.

Результаты прогнозирования международных авиаперевозок

$q_{opt} = 49, n_{opt} = 3; r = 1, M = 119; E = 16,84; Pr = 40$						
	Er(L)					
L	1	5	10	15	20	25
ОБП	3,65	19	24	26	34	34
Экспоненциальное сглаживание	17,5	14	24	23	30	35
ARIMA	7	23	33	37	53	57

Исходя из результатов, можно сделать вывод, что предложенный метод при прогнозировании временных рядов с ярко выраженной сезонной компонентой уступает методу прогнозирования, основанному на экспоненциальном сглаживании.

На *рис. 1* изображен результат прогнозирования методом на основе оптимальных базовых параметров, выполненный с помощью программы Forecast [11-12]. Синей линей изображен прогнозный ряд, красной – исходный ряд.

Для сравнения цен акции ОАО «Газпром» был спрогнозирован с учетом объемов сделок и без учета объема сделок. Результаты экспериментов (оценка точности прогноза рыночной стоимости акции и результаты предобработки) скалярного процесса, выполненных в программе Forecast [11-12], представлен в виде *табл. 3.*

Таблица 3.

Результаты прогнозирования скалярного процесса

$q_{opt} = 125, n_{opt} = 5, \Delta t_{opt} = 15$ *мин*; $r = 1, M = 1681; E = 34,83; Pr = 336$						
	Er(L)					
L	1	5	10	15	20	25
ОБП	0,73	0,48	3,6	4,4	4,3	4,3
Экспоненциальное сглаживание	0,96	1,6	2	2	2	2,1
ARIMA	0,48	0,7	0,6	0,75	0,95	1,05

На *рис. 2* изображен результат прогнозирования цены акции ОАО «Газпром» методом на основе оптимальных базовых параметров, выполненный с помощью программы Forecast [11-12]. Пунктирной линей изображен прогнозный ряд, сплошной – исходный ряд.

Результаты экспериментов (оценка точности прогноза рыночной стоимости акции и результаты предобработки) векторного процесса (с учетом объема сделок), выполненных в программе Forecast, представлен в виде *табл. 4.*

Таблица 4.

Результаты прогнозирования векторного процесса

$q_{opt} = 106, n_{opt} = 3, \Delta t_{opt} = 15$ *мин*; $r = 2, M = 1681; E = 20,18; Pr = 563$						
L	1	5	10	15	20	25
ОБП	0,27	0,46	0,55	0,5	0,7	1,1

9. Заключение

Таким образом, в статье предложен и проверен экспериментально новый метод прогнозирования экономических временных рядов, основанный на предварительной оптимальной дискретизизации («оптимальном загрублении») исходных данных во временные ряды. Данный метод зарекомендовал себя в задачах прогнозирования рядов, в которых априорная информация не позволяет сделать вывод о функциональной зависимости прогнозируемого значения от предыдущих. В тех случаях, когда имеется априорная информация, например, информация о наличии сезонной компоненты, следует использовать «классические», наиболее популярные методы прогнозирования, позволяющие учесть эту информацию.

Метод, основанный на применении ОБП, позволяет при прогнозировании рыночных курсов акций не только учесть объем сделок, но и одновременно спрогнозировать его.

Метод прогнозирования позволяет одновременно осуществлять и его верификацию (оценку достоверности и точности) по участку эталонного временного ряда.

Экспериментально было установлено, что при использовании информации об объеме сделок, возрастает прогнозируемость процесса и увеличивается в несколько раз точность прогноза, что в целом превышает точность прогнозирования ряда, выполненное «классическими» методами.

Предложенный алгоритм может применяться при решении практических задач бизнес-информатики, сводимых к задаче оптимальной дискретизации по уровню и по времени и к задаче прогнозирования временных рядов. ■

Литература

1. Прогностика. Технология. / Под ред. В.И.Сифорова. – М.: Наука, 1990.
2. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. – М.: Мир, 1974.
3. Вапник В. Н. Восстановление зависимостей по эмпирическим данным. – М.: Наука, 1979.
4. Конева Е.С. Выбор моделей для реальный временных рядов // Автоматика и телемеханика. – №6. – 1988. – С. 3-18.
5. Романюк Ю.А. Основы цифровой обработки сигналов. В 3-х ч. Ч.1. Свойства и преобразования дискретных сигналов: Учебное пособие. – М.: МФТИ, 2005.
6. Кирьянов К.Г. Выбор оптимальных базовых параметров источников экспериментальных данных при их идентификации // Идентификация систем и задачи управления: тр. 3-й междунар. конф.– М.: ИПУ РАН, 2004. – С. 187-208.
7. Кирьянов К.Г. Идентификация динамических и информационных характеристик многоканальных систем на основе оптимальной дискретизации данных // Идентификация систем и задачи управления: тр. 9-й междунар. конф.– М.: ИПУ РАН, 2012. – С. 252-265.
8. Кирьянов К.Г., Кузнецов Е.С. Особенности прогнозирования дискретных и аналоговых векторных процессов на основе идентификации их базовых параметров // Труды 14-й Научной конференции по радиофизике.– Н.Новгород: ННГУ, 2010. – С. 278-279.
9. Кирьянов К.Г., Кузнецов Е.С. Модификация метода прогнозирования аналоговых и дискретных процессов в программе forecast 2 // Труды 12-й Научной конференции по радиофизике. – Н.Новгород: ННГУ, 2008. – С. 271-273.
10. Akaike H. A New Look at the Statistical Model Identification. // IEEE Transactions on Automatic Control. – 1974. – Vol. 19. – P. 716-723.
11. Кирьянов К.Г., Кузнецов Е.С. Информационная система прогнозирования векторных временных рядов // Информационные системы и технологии (ИСТ-2010): тез. докл. междунар. науч.-техн. конф.– Н.Новгород: НГТУ, 2010. – С.158-159.
12. Свидетельство о государственной регистрации программы для ЭВМ №2008611799, 09.04.2008

R/S АНАЛИЗ НА ФОНДОВОМ РЫНКЕ

А.В. Зиненко,
кандидат технических наук, доцент кафедры финансов и кредита
Сибирского государственного аэрокосмического университета
им. академика М.Ф.Решетнева

E-mail: anna-z@mail.ru
Адрес: г. Красноярск, просп. имени газеты «Красноярский рабочий», д. 31

В работе описывается алгоритм относительно нового статистического метода – R/S анализа, описанного Гарольдом Херстом. Данный метод анализа временных рядов позволяет определить, является ли временной ряд случайным или персистентным, то есть обладающим долговременной памятью. К временным рядам биржевых котировок применяется алгоритм R/S анализа и делается вывод об их персистентном характере.

Ключевые слова: показатель Херста, самоподобие, степенные законы.

1. Введение

Исследование динамики финансовых рынков является важной частью теории финансовых инвестиций. Сложившиеся в шестидесятых и семидесятых годах прошлого века методы анализа финансовых рынков до сих пор преподаются в экономических вузах, хотя еще с начала 90-х годов прошлого века показано, что они действуют только в периоды стабильного состояния рынка. Речь идет о таких методах анализа, как модель оптимального инвестиционного портфеля Гарри Марковитца, модель CAPM Вильяма Шарпа и модель ценообразования опционов Блека – Шоулза. Эти модели являются базой современной инвестиционной теории. Они основаны на предположении Луи Башелье, сделанном в еще в 1900-м году, о том, что динамика финансовых рынков подчиняется закону нормального или гауссовского распределения.

Вероятностные модели рынка успешно работали вплоть до обвала американского рынка в 1987 году (индекс Доу-Джонса за один день упал на 29,2%). Аналитики уверяли, что это редчайшее событие, согласно кривой Гаусса, в следующий раз нечто подобное может наступить через несколько миллионов лет. Но затем последовали кризисы 1992, 1995, 1998, 2008 года. Это противоречило всем теоретическим постулатам вероятностного подхода. Стало ясно – нужен совершенно другой взгляд на сущность финансового рынка. Ответ дал американский математик, основоположник теории фракталов Бенуа Мандельброт. Он обнаружил явление, которое назвал «толстые хвосты». Его суть заключается в том, что кривая распределения вероятностей изменения рыночных котировок выглядит не так, как гауссовская колоколообразная кривая. А именно, по обеим сторонам «колокола», вероятности гораз-

до выше, чем при нормальном распределении. Эти результаты Мандельброт и его соратники получили, исследуя динамику цен на хлопок, котировки акций крупных американских компаний и фондовых индексов[1]. Они анализировали статистические данные по биржевым котировкам более чем за сто лет, рассматривая различные таймфрэймы. К середине шестидесятых не было таких резких коллапсов, как начиная с конца 80-х, но, тем не менее, соответствие гауссовскому распределению не было обнаружено. Например, согласно анализу Фама промышленного индекса Доу-Джонса, колебания, превышающие пять стандартных отклонений, случались в пять тысяч раз чаще, чем предсказывала нормальная кривая. Кроме того, полученные Мандельбротом кривые распределения вероятностей имели более высокие пики. Мандельброт предположил, что динамика фондовых рынков не является случайной, а подчиняется некому степенному закону.

2. Степенной подход к анализу временных рядов

Степенные законы активно применяются в исследованиях природных социальных процессов. Одним из пионеров исследования степенных законов стал британский гидролог Гарольд Эдвин Херст. Начиная с 1906 года, он работал над новым статистическим методом анализа временных рядов.

Перед Херстом стояла задача изучения разливов Нила с целью строительства водохранилища такого объема, чтобы в засушливые годы население не нуждалась в воде. Для этого ему следовало изучить динамику приливов и отливов Нила за долгие годы и обнаружить в ней какую-либо периодичность. Расход воды Нила изменялся в очень широких пределах, при этом могло быть несколько засушливых и несколько дождливых периодов подряд. Существующие на тот момент методы статистического анализа основывались на том, что временной ряд с большим количеством периодом является случайным и подчиняется гауссовскому закону нормального распределения. Исследуемые события должны быть независимыми друг от друга и иметь одинаковую вероятность. Если временной ряд не подчиняется нормальному закону, то существуют различные методы внесения в него корректировок, чтобы с некоторыми оговорками назвать его случайным и идентично распределенным.

Проблема, которую рассматривал Херст, не допускала такой возможности. С целью выявления какой-либо закономерности он исследовал данные о разливах Нила за 847 лет (древние египтяне вели такие записи).

Следует отметить, что проблемой периодичности разливов Нила озадачивался еще Геродот в 450 году до н.э. Согласно Геродоту, ни жрецы, ни ученые не могли объяснить эту периодичность.

Херстом были обнаружены следующие циклы: за разливами выше среднего в следующем периоде следовали еще большие разливы. Когда направление менялось и наступал засушливый период, за ним следовали еще более засушливые. Мандельброт назвал подобное явление «эффект Иосифа», вспомнив библейскую притчу о семи благодатных и о семи засушливых годах. Согласно Мандельброту, данный эффект имеет место на финансовых рынках.

Помимо статистики разливов Нила Херст изучил еще 51 природное явление. Например, илистые отложения на дне озера, годовые кольца на деревьях. И во всех этих явлениях он исследовал размах, то есть разницу между максимальным и минимальным значением.

Херст вспомнил формулу Эйнштейна, широко используемую при анализе последовательности случайных событий: размах вариации при большом количестве испытаний равен корню из количества испытаний [2].

$$R = n^{0,5}. \qquad (1)$$

Практические исследования Херста природных явлений показали, что размах расширяется несколько быстрее, чем это следует из формулы, приведенной выше. Согласно Херсту показатель степени составил примерно 0,75.

Гидрологи активно использовали формулу Херста. Мандельброт, создатель фрактальной геометрии, обнаружил в формуле Херста законы, которые связаны с фрактальностью временных рядов. Он назвал их «долговременная память»[1].

Применение R/S анализа к фондовым рынкам также является инициативой Бенуа Мандельброта. Согласно его исследованиям ставки по дневным банковским кредитам имеют показатель Херста 0,7, тогда как цены на пшеницу и котировки облигаций являются случайными, то есть имеют H=0,5. Одновременно с Мандельбротом подобные исследования проводил Эдгар Петерс. Им обнаружены

высокие значения показателя Херста для голубых фишек (IBM, Apple, Xerox) и близкие к 0,5 значения для второго эшелона. В своей работе «Фрактальный анализ финансовых рынков» Петерс проводил R/S анализ динамики валютных пар и обнаруживал высокое значение показателя H [2].

Мандельбротом также были исследованы макроэкономические показатели США за несколько сотен лет. Он обнаружил интересный факт: сильная зависимость между предыдущими и последующими значениями со временем уменьшается, но весьма медленно. Аналогичным образом уменьшается зависимость между последовательными природными явлениями. Именно это Мандельброт назвал «долговременной памятью».

Для явлений, изучаемых Херстом, показатель Херста H (изначально он обозначался k, но Мандельброт назвал его в честь автора) не превышал единицы, но был не меньше 0,5. При $H = 0{,}5$ выборка является случайной (таков H например для Броуновского движения). Если показатель Херста колеблется от 0,5 до 1, то процесс характеризуется долговременной памятью. Это означает, что последующие показатели сильно зависят от прошедших. К этому близка характерная для хаоса чувствительность к начальным условиям. Показатель Херста, колеблющийся от 0 до 0,5, означает антиперсистентный процесс. То есть система меняется быстрее, чем случайная. Это означает частые, но небольшие изменения. Можно сказать, что чем выше показатель Херста, тем меньше на временном ряду «зазубрин». Можно предположить, что при показателе Херста равном единице мы будем иметь гладкую прямую (восходящий или нисходящий тренд).

В природе встречаются системы первого типа, с показателем Херста от 0,5 до 1, которые называются персистентными [3] (согласно Мандельброту – инерционными). Персистентные процессы характеризуются независимостью от временного масштаба. Дневные изменения зависят от прошедших дневных изменений в той же степени, что и, например, недельные изменения зависят от прошедших недельных изменений. Очевидно, что это свойство персистентных процессов является фрактальным.

Показатель Херста рассчитывается с помощью R/S анализа, где R – нормированный размах вариации, а S – стандартное отклонение. Формула для расчета выглядит следующим образом:

$$R/S = c \cdot n^H, \qquad (2)$$

где c – константа, n – количество элементов выборки.

Эмпирически Херст рассчитал, что константа c равна приблизительно 0,5. Но это не более чем допущение. Ученые, использующие R/S анализ, берут разные значения c. Вообще, как это будет показано ниже, константу можно вывести как свободный член уравнения регрессии.

Отметим важную особенность: – размах вариации изменяет свой масштаб в зависимости от числа наблюдений по степенному закону. Изменение масштаба по степенному закону – это признак самоподобия и, как следствие, фрактальности временного ряда. Бенуа Мандельброт занялся исследованием степенных законов задолго до открытия и описания фракталов. Он связал одно с другим и по нашему мнению степенные законы стали для Мандельброта одной из ступеней, которая подвела его к фрактальной геометрии. К сожалению, фрактальные модели финансовых рынков не были завершены им при жизни (Мандельброт скончался в 2010 году).

3. Алгоритм R/S анализа

Алгоритм осуществления R/S анализа следующий. Пусть мы имеем выборку из n+1 элементов (n : 2). Данная выборка представляет собой временной ряд, желательно, чтобы количество значений было достаточно высоко (например, Эдгар Петерс при исследовании фондового рынка брал около 5 000 значений) [1]. Преобразуем выборку со помощью логарифмических соотношений, и получим временной ряд со значениями уровней длиной n :

$$log \frac{U_i}{U_{i-1}} \qquad (3)$$

Следующий шаг – это найти наименьший собственный делитель n, не меньший 10. Обозначим его m. Разделим n на $k=n/m$ групп. Элементы в каждой группе обозначим t_i. Очевидно, что количество элементов в каждой группе составляет m. Найдем средние значения в каждой группе

$$\overline{t_k} = \frac{1}{m}\sum_{i=1}^{m} t_i,\ \frac{1}{m}\sum_{i=m+1}^{2m} t_i \ldots \frac{1}{m}\sum_{i=(k-1)m+1}^{n} t_i \qquad (4)$$

и накопленные отклонения от среднего X_i:

$$X_1 = t_1 - \frac{1}{m}\sum_{i=1}^{m} t_i,\ X_2 = (t_2 - \frac{1}{m}\sum_{i=1}^{m} t_i) + X_1 \ldots X_m = \\ = (t_m - \frac{1}{m}\sum_{i=1}^{m} t_i) + X_{m\text{-}1} \qquad (5)$$

Нормированный размах по каждой группе составит

$$R_k - max(X_i) - min(X_i).$$

Также по каждой группе рассчитываем стандартное отклонение S_k по стандартной формуле

$$S_k = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(t_i - \overline{t_k})^2}. \quad (6)$$

Показатель R/S по каждой группе рассчитываем как R_k/S_k. Затем находим средний размах вариации

$$\overline{R/S_j} = \frac{1}{k}\sum_{i=1}^{k} R/S_i \quad (7)$$

Индекс j в данном случае означает, что мы получили средний размах вариации на j-м шаге, который соответствует j-му собственному делителю. Мы повторяем описанную выше процедуру, используя в качестве m все возможные собственные делители. На последнем шаге m будет равно $n/2$.

Таким образом, мы получаем выборку

$$\overline{R/S_j},\ j = 1, \overline{\frac{n}{2}},$$

количество элементов в выборке соответствует количеству собственных делителей n. Теперь можно построить уравнение линейной регрессии, в котором зависимой переменной выступает логарифм показателя R/S, а факторным признаком – логарифм количества элементов в j-й группе k:

$$\log R/S - \log c + H \cdot \log k \quad (8)$$

Параметры этого уравнения линейной регрессии легко находятся методом наименьших квадратов. Можно использовать соответствующую надстройку Microsoft Excel.

Для проверки значимости уравнения регрессии можно использовать статистику Фишера (в Excel рассчитывается с помощью надстройки), которая сравнивается с табличным значением для соответствующего уровня значимости, количества факторов (в нашем случае – один) и количества элементов в выборке. Также следует проверить значимость параметра H с помощью статистики Стьюдента, которая также сравнивается с табличным значением.

В работе «Хаос и порядок на рынках капитала»[3] Петерс предлагает еще один вариант проверки состоятельности расчета показателя Херста. Если мы получили высокое значение данного показателя, то проверить, существует ли долговременная память на самом деле можно, перемешав исходные данные. Действительно, если имеет место эффект долговременной памяти, то порядок данных важен и при перемешивании показатель Херста значительно уменьшится. Если показатель останется неизменным можно говорить о несостоятельности выборки.

4. Реализация алгоритма R/S анализа на примере российских и зарубежных фондовых индексов

Программная реализация алгоритма нахождения показателя Херста предложена Эдгаром Петерсом [3]. В книге «Фрактальный анализ финансовых рынков» им предложен текст программы на языке Gauss. Следует отметить, что сложность в программной реализации вызывает только нахождение нормированного размаха. При найденном нормированном размахе не составит труда воспользоваться надстройкой Excel «Линейная регрессия». В данной надстройке также осуществляется проверка значимости уравнения и параметров. При нахождении показателя R/S сложность вызывает нахождение статистических показателей по k группам. В рамках возможностей Excel реализация описанного алгоритма будет очень трудоемкой. При значительных объемах выборки применять стандартные методы Excel не представляется целесообразным. Но нами для каждого шага алгоритма были написаны макросы на языке VBA, что значительно облегчило задачу и позволило произвести эксперимент по нахождению показателя Херста для российских и зарубежных фондовых индексов.

И Мандельброт и Петерс анализировали подчиненность финансового рынка степенным законам достаточно давно. Мандельброт занимался вопросами наличия степенных законов в разных таймфрэймах рыночного движения еще с начала 60-х, в конце 60-х он начал проводить эксперименты с расчетом показателя Херста. Петерс, брокер по специальности, сотрудничал с Мандельбротом в его исследованиях. Последняя – его монография «Фрактальный анализ финансовых рынков», касающаяся R/S анализа финансовых рынков – была опубликована в 2004 году. В ней Петерс в основном анализировал валютные пары. В предыдущей книге «Хаос и порядок на рынке капитала» Петерсом были рассмотрены котировки акций и биржевых индексов и получены значения показателя Херста от 0,7. При этом Петерс тестирует систему на адекватность путем смешивания исходных данных и подтверждает наличие долговременной памяти.

Также расчет показателя Херста встречается в статьях русского трейдера немецкого происхождения Эрика Неймана. В своей статье «Расчет по-

казателя Херста в целях выявления трендовости (персистентности) финансовых рынков», опубликованной на сайте http://capital-times.com.ua, он, ссылаясь на Мандельброта, утверждает, что фрактальная размерность является обратной величиной к показателю Херста H. Целью данной работы Неймана является показать трендовый, то есть персистентный характер динамики рыночных котировок и опровергнуть гипотезу о ее случайном, гауссовском характере [4]. Очевидно, что для этого достаточно посчитать показатель Херста и убедиться, что он превышает 0,5. При этом Нейман использует несколько упрощенный алгоритм расчета показателя Херста. Во-первых, в качестве исходных данных он берет сами котировки, а не логарифмические преобразования по формуле (3). Логарифмические преобразования необходимы, чтобы привести разные выборки к единому виду (цены одной акции могут измеряться в долларах, а другой – в тысячах долларов). Во-вторых, в расчетах по некоторым валютным парам Нейман берет маленькую выборку >100 элементов. Петерс же брал объем исходной выборки не меньше 5000. Это целесообразно, так как регрессионное уравнение мы строим по количеству собственных делителей.

Мы решили повторить эксперименты Мандельброта, Петерса и Неймана на современных котировках биржевых индексов. В наших расчетах мы четко придерживались приведенного выше алгоритма. В качестве программного инструмента, как уже было сказано, использовались Microsoft Excel и Visual Basic for Applications. В качестве исходных данных были взяты котировки трех мировых индексов: российского индекса ММВБ и индексы двух экономически развитых, но при этом совершенно разных по принципам экономической политики стран: США и Китая. Это индексы Daw Jones Industrial Average и Shanghai Inc. Диапазон исходных данных составил три года: с 5 декабря 2008 г. по 5 декабря 2011 г. Следует отметить, что если бы мы взяли более ранний период, то это могло бы повлиять на чистоту эксперимента, так как в августе 2008 года произошел мировой финансовый кризис. При этом назвать расчеты абсолютно верными и не вызывающими сомнений тоже не представляется возможным, так как за этот период мы получили выборку всего из 728 элементов. Собственных делителей получилось девять – соответственно в логарифмическом регрессионном уравнении было всего девять наблюдений. Источником данных послужил сайт rbc.ru [4]. Итоги регрессионного анализа показаны *в табл. 1.*

Таблица 1.

Итоги регрессионного анализа

Индекс	H (показатель Херста)	t- статистика по H	F- статистика (значимость уравнения регрессии)
DJIA	0,6	31,05	96,7
Shanghai Inc	0,64	19,6	38,4
MICEX	0,57	4,05	164,7

Табличное значение $F_{0,05;\,1;\,7}$ равно 5,59, а $t_{0,05;\,1;\,7}$ – 2,36. Следовательно, все уравнения регрессии и параметры H значимы. Значения показателя Херста говорят о том, что динамика индексов персистентна. Все значения показателя Херста близки к 0,6.

Можно подтвердить данный вывод, построив также графики приращений котировок, а затем сравнив их с приращениями дневного курса GBP/USD, смоделированными Бенуа Мандельбротом еще в 60-х годах [1]. Приведенные ниже графики сам Мандельброт называет «антиквариатом в мире компьютерной графики», однако, они актуальны и сегодня, так как наглядно демонстрируют разницу между персистентной, случайной и антиперсистентной выборкой. Мандельброт смоделировал движение курса британского фунта по отношению к американскому доллару за один год. По оси *X* он отложил временной период в днях, а по оси *Y* дневное приращение курса GBP/USD. Мандельброт смоделировал три случая: случай, когда показатель Херста равен 0,1 (антиперсистентный процесс); 0,5 (случайное движение) и 0,9 (персистентный процесс).

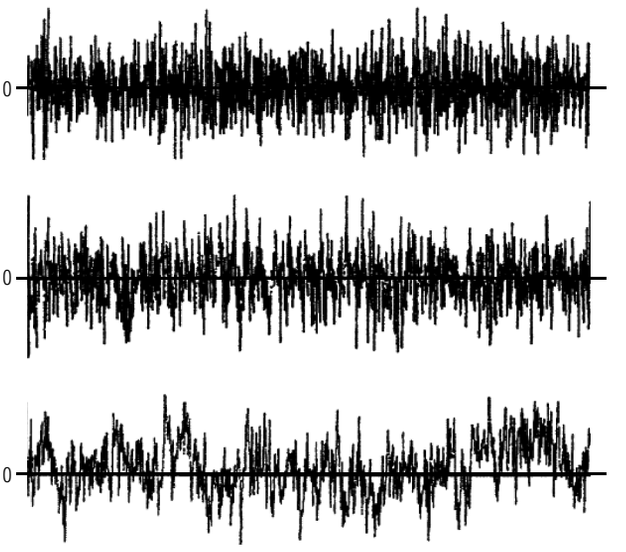


Рис. 1. График приращений «случайной» величины с показателями Херста 0,1(вверху), 0,5 (в середине) и 0,9 (внизу). Источник: [1]

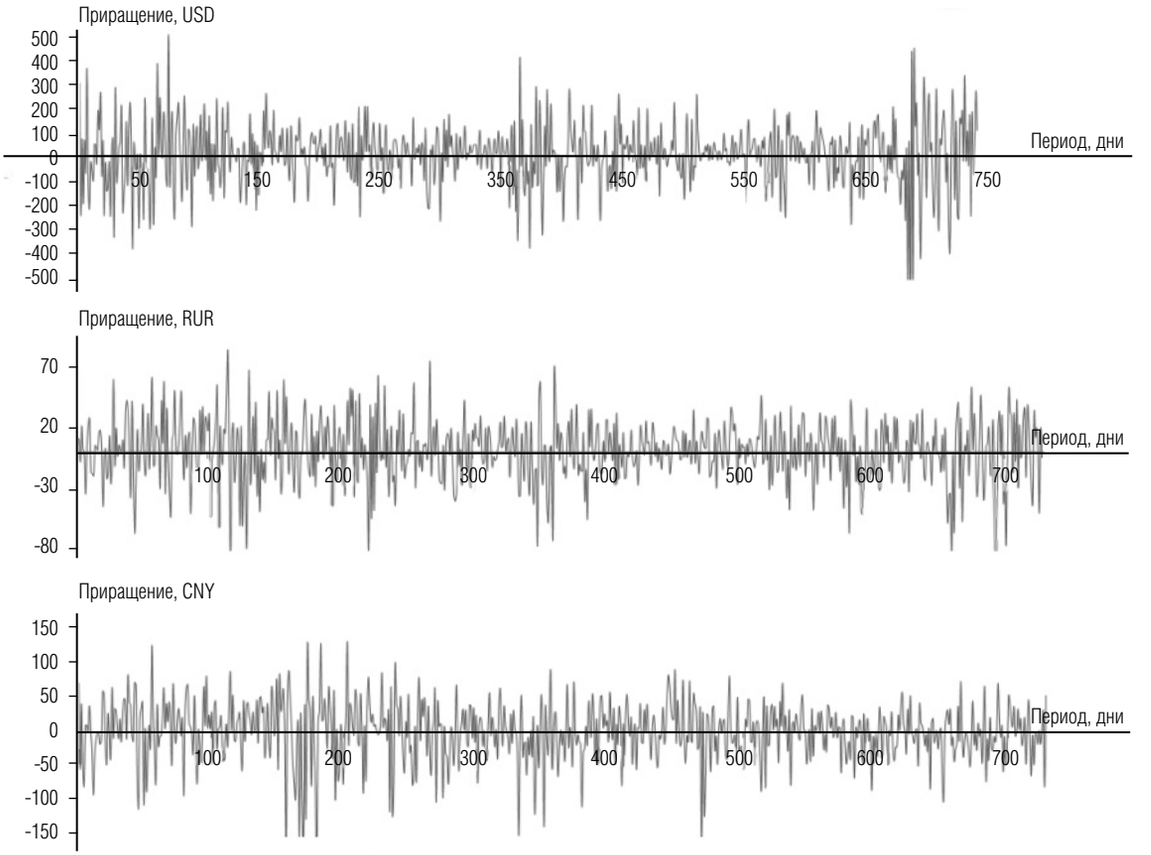


Рис. 2. График приращений котировок биржевых индексов: DJIA (вверху), MICEX (в середине), Shanghai Inc (внизу)

График, представленный на *рис. 1*, демонстрирует небольшие, но часто происходящие изменения, средний – независимые изменения, нижний – трендовость или долговременную память. Теперь изобразим подобные графики изменений анализируемых нами индексов. Мы брали дневные котировки за три года и рассматривали абсолютные приращения, то есть единицы измерения те же, что и в исходных котировках.

Как видно из рисунков, изменения котировок биржевых индексов более всего напоминают иллюстрацию Мандельброта персистентного процесса, что подтверждает наш расчет показателя Херста.

Помимо проверки значимости уравнения регрессии классическими статистическими методами, а также визуального сравнения графиков приращений, мы произвели наиболее надежную проверку правильности гипотезы о персистентности временных рядов биржевых котировок. Логично, что если временной ряд характеризуется долговременной памятью и показывает некую трендовость, то порядок уровней в ряде очень важен. Петерс предлагает перемешать значения уровней временного ряда и повторно посчитать показатель Херста. Если он также будет высок, значит, говорить о правильности выводов о наличии долговременной памяти еще рано.

Мы перемешали исходные временные ряды (всего три – по индексам DJIA, MICEX и Shanghai Inc) , осуществив несколько операций по изменения порядка элементов ряда и перемене элементов местами, используя алгоритмы на VBA. Далее мы полностью повторили алгоритм *R/S* анализа на уже перемешанных данных. Итоги показаны в *табл. 2.*

Таблица 2.

