

ISSN 1998-0663

№3(17)–2011

<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.
Белов В.В.
Грибов А.Ю.
Громов А.И.
Гюнтер Х. (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.
Мулазани М. (Италия)
Пальчунов Д.Е.
Силантьев А.Ю.
Таратухин В.В.
Терзани С. (Италия)
Ульянов М.В.

В ЭТОМ НОМЕРЕ:

**ОБРАЗОВАТЕЛЬНЫЕ
СООБЩЕСТВА:
ОТКРЫТЫЙ КОНТЕНТ**

**КОНКУРЕНТОСПОСОБНОСТЬ
ПРИКЛАДНЫХ
ПРОГРАММ**

**СЕМАНТИЧЕСКИЕ
ФИЛЬТРЫ
И ПРЕОБРАЗОВАНИЯ**

**ИНТЕРНЕТ-ОПЕРАТОРЫ:
ПРАВОВАЯ
КЛАССИФИКАЦИЯ**

*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(17)–2011

СОДЕРЖАНИЕ

Проблемы подготовки специалистов в области ИКТ

Ю.В. Таратухина, О.Г. Чамина

Сетевые сообщества образовательной направленности в поликультурном контексте: метод открытого контента 3

Программная инженерия

Ю.П. Ехлаков, Е.А. Янченко

Методика оценки конкурентоспособности прикладных программных продуктов 10

Анализ данных и интеллектуальные системы

А.В. Кириллов

Метод семантического преобразования обобщенных запросов на основе базы целей 16

В.И. Новицкий

Подход к автоматическому поиску переводных словосочетаний на основе синтаксической информации и многоуровневой фильтрации 24

С.В. Гусев, А.М. Чеповский

Модель для идентификации естественного языка текста 31

Математические методы и алгоритмы решения задач бизнес-информатики

А.А. Незнанов, Ю.В. Старичкова

Программный комплекс для генерации семейств транзитивных графов степени 4 36

Моделирование и анализ бизнес-процессов

А.Н. Бирюков, Д.Ю. Клецких

Метод оценки зрелости организации с использованием модели VRMM и его реализация 45

А.П. Шабанов, М.А. Аракелян

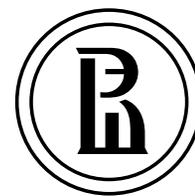
Технология контроля качества обслуживания требований в организационных структурах, предоставляющих услуги массового характера 53

Правовые вопросы бизнес-информатики

А.К. Жарова

О необходимости правовой классификации операторов сети Интернет 60

Annotations 66



БИЗНЕС-ИНФОРМАТИКА

№3(17)–2011

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.

Редакция:
Заместители главного редактора
Горбунов А.Р., Исаев Д.И.

Научный редактор
Лычкина Н.Н.

Технический редактор
Осипов В.И.

Корректор
Витюк Е.Б.

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»

СЕТЕВЫЕ СООБЩЕСТВА ОБРАЗОВАТЕЛЬНОЙ НАПРАВЛЕННОСТИ В ПОЛИКУЛЬТУРНОМ КОНТЕКСТЕ: МЕТОД ОТКРЫТОГО КОНТЕНТА

Ю.В. Таратухина,

*доцент кафедры инноваций и бизнеса в сфере информационных технологий
Национального исследовательского университета «Высшая школа экономики»*

О.Г. Чамина,

*старший преподаватель кафедры прикладной лингвистики
Ульяновского государственного технического университета*

Адрес: г. Москва, ул. Кирпичная, д. 33/5

E-mail: jtaratuhina@hse.ru, o.chamina@ulstu.ru

В статье делается попытка описать общие тенденции организации педагогической коммуникации в образовательных сетевых сообществах разных культур за счет формирования открытого контента, на основании рассмотрения социокультурной специфики образовательных систем.

Ключевые слова: сетевые сообщества, открытый контент, информационная среда, коммуникативно-педагогические инновации, образовательное пространство, кросс-культурная дидактика, мультимедийная дидактика.

Введение

Современное образовательное пространство представляет собой сосуществование оффлайн и он-лайн структур и, собственно, их взаимодействие не только в монокультурном но и в поликультурном формате, что обусловлено интеграционными процессами и вхождением России в международное образовательное сообщество. Безусловно, одной из характерных особенностей

современного образовательного пространства является возникновение сетевых сообществ, многие из которых являются интернациональными. Соответственно, представляется актуальным исследование взаимосвязи когнитивных констант и некоторых особенностей образовательных процессов в сетевых сообществах, в частности, особенностей формирования открытого контента в сетевых сообществах, относящихся к разным культурам.

1. Социокультурная специфика образовательных моделей

В настоящее время мировое образовательное пространство включает в себя несколько национальных моделей образования, зачатую имеющих разноуровневый, неоднородный характер. В первую очередь, это обусловлено дифференцированным социокультурным базисом, различными уровнями развития экономической и социальной систем. Сегодня можно выделить три типа регионов с разной степенью активности реагирующих на новые веяния в образовательных процессах [2]. К числу регионов, очень активно генерирующих образовательные процессы, можно отнести страны Западной Европы, США, Канаду, а также некоторые страны Азиатско-Тихоокеанского региона. Активно реагируют на новые тенденции образовательного характера и Страны Латинской Америки. В то же время некоторые страны Юго-Восточной Азии и центральной Африки остаются совершенно инертными по отношению к интеграционным процессам в образовательном пространстве. При этом само существование образовательного пространства в рамках коммуникативно-культурного пространства будет, так или иначе, заключено в социальные, временные, поведенческие границы, а также в когнитивные мыслеречевые модели.

Существует довольно большое количество классификаций социокультурной коммуникативной специфики. В настоящей работе мы будем опираться на концепцию Г. Хофстеде, который выделил несколько основополагающих аспектов, характеризующих специфику коммуникации в разных культурах [7]. К таким аспектам относятся:

- ◆ индекс индивидуализма/коллективизма, основанный на самоориентации личности;
- ◆ степень иерархической дистанционности, отражающая ориентацию на власть и авторитет;
- ◆ степень избегания неопределенности, характеризующая уровень готовности к риску;
- ◆ мужской или женский стиль деловых взаимоотношений, ориентированный на достижения.

Данную классификацию можно применить при анализе социокультурной специфики образовательных моделей. С. Мясоедов [4] утверждает, что с точки зрения дихотомии критериев индивидуализма и коллективизма в культуре цель обучения в индивидуалистских странах (США, Канада, Австралия,

Великобритания и т.д.) состоит в том, чтобы подготовить индивида к самостоятельной и полноценной жизни в меняющемся мире. В данном контексте даются в основном общие алгоритмы и модели знаний, а также общий инструментарий, который можно применить в зависимости от конкретных задач. Огромная роль уделяется нестандартным, креативным подходам в решении задач, проектной деятельности, умению работать в команде. Дидактической доминантой является самостоятельно выстраивание знаний для решения задач, связанных с разнообразными проектами. Большое внимание уделяется непрерывному обучению, поскольку считается, что информация быстро устаревает.

В коллективистских культурах (Китай, Япония, арабские страны и т.п.) есть предубеждение, что обучение – во многом удел людей молодого возраста. Большой упор делается на получение фундаментальных знаний, что зачастую сопряжено с заучиванием и запоминанием огромного количества информации. Нередко теория не подкреплена практическими навыками. Таким образом, можно сказать, что в коллективистских культурах существует проблема практической применимости фундаментальных теоретических знаний.

С точки зрения критериев «низкой» и «высокой» дистанции власти в разных культурах, в образовательных пространствах есть некоторая специфика. Например, в культурах с низкой дистанцией власти (США, Великобритания, Канада, Австралия, страны Центральной Европы) обучающийся является центральной фигурой, все «крутится вокруг него». Учитель, тьютор, преподаватель, является скорее сопровождающей фигурой. Он не дает знания, а помогает самостоятельно находить необходимую информацию и делать выводы. Поэтому в культурах данного типа высока роль элективных курсов, внеаудиторных занятий. Напротив, в странах с высокой дистанцией власти (Китай, Япония) профессор – это «гуру», центральная фигура. Передаваемая им информация зачастую позиционируется как «бесценная», а занятия носят преимущественно аудиторный характер. С точки зрения критериев «мужественности» и «женственности» культур, «женственные» культуры (Швеция) ориентированы на психологический комфорт личности, социальную адаптацию и среднего обучающегося. В «мужественных» культурах (США) идет ориентация на лучшее, высокая конкуренция, внешние атрибуты академических успехов (портфолио, победы в олимпиадах, конкурсах и т.п.).

С точки зрения критериев избегания неопределенности, в культурах с низкой степенью избегания неопределенности процесс обучения часто ведется по неструктурированным программам, с высоким уровнем вариативности и размытыми критериями оформления. В культурах с высокой степенью избегания неопределенности все подчинено строгому расписанию, инструкциям и учебно-методическим регламентациям. Здесь высока роль образовательных стандартов и типовых рекомендаций, которые постоянно совершенствуются, а в системе образования высока роль контролирующих органов.

Отсюда можно сделать вывод, что в различных культурах будут иметь место различные ценностные ориентиры в образовании, национально-социокультурная специфика работы с информацией, дискурсивные структуры логических построений и т.д.

Что касается ценностной стороны образовательного процесса, то западные (индивидуалистские) культуры ориентированы преимущественно на результативность, в то время как восточные (коллективистские) ориентированы на сам процесс и сохранение гармоничной атмосферы.

Например, представители восточных культур при ответах на вопросы в рамках тестирования, как правило, негативно относятся к вопросам нацеленным на формулирование собственной точки зрения на проблему.

В западных культурах наоборот, вопросы, нацеленные на знание большого количества теоретической информации, не вызывают интерес. Теоретическая информация используется в ограниченных объемах, в то же время используется большое количество кейсов и практических заданий, рассчитанных на формирование навыков гибкости, адаптивности к нестандартным ситуациям, креативности.

Дискурсивные структуры логических построений, используемые в образовательных процессах, также имеют различия: как правило, представители западных культур чаще используют индуктивные речевые модели, восточных — дедуктивные. Безусловно, данные особенности создают проблемы адекватного взаимопонимания и эффективного обучения при реализации образовательных программ поликультурного формата, так как на сегодняшний день кросс-культурные дидактические исследования практически не ведутся.

С учетом обозначенных различий, представляет интерес вопрос о том, каким образом данные

культурологические особенности проявляют себя в формате сетевого образования, в том числе в поликультурном формате.

2. Основные особенности образовательной деятельности в сети

Основной особенностью образовательного процесса в информационном обществе является смена дидактического направления. Главной целью образования больше не является дать обучаемому сумму некоторых знаний, умений и навыков: доминантой становится развитие способностей к самообразованию, гибкое и креативное мышление, высокая мотивация.

Соответственно, для решения данных задач необходима трансформация образовательной среды, в основе которой лежит информационная инфраструктура, на базе которой строится индивидуальная образовательная среда.

Что касается форм обучения, то в данном контексте они также в большинстве своем отходят от традиционных (лекции, семинары) и в основном тяготеют к проектному формату, носящему познавательный-деятельностный характер.

С.В. Бондаренко [1] отмечено, что в сетевых сообществах присутствует новый принцип коммуникации. Согласно традиционной модели педагогической коммуникации в коллективе использовалась модель «от одного ко многим», а при индивидуальном обучении — «от одного к одному». В процессе функционирования сетевых сообществ образовательной направленности действует коммуникативный принцип «от всех ко всем». Данная коммуникативная модель ценна именно тем, что кроме формального и неформального общения на профессиональные темы, обмена опытом и педагогическими практиками, формируется толерантность, групповой и критический подход к решению задач, освоение децентрализованных моделей.

Как правило, в сообществах образовательной направленности можно выделить следующие виды педагогической деятельности: сетевые публикации, конкурсы, консультации, комментарии к материалам, дистанционное обучение, творческие лаборатории, проекты и т.д.

С.В. Бондаренко предлагает классификацию сетевых образовательных сообществ по следующим

критериям: *количественный состав участников, виды совместной деятельности, ориентация на определенные группы обучаемых.*

На основании количественного критерия глобальную социальную общность Интернет можно представить как состоящую из макросоциальных и микросоциальных сетевых сообществ.

По видам *совместной деятельности* автор предлагает следующую классификацию сетевых сообществ:

❖ виртуальное сетевое сообщество преподавателей (решаемые задачи: составление планов, совместная методическая деятельность и т.д.);

❖ виртуальное сетевое сообщество преподавателей и экспертов (решаемые задачи: двусторонний канал между преподавателями и экспертами по определенному предмету, благодаря которому учителя получают подготовку, новости, методическую поддержку и помощь, а эксперты имеют представление о реальном положении дел в сфере их профессиональных интересов);

❖ виртуальное сетевое сообщество учащихся и преподавателей (решаемые задачи: участники сообщества должны выполнить предложенный проект, в данном случае преподаватель воспринимается как один из участников, а не как абсолютный лидер);

❖ виртуальное сетевое сообщество учащихся (решаемые задачи: обсуждение, выполнение проектов, решение общих проблем, взаимопомощь);

❖ виртуальное сетевое сообщество учащихся, преподавателей и экспертов – кластер виртуальных сетевых сообществ (решаемые задачи: учащиеся могут познакомиться с проблемами реального мира, таким образом, появится возможность преодолеть разрыв, который часто разделяет приобретение новых навыков и их применение в реальности.

Данная типология может быть реализована практически во всех видах сетевых сообществ: блогах, вики-сообществах, веб-форумах, чатах и т.п.

Под открытым контентом в настоящем контексте будем понимать методологию генерации и распределения информации как социального достояния в Интернет-среде. По сути, это все, что может быть представлено в цифровом виде (текстовый, музыкальный, видеоконтент и т.п.), исключая программное обеспечение.

Сам по себе феномен открытого контента представляет собой комплексное образование, включающее в себя методологии производства интеллек-

туальных продуктов, собственно сами продукты, находящиеся в свободном доступе, а также некую идеологию, объединяющую участников проектов.

Таким образом, сетевое образовательное сообщество представляется в виде веб-сайтов, на страницах которых присутствуют текстовые и мультимедийные материалы и материалы смежной тематики. Доступ к материалам и переход в гипертекстовом пространстве осуществляется при помощи внутренних и внешних ссылок. В отличие от офф-лайн формата, сетевые сообщества содержат интерактивные элементы в виде тестов, анимаций, игр, онлайн опросов, рассылок, блогов и системы поиска. Как правило, каждому сетевому образовательному сообществу, как веб-ресурсу, присваивается адрес в Интернет-пространстве, по которому его можно найти среди множества других сообществ.

К характерным чертам методологии открытого контента можно отнести ценность знания как такового, а также инновационность, которая обеспечивается за счет аккумуляции знаний участников сообщества.

По мере накопления некоторой критической массы образовательного контента, а также информации методического характера, возникает возможность создания уникальной образовательно-информационной среды для индивидуального обучения. Фактически получается, что во многих ситуациях обучающийся сам формирует индивидуальное образовательное пространство, а преподаватели в лишь помогают ему. Также можно отметить, что в сети имеет место движение за открытое обучение, которое выражается в возможности свободного распространения курсов и методик и создании средств их проектирования.

Стратегия развития образования в информационном обществе, по сути, нацелена на помощь обучающимся быстрее адаптироваться в условиях быстроменяющейся реальности, что выражается в переходе от работы с пассивным, «книжным» контентом к работе с интерактивным контентом.

Определим основные характеристики открытого контента, обязательные для любых сетевых образовательных сообществ:

● наличие «оригинального» контента. Под оригинальным контентом образовательного сообщества подразумевается контент учебно-методического содержания, выполняющий образовательную функцию, который полезен всем пользователям сообщества. В первую очередь, выделяется разви-

тая система рубрикации контента. Контент должен быть независим от дизайна веб-ресурса, что делает возможным участие в его разработке всех пользователей данного сообщества;

- широкие возможности работы с контентом. Возможность редактирования, добавления, изменения, удаления контента. При этом не требуется специальных технических знаний;

- встроенные системы поиска по различным параметрам. Наличие механизмов поиска контента, как расширенного, так и ключевого. Поисковая система должна обеспечивать полнотекстовый поиск информации портала. Таким образом, пользователь должен иметь возможность осуществлять поиск вхождения определенного слова или фразы в текст документов, полей описания ресурсов и т.д. Пользователь составляет поисковый запрос, отправляет его и получает список результатов, ранжированный по степени совпадения или по дате. Тривиальный поисковый запрос может представлять собой просто слово или фразу. Может быть предусмотрена возможность расширенного поиска, где пользователь заполняет поисковую форму;

- интегрированные (внутри системы) способности к адаптации интерфейса в зависимости от требований индивидуальных пользователей. Четкая и продуманная система пользовательских ролей и прав доступа для работы с контентом для пользователей сообщества (роль преподавателя, учителя, студента и т.д.);

- способность к управлению информацией во времени. Возможность взаимодействия пользователей внутри сетевых образовательных сообществ: добавление комментариев, отправка писем, содержащих вопросы по данной теме, возможность общения в форуме и т.д., а также возможность наращивания контента за счет простого подключения дополнительных модулей.

Интерактивность сетевого образовательного сообщества реализуется в следующих режимах:

- ◆ поисковой (инициативной);
- ◆ обучающей (активной);
- ◆ контрольной (тестирующей).

Учитывая вышеуказанные характеристики контента и диалоговый режим информационный среды образовательных сообществ, веб-ресурсы такого типа должны быть представлены системой управления доступом к контенту сайта (Content Management System, CMS).

Существует множество способов управления контентом веб-ресурса, начиная от простого механизма веб-журналов, допускающих ограниченную публикацию содержимого, до полнофункциональной интегрированной прикладной системы, на основе которой можно создать собственную систему управления контентом. При наличии большого количества вариантов в области систем с открытыми исходными кодами выбрать систему, подходящую для создания веб-ресурса подобного типа не так-то просто.

К основным функциональным возможностям CMS относятся:

- ◆ легкая расширяемость с использованием модульной интегрированной среды (отображение таксономии, персональные сообщения, закладки и т.д.);

- ◆ персонализация среды для индивидуального представления контента, основанного на предпочтениях пользователя;

- ◆ ролевая система полномочий, определяющая права доступа для просмотра и редактирования контента;

- ◆ полная индексация контента для поддержки операций поиска;

- ◆ поддержка других форм контента, таких как голосования, тематические комментарии и дискуссии;

- ◆ административная поддержка ведения журналов, анализа и веб-администрирования.

В конечном счете, сетевое образовательное сообщество представляет собой управляемое собрание ресурсов в интегрированной информационной среде. Информационная среда веб-ресурса позволяет конечному пользователю взаимодействовать с его контентом как с единым целым, а не с набором отдельных коллекций и услуг.

Система управления контентом не только создает среды для совместной творческой деятельности (форумы, информационные письма, галереи изображений, системы загрузки файлов и многое другое), но и обеспечивает надежную защиту от несанкционированного доступа, благодаря хорошо спроектированной системе безопасности с использованием технологий PHP, MySQL, JavaScript, HTML.

Для формирования открытого контента в сетевых сообществах используются такие функциональные модули CMS-системы как новости, подписка / рассылка новостей, опросы и голосования,

форумы, обсуждения, доски объявлений, система «медиа-хранилище», вики-модули, блоги, чаты, комментарии, модули видео-чата, гостевые книги и т.д.

Таким образом, веб-ресурс такого типа представляет гибкую и удобную систему работы с контентом и выполняет множество функций, необходимых и достаточных для управления информацией:

- позволяет создавать, удалять и иерархически упорядочивать различные типы контента;
- автоматически формирует удобные средства навигации по сайту (ссылки на документы, различные меню, карту сайта, списки страниц, указатели пройденного пути и т.п.);
- оперирует разделами различных типов (статьи, новостные ленты, форумы, доски объявлений, почтовые формы, опросы, голосования и т.п.);
- наполняет страницы сайта блоками контента разных типов (текст, изображение, список, таблица и т.п.).

3. Специфика организации открытого контента в сетевых сообществах образовательного характера в разных культурах

Использование метода открытого контента в сетевых сообществах разных культурах носит дифференцированный характер.