Итоги проверки гипотезы на персистетность

Индекс	H (показатель Херста)	t- статистика по H	F- статистика (значимость уравнения регрессии)
DJIA	0,21	15,6	48,67
Shanghai Inc	0,225	14,2	67,2
MICEX	0,2	11.02	26,66

Значения *F*- и *t*- статистик превышают табличные, следовательно, уравнение регрессии и параметр *H* значимы. Наша гипотеза о персистентности

временного ряда биржевых котировок оказалась верной, так как прошла успешную проверку: при перемешивании данных показатель Херста для всех индексов стал значительно меньше.

5. Заключение

Классические методы анализа финансовых инвестиций оказались неприменимы в периоды интенсивных колебаний рынка и рыночных коллапсов. Экспериментально было показано, что рыночные котировки не подчиняются нормальному распределению. Диаграмма распределения вероятностей больше похожа на степенной закон, получивший название «Распределение Парето».

Над новой теорией динамики финансового рынка с 1960-х годов работал основатель фрактальной геометрии Бенуа Мандельброт. Затем эстафета была подхвачена профессиональным брокером Эдгаром Петерсом и другими аналитиками. Особый интерес представляет собой расчет показателя Херста, в зависимости от величины которого можно сделать вывод о персистентности (трендовости) либо случайном характере временного ряда. Персистентный ряд при этом обладает свойством самоподобия, из чего можно сделать вывод о его фрактальности и продолжать исследования в фрактальном направлении.

Мы рассчитали показатель Херста для трех индексов наиболее ярких современных мировых держав – российского ММВБ, американского DJIA и китайского Shanghai Inc. Из расчетов стало возможным сделать выводы о персистентности временных рядов.

Наши выводы были проверены тремя способами. Во-первых, была проверена значимость уравнения регрессии и параметра H. Во-вторых, мы сравнили графики приращений котировок с графиками приращений значений временного ряда, представленными Мандельбротом для персистентного, случайного и антиперсистентного процесса. Наиболее значимым является третий способ проверки – расчет показателя Херста на перемешанных данных. При наличии долговременной памяти порядок элементов важен, поэтому перемешанные данные должны показывать более низкое значение показателя Херста. В нашем случае перемешанные данные дали антиперсистентный процесс, следовательно, проверка подтвердила гипотезу. ■

6. Литература

1. Мандельброт Б., Хадсон Р. (Не)послушные рынки. Фрактальная революция в финансах. – М.: Изд. дом «Вильямс», 2006.
2. Петерс Э. Фрактальный анализ финансовых рынков. Применение хаоса в инвестициях и экономике. – М.: Интернет-трейдинг, 2004.
3. Петерс Э. Хаос и порядок на рынках капитала. Новый аналитический взгляд на циклы, цены и изменчивость рынка. – М.: Мир, 2000.
4. Фон Нейман Э. Расчет показателя Херста в целях выявления трендовости (персистентности) финансовых рынков. URL: http://capital-times.com.ua.

МЕТОД АННОТИРОВАННОГО СУФФИКСНОГО ДЕРЕВА ДЛЯ ОЦЕНКИ СТЕПЕНИ ВХОЖДЕНИЯ СТРОК В ТЕКСТОВЫЕ ДОКУМЕНТЫ

Б.Г. Миркин,
доктор технических наук, профессор кафедры анализа данных и искусственного интеллекта Национального исследовательского университета «Высшая школа экономики»

Е.Л. Черняк,
студент магистерской программы «Математическое моделирование» Национального исследовательского университета «Высшая школа экономики»

О.Н. Чугунова,
студент магистерской программы «Математическое моделирование» Национального исследовательского университета «Высшая школа экономики»

E-mail: bmirkin@hse.ru, ktr.che@gmail.com, olya.chug@gmail.com
Адрес: г. Москва, Покровский бульвар, д. 11

Излагается модификация метода аннотированного суффиксного дерева (АСД), разработанного с участием одного из авторов, которая ориентирована на то, чтобы, во-первых, убрать априорное ограничение на глубину конструируемого дерева, во-вторых, сделать более адекватной оценку степени вхождения последовательности букв в текст, и, в-третьих, рассмотреть другие приложения метода. На конкретных примерах описываются методы разработки и использования АСД для двух классов задач анализа текстовой информации: (а) связь корпуса текстов и совокупности ключевых словосочетаний; (б) связь корпуса текстов с таксономией предметной области.

Ключевые слова: анализ текстов, аннотированное суффиксное дерево, интерпретация, концептуальные кластеры.

Введение

Основные работы по автоматизации обработки и анализа текстов идут в разрезе представления текстов как совокупностей слов, как это делается в наиболее популярных методиках «модели мешка слов» и «обработки естественного языка». Значительно реже применяются методики, основанные на представлении текстов как последовательностей символов. Между тем, последние имеют то значительное преимущество, что они не требуют предварительной обработки текстов, на-

пример, выделения таких видов слов как «ключевые слова», «стоп-слова», и пр. В данной работе представлена методика анализа текстов, основанная на понятии суффиксного дерева. Суффиксное дерево – одна из основных структур данных для хранения и поиска фрагментов символических последовательностей – текстов, биологических последовательностей, и т.п. [1]. В последнее время ее все чаще используют для кластеризации текстовых объектов, см. например, [2]. Мы используем суффиксные деревья, аннотированные частотами вхождения фрагментов текстов (АСД), иногда также называемые обобщенными суффиксными деревьями. Этот инструмент позволяет эффективно решать задачи анализа текстов, не связанные с их грамматическими описаниями и не зависящие от языка, на котором они написаны. В частности, в [3] этот аппарат использовался для выделения так называемого «спама» в составе электронной почты. В данной работе метод АСД модифицирован: во-первых, убрано априорное ограничение на глубину конструируемого суффиксного дерева, во-вторых, предложена новая формула оценки степени вхождений, позволяющая убрать ее зависимость от длины анализируемых текстов и, в-третьих, рассматриваются приложения, основанные на предварительном построении так называемой ПС таблицы (см. раздел 3.1). Это: (а) анализ связи данного корпуса текстов и совокупности ключевых словосочетаний путем анализа структуры либо множества текстов, либо множества словосочетаний; (б) анализ связи текстов с таксономией своей предметной области.

Далее в разделе 2 описаны методы построения аннотированного суффиксного дерева (АСД) и наложения строки на АСД. Раздел 3 описывает методики и примеры решения задач типа (а), а раздел 4 посвящен примерам задач типа (б). Раздел 5 заключает статью.

1. Индексирование фрагментов текста с помощью аннотированного суффиксного дерева (АСД)

1.1. Понятие аннотированного суффиксного дерева

Мы рассматриваем текст как последовательность символов. Для уменьшения объема вычислений текст разбивается на короткие фрагменты, в дальнейшем называемые строками. При этом текст рассматривается не как единое целое, а неупорядоченная совокупность строк. Любые фрагменты, содержащие семантически законченные отрывки текста, могут быть использованы в качестве строк. Использование «строкового» представления теряет семантические связи между отрывками, но сохраняет связи внутри них. Однако теряемые нами «дальние» связи обычно теряются и при других подходах к анализу текстов, как, например, при подходе «обработки естественного языка» [4]. Вместе с тем сохраняются «локальные» связи, что существенно отличает данный подход от другого популярного подхода, использующего модель так называемого «мешка слов» [3-5].

АСД организовано как корневое дерево, в котором каждый узел, кроме корня, помечен одним из символов строки [3]. Путь от корня АСД до любого узла кодирует один из фрагментов строки, а путь из корня до листа дерева – один из суффиксов строки. Помимо символьной метки, каждый узел обладает числовой меткой, соответствующей частоте соответствующего фрагмента.

Рис. 1 представляет АСД для строки ‘XABXAC’ (в [1] дается другое, обычное, суффиксное дерево для этой строки).

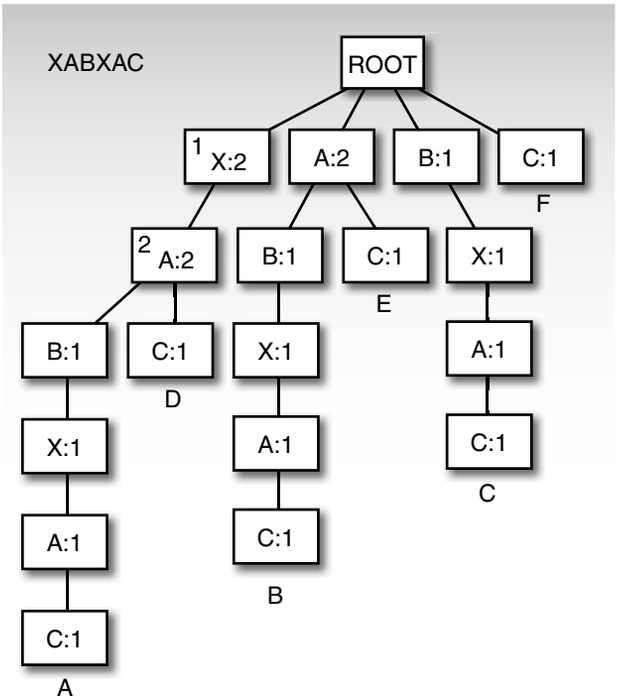


Рис. 1. Аннотированное суффиксное дерево для строки ‘XABXAC’

У строки s = ‘XABXAC’ шесть суффиксов: s[1:] = ‘XABXAC’, s[2:] = ‘ABXAC’, s[3:] = ‘BXAC’, s[4:] = ‘XAC’, s[5:] = ‘AC’, s[6:] = ‘C’. В АСД на Рис. 1 их кодируют пути от корня ROOT до листьев A, B, C, D, E, F, соответственно. Сочетание ‘XA’ дважды

встречается в строке s: это фрагменты s[1:2] = s[4:5] = ‘XA’. На *рис. 1* эти символы представлены узлами, помеченными номерами 1 и 2. Узел 2 имеет двух потомков с частотами 1, соответствующих символам s[3] = ‘B’ и s[6] = ‘C’. Другими словами, ‘XA’ – префикс обоих суффиксов A и D.

АСД для двух и более строк не имеет принципиальных отличий от АСД для одной строки. Оно получается добавлением информации строк к уже построенному дереву и так же представляет все фрагменты коллекции строк и их частоты. Рассмотрим АСД для двух строк: s = ‘XABXAC’ и t = ‘BABXAC’ (*рис. 2*). Строки s и t различаются только первыми символами (s[1] = ‘X’, t[1] = ‘B’). Их суффиксы, начиная со вторых, полностью совпадают. Поэтому АСД на *рис.* 2 отличается от АСД на *рис. 1* только тем, что, во-первых, в нем - новая цепочка узлов G, соответствующая суффиксу t[1:] = ‘BABXAC’, и, во-вторых, изменились частоты, приписанные узлам. Теперь фрагмент ‘XA’ встречается три раза (s[1:2] = s[4:5] = t[4:5] = ‘XA’), поэтому в новом дереве узлам 1 и 2 приписаны частоты 3. Узел 3, помеченный символом ‘B’, является префиксом двух суффиксов: s[3:] = t[3:] = ‘BXAC’, t[1:] = ‘BABXAC’, причем частота первого из них равна 2, второго – 1. Поэтому узлу 3 приписана частота 3.

Важное свойство АСД: частота любого узла равна сумме частот его узлов-детей, так как родительский узел соответствует префиксу нескольких суффиксов и его частота складывается из частот этих суффиксов. Отсюда же следует, что частота родительского узла равна сумме частот листьев, которые он покрывает.

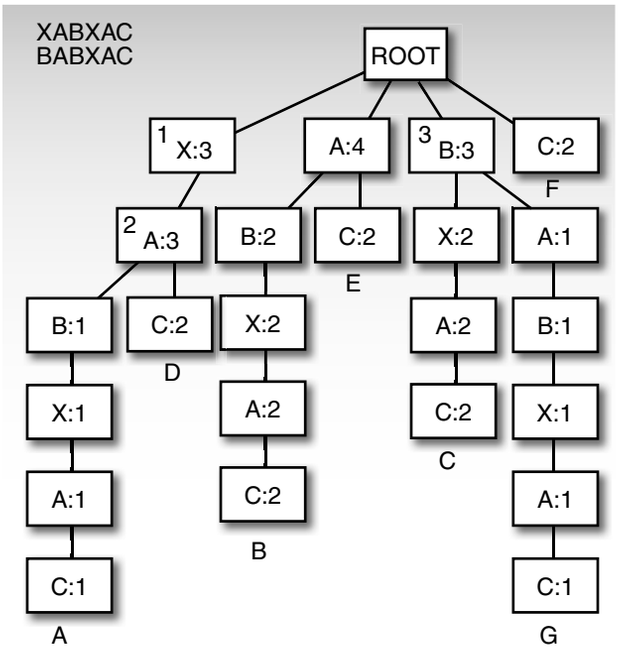


Рис. 2. Аннотированное суффиксное дерево для строк ‘XABXAC’ и ‘BABXAC’

1.2 Алгоритм построения аннотированного суффиксного дерева

Алгоритм построения АСД для коллекции строк основан на последовательном переборе всех суффиксов всех строк из коллекции. Приведем точный алгоритм, модифицированный по сравнению с [3] в связи со строковым представлением текста и измененной мерой качества вхождения.

Построение дерева для коллекции строк $s_1,..., s_N$

Инициализация: создаем пустую структуру, в которой будет храниться АСД. Обозначим ее ast. Далее итеративно будем добавлять в ast подструктуры, соответствующие строкам из входной коллекции.

На i-той итерации алгоритма, i=1, 2, ...,N, для строки $s=s_i$ длины l:

Находим все суффиксы $s[j:]$, где j=1, 2, ..., l.

Для каждого суффикса $s[j:]$ ищем в ast совпадение – путь от корня, совпадающий с максимальным начальным отрезком суффикса $s[j:]$.

Пусть найдено совпадение $m = s[j:k]$, где $k \le l$. Для узлов, попавших в совпадение, увеличиваем частоты на 1.

Если $k < l$, требуется создать новые узлы для фрагмента строки $s[k+1:l]$. Для этого создаем у последнего узла в найденном совпадении нового потомка, помечаем его символом $s[k+1]$ и приписываем ему частоту 1. Таким же образом последовательно создаем узлы для всех оставшихся символов в фрагменте. В результате будет создана новая цепочка узлов, кодирующая текущий суффикс. Если $k = l$, новые узлы не создаются.

Как пример, построим АСД для строки s = ‘XABXAC’ (*рис. 1, 2*). Для первых трех суффиксов, XABXAC, ABXAC, BXAC, совпадений не будет найдено. Поэтому в дерево будут добавлены соответствующие цепочки к узлам A, B, C, у которых частота каждого узла определяется как равная 1. При добавлении следующего суффикса $s[4:]$ = ‘XAC’ будет найдено непустое совпадение ‘XA’, поэтому частота узлов 1 и 2 из совпадения будет увеличена на 1, а для узла 2 - создан новый потомок с меткой ‘C’ и частотой 1. Аналогично, при добавлении суффикса $s[5:]$ = ‘AC’ будет найдено совпадение из одного узла с меткой ‘A’. Следуя алгоритму, частота узла

будет увеличена на 1 и у него будет создан новый потомок с меткой ‘C’ и частотой 1.

Если к уже построенному для строки s=‘XABXAC’ АСД требуется добавить строку t = ‘BABXAC’, то для первого суффикса $t[1:]$ = ‘BABXAC’ будет найдено совпадение из одного узла с меткой ‘B’. Увеличим его частоту и добавим к нему путь из всех остальных символов суффикса. Получим цепочку узлов G на *рис. 2.* Для всех остальных суффиксов строки t будут найдены совпадения, полностью покрывающие суффиксы, поэтому у всех узлов в дереве частоты будут увеличены, но новых узлов создано не будет.

1.3 Процедура наложения строки на АСД

Использование АСД позволяет оценить степень вхождения строки, как последовательности символов, в данный текст.

Введем обозначения: *ast*– АСД, построенное для коллекции строк, $ROOT$ – корень АСД, u – узел АСД, $f(u)$ – его частота, s – строка, степень вхождения которой в коллекцию требуется оценить.

Условной вероятностью узла u назовем величину

$$\hat{p} = \frac{f(u)}{f(parent(u))},$$

где $f(u)$ –частота узла u, $f(parent(u))$– частота узла-родителя u.

Условную вероятность узла на первом уровне АСД считаем равной

$$\frac{f(u)}{\sum_{n:parent(n)=ROOT} f(n)},$$

где n:*parent*$(n)=ROOT$ – множество всех узлов n на первом уровне дерева.

Оценка вхождения строки s в дерево ast получается путем усреднения оценок всех суффиксов строки s. Для каждого суффикса строки s найдем в дереве ast совпадение $m = m_1 \ldots m_k$. Оценка совпадения m – это сумма условных вероятностей узлов, принадлежащих совпадению:

$$score(m) = \sum_{i=0}^{k} \hat{p}(m_i) \qquad (1)$$

Это позволяет агрегировать оценки всех совпадений по формуле:

$$SCORE(s,ast) = \frac{\sum_{i=1}^{l} s([i:]score,ast)}{l}, \qquad (2)$$

где l – длина строки s, $s[i:]$– ее i-тый суффикс. Полная оценка (2) строки s – это средняя оценка совпадений ее суффиксов с последовательностями в дереве.

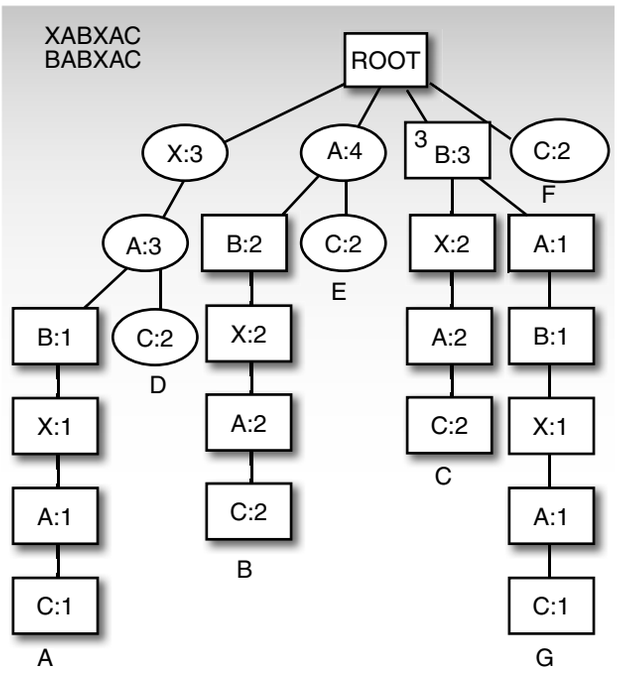


Рис. 3. Наложение строки ‘VXACA’ на АСД коллекции строк {‘XABXAC’,‘BABXAC’}; концы совпадений отображены овалами

Таблица 1.

Оценка результатов наложения всех суффиксов строки ‘VXACA’ на АСД, построенное для коллекции строк {‘XABXAC’,‘BABXAC’}

Суффикс	Совпадение	Score
‘VXACA’	Нет	0
‘XACA’	‘X’->’A’->’C’	3/12 + 3/3 + 2/3=1 11/12
‘ACA’	‘A’->’C’	4/12 + 2/4=5/6
‘CA’	‘C’	2/12
‘A’	A’	4/12

Пример: оценка степени вхождения строки ‘VXACA’ в коллекцию {‘XABXAC’, ‘BABXAC’} (*рис. 3*).

Для того, чтобы найти условную вероятность узлов на первом уровне АСД, будем считать, что частота корня равна сумме частот всех узлов на первом уровне: 3+4+3+2 = 12. Таким образом, для суффикса ‘XAC’ (совпадение по ветви D) получается сумма трех слагаемых, 3/12 для ‘X’, 3/3 для ‘A”, и 2/3 для ‘C’. В *табл. 1* приведены расчеты для всех суффиксов строки. Суммирование этих оценок ‘C’ последующим усреднением дает SCORE = 3.25/ 5 = 0.65.

1.4. Нормировка оценки при сравнении строк с различными АСД

Часто возникает потребность сравнить оценки сходства строк с двумя или более АСД. Получаемые оценки могут сильно зависеть от размеров АСД. Чем больше узлов в АСД, тем больше разброс оценок, получаемых при сличении строк с этим деревом. Для того, чтобы сделать оценки по разным деревьям сравнимыми между собой, модифицируем формулы (1) и (2) так, чтобы нормировать результаты по длине фактических совпадений:

$$score(m) = \frac{\sum_{i=0}^{k} \hat{p}(m_i)}{k}, \qquad (1^*)$$

где k – длина найденного совпадения $m = m_1 ... m_k$. Общая оценка

$$SCORE(s, ast) = \frac{\sum_{i=1}^{l} score(s[i:], ast)}{l} \qquad (2^*)$$

имеет смысл условной вероятности, приходящейся на одну букву суффикса в совпадениях.

Это делает оценки сравнимыми как по документам, так и по словосочетаниям.

1.5 Другие направления использования суффиксных деревьев

Обычные суффиксные деревья часто называют моделью представления текста, альтернативной модели «мешка слов». Мешок слов – пожалуй, самый популярный способ представления текста в компьютере – представляет собой вектор, компоненты которого соответствуют отдельным словам и равны их частотам. Эта модель обладает рядом недостатков и ограничений. Во-первых, в такой модели теряются связи между словами из словосочетаний. Во-вторых, такая модель не всегда удобна. В работе [2], например, утверждается, что в задаче кластеризации текстовых документов модель «мешок слов» не эффективна из-за чрезмерно большой размерности и разреженности векторов частот, представляющих отдельные тексты. С точки зрения авторов этой работы, в задаче кластеризации текстовых документов использование суффиксных деревьев более обосновано. Заметим, что в и в той работе, и в множестве других, суффиксное дерево понимается как иерархическая структура слов, а не символов. Такой поход к представлению суффиксных деревьев впервые предложен в [5] и с алгоритмической точки зрения не отличается от традиционного, изложенного в [1].

Наше представление суффиксных деревьев несколько отличается: во-первых, аннотированные суффиксные деревья имеют другую структуру, во-вторых, мы используем их с другими целями. В рассматриваемых в данной статье задачах аннотированное суффиксное дерево используется в первую очередь для характеристики связей между фрагментами текстов и коллекцией текстов. Эта задача отличается от задачи кластеризации текстовых документов и требует, очевидно, анализа текста не на уровне слов, а на уровне символов.

2. Использование метода АСД для анализа текстов по словосочетаниям

В этом разделе будут описаны два подхода к анализу пары «корпус текстов – совокупность ключевых словосочетаний». Один подход связан с исследованием структуры корпуса в разрезе словосочетаний; другой – с исследованием структуры связей между словосочетаниями согласно данному корпусу.

2.1. ПС таблица

Метод АСД может использоваться для анализа структуры корпуса текстов в разрезе определенных словосочетаний, связанных с этим корпусом. Рассмотрим какой-нибудь корпус текстов, например, набор публикаций о бизнес-процессе в после-кризисной России (2009-2010 годы). Словосочетания могут характеризовать различные типовые события:

1. Изменение организационно-правовой формы
2. Изменение уровня концентрации собственности
3. Повышение эффективности управления затратами
4. Смена генерального директора
5. Участие в судебных разбирательствах
6. Присвоение кредитного рейтинга
7. Выход на международный рынок
8. Публикация финансовой отчетности
9. Реструктуризация кредита
10. Первичное размещение на зарубежной бирже и др.

С помощью АСД метода построим таблицу «публикация-словосочетание» (ПС таблица), в которой строки соответствуют отдельным публикациям (текстам), столбцы – отдельным словосочетаниям, а элементы – оценки степени вхождения строк-словосочетаний в АСД, построенное для соответствующей публикации. Для каждой публикации

строим свое АСД, а затем, с помощью процедуры наложения вычислим оценки степеней вхождения каждого словосочетания публикацию (см. *табл. 2*, представляющую фрагмент одной из наших ПС таблиц). Мы экспериментировали с различными методами разбиения статьи на строки. Хорошие результаты достигаются при разбиении публикации на тройки слов, по-видимому, из-за того, что большинство рассматриваемых словосочетаний тоже состоит из трех слов, так что глубина АСД получается близкой к длине словосочетаний, с ним сличаемых. Построенную ПС таблицу можно использовать как для анализа структуры корпуса публикаций, так и для анализа взаимосвязей между словосочетаниями.

2.2 Анализ структуры корпуса публикаций путем иерархической группировки

ПС таблица позволяет использовать словосочетания как количественные признаки, значения которых она содержит. С ее помощью можно построить иерархическую классификацию публикаций. Мы используем метод иерархической концептуальной кластеризации [6] в модификации Миркина [7]. Концептуальный кластер-анализ отличается от остальных методов кластер-анализа тем, что разделение кластеров в нем осуществляется не по многомерному расстоянию, комбинирующему в себе действие всех рассматриваемых признаков, а по только одному из признаков. Если признак x количественный, то две части, на которые разбивается кластер, отвечают предикатам “x>a” и “x≤a” для некоторого значения a. Если признак – категоризованный, то две части разделения отвечают предикатам “x=a” и “x≠a” для некоторой категории a. В процессе вычислений осуществляется полный перебор всех кандидатов для разбиения – по всем признакам и всем их значениям a – их на самом деле очень немного, линейная функция от числа признаков, и выбирается то разделение, для которого суммарная ассоциация с существующими признаками максимальна. При этом максимально и многомерное расстояние Уорда между центрами разделенных частей [7]. Получаемое «концептуальное» дерево имеет простую интерпретацию и, кроме того, выступает в качестве инструмента отбора информативных признаков – тех, которые действительно участвуют в разделениях. Степень ассоциации иерархического разбиения с признаками измеряется так называемым корреляционным отношением в случае количественных признаков, или коэффициентами, основанными на таблице сопряженности между искомым разбиением и признаками, в случае категоризованных признаков. Оказывается, в последнем случае некоторые известные коэффициенты ассоциации, популярные в построении решающих деревьев, такие как индекс Джини и коэффициент сопряженности Пирсона, эквивалентны специальным случаям критерия квадратичной ошибки в методе к-средних, при условии, что отдельные категории представлены как числовые 1/0 признаки и подходящим образом нормированы [7].

Благодаря специфике метода, каждый кластер в полученной иерархии может быть легко интерпретирован предикатами на пути от корня дерева до листа, соответствующего рассматриваемому кластеру. В силу своей дихотомической структуры, метод может на разных этапах построении дерева иерархии использовать различные значения одного и того же признака, что многократно происходило при расчетах. При этом возникает отдельная задача согласования соответствующих бинарных предикатов — формирование интервалов оценки степени вхождения, соответствующих тому или иному кластеру. Например, кластер {2,4,5} на *рис. 4* описыва-

Таблица 2.

Фрагмент ПС таблицы*

	Доклад Всемирного Банка об экономике России	Международные стандарты финансовой отчетности	Если генеральный директор иностранец
1. Изменение организационно-правовой формы	0.3145	0.3616	0.3644
2. Изменение уровня концентрации собственности	0.5016	0.3148	0.2706
3. Повышение эффективности управления затратами	0.4433	0.2809	0.2445
4. Смена генерального директора	0.2264	0.2351	0.5947

* Столбцы соответствуют публикациям, а строки – словосочетаниям, значения в клетках – степени вхождения словосочетаний в публикации

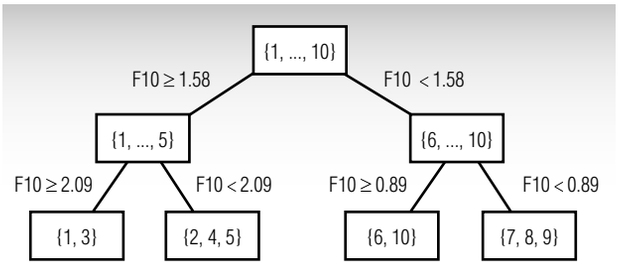


Рис. 4. Пример иерархической классификации

ется условием: степень вхождения словосочетания F10 находится в пределах между 1.58 и 2.09.

В одном из расчетов таким способом было получено дерево иерархии, имеющее 7 уровней и 51 узел, 26 из которых являются листьями (они и есть искомые кластеры публикаций). При этом из нескольких десятков рассматривавшихся словосочетаний в построенном дереве были использованы только те десять, что перечислены в начале раздела 3.1. Этот список характеризует существенные стороны выживания и развития компаний, особенно в после-кризисный период. Он оказался довольно устойчивым относительно различных методов разделения текстов на строки, а также преобразования данных путем обнуления малых значений оценки вхождения словосочетаний в тексты.

2.3. Анализ связей между ключевыми словосочетаниями

Ту же ПС таблицу можно использовать для анализа связей между входящими в нее словосочетаниями. Все публикации разделяются на три типа: (1) статьи, в которых явно выражено только одно словосочетание, так что его оценка по методу АСД значительно превосходит оценки всех остальных словосочетаний; (2) статьи с высокими оценками вхождения двух и более словосочетаний; (3) статьи, где нет ни одного словосочетания с высокой оценкой. Поэтому для каждого словосочетания определены множества публикаций, составляющие его моно- и мульти-ядра, т.е. множества публикаций только типа (1) (моноядро) и типа (2) (мультиядро). Объединение этих двух типов образует множество всех публикаций F(A), в которых оценка вхождения соответствующего словосочетания A превышает заданный порог.