Факторы, обуславливающие воздействие культурных констант на распространение образовательных сервисов веб 2.0, на сегодняшний день являются недостаточно изученными. Очевидно, что люди в разных культурах по-разному относятся к новым технологиям, новым сервисам и новым видам контента. Безусловно, все это должно учитываться при организации международных образовательных сообществ, в том числе и сетевых.

Опираясь на концепцию Г. Хофстеде, можно сделать вывод, что огромное влияние на характер и результативность коммуникации в сетевых образовательных структурах будут оказывать такие критерии, как специфика восприятия информации, открытость к инновациям, характерные особенности коммуникации в конкретной культуре, индексы индивидуализма-коллективизма и т.д. Жители стран с низкой дистанцией власти и низким уровнем избегания неопределенности легче приспосабливаются к новым компьютерным технологиям,

легче и эффективнее обучаются через Интернет и проявляют большую активность в социальных сетях образовательного характера. Например, если рассматривать взаимосвязь между индивидуализмом и проявлением активности при формировании открытого контента в образовательных социальных сетях, то можно заметить, что представителями стран с более высоким уровнем индивидуализма контент чаще дополнялся и менялся. Более того, пользователи чаще пытались производить уникальный контент, а не заниматься ретрансляцией уже существующего.

Интересны наблюдения, связанные с индексом дистанции власти в разных культурах. По сути, обмен информацией в сетевых сообществах образовательного характера представляет собой неиерархический процесс – информация идет «ото всех ко всем». Поэтому в обществах с высокой дистанцией власти, процесс принятия таких коммуникативных моделей в образовании будет требовать больших временных затрат и, возможно, будет менее эффективным, поскольку будет построен по принципу жесткой централизации.

Индекс избегания неопределенности может быть интересен тем, что в странах культурной группы с высоким индексом избегания неопределенности контент скорее будет организовываться в соответствии с существующими правилами и нормами, с воплощением минимума уникальных идей. Сетевые сервисы, характеризующиеся непредсказуемостью, будут более распространены в образовательных сообществах культур с низким уровнем избегания неопределенности.

По нашим наблюдениям, страны с высоким индексом индивидуализма (США, Великобритания) будут, в первую очередь, активными создателями уникального образовательного контента в сетевых сообществах. Интересно и то, что в культурах, имеющих коллективистскую специфику, образовательные сетевые сообщества будут функционировать преимущественно для потребления контента: передачи и обмена знаниями, получения рекомендаций и консультаций (Китай, Греция, Испания).

Более того, в культуры с высоким индексом индивидуализма нацелены на перформативность личности, что обуславливает большое наличие визуальных средств репрезентации знаний. В образовательных сообществах данных культур, например, можно наблюдать большое количество видеоконтента, наряду с текстовым форматом. Например,

в США, так или иначе, преобладает визуальный стиль, во Франции господствует аудиальная система репрезентации информации, Великобритания представляет собой нечто среднее с преобладанием визуальной доминанты. Можно предположить, что во многих образовательных сообществах США будет преобладать визуальный контент, в то время как во Франции – аудиальный.

Таким образом, мы имеем возможность наблюдать влияние когнитивных факторов социокультурного характера на некоторую специфику организации открытого контента в сетевых сообществах образовательного характера разных культурных групп. Все это, несомненно, представляется очень важным в контексте поликультурного образовательного пространства.

Заключение

Таким образом, можно отметить, что сетевое образовательное сообщество представляет собой управляемое собрание ресурсов в интегрированной информационной среде. Основная особенность коммуникации в сетевых образовательных сообществах обусловлена принципом «от всех ко всем». Ценность данной коммуникативной модели состоит в том, что, кроме профессиональной коммуникации теоретического и практического характера,

формируется толерантность, коллективный подход к решению задач и т.п.

Информационная среда веб-ресурса позволяет конечному пользователю взаимодействовать с его контентом как с единым целым, благодаря чему создается качественно иная форма организации педагогической коммуникации в дополнение к существующей традиционной. Применение методологии открытого контента обеспечивает инновационность, обусловленную кумуляцией знаний участников сообщества, что обеспечивает возможность создания уникальной образовательно-информационной среды.

В то же время следует отметить дифференциацию в моделях формирования открытого контента сетевых образовательных сообществ разных культур. Предположительно, это связано с национальными когнитивными моделями коммуникации и специфическими способами работы с информацией. В рамках поликультурных образовательных процессов, несомненно, необходимо также учитывать и прагматическую специфику, которая, в свою очередь, отражается в специфике эргодиизама электронных учебных материалов в разных культурных форматах и влияет на эффективность обучения в целом. В настоящем контексте имеет смысл поставить вопрос о необходимости разработок в области кросс-культурной мультимедийной дидактики. ■

Литература

1. Бондаренко С.В. О типизации виртуальных сетевых сообществ образовательной направленности // Новые инфокоммуникационные технологии в социально-гуманитарных науках и образовании: современное состояние, проблемы, перспективы развития. / Под общ. ред. А.Н. Кулика. -М.: Логос, 2003. С. 399-407.
2. Лиферов А.П. Культурологическая составляющая интернационализации мирового образования. Рязань. Издательство Рязанского государственного педагогического университета, 2006 – 245 с.
3. Меськов В.С., Куликова И.В., Мамченко А.А. Открытый контент как феномен и модель обустройства обществ, базирующихся на знаниях. Открытое образование №5, 2006. С. 71-83.
4. Мясоедов С.П. Управление бизнесом в различных деловых культурах – М: Вершина, 2009. – 320 с.
5. Таратухина Ю.В., Мальцева С.В. Сетевые сообщества: коммуникационные аспекты. Автоматизация и современные технологии. №2. 2008. С. 21-26.
6. Триандис Г. Культура и социальное поведение. М: Форум. 2011. – 378 с.
7. Hofstede G. Culture's Consequences, International Differences in Work Related Values, Sage, 1980
8. Драхлер А.Б. Актуальные проблемы развития сетевых педагогических сообществ. Режим доступа: http://ito.edu.ru/sp/SP/SP-0-2009_11_24.html.

МЕТОДИКА ОЦЕНКИ КОНКУРЕНТОСПОСОБНОСТИ ПРИКЛАДНЫХ ПРОГРАММНЫХ ПРОДУКТОВ

Ю.П. Ехлаков,

доктор технических наук, профессор кафедры автоматизации обработки информации Томского государственного университета систем управления и радиоэлектроники (ТУСУР)

Е.А. Янченко,

студентка кафедры автоматизации обработки информации Томского государственного университета систем управления и радиоэлектроники (ТУСУР)

Адрес: г. Томск, пр. Ленина, д. 40

E-mail: upe@tusur.ru, yea@sibmail.com

В статье рассмотрены вопросы конкурентоспособности прикладного программного продукта. На основе анализа существующих подходов к оценке конкурентоспособности товаров предлагается методика оценки конкурентоспособности прикладного программного продукта, которая включает этапы выбора критериев для оценки, вычисления показателя конкурентоспособности и анализа конкурентоспособности программного продукта.

Ключевые слова: конкурентоспособность, прикладной программный продукт, интегральный показатель конкурентоспособности.

1. Введение

Любая ИТ-компания, разрабатывая программный продукт (ПП) для конкретного заказчика, намерена в дальнейшем адаптировать его под требования других потенциальных потребителей и продвигать на рынке ПП. Прежде чем осуществлять вложение финансовых средств

в развитие ПП, разработчикам необходимо сделать предварительную оценку способности созданного программного продукта конкурировать с продуктами-аналогами.

Предлагаемые в литературе методики оценки конкурентоспособности состоят из этапов: выбора и обоснования характеристик оценки конкурентоспособности, вычисления оценок конкурентоспо-

способности, анализа конкурентоспособности продукта и разработки мероприятий по ее повышению. Главное отличие методик заключается прежде всего в наборе оценочных характеристик конкурентоспособности и методах их свертывания в интегральную оценку.

Так, в [1–3] конкурентоспособность продукции предлагается оценивать на основе двух критериев: полезного эффекта от потребления продукции и цены потребления. Р.А. Фатхутдинов предлагает использовать такие показатели конкурентоспособности, как относительная доля рынка, издержки, отличительные свойства, степень освоения технологии, метод продаж, имидж (известность) [4]. В [5] отмечается, что конкурентоспособность продукции определяют следующие факторы: цена, затраты на эксплуатацию или потребление, предоставляемый сервис, реклама, имидж и авторитет фирмы, соотношение между спросом и предложением, однако в качестве итогового критерия оценки конкурентоспособности продукции предлагается определять долю рынка, занимаемую данной продукцией. В [6] в качестве характеристик оценки конкурентоспособности инновационного продукта предлагается использовать функциональность продукта, технологическую новизну применяемых технологий, затраты на внедрение.

2. Выбор и обоснование показателей оценки конкурентоспособности ПП

В общем случае конкурентоспособность продукта определяется свойствами товара на рынке, характеризующими степень удовлетворения конкретной потребности по сравнению с лучшими аналогичными объектами, представленными на данном рынке. Свойства программного продукта как рыночного товара определяются понятием ПП как самостоятельного отчуждаемого произведения, поставляемого на магнитных носителях либо переданного по сети передачи данных, предназначенного для удовлетворения конкретных потребительских предпочтений пользователей, снабженного программной документацией, сертификатами гарантии качества, рекламными материалами [7].

С учетом вышеназванных особенностей ПП все множество характеристик, отражающих уровень конкурентоспособности ПП, можно условно разбить на три группы: технические, экономические, потребительские. Каждая из характеристик конкурентоспособности ПП может описываться

в виде показателя либо в виде атрибута [3]. Характеристика «показатель» имеет определенную размерность, т.е. может быть выражена в виде некоторой количественной оценки. Характеристика «атрибут» имеет качественное содержание и может быть описана с использованием балльной либо ранговой шкал.

Для оценки *технического уровня программных продуктов* следует использовать ГОСТ 28195-89 «Оценка качества программных средств. Общие положения» [8], ГОСТ 28806-90 «Качество программных средств. Термины и определения» [9], ГОСТ Р ИСО/МЭК 9126-93 «Информационная технология. Оценка программной продукции. Характеристики качества и руководства по их применению» [10]. В этих документах рекомендуется для оценки качества ПП использовать следующие показатели и атрибуты: *функциональные возможности, надежность, практичность, эффективность, сопровождаемость, мобильность*.

В дальнейшем в соответствии со стандартом ГОСТ Р ИСО/МЭК 9126-93 каждый из них может быть декомпозирован. Так, например, характеристика «*функциональные возможности*» может быть описана такими атрибутами, как пригодность, правильность, способность к взаимодействию, согласованность, защищенность.

Экономическая значимость программного продукта может рассматриваться с различных точек зрения: инвестиционной привлекательности ПП как объекта интеллектуальной собственности, затрат на разработку ПП, эффективности практического применения ПП. Все множество показателей, раскрывающих эти аспекты конкурентоспособности ПП, можно разбить на три группы [11–13]:

1) показатели, раскрывающие *совокупную стоимость владения ПП*: приведенная стоимость ПП, стоимость годовой эксплуатации ПП, приведенный фонд оплаты труда ИТ-подразделения, приведенная стоимость общесистемного программного обеспечения, приведенная стоимость аппаратного переоснащения для обеспечения функционирования ПП;

2) показатели, описывающие *инвестиционную привлекательность проекта по разработке ПП*: приведенная стоимость проекта, внутренняя ставка доходности, срок окупаемости проекта, коэффициент доходности инвестиций в нематериальные активы;

3) показатели, отражающие *повышение эффективности деятельности организации*: сокращение

составляющих издержек по реализации бизнес-процессов, сокращение времени потерь, увеличение объемов продаж и т.д.

Оценку *субъективных потребительских предпочтений ПП* формируют непосредственные пользователи программного продукта, специалисты ИТ-служб, отвечающие за установку, адаптацию и техническую поддержку программного продукта, руководители компании [14].

Непосредственные пользователи при оценке конкурентоспособности ПП могут использовать следующие показатели и атрибуты: функциональность ПП, надежность, способность к восстановлению, наличие тестов и контрольных примеров к процедурам обработки, полноту документации, удобство и понятность интерфейса, приемлемую скорость работы ПП.

Специалисты ИТ-службы рассматривают программный продукт с точки зрения внедрения и дальнейшего сопровождения ПП, используя при этом такие показатели и атрибуты, как вид услуги по поставке ПП, возможности эксплуатации на имеющейся программно-аппаратной платформе, простоту инсталляции ПП, степень защищенности от несанкционированного доступа.

Руководители компании, исходя из бизнес-целей проекта, интересуются влиянием экономических показателей приобретаемых ПП на повышение эффективности управления компанией, а также и повышением общей культуры управления, используя для этого такие показатели и атрибуты, как доступность, точность, своевременность, актуальность, полнота, защищенность, глубина ретроспективы информации, усиление имиджевой и инвестиционной привлекательности компании в результате приобретения и внедрения ПП.

Определение конкретного набора показателей и атрибутов из каждой группы должно производиться в зависимости от специфики программного продукта, области применения и значимости продукта для удовлетворения потребности пользователя. Например, надежность наиболее важна для программного обеспечения критичных систем, эффективность — для систем реального времени, а практичность — для конечных пользователей, работающих в диалоговых режимах.

Кроме того, сам выбор показателей и атрибутов оценки конкурентоспособности носит субъективный характер и разными специалистами может осуществляться по-разному. В этом случае

необходимость и достаточность показателей и атрибутов для оценки определяется на основании двух принципов: полноты и существенности. Принцип полноты заключается в том, чтобы, с одной стороны, эксперты на первом этапе выделили как можно большее количество показателей и атрибутов, а с другой — выделенные показатели и атрибуты должны быть доступны для понимания и анализа. Это противоречие снимается при использовании принципа существенности: необходимо использовать только те показатели и атрибуты, которые существенно влияют на конкурентоспособность ПП.

В связи с отсутствием для большинства атрибутов количественных измерителей широкое применение на практике находят методы экспертных оценок. В частности, для определения предпочтений между показателями конкурентоспособности ПП наиболее часто используются методы ранжирования, парного сравнения и непосредственной оценки [14].

3. Вычисление оценок конкурентоспособности

При наличии множества программных продуктов-конкурентов и разнонаправленности влияния отдельных факторов на качество ПП оценка конкурентоспособности исследуемого продукта может производиться как по отдельным, так и обобщенным (интегральным) показателям. Существующие методики расчета интегрального показателя конкурентоспособности можно условно разбить на следующие группы [3]:

- 1) оценка конкурентоспособности на основе одного показателя, например доли рынка занимаемого ПП;
- 2) оценка конкурентоспособности на основе взвешенной суммы отдельных показателей;
- 3) оценка конкурентоспособности на основе средней геометрической отдельных показателей;
- 4) оценка конкурентоспособности на основе взвешенных сумм рейтинговых оценок отдельных показателей.

Так как на практике характеристики конкурентоспособности ПП могут выражаться как через показатели, так и атрибуты, то определение оценок конкурентоспособности следует производить в относительных единицах.

В данной статье для определения оценок конку-

рентоспособности будут использованы первые два метода. Оценка конкурентоспособности на основе одного показателя производится по каждой группе показателей отдельно. Если влияние отдельных показателей на конкурентоспособность ПП не равнозначно, то интегральная оценка должна быть вычислена в виде их взвешенной суммы. С учетом вышеизложенного процедура вычисления оценок конкурентоспособности ПП будет выглядеть следующим образом.

1. Вычисляется оценка конкурентоспособности программного продукта по сравнению с ПП-аналогами по каждому из показателей:

$$Q_i = \frac{p_i}{p_i^k}, \quad i = \overline{1, n}, \quad (1)$$

где Q_i — показатель конкурентоспособности по i -му показателю;

p_i — значение i -го показателя оцениваемого ПП;

p_i^k — значение i -го показателя k -го продукта конкурента;

n — количество параметров.

Если характеристики ПП описываются качественными параметрами, то предварительное определение количественных значений этих характеристик (показателей) производится с использованием метода экспертных оценок:

$$p_i = \sum_{j=1}^m \sum_{g=1}^d x_{igj}, \quad i = \overline{1, n}, \quad (2)$$

где x_{igj} — оценка, выставленная g -м экспертом j -й качественной характеристике i -го показателя;

n — количество параметров;

m — количество качественных характеристик;

d — количество экспертов.

2. Для каждой группы показателей вычисляется абсолютная интегральная оценка конкурентоспособности ПП разработчика по сравнению с ПП-аналогами как взвешенная суммы отдельных показателей из этой группы:

$$K = \sum_{i=1}^n d_i Q_i, \quad (3)$$

где d_i — относительная важность влияния i -го показателя на конкурентоспособность ПП;

n — количество показателей в группе.

3. Вычисляются относительные интегральные оценки конкурентоспособности программного продукта по сравнению с ПП-аналогами:

1) технический уровень / совокупная стоимость владения:

$$K_1 = \frac{K_{TV}}{K_3}, \quad (4)$$

где K_{TV} — групповой показатель технического уровня ПП относительно ПП конкурента;

K_3 — групповой показатель совокупной стоимости владения;

2) потребительские предпочтения / совокупная стоимость владения:

$$K_2 = \frac{K_{PP}}{K_3}, \quad (5)$$

где K_{PP} — групповой показатель потребительских предпочтений ПП относительно ПП конкурента.

4. Анализ конкурентоспособности программного продукта

Анализ абсолютных и относительных интегральных оценок конкурентоспособности ПП разработчика по сравнению с ПП-аналогами позволяет принять один из трех возможных вариантов решений: вывод ПП на целевые рынки; доработка ПП по улучшению показателей, которые уступают показателям ПП-аналога; нецелесообразность вывода ПП на целевые рынки (табл. 1).

Таблица 1.

Анализ конкурентоспособности ПП

Показатель конкурентоспособности	Описание ситуации
K существенно меньше единицы	Степень конкурентоспособности ПП на низком уровне. При выводе на рынок необходимо доработать ПП по показателям, значительно отличающимся от конкурентных продуктов. В случае если значение K меньше критического (задаваемого разработчиком), вывод ПП в выбранном сегменте рынка считается нецелесообразным.
K приблизительно равен единице	Продукты примерно одинаковы по конкурентоспособности. Необходим анализ отдельных показателей, позволяющих повысить конкурентоспособность.
K существенно больше единицы	ПП обладает высокой конкурентоспособностью и превосходит имеющиеся на рынке продукты-конкуренты. Необходима разработка плана продвижения ПП на выбранные целевые рынки.

Кроме того, полученные единичные показатели конкурентоспособности могут быть в дальнейшем положены в основу SWOT-анализа программно-

го продукта, который позволяет разработать план мероприятий по доработке ПП по показателям, существенно уступающим показателям ПП-аналога, а также усилить конкурентные преимущества программного продукта по показателям, превосходящим ПП-аналог.

5. Оценка конкурентоспособности ПП «Контроль организационно-распорядительной деятельности»

Предложенная методика апробирована на примере оценки конкурентоспособности программного продукта «Контроль организационно-распорядительной деятельности» (КОРД), который представляет собой систему электронного документооборота (СЭД) [15].

В качестве целевого сегмента рынка выбраны предприятия среднего и малого бизнеса Сибирского Федерального округа, а в качестве эталона сравнения — программная система «ДЕЛО-Предприятие», занимающая наибольшую долю рынка в выбранном сегменте [16].

Для оценки конкурентоспособности использованы критерии, затрагивающие основные требования всех заинтересованных групп потребителей. Для оценки технического уровня ПП выбран показатель «функциональные возможности». Эту группу в соответствии с ГОСТ Р ИСО 15489-1-2007 составили атрибуты, характеризующие полный жизненный цикл документа: регистрация документов, ведение иерархической структуры документов, ведение журналов работы с документами, ведение регистрационной карточки, ведение номенклатуры дел, ведение истории работы с документами и версиями, работа с поручениями и контроль, поиск документов по различным параметрам, ведение архива электронных документов [17].

Оценка экономической значимости ПП проведена по следующим количественным показателям: стоимость пакета из пяти лицензий, стоимость обновления пакета, стоимость удаленного обслуживания.