Будем считать, что словосочетание A влечет словосочетание B согласно данному корпусу публикаций, если доля множества F(B) в F(A) составляет не менее 60%. Это правило, соответствующее традиционным логико-статистическим построениям, напоминает известный аппарат построения ассоциаций в так называемом майнинге данных [8]. Действительно, в обоих случаях имеется в виду, что одно множество объектов содержит другое, с точностью до определенной ошибки, конечно. Однако имеется и существенная разница. В майнинге данных импликации (ассоциативные правила) ищутся по всему массиву данных, без привязки к каким-либо специфическим утверждениям, что требует задания дополнительного порога на уровень «поддержки» импликации. В нашем случае импликации привязаны к заранее заданным словосочетаниям, и не нуждаются в проверке на уровень поддержки. Данный подход ближе к так называемому детерминационному анализу [9]. Но в детерминационном анализе главное – выявление системы категорий, комбинация которых приводит к максимальной точности импликации. Для нас же главное – выявление структуры связей между заданными словосочетаниями. В частности, в расчете по публикациям о бизнес-процессе в России с использованием нескольких десятков словосочетаний был получен граф, представленный на *рис. 5.* Словосочетания, соответствующие его вершинам:

1. Ввод автоматизированного производства.
2. Выпуск пресс-релизов (с положительными или отрицательными новостями).
3. Изменение размера пакета акций, принадлежащего институциональному инвестору.
4. Изменение уровня концентрации собственности.
5. Повышение квалификации персонала.
6. Проведение вертикального слияния.
7. Проведение операций купли-продажи бренда.
8. Выход на международный рынок.
9. Изменение организационно-правовой формы.
10. Повышение эффективности управления затратами.
11. Публикация финансовой отчетности.
12. Смена финансового директора.

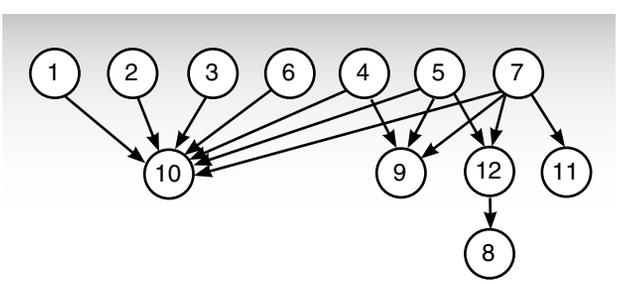


Рис. 5. Граф значимых (на уровне 60%) связей между словосочетаниями по массиву публикаций. Расшифровка номеров приведена в тексте.

Обращает на себя внимание то, что в графе нет контуров и всего два основных уровня, естественным образом интерпретируемых как источники и цели. С содержательной точки зрения, граф допускает разумную интерпретацию, связанную с содержанием процессов развития бизнес-процесса в России в 2009-2010 гг. Показанные в нем цели: уменьшение издержек (10), изменение организационно-правовой формы (9), повышение прозрачности (11), выход на мировые рынки (8) действительно могут помочь бизнесу выжить и развиваться; событие (12) оказывается шагом, ведущим к выходу на мировые рынки. Согласно графу, основными факторами этого процесса являются: купля-продажа брендов (развитие сетевых структур), автоматизация производства, повышение квалификации персонала, а также передача государственных активов в частные руки.

3. Использование метода АСД для анализа связи текстов с таксономией предметной области

В этом разделе будет рассмотрена возможность отображения текстов на таксономию соответствующей предметной области. Под таксономией, или иерархической онтологией, понимается иерархическое представление основных понятий в разрезе таких отношений как «A состоит из B1, B2, ...» или «B — это частный случай A». В настоящее время иерархические онтологии — одно из основных направлений автоматизации хранения, использования и накопления знаний [10-12]. В разделе описаны наши попытки использования метода АСД для отображения текстов в таксономиях математики и информатики, одна – англоязычная (информатика), вторая – русскоязычная (математика).

3.1. Индексирование научных статей таксономическими единицами классификации информатики ACM-CCS

Одно из возможных приложений метода АСД – индексирование научных публикаций по существующим научным классификациям. Например, в журналах международной Ассоциации Вычислительной Техники (Association for Computing Machinery) используется классификационная система вычислительной техники, разработанная этой организацией [13] (ACM-CCS), для индексирования (рубрикации) статей. Авторы вручную приписывают своим статьям две-три таксономические единицы ACM-CCS, наилучшим образом отвечающие им по содержанию. Нас интересует возможность использования метода АСД для автоматизации такого индексирования.

Чтобы реализовать эту идею, для каждой рассматриваемой научной публикации следует:

- ✦ выделить ее ключевые фрагменты, включая заголовок, список ключевых слов и аннотацию (abstract);
- ✦ построить АСД по выделенным фрагментам публикации (по всем или частично);
- ✦ оценить степень вхождения каждой листовой таксономической единицы ACM-CCS в построенное АСД, т.е. построить профиль статьи;
- ✦ выбрать таксономические единицы с максимальными оценками в качестве искомой индексации.

Рассмотрим один из журналов, Journal of the ACM, издаваемый в электронном формате и находящийся в свободном доступе. Для ускорения расчетов использованы только текст аннотации и ключевые слова статей. В *табл. 3* приведены аннотации, таксономические единицы, приписанные авторами, и списки ключевых слов для публикаций [14] и [15]. Это сделано для того, чтобы читатель смог оценить сам, насколько аннотации, полученные с использованием метода АСД и приведенные далее в *табл. 4* и *5*, соответствуют содержанию. Кроме того, приведенные тексты могут быть использованы как данные для тестирования других методов анализа текстов.

Следует иметь в виду, что, несмотря на внешнее сходство с задачей распознавания образов, в данной проблеме нет внешнего учителя. Поэтому не существует объективного измерителя степени успешности работы автоматического индексатора, по крайней мере, в настоящее время.

В левой части *таблиц 4* и *5* представлены АСД-профили, а в правой – авторские аннотации, а также места, которые авторские таксономические единицы заняли в АСД-профиле.

Профиль *табл. 4* представляется вполне удачным, поскольку таксономические единицы из авторского индекса статьи занимают 3 и 4 место в АСД-профиле. Напротив, профиль *табл. 5* – неудачный: авторский индекс статьи содержит таксономические единицы, крайне низко оцененные АСД-профилем. Дело в том, что авторские индексации содержат таксономические единицы, формулировки которых не отражены в тексте аннотации – они передаются другими, синонимичными словами.

Приведенный пример показывает трудности, связанные с методом АСД:

✦ неоправданно высокая оценка общих слов и фраз. Эту проблему отчасти можно решить путем введения списка стоп-слов, включающего все подобные слова, «вручную»;

✦ таксономическая единица получает низкую оценку, если автор предпочитает использовать другое, хотя и близкое по смыслу, понятие.

Эта проблема может быть решена, если с каждой таксономической единицей связать список синонимичных понятий.

3.2. Анализ учебных программ математического цикла НИУ ВШЭ с использованием таксономии РЖ «Математика»

На сайте НИУ ВШЭ имеются в свободном доступе файлы программ различных курсов, относящихся к математическим дисциплинам и читаемым студентам различных специализаций. Естественный вопрос – как эти программы отражают современную математику – может трактоваться как возможность выделения основных кластеров математического знания, содержащихся в этих программах, и их отображения на иерархическую классификацию математики. Поскольку учебные программы написаны на русском языке, в качестве классификации математики мы взяли русскоязычный иерархический рубрикатор реферативного журнала РЖ «Математика» (в настоящее время — на сайте [14]; мы использовали более ранний вариант рубрикатора, доступный в 2010 г., когда проводились расчеты).

Приведем некоторые из полученных результатов. В *табл. 6* представлены таксономические единицы, получившие максимальные оценки вхождения в программу курса «Дискретная математика». Часть таксономических единиц, получивших высокие оценки по методу АСД, оказались не адекватными содержанию учебной программы. Три из представленных в *табл. 6* неадекватных таксономических единиц содержат слово «алгебраический». Это слово или однокоренные с ним часто употребляются в программе «Дискретная математика», поэтому данные таксономические единицы и получили высокие оценки.

Хороший профиль оказался у программы «Дифференциальные уравнения»: он включает в себя большую часть соответствующего раздела таксономии и почти полностью покрывает содержание программы.

Таблица 3.

Аннотации, индексные таксономические единицы и ключевые слова двух публикаций журнала ACM

M. Bojanczyk, A. Muscholl, T. Schwentick, L. Segoufin	M. Grohe, A. Henrich, N. Schweikardt
Two variable logic on data trees and XML reasoning, Journal of ACM, 2009, 56(3), 58 p.	Lower bounds for processing data with few random accesses to external memory, Journal of ACM, 2009, 56(3), 48 p.
Motivated by reasoning tasks for XML languages, the satisfiability problem of logics on data trees is investigated. The nodes of a data tree have a label from a finite set and a data value from a possibly infinite set. It is shown that satisfiability for two-variable first-order logic is decidable if the tree structure can be accessed only through the child and the next sibling predicates and the access to data values is restricted to equality tests. From this main result, decidability of satisfiability and containment for a data-aware fragment of XPath and of the implication problem for unary key and inclusion constraints is concluded.	We consider a scenario where we want to query a large dataset that is stored in external memory and does not fit into main memory. The most constrained resources in such a situation are the size of the main memory and the number of random accesses to external memory. We note that sequentially streaming data from external memory through main memory is much less prohibitive. We propose an abstract model of this scenario in which we restrict the size of the main memory and the number of random accesses to external memory, but admit arbitrary sequential access. A distinguishing feature of our model is that it allows the usage of unlimited external memory for storing intermediate results, such as several hard disks that can be accessed in parallel. In this model, we prove lower bounds for the problem of sorting a sequence of strings (or numbers), the problem of deciding whether two given sets of strings are equal, and two closely related decision problems. Intuitively, our results say that there is no algorithm for the problems that uses internal memory space bounded by $N^{1-\varepsilon}$ and at most $o(\log N)$ random accesses to external memory, but unlimited «streaming access», both for writing to and reading from external memory. (Here N denotes the size of the input and ε is an arbitrary constant greater than 0.) We even permit randomized algorithms with one-sided bounded error. We also consider the problem of evaluating database queries and prove similar lower bounds for evaluating relational algebra queries against relational databases and XQuery and XPath queries against XML-databases.
Primary Classification F.4.1[Mathematical logic and formal languages]:Mathematical logic Additional Classification: F.2.3[Database management]: Languages–Query languages	Primary Classification F.1.1 [Computation by Abstract Devices]: Models of Computation—bounded-action devices; F.1.3 [Computation by Abstract Devices]: Complexity Measures and Classes—relations among complexity classes; relations among complexity measures; Additional Classification: H.2.4 [Database Management]: Systems—query processing; relational databases
General Terms: Theory Key Words and Phrases: First-order logic, data trees, decidability	General Terms: Theory, Languages Key Words and Phrases: complexity, data streams, real-time data, query processing, query optimization, semi-structured data, XML

Таблица 4.

Пример «удачного» АСД профиля

Статья: Bojanczyk M. et al. Two variable logic on data trees and XML reasoning, Journal of the ACM, 2009, Vol. 56(3). 2-48.					
АСД Профиль			Индексационные таксономические единицы (ручное аннотирование)		
ID	TE	ACM–CCS единица	ID	#	ACM–CCS категории
I.6.2	6.1969	Simulation Languages	F.4.3	3	Formal Languages
I.1.3	6.1415	Languages and Systems	H.2.3	4	Languages
F.4.3	6.0796	Formal Languages	H.2.1	13	Logical Design
H.2.3	4.0757	Languages	F.4.1	28	Mathematical Logic
D.4.5	3.7387	Reliability	I.7.2	53	Document Preparation

Таблица 5.

Пример «неудачного» АСД профиля

Статья: Grohe M., et al. Lower bounds for processing data with few random accesses to external memory. Journal of the ACM, 2009, Vol. 56(3), 1-58.					
AST found profile			ACM-CCS index terms (manual annotation)		
ID	TE	ACM–CCS класс	ID	#	ACM–CCS класс
J.1	9.5991	ADMINISTRATIVE DATA PROCESSING	F.1.3	161	Complexity Measures and Classes
I.2.7	7.3757	Natural Language Processing	H.2.4	166	Systems
H.2.5	5.0704	Heterogeneous Databases	F.1.1	220	Models of Computation
H.2.8	4.4196	Database Applications			
C.5.1	4.0146	Large and Medium Computers			

В целом, результаты этого приложения вызывают больше вопросов, чем дают ответов. Это связано, на наш взгляд, не только с вышеотмеченными недостатками метода АСД, но и особенностями использованной таксономии. Использованная версия таксономии РЖ «Математика» (2009) — это дерево неравномерной глубины (от 3 до 8 уровней в разных разделах), содержащее разделы, не сбалансированные между собой по объему и содержанию. Кроме того, в таксономии отсутствуют некоторые современные темы такие как «Дискретная математика» или «Математическая экономика». В разделах, относящихся к современным частям математики, часто опущены важные понятия. Например, категория «Теория игр» содержит таксономические единицы, перечисляющие виды игр, но понятие «равновесие» здесь не представлено. Напротив, имеются разделы, связанные с относительно небольшими частями математики, особенно с точки зрения учебных программ, которым соответствуют глубокие и кустистые поддеревья. Кроме того, в таксономии усложнена система навигации; есть ссылки на удаленные после последней модификации таксономические единицы. Это делает актуальной задачу разработки более адекватной классификации математики, включая, вероятно, информатику и прикладную математику.

Заключение

Методы анализа текстов, основанные на аннотированных суффиксных деревьях (АСД), удобны тем, что не связаны с необходимостью грамматического разбора фраз, т.е. с особенностями того или иного языка, и, более того, вообще позволяют свести до минимума предварительную обработку текстов, составляющую важную и иногда трудоемкую часть других подходов. Путем разбивки текста на строки нам удалось значительно снизить вычислительную трудоемкость метода. Рассмотренные примеры, с одной стороны, показывают эффективность метода в правильно выбранных приложениях и, с другой стороны, показывают пути его дальнейшего совершенствования. Главный недостаток метода – существенная привязка к буквенному содержанию фрагментов текста, от чего, вероятно,

Таблица 6.

Часть профиля программы «Дискретная математика»**

Программа по курсу «ДИСКРЕТНАЯ МАТЕМАТИКА»		
Оценка	Код	Таксономическая единица
25.53	517.986.9	Другие алгебраические структуры в функциональном анализе
18.096	512.562	Частично упорядоченные множества
17.205	512.664.8	Деформации алгебраических структур
16.204	512.732.1	Общие свойства алгебраических пучков
16.025	510.63	Классические логические теории
15.307	512.545.6	Частично упорядоченные группы
12.076	510.6	Математическая логика
11.267	510.64	Неклассические логики
11.253	519.171.2	Представления графов
11.049	519.876.3	Сетевое планирование
10.96	519.712.4	Сложность алгоритмов
10.96	510.52	Сложность алгоритмов
8.31	510.633	Логика высказываний
7.36	519.681.4	Сложность вычислений

можно освободиться путем добавления информации о шумовых словах и словах-синонимах.

Работа частично выполнялась в рамках научно-исследовательского проекта «Учитель-Ученики: 11-04-0019 Разработка и адаптация методов кластер-анализа для автоматизации анализа текстовой информации с использованием онтологии предметной области», поддержанного Программой «Научный фонд НИУ-ВШЭ» в 2011-2012 гг., а также Международной лаборатории по выбору и анализу решений НИУ-ВШЭ. Авторы благодарят руководителя лаборатории Ф.Т. Алескерова за советы и помощь в работе. Часть проекта осуществлялась при финансовой поддержке Министерства образования и науки Российской Федерации в рамках договора N 13.G25.31.0033 от 7 сентября 2010 г., заключенного между Министерством образования и науки Российской Федерации и ЗАО «Авикомп Сервисез». Авторы благодарны анонимному рецензенту, работа над замечаниями которого позволила улучшить изложение. ■

Литература

1. Gusfield D. Algorithms on Strings, Trees, and Sequences. – Cambridge University Press, 1997.
2. Huang C., Yin J., Hou F. Text clustering using a suffix tree measure //Journal of Computers. – 2011. – Vol 10(6). – P. 2180-2186.
3. Pampapathi R., Mirkin B., Levene M. A suffix tree approach to anti-spam email filtering // Machine Learning. – 2006. – Vol. 65(1). – P. 309-338.
4. Manning C., Schütze H. Foundations of Statistical Natural Language Processing. – MIT Press, 1999.
5. Zamir O, Etzioni. O. – Web document clustering: A feasibility demonstration – Proceedings of SIGIR'98, University of Washington, Seattle, USA, 1998.
6. Stepp R., Michalski R.S. Conceptual clustering of structured objects: A goal-oriented approach// AI Journal. – 1986. – Vol. 28. – P. 43-69.
7. Mirkin B. Clustering for Data Mining: A Data Recovery Approach. – Chapman & Hall/CRC, 2005.
8. Han J., Kamber M., Pei J. Data Mining: Concepts and Techniques. – Third Edition. – Morgan Kaufmann, 2011.
9. Чесноков С.В. Детерминационный анализ социально-экономических данных. – 2-е изд. – М.: URSS, 2009.
10. Robinson P.N., Bauer, S. Introduction to Bio-Ontologies. – Chapman & Hall/CRC, 2011.
11. Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – М.: Изд-во Московского университета, 2011.
12. Онтологическое моделирование экономики предприятий и отраслей современной России: Российские исследования и разработки в области онтологического инжиниринга и бизнес-онтологий. – Препринт WP7/2011/08 [Текст] / Ефименко И.В., Хорошевский В.Ф. – Нац. исслед. ун-т «Высшая школа экономики». – М.: Изд. дом Высшей школы экономики, 2011.
13. ACM Computing Classification System, 1998. URL: http://www.acm.org/about/class/1998 (дата обращения: 09.09.2010).
14. Рубрикатор РЖ «Математика». URL: http://www.viniti.ru/russian/math/files/271.htm (дата обращения 15.05. 2012).

** Серым фоном выделены неадекватные таксономические единицы

МЕТОДЫ АНАЛИЗА ИЗМЕНЕНИЙ ГЛОБАЛЬНЫХ И ЛОКАЛЬНЫХ СВОЙСТВ ТЕМПОРАЛЬНЫХ ОРГРАФОВ

В.А. Кохов,
кандидат технических наук, доцент кафедры высшей математики на факультете экономики Национального исследовательского университета «Высшая школа экономики»

В.В. Кохов,
магистрант кафедры прикладной математики Национального исследовательского университета «Московский энергетический институт» (НИУ-МЭИ)

E-mail: viktor-kokhov@ya.ru
Адрес: г. Москва, ул. Красноказарменная, д. 14

Предложены новые методы анализа глобальных и локальных свойств темпоральных орграфов на основе моделей их сложности. Модели сложности позволяют анализировать значимость темпоральных орграфов, их фрагментов и анализировать тенденции их изменения. Приведен сравнительный анализ двух подходов к определению локальных свойств, связанных с вычислением вкладов фрагментов в сложность темпоральных орграфов.

Ключевые слова: темпоральный орграф, модели сложности, сходство, локальный инвариант, глобальный инвариант, значимость фрагмента.

1. Введение

Обычно модели и методы теории графов используются для анализа отношений между элементами сложных структур различной природы [1]. При этом данные отношения между элементами являются постоянными и не меняются во времени. Такие графы в [2] названы «статическими». Если отношения между элементами структуры изменяются во времени, традиционные «статические» графы неприменимы для их описания и моделирования. Как выделено в [3-6] в настоящее время наиболее актуальным направлением является разработка методов анализа графов с изменяемой структурой во времени (темпоральных орграфов (*T*-орграфов)). Работа [2] была началом исследований по графодинамике, т.е. по динамическому описанию структур. В ней выделены базовые классы задач, связанные с определением:

1) равновесного состояния графа и области сходимости к этому состоянию;

2) области циклического изменения состояния графа и длины цикла;

3) расстояния (сходства) между изменяемыми структурами графа;

4) других характеристик.

В качестве наиболее значимой выделялась задача определения расстояния, которое помогает ввести представление об устойчивости изменений структуры графа во времени (графовых траекторий) по отношению к малым возмущениям и о монотонности в смысле этого расстояния процессов в графодинамике. Примерами прикладных задач является анализ изменений:

1) административных структур, обычно описываемых орграфами-деревьями;

2) организации систем связи, снабжения и др.;

3) структуры построения номенклатуры товаров или изделий определенной категории;

4) организации ассоциативной памяти ЭВМ.

В графодинамике ставятся специфические задачи, которые не имеют аналогов в динамике обычно рассматриваемых объектов. В [2] в качестве примеров таких задач выделены:

1) задача определения в графе подграфа, который не меняется или «мало» меняется во времени;

2) задача о «сохранении коллективов», т.е. выделение группы вершин («коллектива»), которые при изменении структуры графа всегда подчинены общему для них «начальнику».

Во всех примерах структура адекватно отображается орграфом. В [3,4] отмечено, что в настоящее время одной из актуальных прикладных задач является задача анализа изменений во времени структур корпоративных социальных сетей (КСС), в особенности сетей коммуникаций сотрудников фирмы. Структурный анализ КСС позволяет:

1) осуществлять мониторинг структуры сети компании и ее изменений;

2) выявлять слабых и сильных акторов в сети;

3) корректировать и моделировать эффективную структуру сети;

4) оперативно выявлять конфликты акторов внутри фирмы;

5) анализировать создание сообществ по формальным и неформальным связям, что способствует созданию единой команды с общей целью;

6) направлять изменения структуры сети с целью повышения обоснованности управленческих решений.

Как выделено в [4], управление КСС, их структурой – новая область менеджмента.

В данной статье предлагаются два подхода к структурному анализу *T*-орграфов. Анализ основан на построении моделей сложности, позволяющих исследовать изменения локальных и глобальных свойств *T*-орграфов, определять сходство *T*-орграфов и выявлять тенденции изменения свойств.

Из анализа работ [7,8] следует, что существуют различные подходы к вычислению значений локальных свойств и связи их с глобальными характеристиками графа. В [7] для характеризации фрагмента $f \subseteq G$ графа с оставшейся частью используется «внешний» топологический индекс (*EFTI*), задаваемый выражением

$$EFTI(f) = TI(G) - [IFTI(f) + \sum_{n} IFTI(G - f)_n],$$

где *TI(G)* – топологический индекс графа *G; IFTI* – «внутренний» топологический индекс фрагмента графа. Суммирование ведется по всем из *n* связным компонентам части $(G - f)$, получаемой удалением из *G* всех вершин, входящих в *f*, и инцидентных им ребер. Для трех видов индексов формулируются естественные требования

$$0 \le EFTI(f) \le TI(G),\ 0 \le IFTI(G - f) \le TI(G).$$

В [8] рассмотрена схема характеризации расположения фрагментов графа метрическими и цепными инвариантами, построенными на основе различных расстояний. Метрические инварианты вычисляются *на единой структуре исходного графа* без разделения его на три части. Таким образом, выделим два основных подхода к характеризации расположения фрагмента в графе:

1) значение инварианта фрагмента вычисляется на единой структуре графа;

2) значение инварианта фрагмента вычисляется с разделением графа на части.

В наиболее общем виде метод вычисления инвариантов, характеризующих расположение фрагментов в графах, и общие требования к их построению по первому подходу приведены в [9].

2. Основные определения

T-орграфом назовем тройку $G = (V^{(t)}, E^{(t)}, T)$, где $|V^{(t)}|$ – множество вершин орграфа в момент времени *t* с числом вершин $|V^{(t)}| = p$, $T = \{1, 2, ..., t, ..., N\}$ – множество натуральных чисел, определяющих (дискретное) время, $|E^{(t)}| = \{\Gamma_t\}$ – семейство соответствий или отображений, множества вершин $V^{(t)}$ в

себя в момент времени $t \in T$, т.е. $(\forall t \in T)\Gamma_t: V^{(t)} \to V^{(t)}$. Через tG обозначим T-орграф в момент времени t. Фрагмент f орграфа tG в момент времени t получается при удалении дуг или вершин и дуг.

Орграф $t_1G = (V^{(t_1)}, E^{(t_1)}, T)$ изоморфен орграфу $t_2G = (V^{(t_2)}, E^{(t_2)}, T)$ $(t_1G \approx t_2G)$, если

$$\exists\varphi:(V^{(t_1)} \leftrightarrow V^{(t_2)}) \,\&\, (\forall v_i, v_j \in V^{(t_1)}[(v_i, v_j) \in E^{(t_1)} \leftrightarrow$$
$$\leftrightarrow (\varphi(v_i), \varphi(v_j)) \in E^{(t_2)}]),$$

где $\varphi(v_i), \varphi(v_j) \in V^{(t_2)}$. Орграф $t_1G = (V^{(t_1)}, E^{(t_1)}, T)$ изоморфно вкладывается в орграф $t_2G = (V^{(t_2)}, E^{(t_2)}, T)$ $(t_1G \subseteq {}^{(f)}t_2G)$, если в t_2G есть фрагмент

$$f = t_2G_* = (V_*^{(t_2)}, E_*^{(t_2)}, T),$$

для которого выполняется условие $t_2G_* \approx t_1G$. Множество всех изоморфизмов орграфа tG на себя образует группу по умножению подстановок φ и обозначается через $Aut(tG)$, порядок группы – через $|Aut(tG)|$. Под числом *канонических изоморфных вложений* tG_* в tG будем понимать величину, определяемую следующим образом:

$$w(tG_*, tG) = W(tG_*, tG) / \left|Aut(tG_*)\right|,$$

где $W(tG_*, tG)$ – число всех изоморфных вложений tG_* в tG.

Абстрактный тип k – произвольный T-орграф, определенный с точностью до изоморфизма. Группу его вершинных автоморфизмов обозначим через $Aut(k)$, множество всех канонических изоморфных вложений абстрактного типа k в орграф tG – через

$$F^{(k)}(tG) = \{tf_1^{(k)}, tf_2^{(k)}, ..., tf_m^{(k)}\},$$

$\left|F^{(k)}(tG)\right|$ – количество фрагментов типа k, а $nk(G)$ – число типов фрагментов в tG. Если на множестве вершин типа фрагмента k и орграфа tG задана нумерация, то фрагмент $tf^{(k)}$ орграфа tG может быть представлен помеченными фрагментами $tf^{(lk)}$, когда каждой вершине типа фрагмента k сопоставляется номер вершины орграфа tG, которой она соответствует при вложении. Число помеченных фрагментов, представляющих один и тот же фрагмент tG, равно $|Aut(k)|$.

Пусть $Aut^{(k)}(tG)$ является индуцированным представлением группы $Aut(tG)$ и определяет симметрию расположения фрагментов типа k в tG. Под k-автоморфизмом орграфа tG будем понимать подстановку g^* на множестве $F^{(lk)}(tG)$, помеченных фрагментов типа k орграфа tG (или канонических изоморфных вложений), индуцированную некоторым вершинным автоморфизмом g орграфа tG. В процессе индуцирования помеченный фрагмент $f_i^{(lk)} \in F^{(lk)}(tG)$, заданный каноническим изоморфным вложением $(v_1, v_2, ..., v_n)$ орграфа tG, переходит в помеченный фрагмент $f_j^{(lk)} \in F^{(lk)}(tG)$, канонический вид которого получен канонизацией вложения $(u_1, u_2, ..., u_n)$, где $u_i = g(v_i)$, $i = 1, 2, ..., n$, n – число вершин фрагмента типа k. *Группой k-автоморфизмов tG* ($Aut^{(k)}(tG)$ или k-группой) будет группа подстановок, носителем которой является все множество k-автоморфизмов для данного k, а групповой операцией – операция произведения подстановок. Тот факт, что множество k-автоморфизмов образует группу, непосредственно следует из свойств $Aut(tG)$. Степень k-группы $|F^{(lk)}(tG)|$ равна числу канонических изоморфных вложений абстрактного типа k в орграф tG, а порядок меньше или равен порядку $Aut(tG)$. Последнее обосновано тем, что два различных нетождественных вершинных автоморфизма могут индуцировать один и тот же k-автоморфизм. Все понятия, связанные с анализом $Aut^{(k)}(tG)$, определяются аналогично понятиям, связанным с анализом $Aut(tG)$ (например, орбиты k-группы). Орбиты k-группы точно характеризуют симметрию расположения фрагментов типа k, в орграфе tG.

Основным инструментом при вычислении глобальных (локальных) характеристик орграфа tG является использование инвариантов орграфа tG (инвариантов, характеризующих расположение фрагментов в tG). Обозначим множество всех T-орграфов через $\Re$. Пусть R – отношение «быть изоморфными» орграфами tG, а Q – непустое множество с отношением эквивалентности τ (множество чисел, векторов, матриц, орграфов и т.д.). Функция IN заданная на множестве $\Re$ и принимающая значения в Q, называется инвариантом орграфа tG, если справедливо услови

$$\forall t_iG, t_jG \in \Re\,[t_iG(R)t_jG \to IN(t_iG)(\tau)IN(t_jG)].$$

Пусть $F^{(lk)}(tG) = \{tf_1^{(lk)}, tf_2^{(lk)}, ..., tf_m^{(lk)}\}$ обозначает множество помеченных фрагментов типа k в орграфе tG. Инвариантом, характеризующим расположение фрагмента $tf^{(lk)}$, называется функция $in(tf^{(lk)})$, заданная на множестве $F^{(lk)}(tG)$, принимающая значения в Q и удовлетворяющая условию

$$\forall tf_i^{(lk)}, tf_j^{(lk)} \in tF^{(lk)}[tf_i^{(lk)}(\xi)\, tf_j^{(lk)} \to in(tf_i^{(lk)})(\tau)in(tf_j^{(lk)})],$$

где ξ – отношение эквивалентности «принадлежать одной и той же орбите k-группы». Все неопределяемые ниже понятия можно найти в [1].

3. Модели структурной спектральной сложности для определения глобальных свойств *T*-орграфов

Концептуальные и математические модели для определения сложности графов, методы построения моделей и результаты их исследования на точность решения задач различения и определения сходства графов приведены в [9-10]. Результаты решения задач различения *орграфов* на основе моделей сложности приведены в [11]. Успех в исследовании структурной сложности и сходства систем неразрывно связан с разработкой методологии построения и исследования структурных инвариантов, позволяющих характеризовать орграф и вычислять вклад фрагментов в его сложность [1,9-11]. Ниже рассмотрены шесть видов инвариантов: 1) структурный спектр; 2) длина структурного спектра; 3) полный структурный спектр; 4) длина полного структурного спектра; 5) индекс сложности; 6) вектор-индекс сложности. Все они характеризуют структурную спектральную сложность (СС-сложность) *tG* в расширяемых наборах (базисах) фрагментов. Индексы и вектор-индексы СС-сложности *tG* характеризуют их количественный и качественный состав в заданном наборе (базисе) фрагментов.