Для описания субъективных пользовательских предпочтений использованы следующие атрибуты:

- ◆ возможность настройки ПП (с коэффициентом относительной важности атрибута - 0,3): дизайнер карточек документов, дизайнер маршрутов документов, создание и редактирование справочников и словарей;

- ◆ степень защищенности от несанкционированного доступа (с коэффициентом относительной важности атрибута - 0,25): шифрование документа, поддержка электронной цифровой подписи (ЭЦП);

- ◆ возможность работы с территориально — распределенными офисами (с коэффициентом относительной важности атрибута - 0,25): интеграция с электронной почтой, веб-доступ, другие средства удаленной работы распределенных офисов;

- ◆ возможность обработки документов (с коэффициентом относительной важности атрибута - 0,2): сканирование документов; совместимость с ПО, обеспечивающим распознавание образов документов; генерация отчетов.

Оценка атрибутов технического уровня и степени пользовательских предпочтений произведена по шкале интервалов, для оценки конкурентоспособности привлекались специалисты-разработчики программного продукта «КОРД». Коэффициенты относительной важности атрибутов также определены методом экспертных оценок.

Групповой относительный показатель конкурентоспособности «КОРД» по сравнению с ПП «ДЕЛО-Предприятие» по техническим характеристикам был равен 1,22, групповой показатель характеристик пользовательских предпочтений — 0,63, групповой показатель по экономическим характеристикам — 1,19. В результате значение интегрального показателя конкурентоспособности «КОРД» составило 1,01.

Расчеты показали, что в целом ПП «КОРД» обладает достаточно приемлемым уровнем конкурентоспособности относительно лидера рынка. По техническим и экономическим показателям он даже превосходит ПП «ДЕЛО-Предприятие», а по характеристикам пользовательских предпочтений несколько уступает ему.

Анализ полученных единичных показателей конкурентоспособности по разделу «Характеристики пользовательских предпочтений» показал, что для успешного продвижения ПП «КОРД» на рынок необходима его доработка по следующим характеристикам:

- ◆ возможности настройки ПП (дизайнер карточек документов, дизайнер маршрутов);

- ◆ возможности обработки документов (совместимость с ПО, обеспечивающим распознавание образов документов);

- ◆ защищенности от несанкционированного доступа (шифрование документа, поддержка ЭЦП). ■

Литература

1. Окрепилов В.В. Управление качеством. – М.: Экономика, 1998. – 200 с.
2. Гличев А.В. Основы управления качеством продукции. – М.: Изд-во стандартов, 1998. – 124 с.
3. Голубков Е.П. Предплановые маркетинговые исследования и стратегический анализ // Маркетинг в России и за рубежом. – 2008. – № 2. – С. 106-125.
4. Фатхутдинов Р.А. Стратегический менеджмент. – М.: ЗАО «Бизнес-школа «Интел-Синтез», 1999. – 416 с.
5. Фомин В.Н. Квалиметрия. Управление качеством. Сертификация. – М.: ЭКМОС, 2000. – 320 с.
6. Буймов А.С., Вайсман Е.Д. Оценка конкурентоспособности инновационного продукта // Маркетинг в России и за рубежом. – 2010. – № 1. – С. 74-80.
7. Ехлаков Ю.П., Ефимов А.А. Функциональные модели и организационно-правовые механизмы продвижения прикладных программных продуктов на рынок корпоративных продаж. – Томск: Изд-во Томск. гос. ун-та систем управления и радиоэлектроники, 2010. – 172 с.
8. ГОСТ 28195-89. Оценка качества программных средств. Общие положения [Электронный ресурс] // Complexdoc.ru: [сайт]. – URL: <http://www.complexdoc.ru/text ГОСТ 28195-89> (дата обращения: 02.04.2011).
9. ГОСТ 28806-90. Качество программных средств. Термины и определения [Электронный ресурс] // Complexdoc.ru: [сайт]. – URL: <http://www.complexdoc.ru/text ГОСТ 28806-90> (дата обращения: 02.04.2011).
10. ГОСТ Р ИСО/МЭК 9126-93. Информационная технология. Оценка программной продукции. Характеристики качества и руководства по их применению [Электронный ресурс] // Complexdoc.ru: [сайт]. – URL: <http://www.complexdoc.ru/text ГОСТ Р ИСО/МЭК 9126-93> (дата обращения: 02.04.2011).
11. Елашкин М. Как оценивать эффективность ИТ? // Открытые системы. – 2004. – № 7. – С. 38-42.
12. Методические рекомендации по определению рыночной стоимости интеллектуальной собственности [Электронный ресурс] // Обзоры книг, находящихся в свободном доступе: [сайт]. [2008]. – URL: http://book-on-line.blogspot.com/2008/02/blog-post_12.html (дата обращения: 04.04.2011).
13. Ехлаков Ю.П. Вывод прикладного программного обеспечения на рынок корпоративных продаж: взгляд разработчика // Маркетинг в России и за рубежом. – 2009. – № 4 (72). – С. 45-50.
14. Евланов Л.Г. Теория и практика принятия решений. – М.: Экономика, 1984. – 176 с.
15. Ехлаков Ю.П., Кириенко В.Е., Сенченко П.В. Методы и технологии документационного обеспечения управленческих решений. – Томск: Томск. гос. ун-т систем управления и радиоэлектроники, 2005. – 180 с.
16. Горячая десятка российских систем электронного документооборота [Электронный ресурс] // Независимый портал о СЭД: [сайт]. – URL: http://www.doc-online.ru/a_id/319/ (дата обращения 24.03.2011)
17. ГОСТ Р ИСО 15489-1-2007. Система стандартов по информатизации, библиотечному и издательскому делу. Управление документами. Общие требования. – М.: Госстандарт России, 2007.

МЕТОД СЕМАНТИЧЕСКОГО ПРЕОБРАЗОВАНИЯ ОБОБЩЕННЫХ ЗАПРОСОВ НА ОСНОВЕ БАЗЫ ЦЕЛЕЙ

А.В. Кириллов,

*аспирант кафедры инноваций и бизнеса в сфере информационных технологий
Национального исследовательского университета «Высшая школа экономики»*

Адрес: г. Москва, ул. Кирпичная, д. 33/5

E-mail: antonv.kirillov@gmail.com

Статья описывает новый метод построения семантических расширений поисковых запросов обобщенного характера для улучшения результатов веб-поиска. Метод базируется на теории К-представлений – теории проектирования семантико-синтаксических анализаторов естественно-языковых текстов с широким использованием формальных средств для представления входных, промежуточных и выходных данных. Изложенный метод реализован на языке программирования Java в экспериментальной поисковой системе AOS (Aspect Oriented Search), эта система в настоящее время тестируется.

Ключевые слова: семантический поиск, семантическое преобразование обобщенного запроса, концептуальный базис, СК-язык, концептуальная база целей, шаблон семантической трансформации.

Введение

Сегодня параллельно с ростом объемов информации в Интернете происходит разработка новых и совершенствование существующих подходов к ее поиску [1]. Все большую актуальность приобретают средства семантического поиска, под которыми понимаются системы, принимающие на вход некоторый запрос, обрабатывающие его с использованием рассуждений над специфичной базой знаний и возвращающие совместимые результаты [2-8]. Входным запросом может являться, например, вопрос на естественном языке (ЕЯ), представление вопроса при помощи триплетов, графическое представление, набор

ключевых слов, отдельные фразы и т.д. В роли базы знаний могут выступать онтологии, аннотированные массивы текста, текстовые документы, веб, XML документы, RDF документы, HTML документы и т.д.

В данной работе предлагается решение для семантически-ориентированного поиска, использующее возможности традиционных поисковых систем, дополняемые семантическим анализом и расширением подаваемых на вход естественно-языковых запросов. С учетом вычислительных мощностей наиболее крупных существующих систем поиска по ключевым словам, предлагается сместить фокус с детального семантического анализа и ин-

дексации содержимого электронных документов на анализ подаваемых на вход системе запросов и последующую генерацию множества семантически расширенных (адаптированных) запросов, которые затем будут передаваться синтаксической поисковой системе. Результаты поиска по каждому из запросов будут анализироваться и сравниваться между собой с целью их фильтрации и ранжирования по степени семантической релевантности.

В работах автора и В.А. Фомичева [8-9] проводился анализ форм ЕЯ-запросов, и были выделены три наиболее интересных, с научной точки зрения, типа вопросов: вопросы достижения целей, аспектно-ориентированные вопросы и вопросы изменения множеств. Рассмотрим данные вопросы на примере информации об успехах компаний, продуктах, выпускаемых ими, и изменениях состава тех или иных организационных единиц компаний, представленных множествами (например, профсоюз или совет директоров). Последовательность поисковых запросов изображена на *рисунке 1*.

Пример. С учетом развития голосовых интерфейсов и средств синтеза и анализа речи пример поиска информации об успешности деятельности некоторой компании X и достигнутых результатах может быть представлен в виде следующего диалога:

Пользователь: Какие успехи были у компании X в этом году?

Система: Компания X выпустила продукт Y, увеличила объем выручки на 7%, открыла новое представительство в Москве.

Пользователь: Каковы особенности продукта Y?

Система: Высокая стабильность, отказоустойчивость, низкая цена и т. д.

Пользователь: Чем продукт Y отличается от продукта Z?

Система: Y более производителен чем Z, Y превосходит Z по следующим показателям ...

Пользователь: Какие изменения были в составе совета директоров в прошедшем году?

Система: В совет директоров компании вошел Петр Иванов.

Таким образом, если пользователь хочет найти информацию о компании, изменениях в составе ее руководства, успехах или неудачах, различных характеристиках продуктов, то полный цикл поиска покрывается предлагаемыми типами вопросов и соответствующими методами преобразования, тем самым увеличивая скорость, удобство и релевантность поиска.



Рис. 1. Схема сценария бизнес-поиска.

Далее в статье будут подробно рассмотрены и структурированы вопросы достижения целей.

Вопросы достижения целей

Успешность функционирования (либо существования) объекта или системы определяется достижением рассматриваемой сущностью целей, которые перед ней поставлены.

Вопросами достижения целей будем называть предложения, в которых запрашивается информация, касающаяся результатов функционирования того или иного объекта или системы. Другими словами, это вопросы, касающиеся успехов либо неудач рассматриваемых объектов или систем. В [9] такие вопросы названы вопросами о достижениях и неудачах.