Под структурным спектром $SS(tG/B)$ орграфа *tG* в базисе $B=<b_1,b_2,...,b_i,...,b_{k1}>$ будем понимать запись следующего вида

$$SS(tG/B)=<\alpha_1,\alpha_2,...,\alpha_i,...,\alpha_{k1}>,$$

где α_i=1, если $b_i \in B$ *изоморфно вкладывается* в *tG* и α_i=0 в противном случае.Через

$$L^{(SS)}(tG/B)=\sum_{i=1}^{k1}\alpha_i$$

обозначим длину структурного спектра *tG* в базисе *B*.

Пусть для орграфа *G* построен его полный структурный спектр (ПСС) в базисе *B:*

$$FSS(tG/B)=<wb_1,wb_2,...,wb_i,...,wb_{k1}>,$$

где b_i – фрагмент базиса; w_i – число канонических изоморфных вложений фрагмента b_i в *tG*; *k*1 число фрагментов базиса *B*, относительно которого характеризуется сложность *tG*. Очевидно, что

$$w(K_1)=p,\ w(K_2)=q.$$

Примем $ISC(K_1)$=1, $ISC(K_2)$=3. Так как для любого фрагмента tf_i орграфа *tG* можно определить его ПСС, а для каждого фрагмента от фрагмента *tG*, можно построить его ПСС и т.д., то рекурсивным методом всегда можно вычислить индекс сложности (*ISC*) и вектор-индекс сложности (*V_ISC*) орграфа *tG* в базисе *B*:

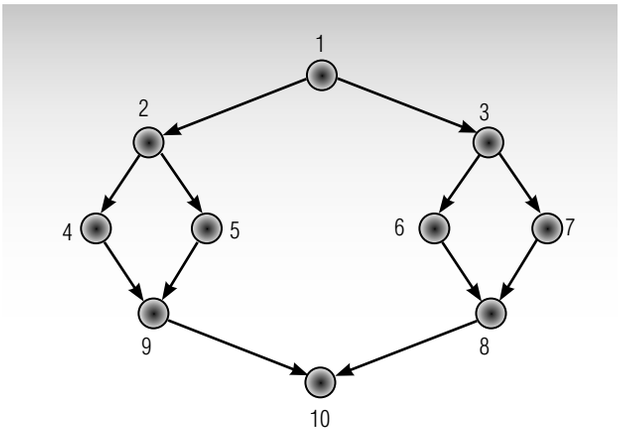


Рис. 1. Пример диаграммы *tG*

$$ISC(tG/B)=w_1ISC(b_1)+w_2ISC(b_2)+...+w_{k1}ISC(b_{k1});\ (1)$$

$$V_ISC(tG/B)=<w_1ISC(b_1);\ w_2ISC(b_2);...;w_{k1}ISC(b_{k1})>.(2)$$

Под длиной ПСС в базисе *B* будем понимать следующую величину

$$L^{(FSS)}(tG/B)=\sum_{i=1}^{k1}w_i\ .$$

Через $V_RSC(tG/B)$ обозначим вектор-индекс относительной сложности орграфа *tG* в базисе *B*, где

$$V_RSC(tG/B)=$$
$$=<w_1ISC(b_1)/ISC(tG);\ ...;w_{k1}ISC(b_{k1})/ISC(tG)>.$$

Вычисление значений $V_RSC(tG/B)$ позволяет определять суммарную относительную значимость фрагментов b_i базиса *B* для каждого типа *k* в общей сложности *tG*. В [9] предложен метод вычисления значений абсолютных и относительных вкладов *для каждого отдельного фрагмента графа* в его сложность с учетом значимости вершин и ребер в разных предметных областях.

Пусть $P=<P_0,P_1,...,P_i,...,P_{(n-1)}>$ обозначает базис путей, где P_i – путь с числом дуг *i*. Тогда для *tG* (см. *рис. 1*) вычислим:

$$SS(tG/B)=<1,\ 1,\ 1,\ 1,\ 1,\ 0,\ 0,\ 0,\ 0,\ 0>;$$

$$FSS(tG/B)=<10,\ 12,\ 12,\ 8,\ 4,\ 0,\ 0,\ 0,\ 0,\ 0>.$$

Согласно формулам (1) и (2) вычислим:

$$ISC(tG/B) = 10ISC(P_0)+12ISC(P_1)+$$
$$+12ISC(P_2)+8ISC(P_3)+4ISC(P_4)=826;$$
$$ISC(P_0/B)=1;\ ISC(P_1/B)=3;$$
$$ISC(P_2/B)=3ISC(P_0)+2ISC(P_1)=9;$$
$$ISC(P_3/B)=4ISC(P_0)+3ISC(P_1)+2ISC(P_2)=31;$$
$$ISC(P_4/B)=5ISC(P_0)+4ISC(P_1)+$$
$$+3ISC(P_2)+2ISC(P_3)=106.$$

В итоге, для *tG* получим:

$$ISC(tG/B)=826;\ V_ISC(tG/B)=<10;36;108;248;424>;$$
$$V_RSC(tG/B)=<0{,}0121;\ 0{,}0436;\ 0{,}1308;\ 0{,}3002;\ 0{,}5133>;$$
$$L^{(SS)}(tG/B)=5;\ L^{(FSS)}(tG/B)=46.$$

Заметим, что проблема определения всех простых путей всех возможных длин между парой вершин орграфа является *NP*-полной проблемой [12]. Выделим, что по разнице значений индексов СС-сложности легко определить *количественную меру сходства орграфов и построить граф их попарного сходства с последующей кластеризацией* [11]. Кроме того, по изменениям значений индексов СС-сложности можно *проводить мониторинг изменения структуры tG* для заданного промежутка времени и определять тенденции изменения сложности структуры.

4. Метод мониторинга изменений интегральных свойств *T*-орграфов на основе анализа моделей сложности

Рассмотрим задачу анализа изменения значений анализируемых моделей СС-сложности *tG* в диапазоне времени от t_1 до t_5 включительно. Пусть *tG* в момент времени t_1 имел структуру t_1G, а в следующие моменты времени $t_2 - t_5$ – соответственно $t_2G - t_5G$ (*рис. 2*). Рассмотрим наиболее сложный случай изменения структуры, когда число вершин и дуг (ребер) остается постоянным, а меняется структура связей вершин.

В *табл. 1* приведены результаты определения значений трех видов моделей СС-сложности.

Таблица 1.

Значения трех видов моделей СС-сложности

t	$L^{(SS)}(tG/B)$	$L^{(FSS)}(tG/B)$	$ISC(tG/B)$
t_1	4	37	1156
t_2	4	38	1112
t_3	4	35	847
t_4	4	33	785
t_5	4	34	644

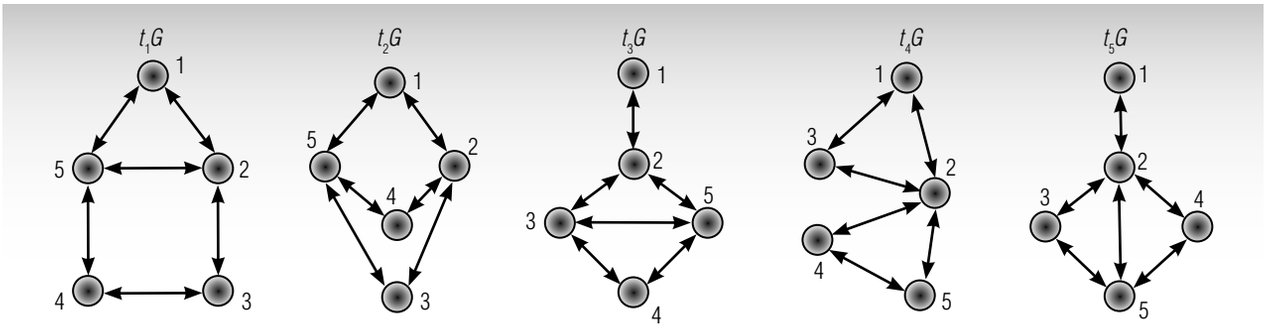


Рис. 2. Диаграммы *tG* в диапазоне времени от t_1 до t_5 включительно

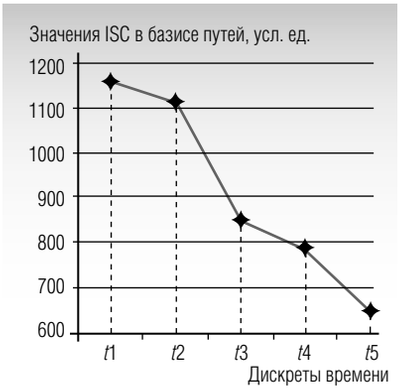


Рис. 3. График изменения значений индекса сложности

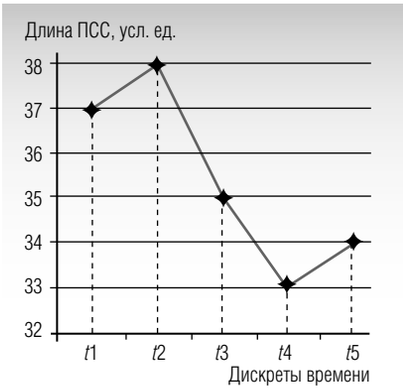


Рис. 4. График изменения значений длины полного структурного спектра

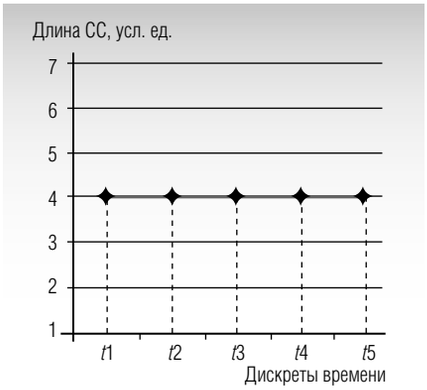


Рис. 5. График изменения значений длины структурного спектра

На *рис. 3-5* показаны графики изменения значений $ISC(tG/P)$, $L^{(FSS)}(tG/B)$, $L^{(SS)}(tG/B)$. Значения $ISC(tG/P)$ характеризуют *интегральную протяженность* и *разветвленность связей структуры tG* в заданный момент времени *t*.

Анализ графиков позволяет определять тенденции изменения значений моделей СС-сложности, а модуль разности дает количественную меру ее изменения и *меру сходства*. На *рис. 6, 7* приведены гистограммы изменения ПСС и значений компонентов вектор-индекса *T*-орграфа в базисе *P*.

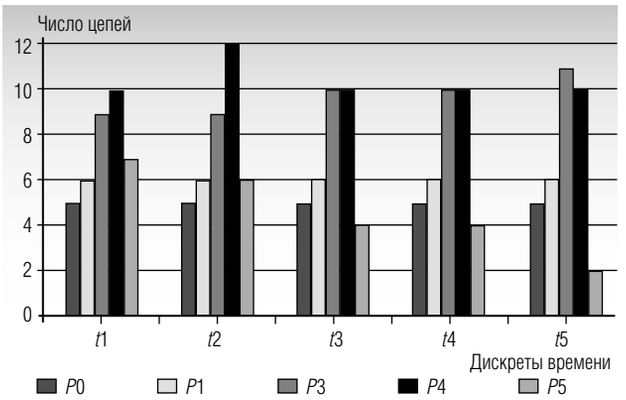


Рис. 6. Гистограммы изменения значений полного структурного спектра *tG* в базисе *P*

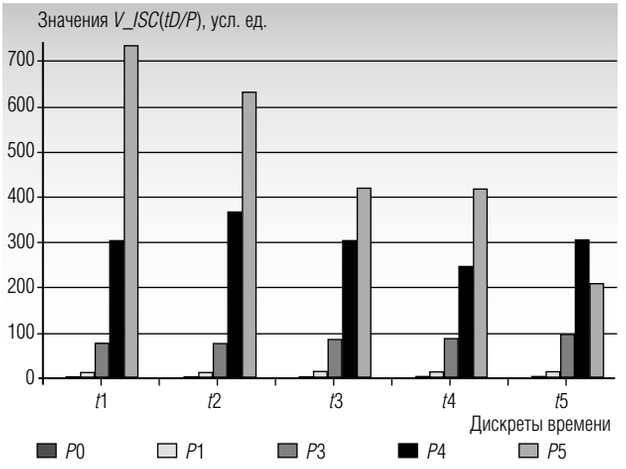


Рис. 7. Гистограммы изменения значений компонентов вектор-индекса *tG* в базисе *P*

Анализ данных (*рис. 7*) приводит к выделению тенденции уменьшения СС-сложности *T*-орграфа. Вклад цепей длины 2 и 3 возрос, что привело к увеличению разветвленности связей вершин, а вклад цепей длины 4, как максимально значимых в вычислении общей сложности, снизился, что привело к уменьшению общей протяженности *T*-орграфа.

Если вместо путей использовать базис контуров, то можно определять и анализировать тенденции изменения диапазона $diapC(tG) = (l_{max} - l_{min})$ охвата контурной связностью вершин *tG*, где l_{max} – максимальная, а l_{min} – минимальная длина контура (цикла). Базис каркасов более точно характеризует разветвленность, а базис клик – плотность *tG*. Чем больше клик в *tG*, тем больше максимально устойчивых по связям подмножеств вершин (акторов в КСС). Наиболее точно характеризует СС-сложность базис всех связных фрагментов *tG*. Однако, как показали вычислительные эксперименты для *tG* с $p > 40$, формирование такого базиса и вычисление значений моделей сложности практически невыполнимо в приемлемое время.

5. Метод мониторинга изменения локальных свойств, вычисленных на неразрывной структуре Т-орграфа

Ниже в качестве локальных свойств рассматриваются вклады вершин ($irc(v^{(t)}/P)$) и вклады дуг (ребер) ($irc(e^{(t)}/P)$) в сложность *tG* , вычисленные по формулам, приведенным в [13]. Значения вкладов вершин и усредненных нормализованных значений вкладов вершин ($irc_{av}(v/P)$) приведены в *табл. 2* и вычислялись на неразрывной структуре *tG*. Для определения тенденций изменения на *рис. 8* приведены гистограммы значений $irc(v^{(t_1)}/P)$ и $irc_{av}(v/P)$. Для вершин v_1, v_3, v_4, v_5 наметилась тенденция увеличения вкладов (значимости), для v_2– уменьшения вклада. Вершины по их значимости получили следующую упорядоченность $v_1 > v_4 > v_3 > v_5 > v_2$. Аналогичные результаты для ребер $e_{ij} = (v_i, v_j)$ ($irc(e^{(t)}/P) = irc1$, $irc(e/P) = irc2$) приведены *в табл. 3* и на *рис. 9*.

Для ребер e_{12}, e_{45} выделилась тенденция увеличения вкладов, для e_{15}, e_{23}, e_{34} – уменьшения вкладов, а для e_{25} – почти стабильности значений вкладов. Ребра по уменьшению значимости получили следующую упорядоченность

$$e_{45} > e_{12} > e_{23} > e_{34} > e_{35} > e_{13} > e_{15} > e_{24} > e_{25} .$$

Таблица 2.

Значения нормализованных вкладов вершин и усредненных нормализованных вкладов вершин в сложность *tG*

V(tG)	$irc(v^{(t_1)})$	$irc(v^{(t_2)})$	$irc(v^{(t_3)})$	$irc(v^{(t_4)})$	$irc(v^{(t_5)})$	$irc_{av}(v)$
v_1	**0,2560**	0,2661	0,3370	0,2420	0,2857	**0,27736**
v_2	**0,1522**	0,1007	0,1009	0,0318	0,0869	**0,09450**
v_3	**0,2197**	0,2661	0,1505	0,2420	0,2437	**0,22440**
v_4	**0,2197**	0,2661	0,2609	0,2420	0,2437	**0,24648**
v_5	**0,1522**	0,1007	0,1505	0,2420	0,1397	**0,15702**

Таблица 3.

Значения нормализованных вкладов ребер и усредненных нормализованных вкладов ребер в сложность *tG*

E(tG)	$irc(e^{(t_1)})$	$irc(e^{(t_2)})$	$irc(e^{(t_3)})$	$irc(e^{(t_4)})$	$irc(e^{(t_5)})$	$irc_{av}(e)$
e_{12}	**0,1928**	0,1666	0,3396	0,0666	0,2887	**0,21086**
e_{13}	**–**	–	–	0,3666	–	**0,07332**
e_{15}	**0,1928**	0,1666	–	–	–	**0,07188**
e_{23}	**0,1602**	0,1666	0,0801	0,0666	0,1056	**0,11582**
e_{24}	**–**	0,1666	–	0,0666	0,1056	**0,06776**
e_{25}	**0,0451**	–	0,0801	0,0666	0,0399	**0,04634**
e_{34}	**0,2484**	–	0,2191	–	–	**0,09350**
e_{35}	**–**	0,1666	0,0617	–	0,2300	**0,09166**
e_{45}	**0,1602**	0,1666	0,2191	0,3666	0,2300	**0,22850**

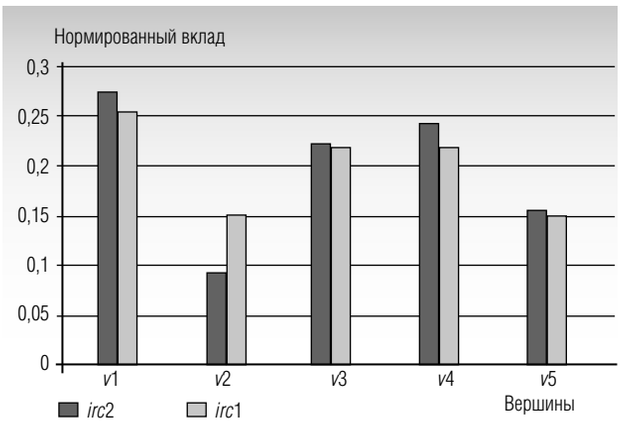


Рис. 8. Гистограммы значений нормализованных вкладов вершин в момент времени t_1 (*irc*1) и усредненных нормализованных вкладов вершин (*irc*2) в сложность *tG*

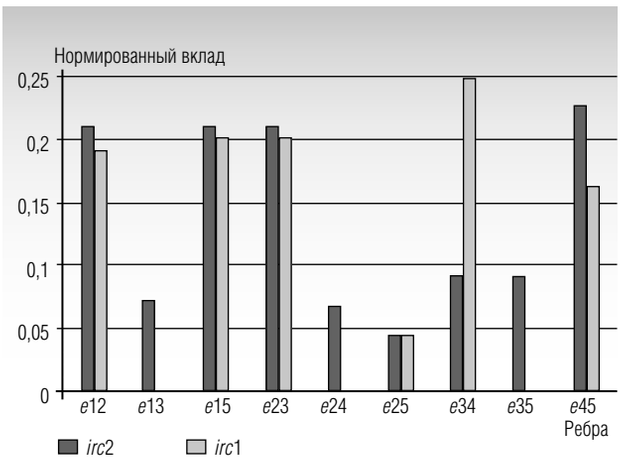


Рис. 9. Гистограммы значений нормализованных вкладов ребер в момент времени t_1 (*irc*1) и в усредненных нормализованных вкладов ребер (*irc*2) в сложность *tG*

6. Метод мониторинга изменения локальных свойств, вычисленных на структуре *T*-орграфа, разделенной на части

Введем новые характеристики, необходимые для проведения исследования динамики изменения значимости вершин и дуг *T*-орграфов по устойчивости к изменениям структуры.

Определение 1. Индексом абсолютного вклада вершины $v \in V$ в сложность орграфа *tG* в базисе *B* называется величина, вычисляемая по формуле $iac(v, tG / B) = ISC((tG - v) / B)$.

Определение 2. Индексом абсолютного вклада всех вершин в сложность орграфа *tG* в базисе *B* называется величина, вычисляемая по формуле

$$IAC(V, tG / B) = \sum_{v \in V} iac(v, tG / B).$$

Определение 3. Индексом относительного вклада вершины $v_j \in V$ в сложность орграфа *tG* в базисе *B* называется величина, вычисляемая по формуле

$$irc(v_j, tG / B) = ISC((tG - v_j) / \sum_{i=1}^{p} ISC(tG - v_i / B).$$

На *рис. 10* приведены гистограммы значений $irc(v, t_1G / B) = irc1$ и средних значений $irc_{av}(v, tG / B) = irc2$, позволяющие выделить тенденции изменения вкладов.

При удалении вершины *v* происходит разрушение орграфа. Чем сильнее разрушение, т.е. чем меньше значение $irc(v, tG / B)$, тем больше значимость вершины *v* в структуре *tG*. Из анализа данных (*рис. 10*) следует, что значимость вершин v_1, v_5

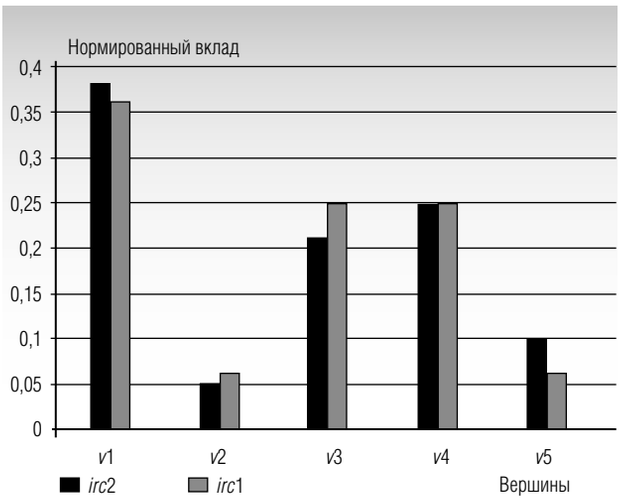


Рис. 10. Гистограммы значений нормализованных вкладов вершин в момент времени t_1 (*irc*1) и усредненных значений нормализованных вкладов вершин (*irc*2) в сложность *tG*

уменьшилась, а остальных вершин увеличилась. Вершины по их значимости получили следующую упорядоченность $v_2 > v_5 > v_3 > v_4 > v_1$. Метод определения $irc(v, tG / P)$ является методом определения классов эквивалентного расположения вершин в *tG*. Предельным числом таких классов является число орбит группы $Aut(tG)$.

Определим новые характеристики для вычисления значимости дуг.

Определение 4. Индексом абсолютного вклада дуги $e_j \in E$ в сложность орграфа *tG* в базисе *B* называется величина, вычисляемая по формуле

$$iac(e_j, tG/B) = ISC((tG - e_j)/B.$$

Определение 5. Индексом абсолютного вклада всех дуг в сложность орграфа *tG* в базисе *B* называется величина, определяемая как

$$IAC(E, tG/B) = \sum_{e_j \in E} iac(e_j, tG/B).$$

Определение 6. Индексом относительного вклада дуги $e_j \in E$ в сложность орграфа *tG* в базисе *B* называется величина, вычисляемая по формуле

$$irc(e_j, tG/B) = ISC(tG - e_j)/\sum_{i=1}^{q} ISC(tG - e_i/B).$$

Метод определения $irc(e, tG/B)$ является методом решения задачи определения классов эквивалентного расположения дуг в *tG*. Предельным числом таких классов является число орбит группы $Aut^{(k)}(tG)$, где $k=p_1$. В *табл. 4* приведены значения $irc(e, tG/P)$ в анализируемые моменты времени, а на *рис. 11* – гистограммы значений $irc(e, t_1G/P)=irc1$ для каждой дуги *tG* и средних значений по интервалу $t_1 - t_5$ нормализованных вкладов $irc_{av}(e, tG/P)=irc2$.

Из анализа данных (*табл. 4, рис. 11*) следует, что для ребер e_{12}, e_{23} выделилась тенденция уменьшения вкладов, для e_{15}, e_{25}, e_{34}, e_{45} – увеличения вкладов. Упорядоченность ребер по их значимости имеет вид $e_{34} > e_{13} > e_{15} > e_{35} > (e_{24}, e_{45}) > e_{12} > e_{25} > e_{23}$.

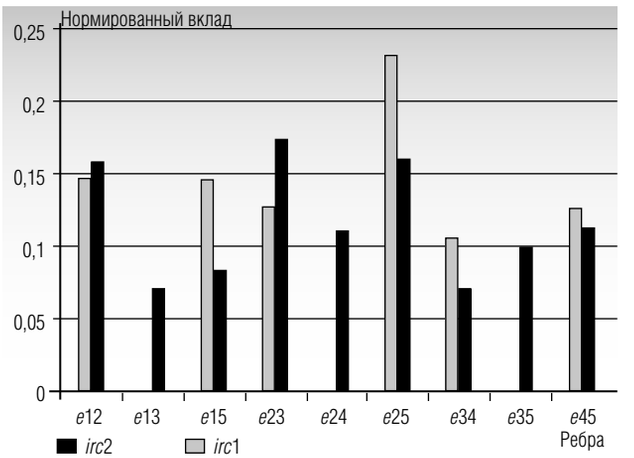


Рис. 11. Гистограммы значений нормализованных вкладов ребер в момент времени t_1 и средних нормализованных вкладов ребер в сложность *tG*

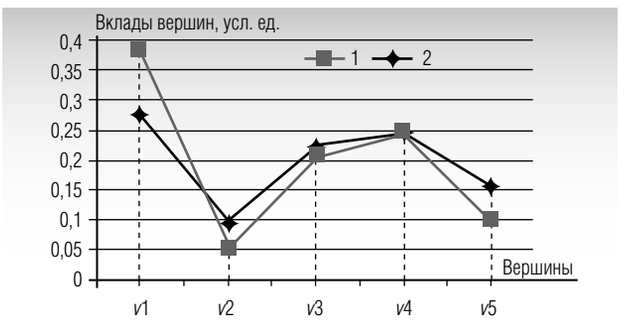


Рис. 12. Графики изменения значений нормализованных вкладов вершин *tG* по двум подходам

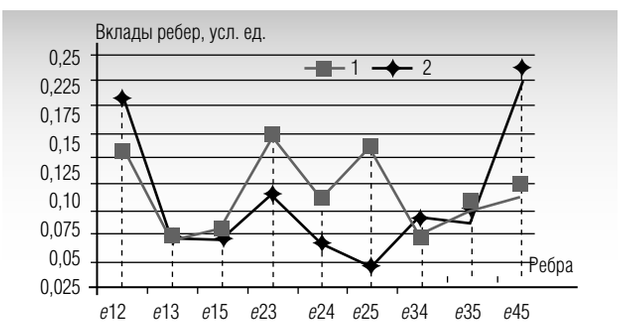


Рис. 13. Графики изменения значений нормализованных вкладов ребер *tG* по двум подходам

Таблица 4.

Значения нормализованных вкладов ребер в анализируемые моменты времени и средних нормализованных вкладов ребер в сложность *tG*

t	e_{12}	e_{13}	e_{15}	e_{23}	e_{24}	e_{25}	e_{34}	e_{35}	e_{45}
t_1	0,147	–	0,147	0,128	–	0,234	0,107	–	0,128
t_2	0,167	–	0,167	0,167	0,167	–	–	0,167	0,167
t_3	0,123	–	–	0,182	–	0,182	0,152	0,209	0,152
t_4	0,198	0,3666	0,104	0,198	0,198	0,198	0,104	–	–
t_5	0,160	–	–	0,197	0,197	0,197	–	0,124	0,124
$irc_{av}(e)$	**0,1589**	**0,0733**	**0,0836**	**0,1743**	**0,1124**	**0,1622**	**0,0728**	**0,0999**	**0,1142**

7. Сравнение двух подходов к вычислению вкладов фрагментов в сложность

На *рис. 12, 13* показаны графики изменения значений вкладов $irc_{av}(v, tG/P)$, вычисленных по двум подходам.

Содержательно, вклад фрагмента по второму подходу характеризует степень значимости разрушения структуры после его удаления, а по первому подходу – степень значимости в общей протяженности и разветвленности по путям, включающим анализируемый фрагмент. В *табл. 5* приведена матрица достроек цепей p_2 до фрагментов базиса P для t_1G и значения вкладов цепей $irc(p_2^{(i)}, tG/P)$, вычисленных по двум подходам.

Таблица 5.

Значения матрицы достроек и значения нормализованных вкладов цепей по двум подходам

№	Цепи $p_2^{(i)}$	Базис			$irc(p_2^{(i)}, tG/P)$	
		P_2	P_3	P_4	Подход 1	Подход 2
1	234	1	3	3	0,1893	0,1143
2	345	1	3	3	0,1893	0,1143
3	215	1	2	2	0,1289	0,1636
4	154	1	2	2	0,1289	0,1069
5	123	1	2	2	0,1289	0,1069
6	125	1	1	1	**0,0684**	**0,0985**
7	152	1	1	1	0,0684	0,0985
8	325	1	3	0	0,0490	0,0985
9	254	1	3	0	**0,0490**	**0,0985**

Анализ результатов (*рис. 12-15, табл. 5*) и результатов объемных вычислительных экспериментов для случайных орграфов привели к следующим выводам:

1) два подхода вместе дают более точный анализ значимости фрагментов, характеризуемых в более широком аспекте;

2) существуют бесконечные семейства орграфов, для которых удаление фрагментов из разных орбит групп $Aut^{(k)}(tG)$, где $k = P_0, P_1, ..., P_{(p-1)}$, приводит к изоморфным частям орграфов и, следовательно, значения вкладов по второму подходу одинаковые там, где они должны быть разными;

3) результаты вычисления вкладов фрагментов tG по первому подходу являются более точными, так как чаще приводят к определению орбит групп $Aut^{(k)}(tG)$, где $k = P_0, P_1, ..., P_{(p-1)}$;

4) время решения задач по второму подходу больше, чем по первому подходу. Так, например, при

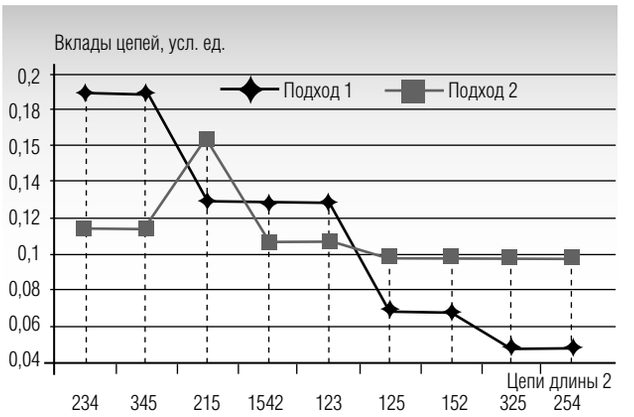


Рис. 14. Значения нормализованных вкладов цепей по двум подходам

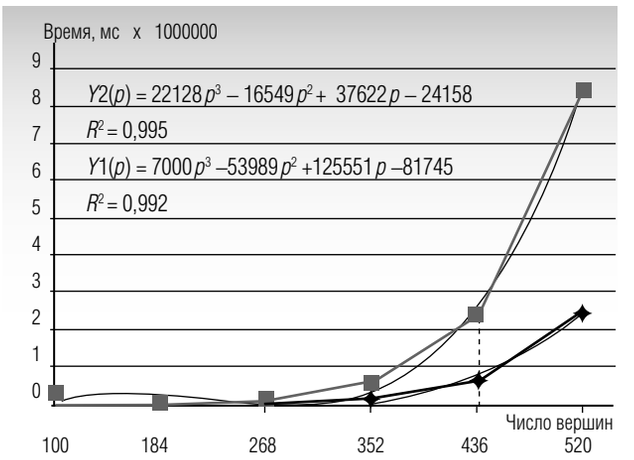


Рис. 15. Экспериментальные оценки вычислительной сложности решения задач по двум подходам

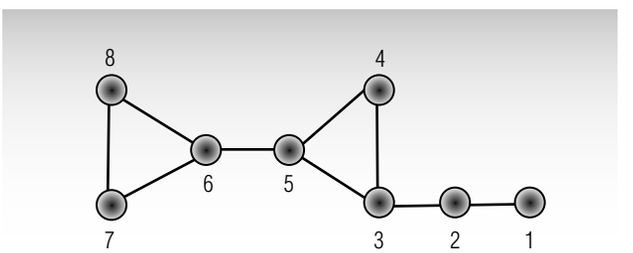


Рис. 16. Пример графа с изоморфными примарными подграфами для неавтоморфных вершин

анализе орграфов с числами $100 \le p < 600$ и средней степенью (суммой полустепеней исхода и захода) равной четырем, экспериментальные оценки времени вычисления $irc(v, tG/P_{0\text{-}6})$ имеют вид, представленный на *рис. 15.*

На *рис. 16* приведен пример графа, удаление вершин v_3 или v_6, принадлежащих разным орбитам $Aut(G)$, приводит к изоморфным примарным под-

графам. Для графа t_1G (*рис. 2*) удаление цепей 125 или 254, принадлежащих разным орбитам группы $Aut^{(P_2)}(tG)$, приводит к изоморфным частям, а значит и к одинаковым значениям вкладов этих цепей (*см. табл. 5*).

8. Заключение

В работе рассмотрены новые методы структурного анализа *T*-орграфов. Они позволяют исследовать как локальные, так и глобальные структурные свойства *T*-орграфов, выявить тенденции их изменения во времени. Сравнение двух подходов к вычислению вкладов фрагментов в общую сложность привело к выделению первого подхода, как наиболее точного. Предложенные модели сложности можно рассматривать как систему иерархического уточняющего анализа, позволяющего решать задачи исследования динамических свойств *T*-орграфов, в частности, монотонности их изменения. Методы анализа *T*-орграфов реализованы в АСНИ «Мастерская граф-моделей» [13] и используются в учебном процессе НИУ ВШЭ и НИУ МЭИ. ■

Литература

1. Алгоритмы и программы решения задач на графах и сетях / М.И.Нечепуренко, В.К.Попков, В.А.Кохов и др. – Новосибирск: Наука, 1990.
2. Динамический подход к анализу структур, описываемых графами (основы графодинамики) / М.А.Айзерман, Л.А.Гусев, С.В.Петров, И.М.Смирнова // АиТ. – 1977. – № 7. – С. 136-151.
3. Wang H., Aggarwal C.C. Managing and Mining Graph Data. – Springer, 2010.
4. Jackson M.O. Social and Economic Networks. – Princeton University Press, 2008.
5. Берштейн Л.С., Боженюк А.В. Использование темпоральных графов как моделей сложных систем // Известия ЮФУ. Технические науки. – 2010. – № 4 (105). – С. 198-203.
6. Bramsen P.J. Doing Time: Inducing Temporal Graphs. Technical report. – Massachusetts Institute of Technologies, 2006.
7. Mekenyan O., Bonchev D. Topological indices for molecular fragments and new graph invariants // Journal of Mathematical Chemistry. – 1988. – № 3. – P. 347-375.
8. Skorobogatov V.A., Dobrynin A.A. Metric analysis of graphs // MATCH – Communications in Mathematical and in Computer Chemistry. – 1988. – № 23. – P. 105-151.
9. Кохов В.А. Концептуальные и математические модели сложности графов. – М: Изд-во МЭИ, 2002.
10. Кохов В.А. Граф-модели для анализа сходства структур на основе их сложности / Одиннадцатая национальная конференция по искусственному интеллекту с международным участием. КИИ-2008: Труды конференции. В 3-х т. – Том 2. – М.: ЛЕНАНД, 2008. – С. 70-78.
11. Кохов В.А., Кохов В.В. Метод решения задачи различения орграфов на основе сложности // Бизнес-информатика. – 2011. – № 1 (15). – С. 11-23.
12. Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи: Пер. с англ. – М.: Мир, 1982.
13. Kokhov V.A., Neznanov A.A., Tkachenko S.V. Integrated research environment «Graph Model Workshop» // Fourth conceptual structures tool interoperability workshop (CS-TIW'09). – Moscow, Russia, 2009. – P. 7-12.

ПРОГРАММНЫЙ КОМПЛЕКС ДЛЯ УПРАВЛЕНИЯ СТРУКТУРОЙ ИНВЕСТИЦИОННОГО ПОРТФЕЛЯ

А.Г. Исавнин,
доктор физико-математических наук, профессор кафедры математических методов в экономике Казанского федерального университета, филиал в г. Набережные Челны

Д.Р. Галиев,
ассистент кафедры математических методов в экономике Казанского федерального университета, филиал в г. Набережные Челны

Адрес: г. Набережные Челны, бульвар Строителей, д. 1 (5/10)
E-mail: isavnin@mail.ru, damir.galiev@mail.ru

В настоящей статье рассматривается разработанное программное обеспечение для управления структурой портфеля ценных бумаг. Описаны реализованные модели и алгоритмы, а также приведены результаты экспериментов на данных российского фондового рынка. Рассмотрены преимущества и недостатки, альтернативные методы измерения риска и доходности. Разработанное приложение позволяет подключаться к торгам посредством взаимодействия с популярными торговыми терминалами.

Ключевые слова: инвестиционный портфель, моделирование, генетические алгоритмы, нейронные сети.

1. Математические модели выбора оптимального инвестиционного портфеля

Формирование оптимального инвестиционного портфеля является весьма актуальной задачей, как для представителей бизнеса (институциональных инвесторов), так и для частных инвесторов. Применение количественных методов позволяет повысить качество инвестиционного портфеля. В большинстве современных аналитических программ отсутствует возможность создавать продвинутые количественные модели выбора оптимального инвестиционного портфеля. В настоящей статье описывается разработанный нами программный комплекс для построения оптимальных инвестиционных портфелей с использованием нескольких оптимизационных моделей.

Рассмотрим некоторые математические модели выбора оптимального инвестиционного портфеля, представленные в программном обеспечении. Пусть $x=(x_1,x_2,...,x_n)$ – вектор, определяющий структуру инвестиционного портфеля, n – количество активов в портфеле. При этом

$$\sum_{i=1}^{n} x_i = 1 .$$

Доходности активов являются случайными величинами и характеризуются вектором

$$\overline{R}=(R_1,R_2,...,R_n).$$

Очевидно, что доходность инвестиционного портфеля

$$R=\sum_{i=1}^{n} R_i x_i .$$

Основной идеей большинства моделей выбора оптимального инвестиционного портфеля является максимизация доходности при ограниченном риске или минимизация риска при ограниченной доходности. В качестве мер риска могут использоваться такие показатели как [1]:

1. Value-at-Risk (*VaR*)
2. Conditional Value-at-Risk (*CVaR*)
3. Mean Semi-Absolute Deviation (*MSAD*)
4. Нижние частные моменты (*LMP*) и, как частный случай, полудисперсия – нижний частный момент порядка 2
5. Просадка как мера риска (*CDaR*)

В одной из наших предыдущих работ были описаны результаты численных экспериментов на российском фондовом рынке с вышеперечисленными мерами риска [2]. Итак, рассмотрим следующие модели:

✦ 1. Классическая модель Г. Марковица.

$$\begin{aligned} &\min_x \sum_{i=1}^{n}\sum_{j=1}^{n} x_i V_{ij} x_j , \\ &\sum_{i=1}^{n} x_i MR_i = m_p , \\ &\sum_{i=1}^{n} x_i = 1, \end{aligned} \qquad (1)$$

где V_{ij} – матрица ковариаций доходностей, MR_i – средняя доходность актива i, m_p – заданная пользователем ожидаемая доходность. Целевая функции представляет собой среднее квадратическое отклонение (СКО) доходности портфеля. Минимизируя СКО, можно снизить и возможное отклонение доходности в положительную сторону, что является одним из недостатков данной модели.

✦ 2. Модель максимизации доходности при ограниченном Value-at-Risk (*VaR*).

$$\begin{aligned} &\max_x \sum_{i=1}^{n} R_i x_i , \\ &VaR_\alpha = -\mu_p - z_\alpha \sigma_p , \\ &\sum_{i=1}^{n} x_i = 1, \\ &l_i \le x_i \le u_i , \ i=\overline{1,n} , \end{aligned} \qquad (2)$$

где l_i и u_i отражают дополнительные ограничения на долю i-го актива в портфеле, VaR_α – заданный пользователем уровень Value-at-Risk. Значение *VaR* уровня $1-\alpha$ (VaR_α) определяется как

$$P\left(R_p \le -VaR_\alpha\right) = \alpha \qquad (3)$$

Совершив ряд преобразований, получим:

$$P\left(R_p \le -VaR_\alpha\right) = \alpha \Leftrightarrow P\left(\frac{R_p-\mu_p}{\sigma_p} \le \frac{-VaR_\alpha-\mu_p}{\sigma_p}\right) = \alpha \Leftrightarrow$$

$$\frac{-VaR_\alpha-\mu_p}{\sigma_p} = z_\alpha \Leftrightarrow VaR_\alpha = -\mu_p - z_\alpha \sigma_p , \qquad (4)$$

где R_p – доходность портфеля, μ_p – средняя доходность портфеля, σ_p – риск портфеля (СКО), z_α – квантиль стандартного нормального распределения порядка α. При этом полагается, что доходность портфеля имеет распределение, близкое к нормальному. Так, *VaR* уровня 95% (т.е. при $\alpha = 0{,}05$) будет определяться как $VaR_{0,05} = -\mu_p + 1{,}6449\,\sigma_p$.

✦ 3. Модель минимизации *MSAD* при ограниченной снизу доходности и наличии дополнительных ограничений на структуру портфеля.

$$\begin{aligned} &\min_x M\left(\left|min\{0, R - MR\}\right|\right), \\ &\sum_{i=1}^{n} R_i x_i \ge R_0 , \\ &\sum_{i=1}^{n} x_i = 1, \\ &l_i \le x_i \le u_i , \ i=\overline{1,n}, \end{aligned} \qquad (5)$$

где R_0 – константа, которая задается пользователем модели и означает нижний предельный уровень доходности портфеля; l_i и u_i отражают дополнительные ограничения на долю i-го актива в портфеле; M – знак математического ожидания. Целевая функция $M\left(\left|min\{0, R\text{-}MR\}\right|\right)$ определяет среднее полуабсолютное отклонение доходности портфеля (Mean Semi-Absolute Deviation – *MSAD*).

✦ 4. Модель максимизации асимметрии доходности при ограниченной снизу доходности и наличии дополнительных ограничений на структуру портфеля.

$$\max_x \frac{M\left[(R-MR)^3\right]}{\sigma_R^3},$$
$$\sum_{i=1}^{n} R_i x_i \geq R_0,$$
$$\sum_{i=1}^{n} x_i = 1, \qquad (6)$$
$$l_i \leq x_i \leq u_i, \quad i=\overline{1,n},$$

где R_0 – константа, задается пользователем модели и означает нижний предельный уровень доходности портфеля; l_i и u_i отражают дополнительные ограничения в портфеле на актив i; σ_R – СКО доходности портфеля. Целевая функция

$$\frac{M\left[(R\text{-}MR)^3\right]}{\sigma_R^3}$$

определяет коэффициент асимметрии доходности портфеля, который обладает определенным экономическим смыслом. Если коэффициент асимметрии имеет положительное значение, то высокие значения доходности являются более вероятными (*рис. 1а*) и наоборот (*рис. 1б*). Следовательно, чем выше значение положительного коэффициента асимметрии, тем лучше сформированный портфель.

✦ 5. Модель максимизации эксцесса доходности при ограниченной снизу доходности и наличии дополнительных ограничений на структуру портфеля.

$$\max_x \left(\frac{M\left[(R-MR)^4\right]}{\sigma_R^4} - 3 \right),$$
$$\sum_{i=1}^{n} R_i x_i \geq R_0,$$
$$\sum_{i=1}^{n} x_i = 1, \qquad (7)$$
$$l_i \leq x_i \leq u_i, \quad i=\overline{1,n},$$

где R_0 – константа, которая задается пользователем модели и означает нижний предельный уровень доходности портфеля; l_i и u_i отражают дополнительные ограничения в портфеле на актив i; M – математическое ожидание; σ_R – СКО доходности портфеля. Целевая функция

$$\gamma = \left(\frac{M\left[(R-MR)^4\right]}{\sigma_R^4} - 3 \right)$$

определяет коэффициент эксцесса доходности портфеля. Эксцесс характеризует меру остроты пика распределения доходности портфеля. Острый

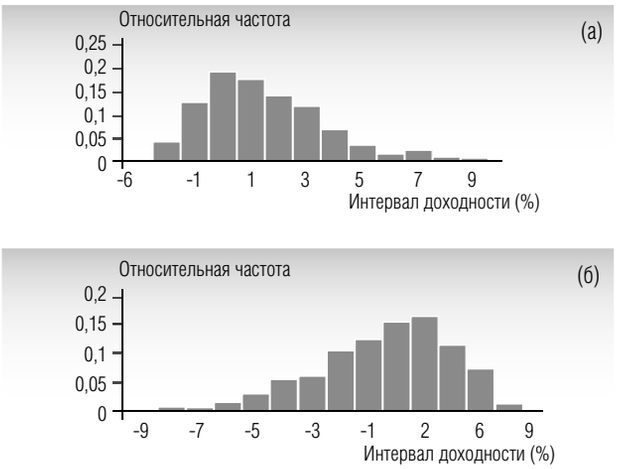


Рис. 1. Гистограммы доходности с разными коэффициентами асимметрии: (а) положительная асимметрия; (б) асимметрия отрицательна

пик $\gamma > 0$ говорит о преимущественном стремлении доходности к своему среднему значению (*рис. 2а*), что, как правило, говорит о меньшем значении волатильности портфеля σ_R, а, следовательно, о меньшем риске. Гладкий пик $\gamma < 0$ свидетельствует о тяжелых хвостах распределения и, следовательно, о повышенных рисках (*рис. 2б*).

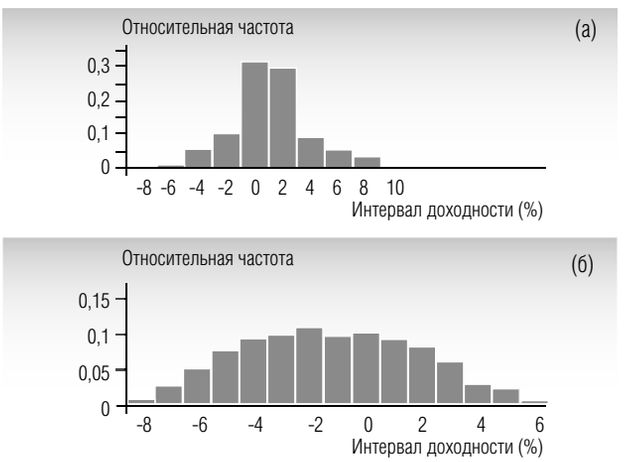


Рис. 2. Гистограммы доходности с разными коэффициентами эксцесса: (а) положительный коэффициент эксцесса; (б) отрицательный коэффициент эксцесса

В моделях (2-5) используется классический показатель эффективности – средняя доходность и прогнозы по нейросетевым и эконометрическим моделям.

Для решения оптимизационных задач применен метод генетических алгоритмов [3]. Генетические алгоритмы (ГА) относятся к классу эволюционных алгоритмов. В теории генетических алгоритмов принято использовать биологические термины. Например, хромосома – конкретное решение, представ-

ленное в виде битовой строки. Индивид, родитель, потомок являются хромосомами. Набор хромосом образует популяцию. Общий порядок работы генетического алгоритма выглядит следующим образом:

1. Инициализация популяции решения случайным образом.
2. Выбор части популяции (родителей) для порождения потомков. Процедура выполняется с помощью оператора селекции.
3. Применение оператора скрещивания.
4. Мутация новых решений (потомков).
5. Формирование новой популяции путем выбора лучших решений из множества родителей и потомков.
6. Повторение шагов 2-5 до тех пор, пока не будет выполнен критерий остановки.

На первом шаге инициализации определяются значения параметров генетического алгоритма: длина хромосомы, размер популяции и вероятность мутации. Для осуществления процедуры селекции можно использовать метод турнирной селекции [4]. Сначала пользователь определяет, так называемый, размер турнирной группы N_t и количество групп (обычно создают 5-6 групп). Далее случайным образом в каждую группу из популяции индивидов выбираются N_t индивидов. Из выбранных индивидов в каждой группе отбирается индивид с лучшим значением функции приспособления (например, максимальная доходность или минимальный риск). В качестве функции приспособления (фитнесс-функции) используется целевая функция оптимизационной задачи. Для учета ограничений можно применять метод штрафных функций. После чего лучшие индивиды от каждой группы подвергаются скрещиванию.

Для скрещивания индивидов используют следующий подход [5]. С помощью оператора селекции отбирают 2 индивида. Одноточечное скрещивание

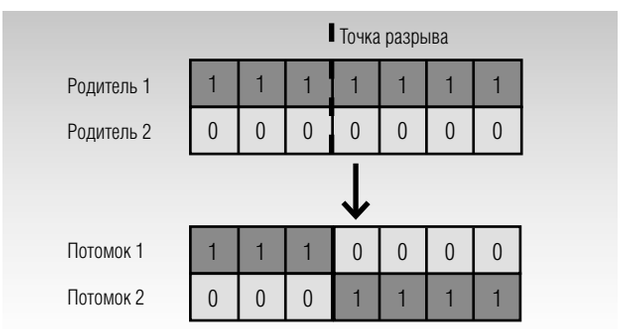


Рис. 3. Пример точечного скрещивания

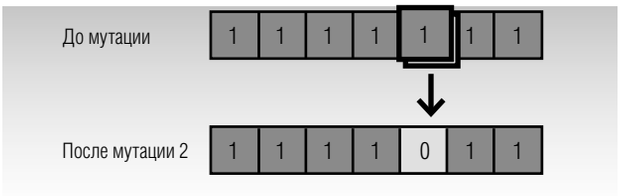


Рис. 4. Пример мутации

позволяет получить два новых индивида. Точка скрещивания выбирается случайно. Пример показан на *рис. 3*. Мутация осуществляется следующим образом: с выбранной вероятностью P, каждый бит индивида, полученного после процедуры скрещивания, изменяется на противоположный. Пример представлен на *рис. 4*.

Новая популяция формируется путем сортировки. То есть формируется промежуточная популяция путем объединения родителей и потомков. N лучших, с точки зрения функции приспособления, индивидов перейдут в новое поколение. Здесь N – размер популяции. Алгоритм останавливают после того, как будет выполнено заданное количество итераций. Обычно принято проводить $N_1 \times N_2$ итераций, где N_1 – максимальное число популяций, N_2 – размер популяций, изначально задаваемые пользователями.

2. Использование прогнозов доходностей активов при моделировании

В разработанном нами программном обеспечении были использованы два типа моделей прогнозирования временных рядов: модель авторегрессии-скользящего среднего (*autoregressive moving-average model, ARMA*) и нейросетевая модель прогнозирования временного ряда. Прогнозы доходности по этим моделям могут быть учтены при формировании оптимального инвестиционного портфеля. Рассмотрим их основную суть. Моделью *ARMA (p, q)*, где p и q целые числа, задающие порядок модели, называется следующий процесс генерации временного ряда y_t:

$$y_t = c + \sum_{i=1}^{p} \alpha_i y_{t-i} + \varepsilon_t - \sum_{j=1}^{q} \beta_j \varepsilon_{t-j}, \quad (8)$$

где c – константа, $\{\varepsilon_t\}$ – белый шум, т.е. последовательность независимых и одинаково распределённых случайных величин с нулевым средним, а α_i и β_j – действительные числа (авторегрессионные коэффициенты и коэффициенты скользящего среднего, соответственно) [6]. Для идентификации модели (нахождения значений p и q) анализируются автокорреляционная и частная автокорреляцион-

ная функции. Для оценки параметров модели α_i, β_j и c можно применить алгоритм Бокса-Дженкинса [7] или метод максимального правдоподобия [8]. В нашем приложении реализован последний метод.

Нейронные сети являются мощным эвристическим инструментом для решения широкого класса задач. В нашем программном комплексе алгоритм нейронных сетей применен для прогнозирования временного ряда. Математическая модель нейрона выглядит следующим образом (*рис. 5*). Эта модель была предложена Макколом и Питтсом [9] еще в 1943 г.

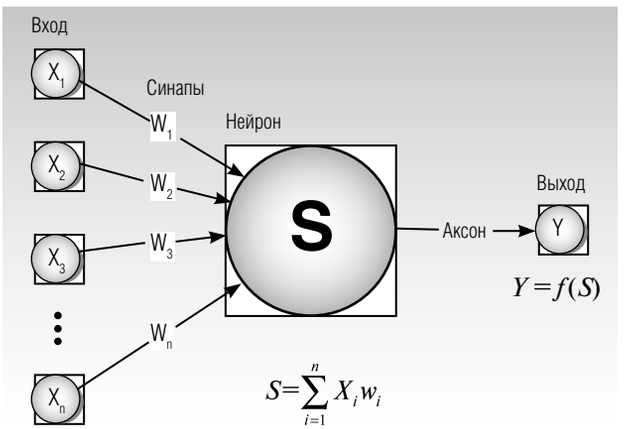


Рис. 5. Математическая модель нейрона

Сигналы X_1, X_2, ..., X_n, поступая на синапсы, преобразуются линейным образом

$$S=\sum_{i=1}^{n} X_i w_i ,$$

где w_i – веса соответствующих синапсов, n – размерность пространства входных сигналов. Далее в действие вступает нелинейная функция активации $Y=f(S)$. В нашем случае, в качестве функции активации была выбрана сигмоидная функция:

$$f(S)=\frac{1}{1+e^{(-bS)}}. \qquad (9)$$

Для обучения нейронной сети был использован алгоритм обратного распространения ошибки – это итеративный градиентный алгоритм обучения, который используется с целью минимизации среднеквадратичного отклонения текущего выхода и желаемого выхода многослойных нейронных сетей [10].

Архитектура двухслойной сети для прогнозирования временного ряда представлена на *рис. 6*. У рассматриваемой сети n входов и один выход. Несмотря на свою простоту, подобная архитектура широко используется для решения прикладных задач [11]. Каждый i-й нейрон первого слоя ($i=\overline{1,m}$) имеет n входов с весами w_{1i}, w_{2i}, ... w_{ni}. Заметим, что сумма весов не обязательно равна единице. После получения входных сигналов, нейрон суммирует их с соответствующими весами, затем применяет к этой сумме функцию активации и пересылает результат на вход нейрона второго (выходного) слоя. В свою очередь, нейрон выходного слоя суммирует полученные от второго слоя сигналы с некоторыми весами v_j, после чего результат вновь проходит через функцию активации.

Алгоритм обратного распространения ошибки можно описать следующим образом.

1. На этапе инициализации весам сети присваиваются начальные значения – некоторые малые случайные числа.

2. Выбирается очередная обучающая пара (X; Y) из обучающего множества; вектор X подается на вход сети. Значения входа X формируются методом «скользящего окна». К примеру, для временного ряда X_t обучающая пара с размером окна, равным 3, может выглядеть следующим образом: ((X_{t-1},X_{t-2},

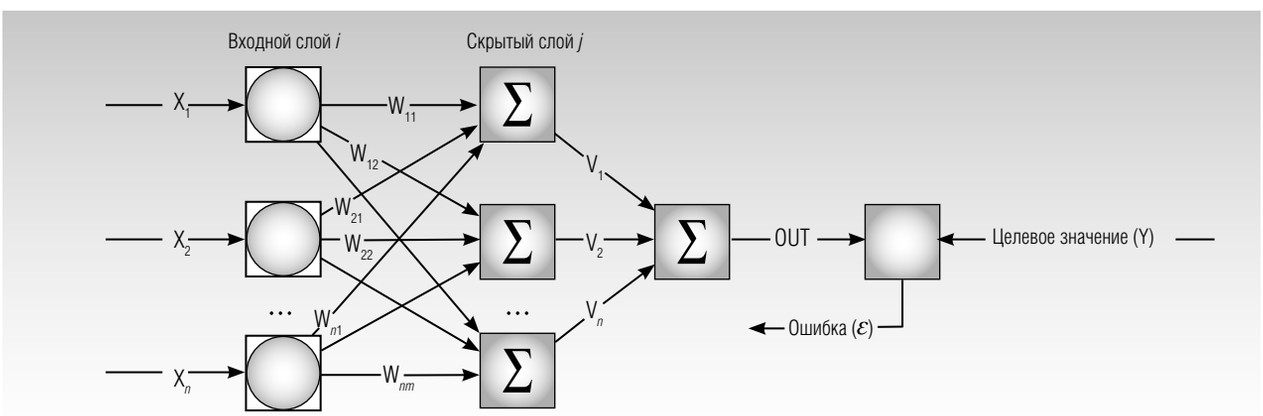


Рис. 6. Двухслойная нейронная сеть

X_{t-3}); X_t). Здесь в качестве целевого значения (Y) выступает реальное значение котировки (X_t) в момент t, а в качестве входа (X) «предыстория» в виде вектора (X_{t-1}, X_{t-2}, X_{t-3}).

3. Вычисляется выход сети.

4. Вычисляется разность между требуемым (целевым, Y) и реальным (вычисленным) выходом сети.

5. Веса сети корректируются так, чтобы минимизировать ошибку.

6. Шаги со 2-го по 5-й повторяются для каждой пары обучающего множества до тех пор, пока ошибка на всем множестве не достигнет приемлемой величины или не будет произведено допустимое число итераций.

Главной проблемой большинства моделей прогнозирования временных рядов является подбор горизонта обучения. Нейронные сети могут подгоняться под обучающую выборку и вне пределов этой выборки часто обладают не очень качественными предикативными свойствами. Рассмотренные и реализованные в программе модели прогнозирования временных рядов не являются исчерпывающими и лучшими в своем роде. Нашей основной целью является показать, как наличие более-менее качественных оценок будущих доходностей может повлиять на результат и качество инвестиционного портфеля.

3. Описание разработанного программного комплекса

Приложение разрабатывалось в среде Microsoft Visual Studio 2010 (Express edition) на языке C# 4.0. Продукт требует установленной среды Microsoft .Net 4, которая, в случае отсутствия, загружается автоматически с узла Microsoft во время установки. Приложение тестировалось и корректно работало на ОС Windows XP, Vista, Seven. Структура программы представлена на схеме.