Примеры вопросов достижения целей: «Какие успехи были у сборной России по футболу в 2009 году?», «Какие неудачи испытала компания Sun в 2010 году?», «Каковы успехи компании Intel за 2010 год?». Из представленных примеров видно, что если подать их на вход поисковой системе в таком виде, то результаты поиска будут низкого качества и не будут содержать в себе ответов на поставленный вопрос.

Для дальнейшего анализа была выбрана область предпринимательской деятельности: анализироваться будут успехи и неудачи организаций (компаний). Под целью компании будем понимать конечный желаемый результат, который определяется в процессе планирования и регулируется функциями управления. Рассмотрим фрагмент дерева целей отдельно взятой компании (*рисунок 2*):

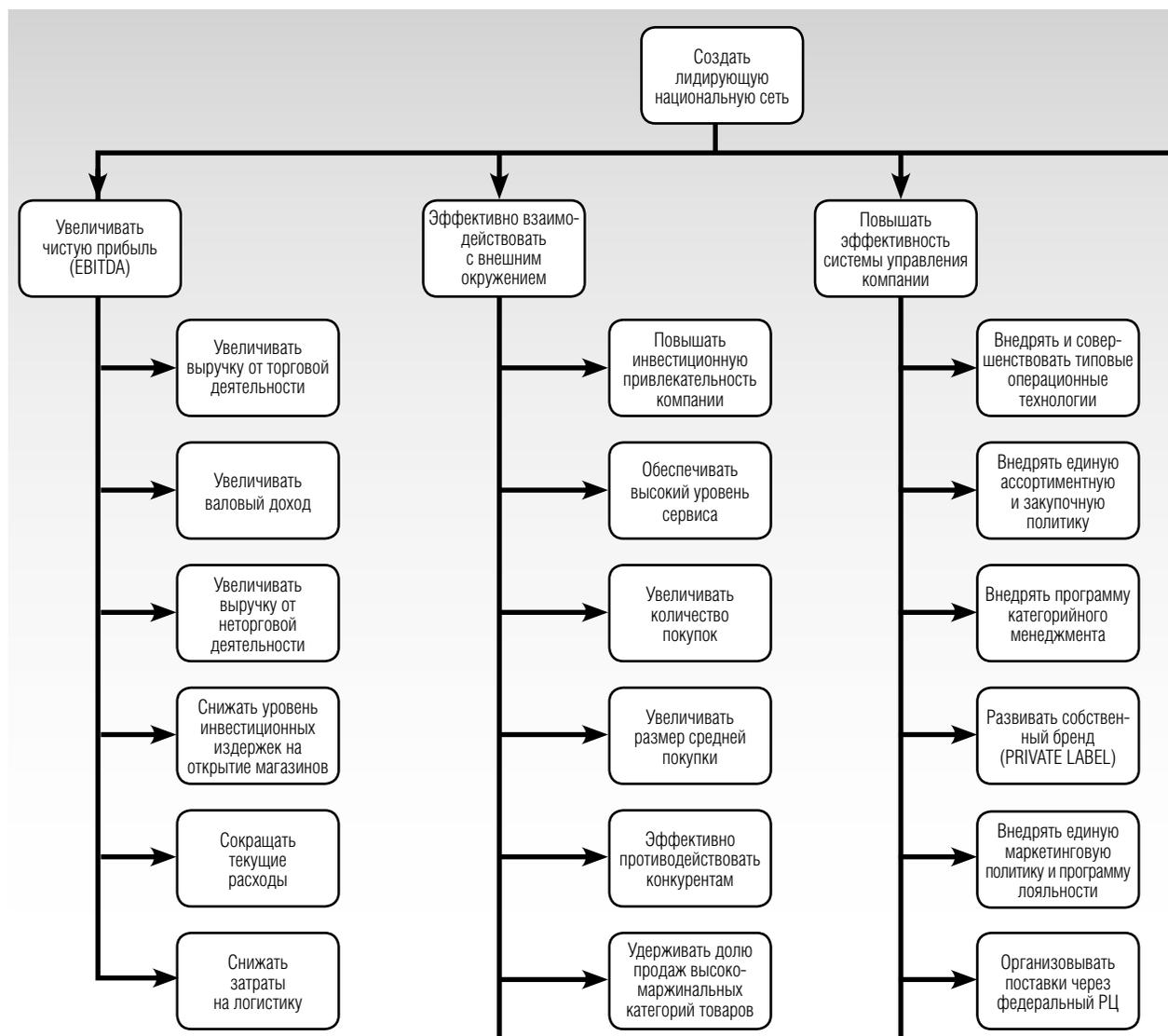


Рис. 2. Фрагмент дерева целей реально существующей компании.

Анализ данного дерева целей, а также деревьев целей других компаний показал, что успешность компании определяется фактами достижения компанией поставленных целей. Дальнейшая детализация целей и формирование фраз-индикаторов, определяющих нахождение в тексте указания на достижение цели, продемонстрировали, что, действительно, указания на детализированные цели встречаются в электронных документах гораздо чаще, чем описание непосредственно успехов или неудач каждой конкретной компании. Примеры текстов, извлеченных поисковой системой и указывающих на достижение некоторых из рассматриваемых показателей, представлены в табл. 1.

Таким образом, чтобы обнаружить документы, содержащие в себе информацию об успехах либо

неудачах компании, необходимо декомпозировать первоначальный ЕЯ-запрос на множество словосочетаний – индикаторов: ЕЯ-выражений, наличие которых в тексте документов позволяет определить соответствие этого документа первоначальному запросу. Анализ целей нескольких компаний позволил обобщить эти цели и сформировать факты, указывающие на достижение компаниями этих целей (табл. 2). В данной таблице обозначение *pos* соответствует тому факту, что данное событие является индикатором успешности деятельности компании, *neg* соответствует неудаче. Пара (+, *pos*) обозначает, что рост данного показателя является позитивным фактором, (-, *pos*) обозначает, что снижение данного показателя является позитивным фактором.

Таблица 1.

Примеры текстов,
извлеченных поисковой системой

Цели	Примеры текстов
Увеличивать выручку	Выручка индийской компании в отчетный период увеличилась в 3,2 раза - до 31,5 млн долл. Выручка РБК увеличилась на 46% благодаря уверенному росту основного бизнеса компании .
Увеличивать доход	Компания «Визард» после внедрения МАНХ увеличила доход более чем в два раза по сравнению с докризисными показателями. Доход компании Nусomed в I квартале увеличился на 30 % .
Увеличивать прибыль	Чистая прибыль АФК «Система» по US GAAP за 9 месяцев 2010г. составила 471 млн 452 тыс. долл. против 2 млрд 108 млн 46 тыс. долл. годом ранее. Чистая прибыль компании ЛУКОЙЛ за первое полугодие 2006 года увеличилась на 55 процентов.
Снижать затраты	Совет директоров ЛУКОЙЛа принял решение, что капитальные затраты компании в 2009 году будут снижены почти вдвое.
Сокращать расходы	Компания Intel планирует в текущем году сократить свои расходы на несколько сотен миллионов долларов.
Применять новейшие технологии и ресурсы	Компания Softline внедрила технологии управления лицензиями на продукты Adobe Systems в ЗАО «Международный центр финансово-экономического развития». Компания Lenovo внедрила в линейку бизнес ноутбуков ThinkPad технологии NVIDIA Optimus.
Эффективно противодействовать конкурентам	Компания Google поглотила фирму VmрrTop, занимающуюся разработкой инновационного трехмерного интерфейса для управления рабочим столом с активным использованием мультитач навигации.
Контролировать соблюдение стандартов	Производство компании «Союз-Виктан» соответствует стандартам ISO (ИСО) 14000.
Выходить на новые рынки	Компания WELHOME вышла на рынок инвестиционных услуг с новыми предложением. Компания ZTE вышла на рынок планшетных компьютеров.

Таблица 2.

Сгруппированные факторы достижения целей компаний

КОМПАНИЯ								
Внешние факторы	Внутренние факторы	Изменения						
поглощение компании (pos) поглощение компанией (neg) лидер рынка\сферы\отрасли\ сегмента\направления (pos) выпуск продукта (pos) открытие нового офиса\представительства (pos) заключение выгодной\успешной сделки\контракта (pos) создание нового предложения (pos) выход на новый рынок (pos) завоевание рынка (pos) медаль\приз (pos) победа\выигрыш в конкурсе (pos) работодатель года (pos) компания года (pos) заключить контракт (pos) расторгнуть контракт (neg)	<table border="1"> <tr> <td>автоматизация оптимизация улучшение модернизация</td> <td>производство процесс оборудование технология</td> </tr> <tr> <td>соответствие</td> <td>требования стандарты</td> </tr> <tr> <td>внедрение</td> <td>система методика методология технология</td> </tr> </table>	автоматизация оптимизация улучшение модернизация	производство процесс оборудование технология	соответствие	требования стандарты	внедрение	система методика методология технология	увеличение уменьшение повышение снижение расширение
автоматизация оптимизация улучшение модернизация	производство процесс оборудование технология							
соответствие	требования стандарты							
внедрение	система методика методология технология							
		объем продаж (+, pos) рейтинг (+, pos) привлекательность (+, pos) выручка (+, pos) доход (+, pos) прибыль (+, pos) уровень сервиса (+, pos) стоимость бренда (+, pos) реализация продкции (+, pos) число офисов (+, pos) расходы (-, pos) затраты (-, pos) себестоимость продукции (-, pos) количество партнеров (+, pos) число поставщиков (+, pos) ассортимент (+, pos)						