Ядро системы состоит из модуля общих вычислений, модуля генетических алгоритмов оптимизации и модуля прогноза временных рядов. Информация об активах хранится в базе данных. Пользовательский интерфейс позволяет работать с данными, отображать графики и указывать настройки алгоритмов. Возможность подключения к реальным торгам осуществлена благодаря использованию свободно распространяемой библиотеки Stock# [12]. В разработанном приложении реализовано подключение к торгам посредством биржевого торгового терминала Quik [13]. Однако библиотека Stock# позволяет с малейшими изменениями в коде связать разрабатываемое приложение с другими торговыми терминалами: SmartCOM, Alfa-Direct, PlazaII и др. [14]. Механизм взаимодействия приложения с Quik представлен на *рис. 8.* Dynamic

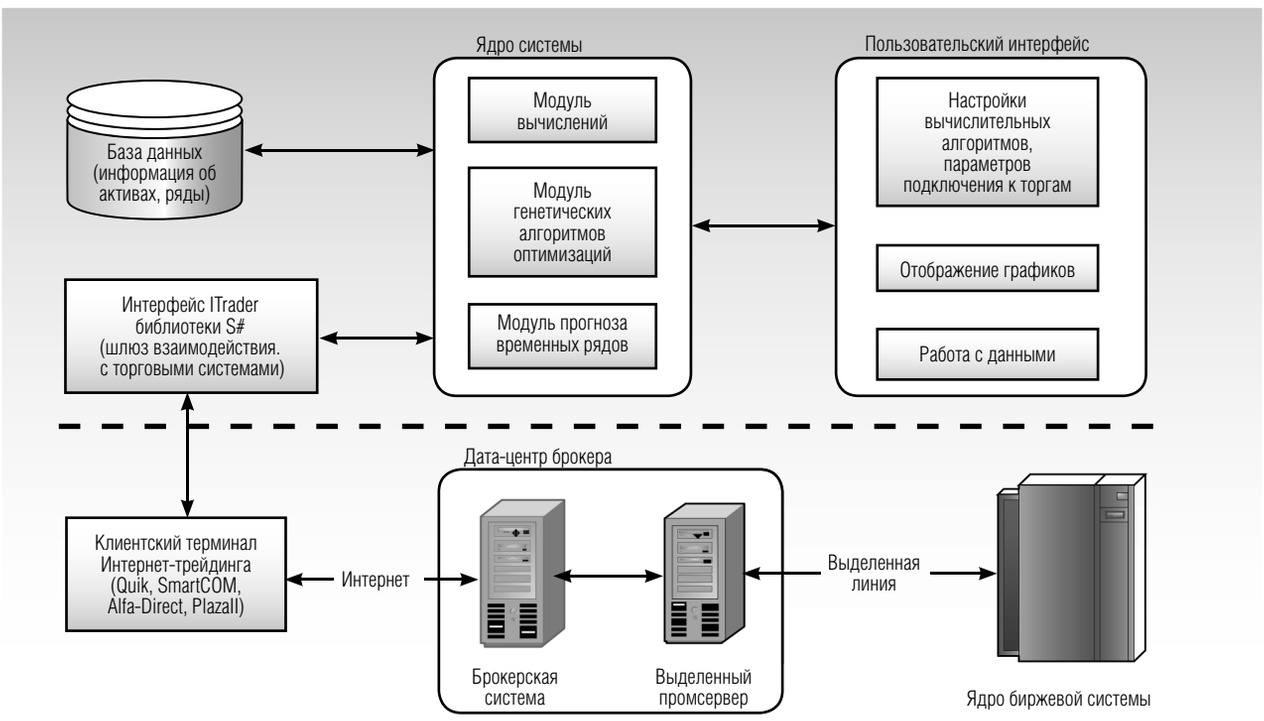


Рис. 7. Схема структуры программы



Рис. 8. Механизм взаимодействия Quik и приложения

Data Exchange (DDE) – это технология быстрой передачи данных, которую поддерживает Quik.

Рассмотрим процесс работы с программой. Главное окно предоставлено на *рис. 9*. Пользователю выводится список активов, из которых будет составлен инвестиционный портфель. Панель инструмента состоит из трех блоков. Первый блок отвечает за стандартные действия с файлами: создание нового файла, открытие, сохранение. Второй блок отвечает за управление списком активов: добавление, редактирование и удаление. В третьем блоке находятся средства редактирования настроек, средства расчета структуры портфеля по определенным моделям и средство вывода результатов вычислений.

При добавлении и редактировании активов отображается окно примерно одинаковой структуры (*рис. 10*). Пользователь вручную вводит название и тикер актива, а также временной ряд цен, который также можно вставить из буфера обмена. После чего рассчитывается временной ряд доходности, базовая статистика доходности актива, а также отображаются график цен и гистограмма распределения доходности. В меню редактирования можно осуществлять прогноз временного ряда по нейросетевой и ARMA моделям. Предусмотрена возможность ручного ввода стороннего прогноза (прогноза, полученного из других источников) посредством отдельного окна.

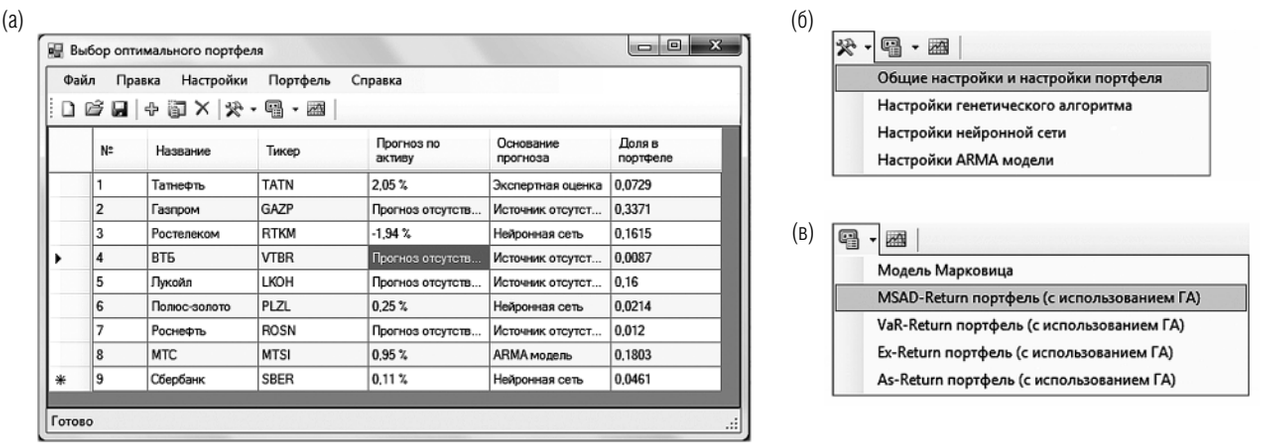


Рис. 9. Главное окно программы; (а) – главное окно; (б) – настройка программы; (в) – выбор модели

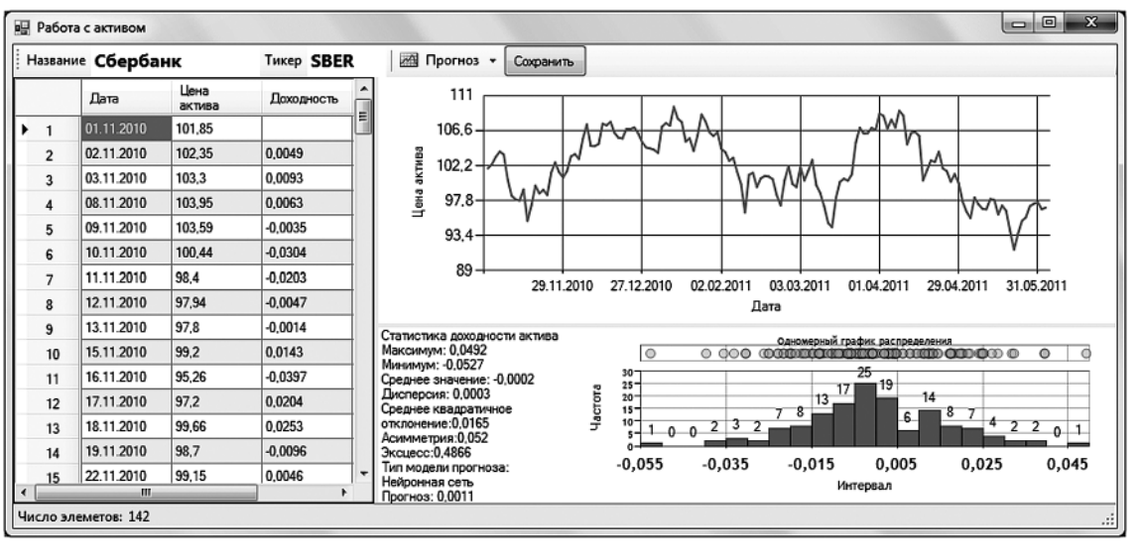


Рис. 10. Меню редактирования актива

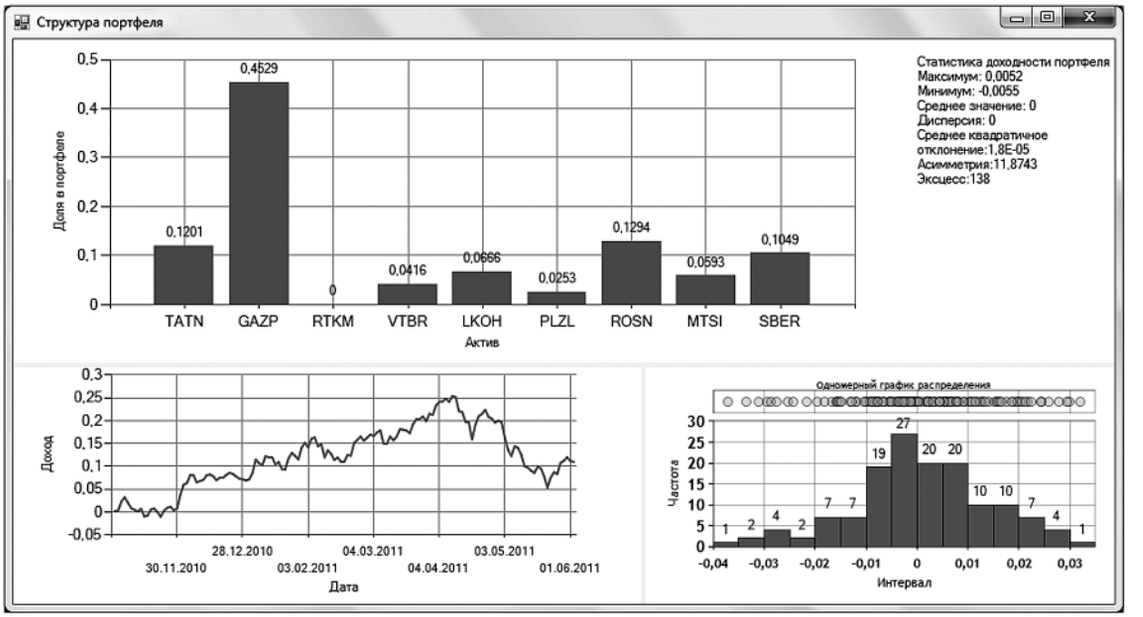


Рис. 11. Отображение результатов оптимизации

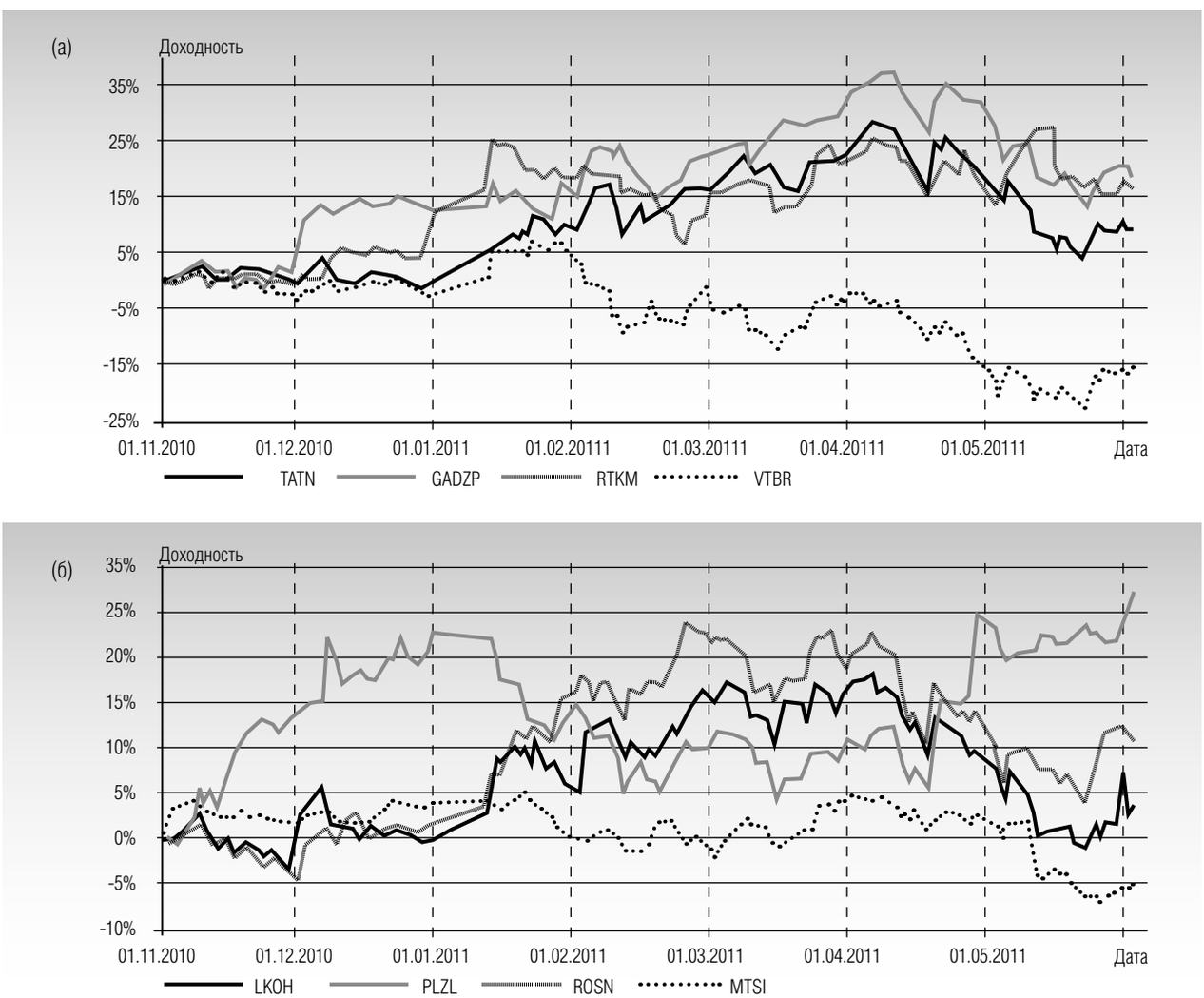


Рис. 12. Кривые доходностей активов

Результаты оптимизации структуры портфеля выводятся в отдельном окне (*рис. 11*). В верхней части окна отображена гистограммы структуры портфеля и базовая статистика исторической доходности портфеля. В левом нижнем углу отображается кривая доходности портфеля при вычисленной структуре. В правом нижнем углу выводится гистограмма распределения исторической доходности портфеля.

Структуру портфеля и информацию об активах можно экспортировать в Microsoft Excel.

4. Эксперименты на данных российского фондового рынка

Эксперименты проводились на данных российского фондового рынка – объединенной биржи ММВБ-РТС. Активы, с которыми проводились эксперименты: обыкновенные акции ОАО «Татнефть» (TATN), ОАО «Газпром» (GAZP), ОАО «Ростелеком» (RTKM), ОАО «ВТБ» (VTB), ОАО «Лукойл» (LKOH), ОАО «Полюс-Золото» (PLZL), ОАО «Роснефть» (ROSN), ОАО «МТС» (MTSI), ОАО «Сбербанк» (SBER). Временной период проведения экспериментов: 01.11.2010 – 01.06.2011. Кривые доходностей активов на рассматриваемом промежутке представлены на *рис. 12* (*а*) и (*б*).

При проведении экспериментов с реальными активами необходима была историческая информация о доходности активов. На каждом этапе (всего 14 этапов) рассматриваемого промежутка нейронная сеть и ARMA предварительно обучались на основе информации о доходности активов за предыдущие 3 месяца. Прогнозы моделей использовались в течение 2-х недель, после чего происходило их переобучение.

Как уже отмечалась, для оценки доходности кроме исторической средней использовались прогнозы на основе нейросетевой и ARMA моделей. Результаты прогнозирования с использованием модели ARMA представлены в *табл. 1.* Результаты для нейронной сети приведены в *табл. 2.* Информация в таблицах представлена по данным вне обучающей выборки. Здесь MSE (Mean Squared Error) - среднеквадратическая ошибка:

$$MSE=\frac{1}{n}\sum_{t=1}^{n} e_t^2, \qquad (10)$$

где e_t – разность фактического и реального значения для момента t, n – объемы выборки. Как видно из результатов, практически для всех активов нейронные сети демонстрируют более предпочтительные результаты по сравнению с ARMA моделью.

Таблица 1.

Результаты обучения ARMA моделей

Актив	Макс. ошибка, (%)	Мин. ошибка, (%)	Средняя ошибка, (%)	MSE
TATN	20.02	2.10	13.93	6.85e-2
GAZP	17.01	1.84	14.70	9.91e-2
RTKM	16.71	3.44	15.22	1.14e-2
VTBR	19.62	2.09	14.70	5.75e-2
LKOH	21.70	3.66	13.34	8.51e-2
ROSN	18.64	2.44	17.23	2.74e-2
MTSI	19.50	1.37	15.44	5.81e-2

Таблица 2.

Результаты обучения нейросетевых моделей

Актив	Макс. ошибка, (%)	Мин. ошибка, (%)	Средняя ошибка, (%)	MSE
TATN	18.33	0.71	11.97	5.76e-2
GAZP	18.38	0.95	8.98	8.71e-2
RTKM	14.97	0.45	10.10	1.64e-2
VTBR	17.35	1.46	14.77	4.63e-2
LKOH	16.05	1.04	15.39	7.15e-2
ROSN	18.32	0.69	12.53	1.93e-2
MTSI	12.89	2.35	12.06	4.34e-2

Для оценки соотношения риска и доходности портфелей использовался коэффициент Шарпа [15]. Этот показатель по-другому называют коэффициент «доходность-разброс» (reward-to-variability ratio):

$$RVAR=\frac{r_p-r_f}{\sigma_p}, \qquad (11)$$

где r_p – средняя доходность портфеля за рассматриваемый промежуток времени, r_f – среднее значение безрисковой ставки, σ_p – стандартное отклонение доходности портфеля (общий риск). В качестве безрисковой ставки использовалась доходность государственных облигаций. Мы также приводим статистику максимальной просадки доходности портфеля. Этот показатель в последнее время становится все более популярным. В *табл. 3* приведена статистика портфелей. Здесь характеристики портфеля в рамках модели Марковица рассчитывались только для случая с использованием средней исторической доходности (для сравнения методов оценки риска).

Что касается близости ожидаемых характеристик портфелей (риска и доходности) и реальных значений, то здесь результат напрямую зависит от использованной модели прогнозирования временных рядов доходностей. Особенно это существенно для доходности портфеля. Ожидаемые значения коэффициентов асимметрии и эксцесса отличались несильно от ожидаемых значений (максимальное отличие – 20,33%, минимальное – 2,87%, среднее

– 15,82%). В то же время значения мер риска VaR и MSAD отличались сильнее (максимальное отличие – 35,03%, минимальное – 3,11%, среднее – 21,23%).

Таблица 3.

Статистика портфелей

Модель	Доходность на рассматриваемом периоде, (%)	Максимальная просадка, (%)	Коэффициент Шарпа
1. Использование средней доходности			
1. Модель Марковица	7,8	5,10	1,25
2. VaR-Доходность	10,1	3,52	1,27
3. MSAD-Доходность	11,8	3,61	1,31
4. Асимметрия-Доходность	12,3	4,61	1,29
5. Эксцесс-Доходность	12,1	4,94	1,28
2. Использование прогноза доходности по модели ARMA			
1. VaR-Доходность	11,5	3,44	1,36
2. MSAD-Доходность	12,8	3,51	1,40
3. Асимметрия-Доходность	13,3	4,41	1,41
4. Эксцесс-Доходность	12,9	4,81	1,45
3. Использование прогноза доходности по нейросетевой модели			
1. VaR-Доходность	13,4	3,45	1,39
2. MSAD-Доходность	13,2	3,49	1,42
3. Асимметрия-Доходность	13,43	4,45	1,41
4. Эксцесс-Доходность	13,5	4,78	1,43

Динамика доходности некоторых портфелей представлена на *рис. 13* (*а,б*). В скобках указаны модели, которые были использованы для прогнозирования доходности.

По результатам можно сделать основной вывод, что использование прогнозов доходностей, особенно на основе нейронных сетей, позволяет повысить качество инвестиционных портфелей. Модели, в которых были использованы прогнозы доходностей активов, в большинстве случаев, обладают лучшим соотношением риска и доходности, по-сравнению с моделями, где в качестве доходности была использована историческая средняя доходность. Результативность напрямую зависит от качества выбранной модели прогнозирования. Что касается мер риска, то практически сохранились результаты, которые были ранее полученные в [2]. То есть мы ещё раз можем наблюдать, что портфели, оптимизированные по *VaR* и асимметрии, обладают лучшим соотношением риска и доходности. Также можно отметить тот факт, что модели «эксцесс-доходность» продемонстрировали довольно приемлемые результаты.

5. Заключение

Разработанное программное обеспечение позволяет решать задачу выбора оптимального инвестиционного портфеля с использованием современных специализированных моделей выбора оптимального инвестиционного портфеля. В приложение встроен модуль по прогнозу доходности рассматриваемых активов на основании ARMA- и нейросетевых моделей. Также есть возможность учитывать прогнозы экспертов и сторонних моделей. Комплекс обладает возможностью подключения к биржевым торгам, благодаря использованию библиотеки Stock#. Программа обладает достаточной гибкостью: можно реализовывать новые модели выбора портфеля, до-

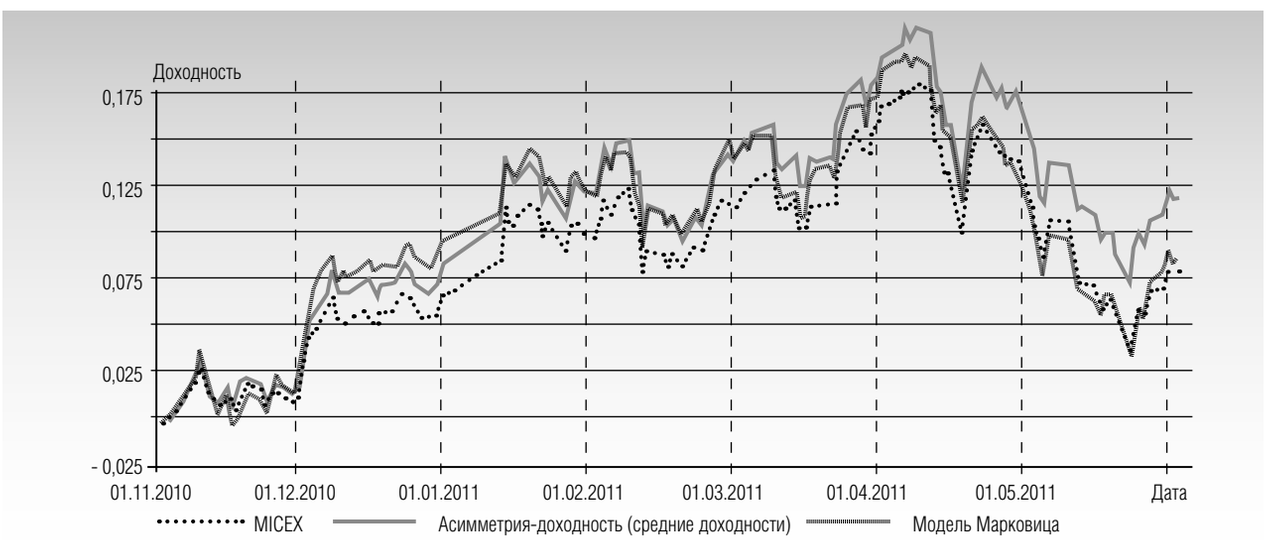


Рис. 13 а. Динамика доходности некоторых портфелей

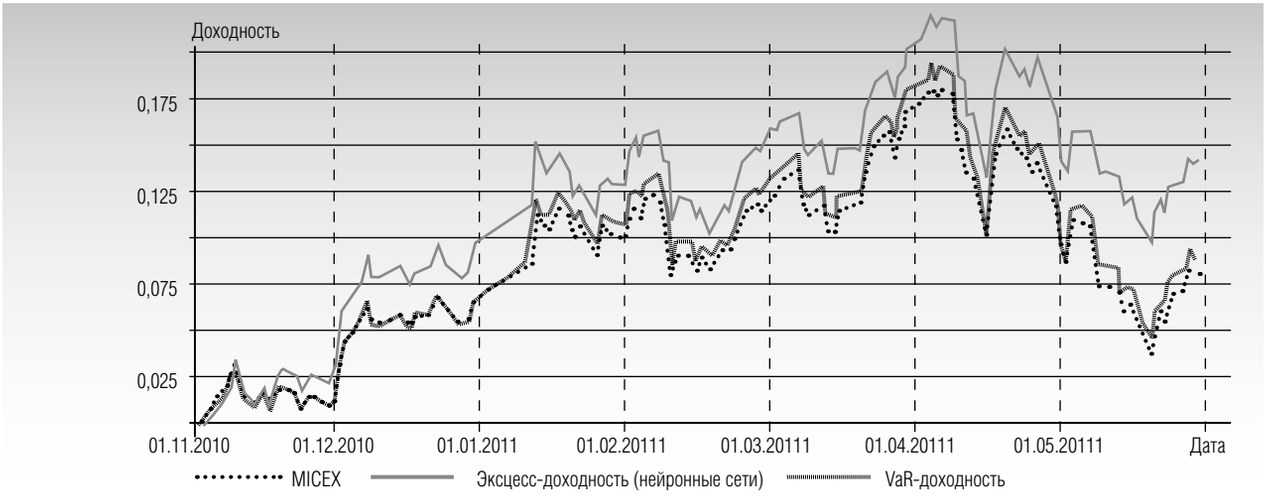


Рис. 13 б. Динамика доходности некоторых портфелей

бавить модели прогнозирования временных рядов, улучшить взаимодействие программы с биржевыми терминалами. Результаты экспериментов на российском фондовом рынке позволяют говорить о возможности использования разработанного приложения на практике. Результаты моделей в большой степени зависят от качества введенных экспертных оценок или методов прогноза временных рядов. Эксперименты с данными российского фондового рынка с использованием разработанного приложения позволили убедиться в приемлемости некоторых моделей и их преимуществе перед стандартными подходами. Модели, в которых рассчитываются показатели на основе исторических данных, являются адекватными в течение относительно короткого срока и только в рамках сложившейся рыночной ситуации. ■

Литература

1. Lohre H., Neumann T., Winterfeldt T. Portfolio construction with Asymmetric Risk Measures. – Frankfurt am Main: Union Investment, 2007.
2. Исавнин А.Г., Галиев Д.Р. Модели портфельного инвестирования с применением асимметричных мер риска и генетических алгоритмов // Финансовая аналитика. – 2011. – №48. – С. 32–38.
3. Курейчик В.В., Голодков Л.А., Курейчик В.Н. Генетические алгоритмы. – М.: Физматлит, 2010.
4. Еремеев А.В. Генетические алгоритмы и оптимизация. – Омск: Институт математики им. С.Л. Соболева, 2010. – URL: http://iitam.omsk.net.ru/~eremeev/PAPERS.SK/Method.pdf (дата обращения 17.03.2012)
5. Эволюционные методы моделирования и оптимизации сложных систем / Е.С.Семенкин, М.Н.Жукова, В.Г.Жуков, И.А.Панфилов, В.В.Тынченко. – Красноярск, 2007. – URL: http://library.krasu.ru/ft/ft/_umkd/22/u_lectures.pdf (дата обращения 17.03.12)
6. Дуброва Т.А. Статистические методы прогнозирования в экономике. – М.: Московский международный институт эконометрики, информатики, финансов и права, 2003.
7. Бокс Д., Дженкинс Г. Анализ временных рядов. Прогноз и управление. Вып. 2. – М.: Мир, 1974.
8. Эконометрия / В.И.Суслов, Н.М.Ибрагимов, Л.П.Талышева, А.А.Цыплаков. – Новосибирск : Изд. Новосибирского государственного университета, 2005.
9. Хайкин С. Нейронные сети: полный курс. 2-е издание.: Пер. с англ. – М.: Изд. дом «Вильямс», 2006.
10. Каллан Р. Основные концепции нейронных сетей.: Пер. с англ. – М.: Изд. дом «Вильямс», 2001.
11. Каширина И.Л. Нейросетевые технологии: учебно-методическое пособие для вузов. – Воронеж : Изд. Воронежского государственного университета, 2008.
12. Официальный сайт библиотеки Stock#. URL: http://stocksharp.com/ (дата обращения: 17.03.2012)
13. Официальный сайт системы Quik. URL: http://quik.ru/about/ (дата обращения: 17.03.2012)
14. Справочная система библиотеки Stock#. URL: http://stocksharp.com/doc/html/87d2cacd-5492-4bca-9140-7d7c3f5218d7.htm (дата обращения: 17.03.2012)
15. Le Sourd V. Performance Measurement for Traditional Investment // EDHEC, 2007. URL: http://www.edhec-risk.com/performance_and_style_analysis/perf_measurement (дата обращения: 17.03.2012)

МОНИТОРИНГ И ПЛАНИРОВАНИЕ РАЗВИТИЯ СИСТЕМ ИНФОРМАЦИОННОЙ ПОДДЕРЖКИ КОРПОРАТИВНОГО УПРАВЛЕНИЯ И СТРАТЕГИЧЕСКОГО МЕНЕДЖМЕНТА

Д.В. Исаев,
кандидат экономических наук, доцент кафедры бизнес-аналитики Национального исследовательского университета «Высшая школа экономики»

E-mail: disaev@hse.ru
Адрес: г. Москва, ул. Кирпичная, д. 33/5

В статье рассматриваются вопросы мониторинга и планирования развития систем информационной поддержки корпоративного управления и стратегического менеджмента (СИП КУСМ). Раскрывается роль функций мониторинга и планирования в общей системе управления развитием, формулируются основные требования к этим функциям и принципы их практической реализации. Описаны основные элементы системы мониторинга и планирования развития СИП КУСМ, обозначены направления дальнейших исследований в данной области.

Ключевые слова: корпоративное управление, стратегический менеджмент, система управления эффективностью, информационная поддержка, мониторинг, планирование.

Введение

В настоящее время многие крупные предприятия и организации сталкиваются с существенными проблемами, лежащими в плоскости корпоративного управления и стратегического менеджмента. К таким проблемам, в частности, относятся отсутствие целостности системы управления, недостаточная гибкость и адаптивность стратегий и планов, недостаточная прозрачность для внешних и внутренних заинтересованных лиц. Попытки решения перечисленных проблем привели к появлению концепции систем управления эффективностью (*Performance Management Systems, PMS*). В соответствии с определением А.Феррейры и Д.Отли, PMS охватывает «формальные и неформальные механизмы, процессы, системы и коммуникации, используемые организациями для достижения ключевых целей, определенных руководством, для поддержки стратегического процесса и текущего управления посредством анализа, планирования, формирования метрик, контроля, стимулирования и управления эффективностью в широком смысле, а также для поддержки и содействия обучению организации и изменениям» [14, с. 264].

Важная роль в системах управления эффективностью отводится вопросам информатизации: одним из элементов концептуальной модели PMS, предложенной А.Феррейрой и Д.Отли [14], являются информационные потоки, системы и сети, поддерживающие управленческие процессы стратегического уровня. В целях дальнейшей конкретизации информационного аспекта корпоративного управления и стратегического менеджмента автором предложена концепция системы информационной поддержки корпоративного управления и стратегического менеджмента (СИП КУСМ), которая может считаться одной из подсистем «глобальной» PMS. Под СИП КУСМ понимается «комплекс средств, направленных на решение задач сбора, хранения, аналитической обработки и представления информации, являющейся ключевой для обеспечения информационной прозрачности организации и поддержки принятия стратегических управленческих решений внешними и внутренними заинтересованными лицами» [4, с.103].

Управление развитием СИП КУСМ имеет иерархическую структуру и включает три уровня: верхний (уровень СИП КУСМ в целом), средний (уровень подсистем) и нижний (уровень проектов), при этом на каждом из уровней имеют место функции проектирования, планирования и контроля [5; 6]. Наиболее важную роль в определении стратегии развития системы играет верхний уровень управления, но если вопросы проектирования СИП КУСМ проработаны достаточно детально [4], то подходы к мониторингу и планированию развития таких систем на сегодняшний день изучены недостаточно и поэтому представляют собой актуальное направление исследований. В качестве первого шага в этом направлении предлагается описание общих принципов и основных элементов системы мониторинга и планирования развития СИП КУСМ.

Существующие подходы к оценке программ развития систем управления и информационных систем

Методы, применяемые для оценки инвестиций в развитие систем управления и информационных систем, можно разделить на две обширные группы – финансовые и нефинансовые [19].

К финансовым методам относятся различные вариации известных методов инвестиционного анализа – бухгалтерская норма прибыли, период окупаемости, а также методы, основанные на дисконтировании денежных потоков [13]. Все эти методы, так или иначе, основаны на сопоставлении объемов инвестирования и экономических выгод, получаемых в результате этих инвестиций. Иначе говоря, для практического применения этих методов требуется, чтобы и затраты, и экономические выгоды были с достаточной степенью надежности оценены в финансовом выражении. Что касается управления развитием СИП КУСМ, то здесь затратная часть проектов развития действительно может быть оценена в финансовых терминах. Однако финансовая оценка экономических выгод, получаемых организацией в долгосрочной перспективе за счет совершенствования информационной поддержки заинтересованных лиц, выглядит проблематичной, поскольку такие выгоды в подавляющем большинстве случаев являются нематериальными и лишь опосредованно влияют на финансовые показатели. Сказанное позволяет сделать вывод о невозможности практического применения финансовых методов для решения задач оценки текущего уровня СИП КУСМ и обоснованного выбора программы ее дальнейшего развития.

Что касается нефинансовых методов оценки проектов и программ развития систем управления и информационных систем, то круг таких методов весьма широк: к ним, в частности, относятся многокритериальные методы принятия решений, бенчмаркинг и некоторые другие [19]. Среди них наиболее привлекательными с точки зрения задач развития СИП КУСМ представляются методы, основанные на сбалансированных системах показателей и методы принятия решений на основе экспертных оценок.

Методология сбалансированных систем показателей (*Balanced Scorecard, BSC*), получила широкое распространение, не в последнюю очередь, благодаря трудам Р.Каплана и Д.Нортона [7; 8; 9; 10]. Достоинством данной методологии является то, что она позволяет организациям описывать свои цели в виде совокупности финансовых и нефинансовых метрик, а также счетных карт, формируемых для разных уровней управления организацией. Универсальность данного подхода позволила адаптировать его к разным управленческим задачам, включая задачи планирования развития информационных систем и обоснования инвестиций в информационные технологии [19]. В настоящее время перспективным направлением является сочетание BSC с методом системно-динамического моделирования, который был впервые предложен Дж.Форрестером [15] и впоследствии

нашел широкое практическое применение для решения целого ряда задач, включая задачи стратегического управления развитием сложных экономических систем [1; 2; 3]. Такая интеграция, получившая название динамической сбалансированной системы показателей (*Dynamic Balanced Scorecard, DBSC*), позволяет преодолеть ограниченность «классической» теории Р.Каплана и Д.Нортона в части описания причинно-следственных связей между показателями [17].

Что касается методов принятия решений на основе экспертных оценок, то они весьма эффективны именно в тех случаях, когда точное измерение параметров объекта управления по тем или иным причинам не представляется возможным (а именно эта ситуация весьма характерна для задач развития СИП КУСМ). В таких случаях оценка объекта управления и принятие решений в отношении его развития основывается на мнениях экспертов и соответствующих методах обработки полученных экспертных оценок [11; 16; 18].

Таким образом, именно эти две группы методов могут быть положены в основу разработки методических рекомендаций в области мониторинга и планирования развития СИП КУСМ.

Требования к мониторингу и планированию развития СИП КУСМ

В общем случае, под мониторингом понимается регулярный процесс сбора, структурирования и обобщения информации о характеристиках, событиях и явлениях, относящихся к некоторому объекту – объекту мониторинга. В данном случае в качестве объекта мониторинга может выступать как СИП КУСМ в целом, так и ее отдельные элементы (подсистемы). Информация, полученная в процессе мониторинга, может подвергаться более углубленной аналитической обработке, либо с использованием специальных предметно-ориентированных процедур, либо с привлечением экспертов. В любом случае, конечной целью мониторинга является формирование мнения о текущем состоянии объекта мониторинга и динамике его развития. Планирование относится к тем же объектам и также предусматривает оценку динамики развития этих объектов. Однако, в отличие от мониторинга, ориентированного на исторические данные, здесь речь идет о состоянии исследуемых объектов в будущем и формировании календарных планов развития этих объектов. При этом принимается во внимание возможность субъекта управления влиять на динамику развития объекта управления посредством некоторых управляющих воздействий. Мониторинг и планирование являются основой для контроля исполнения планов, предусматривающего констатацию отклонений фактически достигнутых результатов от запланированных и выявление причин этих отклонений. Таким образом, мониторинг и планирование являются тесно взаимосвязанными функциями, что позволяет применять для их реализации одни и те же методы и подходы.

Принципы мониторинга и планирования вытекают из общих принципов, относящихся к системе управления развитием СИП КУСМ [6]; по сути, имеет место проецирование общих принципов на функции мониторинга и планирования, с учетом роли этих функций в системе управления. В частности, принцип измеримости предусматривает описание объектов мониторинга и планирования в терминах количественных показателей, характеризующих различные аспекты текущего, ретроспективного или перспективного состояния системы. Принцип декомпозиции связан с возможностью рассмотрения СИП КУСМ не только в целом, но и в разрезе входящих в ее состав элементов (подсистем), а также отдельных аспектов, определяющих уровень развития СИП КУСМ. Принцип динамики означает рассмотрение изменений в состоянии системы и ее подсистем, происходящих с течением времени в прошлом или будущем. Принцип целевой ориентации означает оценку динамики развития СИП КУСМ (фактической или планируемой) в сопоставлении с некоторым целевым уровнем развития системы, который, в свою очередь, определяется исходя из принципа соответствия корпоративной стратегии организации. Наконец, принцип реалистичности предусматривает формирование планов развития СИП КУСМ с учетом доступных ресурсов, включая ресурс времени.

Учитывая особенности применения принципов управления развитием СИП КУСМ по отношению к рассматриваемым функциям, можно сформулировать общие требования, предъявляемые к системе мониторинга и планирования:

✦ уровень развития СИП КУСМ должен рассматриваться с учетом всех существенных аспектов, находящихся в рамках нескольких укрупненных перспектив и в совокупности определяющих ценность системы;

✦ аспекты развития СИП КУСМ должны описываться в терминах измеримых показателей, кото-

рые должны рассматриваться в динамике, с учетом причинно-следственных связей между ними;

✦ уровень развития СИП КУСМ оценивается на основе консолидации частных показателей в обобщенный интегральный показатель, динамика которого определяет траекторию развития системы;

✦ система мониторинга и планирования должна позволять детализировать уровень развития СИП КУСМ в разрезе аспектов, а также подсистем, входящих в ее состав;

✦ для оценки уровня развития СИП КУСМ конкретной организации относительно практик, существующих в данной области, в системе мониторинга и планирования должны быть предусмотрены уровни зрелости;

✦ для обеспечения целевой ориентации развития системы должны быть предусмотрены целевые уровни, отражающие намерения организации в части развития своей СИП КУСМ;

✦ каждой из возможных программ развития системы соответствует своя прогнозная траектория развития, выбор одной из программ для практической реализации производится на основе сопоставления этих траекторий с целевыми уровнями развития;

✦ для обоснования выбора программы развития из множества возможных программ должны применяться метрики ожидаемой эффективности (качества) программ, а также соответствующие процедуры выбора. Эти же метрики могут применяться для анализа фактической эффективности развития системы;

✦ для анализа отклонений фактической траектории развития от запланированной должны быть предусмотрены специальные индикаторы;

✦ система мониторинга и планирования не должна быть догматичной: в случае необходимости она должна допускать внесение изменений, на основе специальных правил уточнения и пересмотра параметров модели.

Перечисленные требования служат основой для более детальной проработки отдельных элементов системы мониторинга и планирования развития СИП КУСМ.

Основные элементы системы мониторинга и планирования развития СИП КУСМ

Сформулированные требования позволяют более детально описать основные элементы системы мониторинга и планирования развития СИП КУСМ, к числу которых относятся: перспективы, аспекты, частные и интегральные показатели, траектории развития, уровни зрелости и целевые уровни развития, программы развития и метрики их эффективности, а также индикаторы отклонений фактических траекторий от запланированных.

Перспективы представляют собой результат укрупненной декомпозиции СИП КУСМ. Каждая из них представляет собой взгляд на систему с определенной точки зрения и, таким образом, служит для лучшего понимания системы в целом. В качестве перспектив могут выступать заинтересованные лица (аналог клиентской перспективы в методологии BSC), методы и процессы управления (аналог перспективы внутренних процессов), управленческий персонал (аналог перспективы обучения и роста), а также информационные системы (в BSC аналога не имеет). Заметим, что финансовая перспектива в данном случае не рассматривается, поскольку деятельность СИП КУСМ не приносит каких-либо непосредственных финансовых результатов, а лишь создает условия для их получения.

Аспекты представляют собой качественные характеристики СИП КУСМ, в совокупности определяющие уровень развития системы с точки зрения ее ценности для организации. Аспекты формируются в рамках перспектив: каждой перспективе соответствуют один или несколько аспектов, обеспечивающих наиболее полную конкретизацию сущности данной перспективы. Заметим, что аспекты развития СИП КУСМ не могут считаться аналогами целей в традиционной системе BSC: если цели BSC носят временный характер (достижение цели делает ее неактуальной), то аспекты СИП КУСМ носят не временный, а перманентный характер. Что касается целевой ориентации развития системы, то ее необходимость, безусловно, не подлежит сомнению, однако обеспечивается она иным образом – через целевые уровни развития, которые будут рассмотрены ниже.

Аспекты связаны между собой причинно-следственными связями, поскольку в любой достаточно сложной системе (а СИП КУСМ, несомненно, относится к сложным системам) изменение в какой-либо одной области влечет за собой изменения в других областях. На уровне аспектов причинно-следственные связи могут быть описаны лишь качественным образом (при помощи вербальных конструкций «если... то...»), а более точное моделирование становится возможным только после детального описания аспектов в терминах измеримых показателей.

Частные показатели служат для описания рассматриваемых аспектов в количественном выражении, при этом каждому аспекту соответствуют один или несколько частных показателей. Среди частных показателей можно выделить основные, в той или иной мере характеризующие уровень развития СИП КУСМ, и вспомогательные, используемые для промежуточных расчетов. Частные показатели могут иметь как фактические, так и плановые значения. В любом случае значения показателей рассматриваются в динамике, характеризуя их изменение с течением времени.

Между показателями имеют место причинно-следственные связи, которые являются производными по отношению к упомянутым выше причинно-следственным связям между аспектами. С точки зрения причинно-следственных связей среди частных показателей можно выделить первичные и вторичные. Первичные показатели подвержены непосредственному влиянию программы развития СИП КУСМ, а вторичные формируются опосредованно – через первичные и, возможно, другие вторичные показатели. Иначе говоря, мероприятия программы развития СИП КУСМ (завершение проектов, этапов, работ) оказывают влияние на первичные показатели, а через них воздействие передается вторичным показателям. Воздействие мероприятий на первичные показатели и последующие взаимосвязи между показателями могут описываться нелинейными зависимостями, носить вероятностный характер, быть связанными с временными лагами и обратными связями. Все это создает предпосылки для описания динамики показателей с применением системно-динамического моделирования.

Интегральные показатели строятся на основе частных. Необходимость интегральных показателей объясняется тем, что количество частных показателей, характеризующих уровень развития СИП КУСМ, может быть довольно большим, что делает их обобщенную интерпретацию весьма затруднительной.

Основным интегральным показателем является интегральный показатель уровня развития СИП КУСМ в целом. На его основе можно рассматривать **траекторию развития СИП КУСМ**, представляющую собой изменение значения основного интегрального показателя с течением времени. Ретроспективная траектория развития СИП КУСМ (относящаяся к прошлому) является единственной. Что касается перспективной траектории (описывающей ситуацию в будущем), то она определяется программой развития, поскольку именно мероприятия программы развития опосредованно (через первичные и вторичные частные показатели) оказывают влияние на интегральный показатель. Более того, поскольку причинно-следственные связи между показателями часто носят вероятностный характер, программе развития может соответствовать не одна траектория, а семейство траекторий, у каждой из которых есть своя вероятность.

Помимо траектории развития СИП КУСМ в целом, аналогичные траектории могут строиться для отдельных аспектов и подсистем, на основе соответствующих интегральных показателей. Это полезно для детализации основного интегрального показателя (и, соответственно, основной траектории), а также для обеспечения преемственности между верхним уровнем управления развитием СИП КУСМ (уровнем СИП КУСМ в целом) и средним уровнем (уровнем подсистем). Таким образом, помимо интегрального показателя уровня развития СИП КУСМ в целом, могут рассматриваться интегральные показатели каждого из аспектов развития СИП КУСМ, интегральные показатели каждой из подсистем СИП КУСМ, а также (для всех подсистем) интегральные показатели для каждого из аспектов развития соответствующей подсистемы. Дополнительные интегральные показатели являются субординированными по отношению к основному интегральному показателю – показателю уровня развития СИП КУСМ в целом.

Интегральные показатели уровня развития СИП КУСМ представляют собой аналог счетных карт в методологии BSC и также могут строиться с применением скоринговых моделей.

Уровни зрелости служат для того, чтобы иметь возможность оценивать уровень развития СИП КУСМ конкретной организации (как фактически достигнутый, так и планируемый в будущем) по отношению к некоторой шкале, характеризующей современные представления о возможных качественных состояниях системы. Основой для построения такой шкалы могут служить результаты обобщения прогрессивных методологических и технологических разработок, а также опыта их практического применения. Определение уровней зрелости в значительной мере основано на идеях бенчмаркинга, предусматривающего сравнение организационно-управленческих практик объекта управления с

аналогичными практиками других организаций, а также с лучшими достижениями в соответствующей области [12]. Таким образом, уровни зрелости носят объективный характер и не зависят от особенностей развития СИП КУСМ конкретной организации. Для обеспечения сопоставимости уровни зрелости формируются в терминах тех же интегральных показателей, которые используются для описания траекторий развития.

Целевые уровни служат для обеспечения целевой ориентации развития СИП КУСМ конкретной организации. Они формируются на основе уровней зрелости, но если уровни зрелости определяются объективными критериями, то целевые уровни отражают субъективные намерения руководителей организации в отношении развития своей системы управления. Целевой уровень может быть единственным, но также допускается наличие нескольких целевых уровней (например, «уровень необходимости» и «уровень достаточности»). Поскольку требования организации к своей системе управления изменяются с течением времени, целевые уровни также должны рассматриваться в динамике. Как и в случае с уровнями зрелости, в целях сопоставимости целевые уровни описываются теми же интегральными показателями, что и траектории развития.

Критерии и метрики эффективности программ развития применяются для решения задачи выбора одной из программ развития из множества потенциальных программ, для последующей практической реализации. Поскольку каждой потенциальной программе соответствует своя прогнозная траектория, выбор должен производиться на основе сопоставления потенциальных траекторий с целевыми уровнями развития. Такое сопоставление основывается на определенных критериях, отражающих результативность, риски и затраты, связанные с программой развития. При этом применяются метрики эффективности программы развития – измеримые показатели, позволяющие описывать критерии эффективности в количественном выражении (каждому критерию соответствуют одна или несколько метрик).

Критерии и метрики эффективности программ развития могут использоваться не только для решения задачи выбора из нескольких альтернативных программ, но и для оценки успешности реализации выбранной программы на основе сопоставления фактической траектории развития с целевыми уровнями.

Индикаторы отклонений фактической траектории от запланированной применяются для контроля динамики развития системы. Причинами таких отклонений могут быть как неисполнение программы развития по объемам и/или срокам, так и неточности, допущенные при описании прогнозной траектории, соответствующей выбранной программе. Это означает, что даже в случае полного выполнения всех мероприятий выбранной программы могут возникать (а скорее всего – обязательно возникнут) отклонения от запланированной траектории, просто в силу невозможности безошибочного построения прогнозной траектории для заданной программы развития, главным образом – за счет неизбежных неточностей в описании причинно-следственных связей. Так или иначе, индикаторы отклонений дают важную информацию, необходимую для своевременного осуществления управленческих воздействий, включая внесение в программу развития необходимых корректировок.

Заключение

Предлагаемый подход к мониторингу и планированию развития СИП КУСМ основан на идеях Balanced Scorecard и имеет с этой методологией много общего. Главное сходство – целевое управление, основанное на описании определенных аспектов развития системы в терминах измеримых показателей и причинно-следственных связей, а также на оценке стратегических инициатив (программ развития) с точки зрения их влияния на состояние системы и достижение целей.

Однако есть и различия, отражающие специфику управления развитием СИП КУСМ. Главное отличие состоит в том, что если в BSC сначала определяются цели, и только потом – инициативы, которые призваны обеспечить достижение этих целей, то в методике мониторинга и планирования развития СИП КУСМ цели и инициативы (программы развития) формируются параллельно, и только впоследствии сравниваются между собой. Это объясняется тем, что мероприятия по развитию СИП КУСМ не могут считаться независимыми, они выстраиваются в логические последовательности, которые в совокупности и образуют потенциальные стратегии развития СИП КУСМ. Поэтому здесь не возникает задачи формирования портфеля из некоторого множества независимых проектов, вместо этого происходит формирование программ развития в целом.

Предлагаемый подход может быть принят за основу для дальнейших исследований, к основным направлениям которых можно отнести разработку типовой (референтной) модели системы мониторинга и планирования, моделирование причинно-следственных связей (с применением системно-динамического подхода), разработку рекомендаций в области выбора программы развития (с применением методов принятия решений), а также разработку рекомендаций в области перехода от референтной модели мониторинга и планирования развития СИП КУСМ к частным моделям. ■

Литература

1. Акопов А.С. Использование средств динамического имитационного моделирования для подготовки управленческих решений в ТЭК // Системы управления и информационные технологии. – 2004. – №2. – С. 72-77.
2. Акопов А.С. Системно-динамическое моделирование стратегии банковской группы // Бизнес-информатика. – 2012. – №2 (20). – С. 10-19.
3. Горбунов А.Р. Системная динамика: моделирование принятия стратегических и оперативных решений (метод функции капитального класса) // Бизнес-информатика. – 2008. – №2 (4). – С. 25-34.
4. Исаев Д.В. Корпоративное управление и стратегический менеджмент: информационный аспект. – М.: Изд. дом ГУ-ВШЭ, 2010.
5. Исаев Д.В. Процесс управления развитием систем информационной поддержки корпоративного управления и стратегического менеджмента // Аудит и финансовый анализ. – 2012. – №4. – С. 276-281.
6. Исаев Д.В. Развитие систем информационной поддержки корпоративного управления и стратегического менеджмента // Бизнес-информатика. – 2011. – №2 (16). – С. 56-62.
7. Каплан Р., Нортон Д. Сбалансированная система показателей. От стратегии к действию. – М.: Олимп-бизнес, 2003.
8. Каплан Р., Нортон Д. Организация, ориентированная на стратегию. Как в новой бизнес-среде преуспевают организации, применяющие сбалансированную систему показателей. – М.: Олимп-бизнес, 2004.
9. Каплан Р., Нортон Д. Стратегическое единство. Создание синергии организации с помощью сбалансированной системы показателей. – М.: Вильямс, 2006.
10. Каплан Р., Нортон Д. Стратегические карты. Трансформация нематериальных активов в материальные результаты. – М.: Олимп-Бизнес, 2007.
11. Кравченко Т.К. Экспертная система поддержки принятия решений // Открытое образование. – 2010. – №6 (83). – С. 147-156.
12. A Guide to the Business Analysis Body of Knowledge (BABOK Guide). Version 2.0. – Toronto: International Institute of Business Analysis, 2009.
13. Drury C. Management and Cost Accounting. 7th edition. – London: Cengage Learning EMEA, 2008.
14. Ferreira A., Otley D. The design and use of performance management systems: An extended framework for analysis // Management Accounting Research. – 2009. – 20. – P. 263-282.
15. Forrester J.W. Industrial dynamics. – Cambridge, MA: MIT Press, 1961.
16. Multiple criteria decision analysis: State of the art surveys / Edited by J.Figueira, S.Greco, M.Ehrgott. – Boston: Springer, 2005.
17. Rydzak F., Magnuszewski P., Pietruszewski P., Sendzimir J., Chlebus E. Teaching the Dynamic Balanced Scorecard // Proceedings of the 22nd International Conference of The System Dynamics Society, July 25-29, 2004, Oxford, England, UK.
18. Saaty T.L., Peniwati K. Group Decision Making: Drawing out and Reconciling Differences. – Pittsburgh: RWS Publications, 2008.
19. Schniederjans M.J., Hamaker J.L., Schniederjans A.M. Information technology investment: Decision-making methodology. – NY: World Scientific Publishing, 2005.

ОБ ИННОВАЦИОННОМ ПОДХОДЕ К СОЗДАНИЮ И ИСПЫТАНИЯМ КОНТРОЛЬНО-УЧЁТНЫХ МОДУЛЕЙ ДЛЯ КОРПОРАТИВНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

М.А. Аракелян,
заместитель директора департамента, ООО «ИБС Экспертиза»

Е.В. Чепин,
кандидат технических наук, ведущий научный сотрудник ИПИ РАН, заместитель заведующего кафедрой МФТИ

А.П. Шабанов,
доктор технических наук, старший научный сотрудник, главный эксперт, ООО «ИБС Экспертиза»

Адрес: г. Москва, Дмитровское шоссе, д. 9Б
E-mail: marakelian@ibs.ru, echepin@mail.ru, ashabanov@ibs.ru

В статье рассматривается методический подход к реализации проектов создания контрольно-учётных модулей корпоративных информационных систем. Данный подход основан на свойстве повторяемости (цикличности) выполнения предприятием – разработчиком однотипных проектов, а также на использовании инновационных системно-технических решений в сочетании с готовыми изделиями, в качестве которых выступают специализированные программы. Областью применения данного подхода являются системы поддержки услуг, связанных с информационными технологиями, а также инвентаризация и контроль состояния технических и программных средств корпоративной компьютерной сети и контроль её транспортных средств.

Ключевые слова: информационная система, управление проектом, контрольно-учётн-ый модуль, системные и технические решения, программные продукты, контрольно-учётная система, системные и технические решения, программные продукты.

Введение

Одним из направлений модернизации корпоративных информационных систем (КИС) для организационных систем, практически любого масштаба в последние годы является создание и внедрение в их составе контрольно-учётных модулей различного назначения, объединённых общей задачей – повышением эффективности контроля, а также повышением эффективности управления ресурсами КИС. Имеются в виду вычислительные, информационные, транспортные,

людские ресурсы, а также ресурсы систем хранилищ данных, центров обработки данных, виртуальных вычислительных систем. Цель, которую стремятся достичь при реализации таких проектов, состоит в обеспечении необходимого уровня услуг, обеспечиваемых информационными технологиями (ИТ услуг). Логико-процессной основой функционирования контрольно-учётных модулей и, одновременно, программами-конструкторами информационных форм и интерфейсов, как правило, являются программные продукты, например, из состава программных комплексов *BMC Business Service Management* [1] и *HP Software Manager* [2]. В то же время, реализация подобных проектов, несмотря на наличие готовых изделий, требует от предприятий-разработчиков значительных усилий на выполнение проектных, инсталляционных и настроечных работ, на проведение испытаний и опытную эксплуатацию, на сопровождение модулей в ходе их применения (для актуализации программ, технологической информации и др.). Поэтому эти проекты относятся к классу сложных проектов. Анализ сроков их выполнения, а также сил, задействованных для этого, показывает, что нередки случаи, когда они превышают изначально запланированные. Такое положение, в значительной степени можно объяснить иллюзиями, существующими на предприятиях-заказчиках в отношении возможностей программных продуктов, приобретаемых для построения модулей, кажущейся «простоты» их настройки и сопровождения эксплуатации. При этом предприятия-разработчики, в свою очередь, не в полной мере осознают требования и степень готовности/зрелости предприятия-заказчика. В этих случаях проекты стартуют и проводятся по упрощённой схеме. Она заключается в том, что

✦ на стадии замысла жизненных циклов контрольно-учётных модулей для КИС предприятий не учитываются общие методологические принципы системного подхода, не определяются критерии эффективности;

✦ на стадии разработки не задействуются средства аналитического и имитационного моделирования, не проводятся натурные испытания; не проводятся исследования системных и технических решений, в том числе патентные исследования; выбор специализированных программных продуктов осуществляется без их привязки к оптимальным системным решениям.

В результате такого упрощённого подхода к реализации проектов по созданию контрольно-учётных модулей для КИС существуют следующие риски:

✦ для предприятия-разработчика имеются риски упущенной прибыли и непредвиденных расходов;

✦ для предприятия-заказчика имеются риски обесценивания капиталовложений из-за изменения запланированных сроков ввода системы в эксплуатацию и непредвиденных дополнительных инвестиций.

С целью предотвращения указанных рисков, разработан инновационный подход к реализации сложных проектов по созданию в КИС контрольно-учётных модулей, который и рассматривается в настоящей статье.

Основные принципы инновационного подхода

Инновационный подход к реализации проектов учитывает такое свойство рассматриваемого класса модулей, как массовость применения, по существу, одних и тех же системно-технических решений и интерфейсов для КИС, относящихся к разным предприятиям или подразделениям крупных предприятий. Данное свойство позволяет говорить о наличии типовых контрольно-учётных модулей, имеющих отношение к одному и тому же виду деятельности, например: (а) поддержка ИТ услуг; (б) аудит и контроль состояния технических и программных средств КИС; (в) контроль над транспортными средствами передачи данных. С учётом отмеченного свойства, отличительными принципами циклического подхода к реализации проектов выступают:

✦ создание на предприятии-разработчике системы *производственного стенда*; цель – разработка и функциональная настройка модулей. Стенд создаётся, например, в рамках первого по очереди проекта от даты принятия решения о создании такого модуля. Предполагается цикличность использования стенда в однотипных проектах; наращивание функциональности стенда для разных типов контрольно-учётных модулей;

✦ включение в состав команды очередного проекта в стадии замысла жизненного цикла системы соответствующих штатных технических специалистов с целью последующей работы на стенде;

✦ создание и сопровождение базы известных модулей и системно-технических решений для КИС



Рис. 1. Основные процессы реализации проекта при инновационном подходе

различного назначения (типов); преемственность и унификация модулей.

Основные процессы инновационного подхода к реализации сложных проектов по созданию совокупности модулей контрольно-учётной системы для КИС показаны ниже на *рис. 1.*

В стадии замысла реализуются процессы: (а) обследование КИС и (б) обоснование требований к создаваемой контрольно-учётной системе.

В стадиях разработки и производства (внедрения) реализуются процессы: (а) разработка модулей системы; (б) настройка модулей; (в) внедрение модулей системы в КИС предприятия; (г) проведение интеграционных настроек модулей и интерфейсов пользователей системы.

В стадии применения реализуются процессы: (а) актуализация программной платформы, информационного обеспечения; (б) анализ эффективности функционирования системы.

Инновационный (циклический) подход к реализации проектов, в зависимости от частоты ведения проектов однотипных контрольно-учётных систем и от величины инвестиций в стенд, по сравнению с известным подходом, характеристика которого приведена во введении,

✦ позволяет предотвратить указанные выше риски;

✦ сократить сроки однотипных проектов;

✦ получить ожидаемый эффект, в т.ч. финансовую прибыль.

В *табл. 2* и *табл. 3* в качестве примера приведён график выполнения одного и того же проекта при известном подходе (Вариант I) и при циклическом подходе (Вариант II). При Варианте I все работы выполняются на объектах заказчика.

На графике (*табл. 2* и *табл. 3*) приведены наименования и сроки выполнения основных работ по созданию контрольно-учётных систем на основе готовых продуктов – прикладных программ, специализированных для поддержки следующих видов деятельности:

✦ управление подразделениями информатизации ведомств, учреждений, предприятий, их филиалов и подразделений;

✦ аудит (инвентаризация) состава технических и программных средств КИС и контроль состояний этих средств в их привязке к ИТ услугам, которые обеспечиваются этими средствами;

✦ анализ данных аудита и контроля, принятие на основе этого анализа решений по управлению информационными, вычислительными, транспортными и людскими ресурсами КИС.