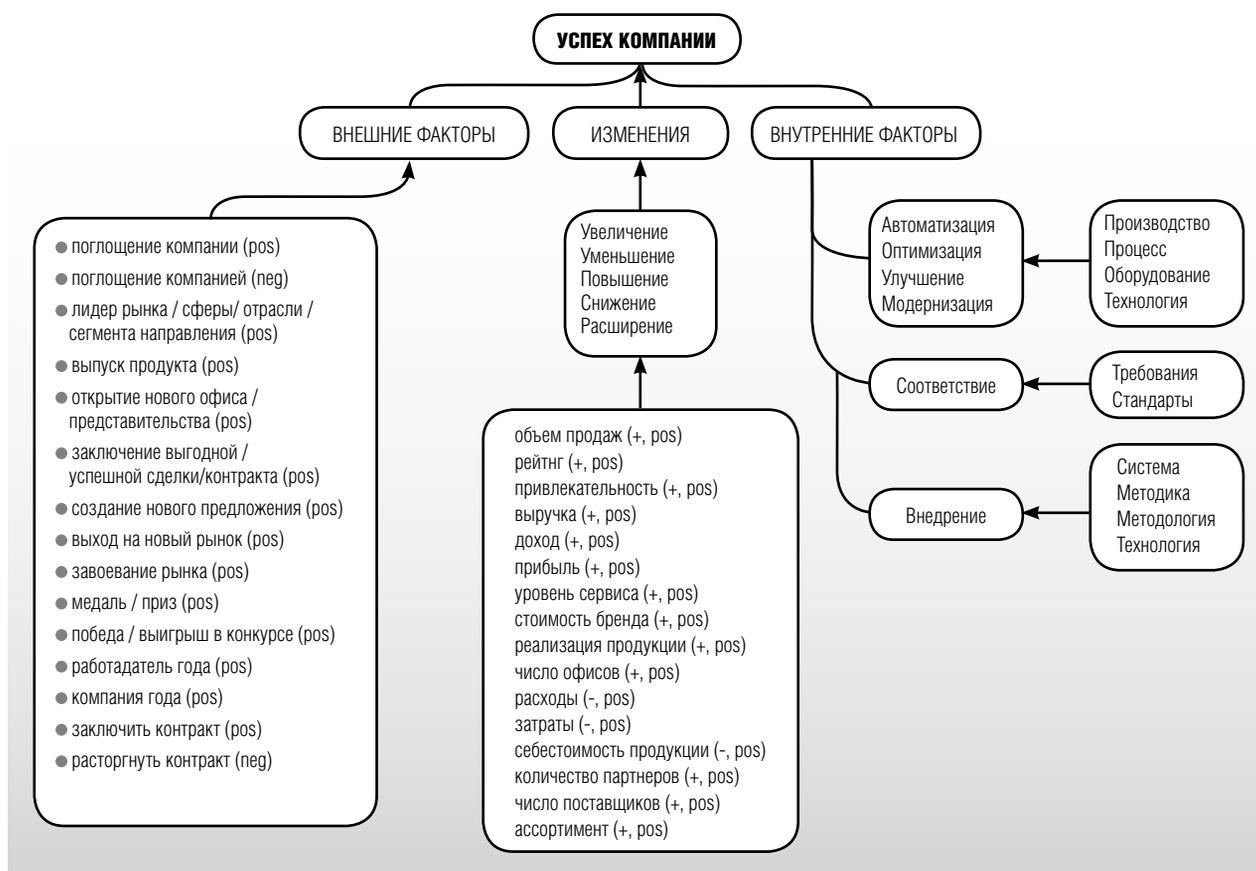


Рис. 3. Графическое представление дерева факторов достижения целей.

Формальная модель базы знаний для представления целей

Представим таблицу 2 в виде графической схемы (рис. 3) для большей наглядности:

На представленной схеме видно, что факторы достижения успеха имеют общую форму.

В монографиях В.А. Фомичева [10-12] определен класс формальных объектов, названных концептуальными базисами (к.б.) и задающих базовые сведения о системе первичных концептуальных единиц, используемой прикладной интеллектуальной системой.

Определение 1.

Упорядоченная тройка вида $(B, цел, рац)$ называется концептуальным базисом с числовой разметкой \Leftrightarrow когда B – произвольный концептуальный базис, $цел$ и $рац$ – два различных сорта из $Sf(B)$, и выполняются следующие условия:

1) первичный информационный универсум $X(B)$ включает подмножества $Natural, Pos-rational, Z1, Z2, Numbers$, где $Natural$ – множество всех цепочек

вида d_1, \dots, d_n , где $n \geq 1$, и для $k=1, \dots, n$ d_k – цифра из множества $\{‘0’, ‘1’, \dots, ‘9’\}$, причем из $d_1 = 0$ следует, что $n = 1$; $Pos-rational$ – множество всех цепочек вида b, c , где $b, c \in Natural$; $Z1$ – объединение множества $Natural$ и множества всех цепочек вида $-nb$, где nb – цепочка из множества $Natural$; $Z2$ – объединение множества $Pos-rational$ и множества всех цепочек вида $-numb$, где $numb$ – цепочка из множества $Pos-rational$; $Numbers$ – объединение множеств $Z1$ и $Z2$;

2) для каждого элемента d множества $Z1$ типом $tp(d)$ является сорт $цел$, и для каждого элемента h множества $Z2$ типом $tp(h)$ является сорт $рац$;

3) сорт $цел$ является конкретизацией сорта $рац$ для отношения общности Gen .

Пример.

Множество $Natural$ включает цепочки 123 и 4125; множество $Numbers$ включает, в частности, цепочки 12,78, -0,315 и -542.

Определение 2.

Концептуальной базой целей (к.б.ц.) называется набор $Gbase$ вида

$(B, \text{цел, рац, событие, Goals}), \quad (1)$

где тройка $(B, \text{цел, рац})$ является концептуальным базисом с числовой разметкой, *событие* является выделенным сортом из множества $St(B)$, и выполняются следующие условия:

1) множество $Acts(B) = \{y \text{ из } X(B) \mid tp(y) = \uparrow \text{событие}\}$ не пусто и конечно;

2) первичный информационный универсум $X(B)$ включает такой элемент *#Объект-интереса*, что тип $tp(\text{\#Объект-интереса})$ является конкретизацией базового типа [*объект*];

3) универсум $X(B)$ включает элементы (обозначения тематических ролей) *Агент* и *Роль2*, причем $tp(\text{Агент}) = \{(\text{событие}, s)\}$, $tp(\text{Роль2}) = \{(\text{событие}, t)\}$, где s и t – элементы множества сортов $St(B)$, причем ни один из этих сортов не является конкретизацией сорта P (сорт «смысл сообщения») или сорта *событие*;

4) *Goals* является некоторым конечным множеством выражений СК-языка $Ls(B)$ вида

$\langle \text{event-concept} * (\text{Агент}, d1)(\text{Роль2}, d2), \text{estimation} \rangle,$

где *event-concept* принадлежит множеству $Acts(B)$, $d1$ и $d2$ – различные элементы универсума $X(B)$, причем какой-либо из этих элементов является символом *#Объект-интереса*, *estimation* – элемент множества *Numbers*, обозначающий рациональное число от -1 до 1, отличное от 0.

Пример.

Множество *Goals* может включать цепочку $\langle \text{поглощение-орг} * (\text{Агент}, \text{\#Объект-интереса})(\text{Роль2}, \text{нек компания}), 1 \rangle.$

Данная модель позволит хранить записи о фактах успехов либо неудач какой-либо компании. Хранение информации в базе знаний именно в такой форме позволяет с легкостью генерировать выражения, которые послужат в качестве новых поисковых запросов, а также расширять спектр данных, хранящихся в базе. База целей может постоянно расширяться, позволяя дополнять ее новыми фактами, относящимися к успехам либо неудачам компаний.

Однако для семантического расширения поисковых запросов этих данных недостаточно. Необходимо иметь возможность преобразовывать данные факты в словосочетания, которые наиболее вероятно встречаются в электронных документах. Следует учитывать, что преобразование должно учитывать морфологические свойства слов и синтаксис фраз,

чтобы наиболее гибко трансформировать факты в ЕЯ-выражения. Для этих целей введем понятие шаблона семантической трансформации.

Определение 3.

Пусть *Gbase* – концептуальная база целей (к.б.ц.) вида (1), *Morph-values* – конечное множество символов, интерпретируемых как значения различных морфологических признаков (*существит, глагол, прош-время, наст-время, пассив-зalog* и т.д.). Тогда шаблоном семантической трансформации, порожденным к.б.ц. *Gbase* и множеством *Morph-values*, называется произвольный упорядоченный набор вида

$(\text{sem-pattern}, X, Y, Z, \text{prop-chain}), \quad (2)$

где *sem-pattern* – элемент множества *Goals*, (X, Y, Z) – произвольная перестановка без повторений из символов *#A#*, *#Pred#*, *#B#*, и *prop-chain* – цепочка вида $v[1] * v[2] * \dots * v[k]$, где $1 \leq k$, $v[1], \dots, v[k]$ – элементы множества *Morph-values*.

Пример.

Пусть *sem-pattern* – цепочка $\langle \text{поглощение-орг} * (\text{Агент}, \text{\#Объект-интереса})(\text{Роль2}, \text{нек компания}), 1 \rangle.$ Тогда набор $(\text{sem-pattern}, \text{\#A#}, \text{\#Pred#}, \text{\#B#}, \text{глаг} * \text{прош-время})$ является одним из возможных шаблонов семантической трансформации.

Методика преобразования вопросов достижения целей к расширенному виду

Преобразование происходит при помощи правил трансформации, уникальных для каждого из фактов. В правиле трансформации указываются порядок слов в результирующем словосочетании, а также их формы. На основании хранимых в базе фактов и соответствующих правил трансформации генерируются словосочетания, которые позволят традиционной системе поиска по ключевым словам найти документы, содержащие в себе упоминания данных фактов. После того, как множество словосочетаний-индикаторов сформировано, оно передается на вход синтаксической поисковой системе, которая осуществит отдельный поиск по каждому из них. Множество результирующих документов, возвращенных поисковой системой будет проанализировано с точки зрения количества вхождений различных словосочетаний-индикаторов, т.е. указателей на присутствие упоминания того или иного факта в документе. Релевантность документов будет опре-

деляться во-первых, по количеству вхождений различных фактов, а во-вторых, по рейтингу документа, определяемому по алгоритму *PageRank*. Отсортированные по релевантности первоначальному поисковому запросу документы затем будут переданы пользователю.

Метод порождения результирующих индикаторов.

1. По входному запросу находится информационная единица *Studied-object*, обозначающая объект интереса запроса. Эта информационная единица интерпретируется как значение символа *#Объект-интереса*.