В данном примере эффект от применения инновационного подхода к реализации однотипных проектов по созданию контрольно-учётных систем заключается в сокращении на четыре недели времени реализации одного проекта.

Область применения инновационного подхода

Областью применения циклического подхода являются системы поддержки ИТ услуг, которые включают подсистемы инвентаризации и контроля состояния технических и программных средств

Таблица 1.

График выполнения одного проекта. Вариант I — известный подход

стадия проекта \ неделя	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
подготовка договора, обследование, заказ и поставка тех.средст	●	●	●	●	●	●	●	●	●	●														
проектные работы и настройка ПО						●	●	●	●	●	●	●	●	●	●	●	●	●						
испытания; ввод в эксплуатацию																			●	●	●	●	●	●

Таблица 2.

График выполнения одного проекта. Вариант II — циклический подход

стадия проекта \ неделя	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
подготовка договора, обследование, заказ и поставка тех.средст	●	●	●	●	●	●	●	●	●	●														
настройка ПО на стенде	●	●	●	●	●	●	●	●																
проектные работы; привязка ПО к инфраструктуре заказчика						●	●	●	●	●	●	●	●	●										
испытания; ввод в эксплуатацию															●	●	●	●	●	●				

КИС, контроля над её транспортными средствами передачи данных, в том числе в различных физических средах и управление параметрами передачи в зависимости от параметров физической среды Системы поддержки ИТ услуг, системы инвентаризации и мониторинга достаточно хорошо представлены в публикациях. Они реализованы в существующих КИС, например [1-3]. Краткий обзор инновационных (запатентованных) системно-технических решений, относящихся к классу адаптивных информационных систем, приведён в работе [4]. Эти решения предназначены для управления производительностью трактов КИС и обеспечивают значительное повышение их производительности. Поэтому такие решения необходимо использовать в проекте в стадиях разработки и производства (внедрения). Каждое из приведённых в обзоре решений можно рассматривать или в целом, как аппаратно-программный компонент создаваемой для КИС контрольно-учётной системы, или использовать алгоритм, положенный в основу этого решения, для разработки аналогичной программы.

При этом данная программа, в соответствии с рассматриваемым подходом должна разрабатываться на стенде с учётом последующего размещения и функционирования в составе КИС заказчика, например, в одном из существующих физических серверов или в виртуальной вычислительной среде. Разработанная программа должна обеспечивать программное сопряжение с программами, построенными на основе программных продуктов, например [1, 2], которые реализуют типовые функции контроля, учёта и управления.

В качестве примеров, иллюстрирующих описанный выше подход к разработке на стенде составной компоненты для разрабатываемой системы, рассмотрим следующие нетривиальные системно-технические решениия по поддержке ИТ услуг:

✦ Поддержка услуг по обеспечению доступа к ИТ услугам по линиям связи с различным временем распространения сигналов [5, 6];

✦ Поддержка услуг по обеспечению доступа к ИТ услугам через двухпроводную соединительную телефонную сеть [6].

Обеспечение доступа к ИТ услугам по линиям связи с различным временем распространения сигналов

При организации доступа субъектов управления к ИТ услугам одновременно по нескольким коммутируемым линиям городских и междугородных телефонных сетей необходимо применять решения, исключающие потери информации в связи с различным временем распространения сигналов в коммутируемых линиях. С целью обеспечения своевременного доступа к информационно-телекоммуникационным услугам по каналам связи с различным временем распространения сигналов разработано специализированное устройство приёма данных [5]. Схема устройства приёма данных приведена ниже на рисунке (*рис. 2*).

Устройство приёма данных работает следующим образом:

✦ на информационные входы коммутаторов и анализатора из разных каналов поступают информационные сигналы (например, совокупность всех информационных сигналов представляет собой требование на доступ к услуге). Каждой последовательности (пакету) информационных сигналов предшествует служебный сигнал; моменты поступления служебных сигналов неизвестны и в общем случае отличаются друг от друга; в коммутаторах служебные сигналы записываются;

✦ на выходах анализатора появляются сигналы установки; между появлением первого и последнего сигналов установки в коммутаторах, в которых уже записаны служебные сигналы, производится

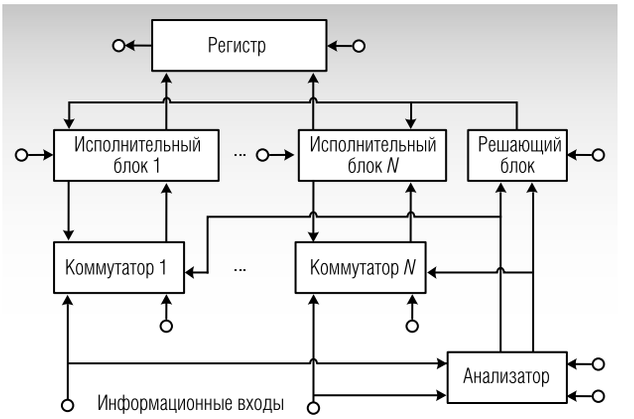


Рис. 2. Структурно-функциональная схема устройства приёма данных

запись поступающих информационных сигналов; считывание этих сигналов запрещено сигналами с соответствующих исполнительных блоков;

✦ в то время, когда выделяется служебный сигнал, с выхода анализатора сигнал установки поступает в соответствующий коммутатор и в решающий блок; коммутатор устанавливается в состояние готовности;

✦ при поступлении в решающий блок сигналов установки со всех выходов анализатора, с выхода решающего блока поступает разрешающий потенциал в исполнительные блоки; исполнительные блоки переключаются и пропускают тактовые импульсы в коммутаторы для считывания информационных сигналов в регистр и далее на выход устройства, например, в центр обработки данных предприятия.

На основе описанного выше устройства приёма данных разработано устройство для сопряжения вы-

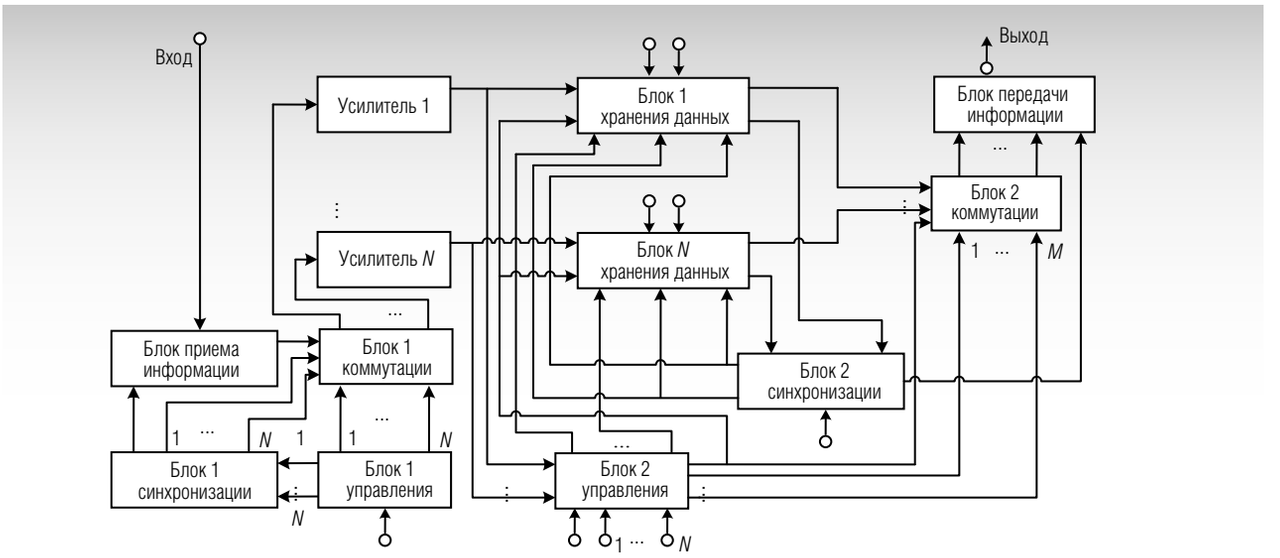


Рис. 3. Структурно-функциональная схема устройства для сопряжения вычислительных ресурсов

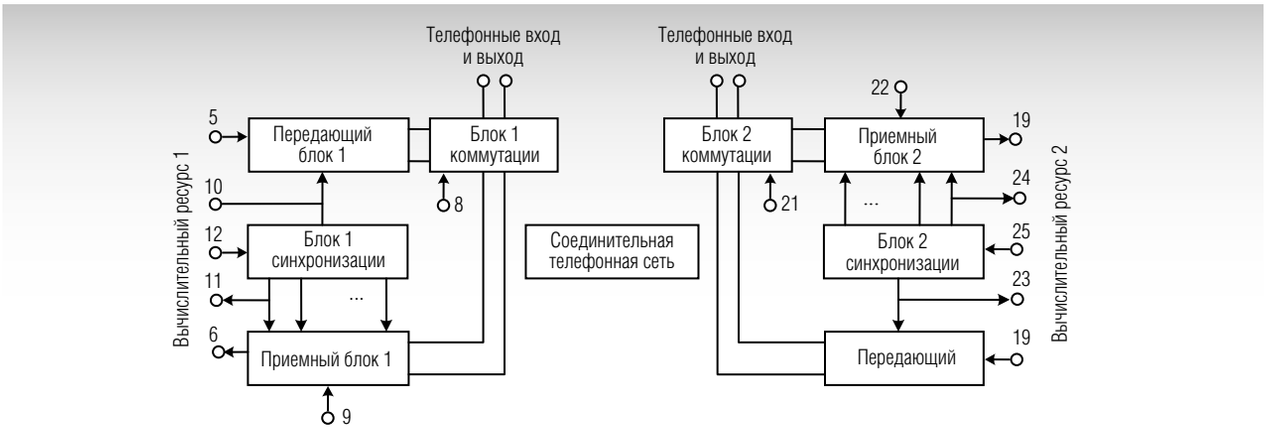


Рис. 4. Структурно-функциональная схема устройства для сопряжения вычислительных ресурсов

числительных ресурсов, предназначенное для достоверной (без потери информации) передачи и приёма информации в многоканальных трактах, образованных в КИС. Данное устройство обеспечивает:

✦ приём пакета информационных сигналов (требование на доступ к услуге) от передающего вычислительного ресурса (например, персонального компьютера субъекта управления);

✦ распределение информационных сигналов этого пакета в разные каналы связи с целью одновременной их передачи в направлении приёмного вычислительного ресурса (например, центра обработки данных);

✦ приём этих сигналов, сборка их в едином пакете и передача пакета информационных сигналов в центр обработки данных.

Схема устройства для сопряжения вычислительных ресурсов приведена на *рис. 3.*

Таким образом, рассмотренные выше системно-технические решения по сопряжению вычислительных ресурсов позволяют обеспечить быстрый доступ к ИТ услугам по каналам с нестабильным временем распространения сигналов.

Подробное описание устройства для сопряжения вычислительных ресурсов и его составных частей — блоков хранения данных, коммутации, управления и синхронизации, приведено в описании изобретения [6].

Обеспечение доступа к ИТ услугам по двухпроводной соединительной телефонной сети

Для организации доступа субъектов управления к услугам КИС предприятия (организации, ведомства) через двухпроводную соединительную телефонную сеть разработано специализированное устройство для сопряжения вычислительных ресурсов. Особенностью данного устройства, по сравнению с другими, является высокая пропускная способность образующихся в телефонной сети трактов передачи данных. Схема устройства для сопряжения вычислительных ресурсов приведена на *рис. 4.*

На схеме цифрами обозначены следующие входы и выходы устройства

✦ 5 и 6 – выход и вход первого вычислительного ресурса;

✦ 18 и 19 – выход и вход второго вычислительного ресурса;

✦ 10 и 11 – выходы синхронизации соответственно передачи и приёма первого вычислительного ресурса;

✦ 23 и 24 – выходы синхронизации соответственно передачи и приёма второго вычислительного ресурса;

✦ 12 и 25 – входы записи «1» в блоки синхронизации;

✦ 8 и 9 – соответственно вход управления и вход установки эталона на стороне первого вычислительного ресурса;

✦ 21 и 22 – соответственно вход управления и вход установки эталона на стороне второго вычислительного ресурса.

Устройство для сопряжения вычислительных ресурсов работает следующим образом.

В исходном состоянии подачей операторами обоих вычислительных ресурсов управляющих сигналов на входы управления блоков коммутации с входов управления 8 и 21 устройство переводится в телефонный режим. При этом телефонные входы и выходы подключаются через блоки коммутации к соединительной телефонной сети. Операторы

обоих вычислительных ресурсов по телефону договариваются о сеансе передачи данных и о порядке смены направлений передачи данных. Например, устанавливается время начала сеанса 12.00, продолжительность сеанса 12 минут, смена направления передачи данных через каждые 2 минут, первое направление передачи от первого вычислительного ресурса. После этого операторы первого и второго вычислительных ресурсов подачей управляющих сигналов на входы управления соответствующих блоков коммутации переводят устройство в режим передачи данных от первого вычислительного ресурса ко второму.

В режиме передачи данных устройство работает следующим образом. С выхода первого вычислительного ресурса информационные сигналы подаются на вход передающего блока 1 с частотой следования синхронизирующих импульсов, поступающих из блока 1 синхронизации. Передающий блок 1 настроен таким образом, чтобы автоматическая телефонная станция в соединительной телефонной сети не принимала интервалы времени прохождения информационных сигналов, равных логическому «0», за паузы между импульсами набора номера. В тоже время интервалы времени прохождения информационных сигналов, равных логической «1» не должны приниматься за импульсы набора номера. На приёмной стороне в приёмном блоке 2 происходит преобразование аналогового сигнала, приходящего из соединительной телефонной сети и подвергшегося искажениям, в цифровые комбинации числом m для каждого логического «0» или «1» и по n бит в каждой комбинации. Например, при реализации аналого-цифрового преобразователя в приёмном блоке в виде ИКМ преобразователя параметр n=8, а параметр m можно выбрать равным трём. Под воздействием двух последовательностей синхронизирующих импульсов соответственно с частотами mnf и mf, поступающих из блока 2 синхронизации, происходит:

✦ поочерёдное накопление в приёмном блоке 2 кодов m дискретных отсчётов одного и того же принятого аналогового сигнала;

✦ поочерёдное сравнение этих кодов с эталонным кодом, подаваемым с входа 22 установки эталона;

✦ по результатам сравнения восстанавливаются информационные сигналы, переданные из первого вычислительного ресурса, и затем передаются во второй вычислительный ресурс.

Аналогичным образом осуществляется передача информационных сигналов в обратном направлении.

Таким образом, рассмотренное устройство для сопряжения вычислительных ресурсов обеспечивает высокую пропускную способность информационного тракта между двумя вычислительными ресурсами в КИС при использовании двухпроводной соединительной телефонной сети для обмена пакетами информационных сигналов. Данный эффект достигается за счёт того, что автоматическая телефонная станция в соединительной телефонной сети воспринимает процесс передачи данных как обычный аналоговый процесс, происходящий при разговоре двух абонентов. Тем самым исключаются возможные прерывания процесса передачи данных при совпадении этих комбинаций с комбинациями специальных кодов автоматической телефонной станции. Тракт передачи данных становится «прозрачным».

Подробное описание устройства для сопряжения вычислительных ресурсов и его составных частей, — передающего и приёмного блоков, блоков синхронизации и коммутации, приведено в описании изобретения [7].

Заключение

В настоящей работе рассмотрен методический подход к реализации проектов по созданию контрольно-учётных систем для КИС. Данный подход основан: на свойстве повторяемости, т.е. цикличности выполнения предприятием-разработчиком однотипных проектов; на использовании инновационных системно-технических решений в сочетании с готовыми изделиями, — специализированными программами. Областью применения данного подхода являются системы поддержки услуг информационных технологий, инвентаризация и контроль состояния технических и программных средств корпоративной компьютерной сети, контроль над её транспортными средствами. Приведены примеры, которые иллюстрируют описанный выше подход, в части разработки на базе инновационных системно-технических решений отдельных компонент проектируемой системы. Эти компоненты обеспечивают быстрый доступ пользователей к ИТ услугам по каналам с нестабильным временем распространения сигналов и высокую пропускную способность информационного тракта между двумя вычислительными ресурсами в КИС при использовании для обмена пакетами информационных сигналов двухпроводной соединительной телефонной сети. ■

Литература

1. Программный комплекс «Business Service Management (BSM)».URL: http://www.bmc.com/solutions/bsm?cmp=redirect_bsm?intcmp=home_bsm_initiatives (дата обращения: 01.08.2012).
2. Программный комплекс «HP Software Manager». URL: http://www8.hp.com/ru/ru/software/software-product.html?compURI=tcm:172-937082&jumpid=reg_r1002_ruru_c-001_title_r0005 (дата обращения: 01.08.2012).
3. ITIL® V3 Glossary Russian Translation v0.92 // ITIL® V3 Translation Project. 30 Apr 2009.
4. Шабанов А.П. Ось адаптивного управления: «информационные системы – организационные структуры массового обслуживания // Бизнес-Информатика, 2010, № 3(13), С. 19-26.
5. Шабанов А.П., Ладиков В.П., Булдаков М.В. Устройство приёма данных // Описание изобретения по авторскому свидетельству SU 1478360 A1, кл. H 04 J 3/16, опубл. 07.05.89 г. в бюл. № 17.
6. Шабанов А.П., Просалков А.С., Ладиков В.П. Устройство для сопряжения ЦВМ // Описание изобретения по авторскому свидетельству SU 1494008 A1, кл. G 06 F 13/00, опубл. 15.07.89 г. в бюл. № 26.
7. Шабанов А.П., Бабичев В.Ю. Двухканальное устройство для сопряжения ЭВМ // Описание изобретения по авторскому свидетельству SU 1735860 A1, кл. G 06 F 13/00, опубл. 23.05.92 г. в бюл. № 19.

ЖУРНАЛ «БИЗНЕС-ИНФОРМАТИКА»
ОСУЩЕСТВЛЯЕТ РАЗМЕЩЕНИЕ РЕКЛАМНЫХ
И РЕКЛАМНО-ИНФОРМАЦИОННЫХ МАТЕРИАЛОВ

Расценки:

Обложка: 2, 3, 4 страница обложки, полноцветная печать, полоса 210×290 мм (А4) – 40 тыс. руб.

Текстовый блок, чёрно-белая печать:
- ✦ полоса – 20 тыс. руб.;
- ✦ 1/2 полосы – 15 тыс. руб.;
- ✦ 1/4 полосы – 10 тыс. руб.;
- ✦ меньший объём – 7 тыс. руб.

Вставка (4 полосы, полноцветная печать – 60 тыс. руб.).

Рекламно-информационный блок (8 полос, полноцветная печать) – 80 тыс. руб.
Рекламно-информационный блок (16 полос, полноцветная печать) – 90 тыс. руб.

Корпоративный специальный выпуск – по договоренности.

Материалы принимаются с учётом следующих параметров:
- ✦ дообрезной формат – 215×300 мм;
- ✦ обрезной формат – 210×290 мм;
- ✦ поле набора полосной рекламы – 190×270 мм – с отступом от границ обрезного формата по 10 мм с каждой стороны;
- ✦ файл TIF, EPS, PDF – разрешение не менее 300 dpi.

STRATEGY OF DEVELOPMENT OF INFORMATION-COMMUNICATION TECHNOLOGIES IN THE REPUBLIC OF AZERBAIJAN. THE FORESIHGT-2020

R. Karayev

Annotation

The analysis of National strategy of informatization for the period till 2015 is given. The limited opportunities of rating doctrine of ITU at the decision of a question of the further development of branch connected to the Concept of development of the country up to 2020 are marked. The question of development of the market strategy reflecting mission of ICT at a stage of transition of Republic from resource-export to resource-innovative («not oil») economy is rises.

Key words: ICT, development strategy, «not oil» economy, mission of ICT, Azerbaijan.

THE ROLE OF INNOVATION AND MOBILE TECHNOLOGIES IN THE DEVELOPMENT OF SMALL AND MEDIUM ENTERPRISES IN EMERGING MARKETS

V. Taratoukhine, E. Bazhenova

Annotation

The article contains the analysis of the impact of information and communication technologies as well as mobile technologies for the conduction of small and medium business in the emerging countries. Special approach is suggested for the development of the concept of business-processes management on small and medium enterprises.

Key words: information and communication technologies, mobile technologies, business—process, enterprise mana-gement, small business, software-as-a-service, cloud techno-logies.

FORECASTING OF ECONOMIC PROCESSES ON THE DEFINITION BASIS OF THEIR OPTIMUM BASE PARAMETERS

Y. Kuznetsov

Annotation

Features of a new method and the algorithms of forecasting based on definition of optimum base parameters of processes and signals are considered, is optimum digitization on level and time in time numbers. It is shown that information systems of forecasting, developed on the basis of this method, allow predicting effectively economic processes.

Key words: Model of the initial data, optimum digitization of the data, optimum data base parameters, predicting operators, forecast.

R/S ANALYSES OF STOCK MARKET

A. Zinenko

Annotation

The paper describes a new statistic approach developed by Harold Herst. This temporal series analyses method makes it possible to define the character of temporal series: whether it is persistent (trend), random or anti-persistent. We apply the R/S analyses algorithm to stock indexes and make a conclusion about the character of temporal series.

Key words: Hurst Index, self-similarity, power functions.

METHOD OF ANNOTATED SUFFIX TREE FOR SCORING THE EXTENT OF PRESENCE OF A STRING IN TEXT

B. Mirkin, E. Chernyak, O. Chugunova

Annotation

A modification of the method of annotated suffix tree (AST) is presented. The modified method does not need pre-specifying the depth of the tree to be built

and involves an intuitive scoring function. The method applies to build a publication-to-keyphrase table, which further on is used to advance in the analysis: (a) interrelation between a text corpus and a pre-specified set of key phrases, (b) relation between a text corpus and a taxonomy of the domain. This is illustrated with real-world examples.

Key words: text analysis, annotated suffix tree, interpretation, conceptual cluster.

✦

METHODS OF THE CHANGES ANALYSIS OF GLOBAL AND LOCAL PROPERTIES OF TEMPORAL DIGRAPHS

V. A. Kokhov, V.V. Kokhov

Annotation

The new methods for the structural analysis of global and local properties of temporal digraphs are offered. The methods use models of digraphs complexity and to allow analyze the tendencies of change of the global and local properties of temporal digraphs. Comparative analysis method of two approaches to a problem determination fragments significance of temporal digraphs are considered.

Key words: Temporal digraph. Complexity of digraph. Local invariant of digraph. Global invariant of digraph. Trend analysis. The importance of a fragment.

✦

SOFTWARE APPLICATION FOR INVESTMENT PORTFOLIO STRUCTURE MANAGEMENT

A. Isavnin, D. Galiev

Annotation

The article presents a program product for portfolio selection management. Models and algorithms are described. We also describe the results of experiment with real assets traded at Russian stock exchange market. The strengths and weaknesses of different approaches are reviewed. We describe development potential of the product as well. The application of the software may be associated with exchange trades by interaction with popular trade terminals.

Key words: investment portfolio; modeling; genetic algorithms, neural networks.

✦

MONITORING AND PLANNING OF DEVELOPMENT OF INFORMATION SUPPORT SYSTEMS FOR CORPORATE GOVERNANCE AND STRATEGIC MANAGEMENT

D. Isaev

Annotation

In the paper the questions of monitoring and planning of Performance Management Information Support Systems (PMISS) development are discussed. The role of the monitoring and planning functions in the general development management system is disclosed, basic requirements to these functions and general principles of their practical implementation are formulated. The main elements of the monitoring and planning system are described, the perspectives of further research are proposed.

Key words: corporate governance; strategic management; performance management system; information support; monitoring; planning.

✦

THE INNOVATIVE APPROACH TO THE CREATION AND TESTING OF CONTROL AND ACCOUNTING MODULES FOR CORPORATE INFORMATION SYSTEMS

M. Arakelyan, E. Chepin, A. Shabanov

Annotation

In the paper a methodological approach to development of control-accounting modules of corporate information systems is discussed. The approach is based on cyclic recurrence of similar projects performed by a developing company, as well as on innovative technological solutions combined with existing specialized programs. The approach may be applied in the fields of information technology service support systems, as well as inventory and monitoring of hardware and software of corporative computer networks and control of its vehicles.

Key words: corporate information systems, project management, control and accounting systems, systems and technical solutions, software products.

Тематичекие рубрики журнала «БИЗНЕС-ИНФОРМАТИКА»

№	Рубрика
1	Математические модели социальных и экономических систем
2	Программная инженерия
3	Анализ данных и интеллектуальные системы
4	Математические методы и алгоритмы решения задач бизнес-информатики
5	Моделирование и анализ бизнес-процессов
6	Информационные системы и технологии в бизнесе
7	Электронный бизнес
8	Интернет-технологии
	Дополнительные рубрики вне номенклатуры
9	Тематические обзоры
10	Правовые вопросы бизнес-информатики
11	Стандартизация, сертификация, качество, инновации
12	Дискуссионный клуб / Опыт бизнеса

К СВЕДЕНИЮ АВТОРОВ

Редакция просит авторов при оформлении статей и тематических обзоров придерживаться следующих правил и рекомендаций:

1. Предоставляемый авторами материал должен соответствовать рекомендуемой структуре статей журнала.

2. Статья направляется в редакцию в электронном виде (в формате MS WORD версия 2003) и в виде бумажной копии, распечатанной на одной стороне листов А4. Первая страница оригинала подписывается всеми авторами статьи.

3. Ориентировочный объём статьи, предлагаемой к публикации, – 20–25 тыс. знаков (с пробелами) или 30–35 тыс. знаков – для обзорных статей по направлениям.

4. Кегль набора – 12 пунктов с полуторным интервалом. Нумерация страниц – вверху по центру. Поля: левое – 2,5 см, верхнее, нижнее и правое – по 1,5 см.

5. При наборе выключных и строчных формул должен быть использован редактор формул MS Equation. В формульных и символических записях греческие (русские) символы, а также математические функции записываются прямыми шрифтами, переменные аргументы функций в виде английских (латинских) букв записываются наклонным начертанием (курсивом), например, «cos *a*», «sin *b*», «min», «max».

6. Формулы, таблицы и сноски (не концевые) оформляются стандартными средствами редактора MS WORD. Нумерация формул, рисунков и таблиц – сквозная, по желанию авторов допускается двойная нумерация формул с указанием структурного номера раздела статьи и – через точку – номера формулы в разделе.

7. Рисунки (графики, диаграммы и т.п.) оформляются средствами Word, Excel, Illustrator. Ссылки на рисунки в тексте обязательны и должны предшествовать позиции размещения рисунка. Допускается использование графического векторного файла в формате wmf/emf или cdr v.10. Фотографические материалы предоставляются в формате TIF или JPEG с разрешением не менее 300 dpi.

8. Библиографический список составляется в соответствии с требованиями ГОСТ. Нумерация библиографических источников – в порядке цитирования. Ссылки на иностранную литературу – на языке оригинала без сокращений.

Структура статей строится по правилам, рекомендованным журналом «Бизнес-информатика».

Плата с аспирантов за публикацию рукописей не взимается.

Журнал «БИЗНЕС-ИНФОРМАТИКА» Рекомендуемая структура статей

Журнал публикует исследовательские научные статьи, размещаемые в рубриках журнала, тематические обзоры, отражающие современное состояние проблем в области бизнес-информатики и сообщения, размещаемые в рубриках «Дискуссионный клуб» и «Опыт бизнеса».

Титульный лист рукописи начинается с указания Ф.И.О. авторов публикации с обязательным указанием учёной степени, учёного звания, должности, основного места работы и e-mail. Титульный лист должен быть подписан всеми авторами статьи.

I. Исследовательские научные статьи (для размещения в тематических рубриках)

Редколлегия рекомендует авторам после названия статьи приводить ***аннотацию***, в которой излагается краткое содержание статьи, её основные результаты и область применения. Авторам рекомендуется структурировать статью, выделяя ***введение***, содержащее описание проблемы или задачи, обзор существующих подходов или методов решения, их недостатки, и основную цель статьи; ***постановку задачи***, включающую допущения и ограничения; ***содержательную часть*** статьи, в которой предлагаемые решения должны быть аргументированы и сравниваться с существующими подходами или решениями; ***заключение***, содержащее краткое изложение новых результатов, полученныхв статье и область их применения; ***библиографический список***, оформленный в соответствии с ГОСТ. Текст статьи должен содержать нумерованные ссылки на все указанные библиографические источники. Структурирование статьи и нумерация её разделов проводится по усмотрению авторов.

Возможный вариант структуры статьи:

- Ф.И.О;
- учёная степень, учёное звание, должность, основное место работы, e-mail;
- название статьи.
- аннотация;
- 1. Введение.
- 2. Постановка задачи.
- 3. Основная содержательная часть статьи.
- 4. Экспериментальные результаты (опционально).
- 5. Заключение.
- 6. Библиографический список.

II. Тематические обзоры по направлениям

Редколлегия рекомендует авторам структурировать обзор, выделяя аннотацию, содержащую тематику, краткое содержание обзора и область применения; ***введение***, в котором даётся краткий исторический обзор тематики; ***содержательную часть*** обзора с критическим анализом существующих направлений; ***заключение***, в котором отражаются перспективы развития в рамках обозреваемой тематики и наиболее интересные направления с точки зрения научных и практических разработок и методов; ***библиографический список***, оформленный в соответствии с ГОСТ.

Текст обзора должен содержать нумерованные ссылки на все указанные библиографические источники. Структурирование обзора и нумерация его разделов проводится по усмотрению авторов.

Возможный вариант структуры обзора:

- Ф.И.О;
- учёная степень, учёное звание, должность, основное место работы, e-mail;
- название обзора;
- аннотация;
- 1. Введение.
- 2. Основная содержательная часть обзора.
- 3. Заключение.
- 4. Библиографический список.

Редколлегия журнала проводит обязательное рецензирование рукописей. Статья принимается к публикации только после получения положительного заключения рецензента и одобрения на заседании редакционной коллегии журнала.