2. По шаблону вида (2) строится цепочка *X Y Z*.

3. В цепочке *X Y Z* символ *#A#* заменяется на произвольную лексическую единицу, которой соответствует информационная единица *Studied-object*.

4. Пусть *sem-pattern* – цепочка вида *<event-concept*(Агент, d1)(Роль2, d2), estimation>*. Тогда в цепочке, полученной на Шаге 3, символ *#Pred#* заменяется на произвольную лексическую единицу *pred-word-form*, которой соответствует информационная единица *event-concept*, причем словоформа *pred-word-form* должна обладать значениями морфологических признаков, задаваемыми цепочкой *prop-chain*.

5. В цепочке, полученной на Шаге 4, символ *#B#* заменяется на произвольную лексическую единицу, соответствующую той из семантических единиц *d1* и *d2*, которая отлична от символа *#Объект-интереса* в исходном шаблоне.

Пример 1.

Пусть *Studied-object = firm-Oracle*, *sem-pattern* – цепочка *<поглощение-орг * (Агент, #Объект-интереса)(Роль2, нек фирма1), 1>*. Тогда по шаблону семантической трансформации вида (*sem-pattern, #A#, #Pred#, #B#, глаг * ПрошВр*) в соответствии с данным алгоритмом может быть построен результирующий индикатор *Oracle поглотил компания*.

Пример 2.

Пусть начальный запрос *w = «Каковы успехи компании Oracle?»*. После определения типа запроса и объекта интереса [8] как пары (успех, Oracle) можно переходить к построению расширенного множества запросов. Примеры записанных фактов, шаблонов трансформации и результирующих словосочетаний-индикаторов представлены в таблице 4.

Запись факта	Правило трансформации	Результирующий индикатор
<поглощение-орг*(Агент, #Объект-интереса)(Роль2, компания), 1>	#A+#Pred + #B + глаг * ПрошВр	Oracle поглотил компания
<выпуск*(Агент, #Объект-интереса)(Роль2, нек продукт1), 1>	#A+#Pred + #B + глаг * ПрошВр	Oracle выпустил продукт
	#B+#Pred + #A + глаг * ПрошВр *ЗалогПасс	Продукт выпущен Oracle
	#A+#Pred + #B + глаг * ПрошВр * ЗалогПасс	Oracle выпущен новый продукт
<открытие*(Агент, #Объект-интереса)(Роль2, нек офис), 1>	#A+#Pred + #B + глаг * ПрошВр	Oracle открыл новый офис
<увеличение-знач-парам*(Агент, #Объект-интереса)(Роль2, объем-продаж), 1>	#A+#Pred + #B + глаг * ПрошВр	Oracle увеличил объем продаж
		Oracle повысил объем продаж
	#B+#A + #Pred + глаг * ПрошВр	Объем продаж Oracle увеличился
		Объем продаж Oracle возрос
		Объем продаж Oracle вырос
<заклучение1 * (Агент, #Объект-интереса) (Роль2, нек сделка), 1>	#A+#Pred + #B + глаг * ПрошВр	Oracle заключил сделку
	#B+#Pred + #A + глаг * ПрошВр *ЗалогПасс	Сделка заключена Oracle
	#B + #A	Успешная сделка Oracle

Заклучение

Часто для вопросов на естественном языке отсутствуют методы их анализа для повышения качества поиска, особенно это характерно для вопросов определенных типов. Рассмотренные вопросы достижения целей представляются интересными для исследования и анализа. Для адекватного поиска ответов на такие вопросы был предложен метод, основанный на декомпозиции целей организаций и генерации множества детализирующих запросов, содержащих индикаторы, позволяющие определить соответствие текста документов первоначальному запросу. В результате работы семантического анализатора часть документов будет отсеяна, будет учтено

дублирование документов и количество вхождений индикаторов в документ. На выходе пользователь получит набор документов, упорядоченный по количеству фактов успехов либо неудач той или иной компании. Факторам могут быть присвоены веса, которые будут учитываться при ранжировании результирующего множества документов.

Предложенные модели базы целей и шаблонов семантических трансформаций могут легко быть дополнены и расширены, а использование выражений СК-языков для представления знаний позволяет очень гибко манипулировать описаниями информации, хранимой в базе целей.

Использование данного подхода значительно расширяет возможности поиска, потому что дополняет его возможностями рассуждений. Ведь методика, основанная на детализации целей и генерации словосочетаний-индикаторов, выходит за рамки построения обычных отношений вида is-a, применяемых в системах поиска, основанных на семантических сетях, либо использующих как базу знаний какую-либо таксономию. Изложенный метод реализован на языке программирования Java в экспериментальной поисковой системе AOS (Aspect Oriented Search), эта система в настоящее время тестируется. ■

Литература

1. Кириллов А.В. Поисковые системы: компоненты, логика и методы ранжирования // Бизнес-информатика. 2011. № 4(10). – С. 51-59.
2. Halpin H., Lavrenko V. Relevance Feedback Between Hypertext and Semantic Search // Proc. Conference WWW2009 (April 20-24, 2009, Madrid, Spain).
3. Lei Y., Uren V., Motta E.. Semsearch: A search engine for the semantic web // Proc. 5th International Conference on Knowledge Engineering and Knowledge Management Managing Knowledge in a World of Networks, Lect. Notes in Comp. Sci. – Springer, Podybrady, Czech Republic, 2006. – pp. 238-245.
4. Bhagdev R., Chapman S., Ciravegna F., Lanfranchi V., Petrelli D. Hybrid search: Effectively combining keywords and semantic searches // The Semantic Web: Research and Applications. – Springer, Berlin / Heidelberg, 2008. – pp. 554-568
5. d'Aquin M., Baldassarre C., Gridinoc L., Angeletou S., Sabou M., Motta E. Characterizing knowledge on the semantic web with watson // EON, 2007. – pp. 1-10.
6. Kaufmann E., Bernstein A., Zumstein R. Querix: A Natural Language Interface to Query Ontologies Based on Clarification Dialogs // 5th International Semantic Web Conference (ISWC 2006). – Springer, 2006. – pp. 980-981.
7. Bernstein A., Kaufmann E., Gohring A., Kiefer C. Querying Ontologies: A Controlled English Interface for End-users // 4th International Semantic Web Conference (ISWC 2005), November 2005. – pp. 112-126.
8. Кириллов А.В., Фомичев В.А. О новом подходе к семантическому преобразованию естественно-языковых запросов поисковых систем // Бизнес-информатика. № 1 (15). С. 19-26.
9. Fomichov V.A., Kirillov A.V. Semantic Transformation of Search Requests for Improving the Results of Web Search // Pre-Conference Proceedings of the Focus Symposium on Knowledge Management Systems (August 2, 2011, Focus Symposia Chair: Jens Pohl) in conjunction with InterSymp-2011, 23rd International Conference on Systems Research, Informatics and Cybernetics, August 1 – 5, 2011, Baden-Baden, Germany), Collaborative Agent Design Research Center, California Polytechnic State University. – San Luis Obispo, CA, USA, 2011. – pp. 37-43.
10. Фомичев В.А. Формализация проектирования лингвистических процессоров. – М.: МАКС Пресс, 2005. – 368 с.
11. Фомичев В.А. Математические основы представления содержания посланий компьютерных интеллектуальных агентов. – М.: Изд. дом ГУ-ВШЭ, изд. «ТЕИС», 2007. – 176 с.
12. Fomichov V.A. Semantics-Oriented Natural Language Processing: Mathematical Models and Algorithms / Series: IFSR International Series on Systems Science and Engineering, Vol. 27. – Springer: New York, Dordrecht, Heidelberg, London, 2010. – 354 p.

ПОДХОД К АВТОМАТИЧЕСКОМУ ПОИСКУ ПЕРЕВОДНЫХ СЛОВСОЧЕТАНИЙ НА ОСНОВЕ СИНТАКСИЧЕСКОЙ ИНФОРМАЦИИ И МНОГОУРОВНЕВОЙ ФИЛЬТРАЦИИ

В.И. Новицкий,

сотрудник компании АВВУУ, аспирант кафедры распознавания изображений и обработки текста Московского физико-технического института (МФТИ)

Адрес: г. Москва, ул. Отрадная, д. 2Б, стр. 6

E-mail: nov.valerij@gmail.com

В работе описывается подход к автоматическому построению списка словосочетаний по корпусу выровненных параллельных текстов (текстов и их переводов, сопоставленных по предложениям). Подход основан на сопоставлении деревьев синтаксического разбора предложений. Ключевой особенностью данной работы является набор фильтров для удаления нерелевантных словосочетаний.

Ключевые слова: обработка естественного языка, переводные словосочетания.

1. Введение

Данная работа посвящена разработке алгоритма поиска словосочетаний вместе с их переводами на другой язык по корпусу параллельных текстов. Нас будет интересовать возможность получения нетривиальных переводов (т.н. переводных словосочетаний).

Такие словосочетания представляют собой ценный лингвистический ресурс. Например, они могут использоваться для машинного перевода текста, в качестве статистических данных в других задачах или как справочный материал для лингвистов.

Лингвистические технологии по автоматическому анализу и синтезу текстов, в свою очередь, являются полезным инструментом для автоматизации и оптимизации различных бизнес-процессов. Это может быть семантический поиск, автореферирование, машинный перевод и другие задачи, связанные с обработкой больших текстовых баз.

Особенностью данной работы является предложенный набор эвристических фильтров для выделения семантически значимых словосочетаний среди всех встретившихся в корпусе.

1.1. Постановка задачи

Дан корпус параллельных текстов (текстов и их переводов на другой язык). Тексты выровнены по предложениям. Требуется найти по ним словосочетания и их переводы.

Словосочетание — это устойчивое выражение из двух или более слов, связанных семантически (по смыслу) и грамматически.

Оба этих свойства будут использованы для поиска словосочетаний. Семантическую связь будем оценивать по частотности — слова должны достаточно часто встречаться вместе. Грамматическую связанность будем определять, анализируя дерево синтаксического разбора предложения.

Для выделения семантически-значимых словосочетаний используется ряд фильтров, описанных далее и позволяющих убрать «шумовые» высокочастотные элементы. На получаемые словосочетания дополнительно накладываются ограничения, такие, как разница в длине словосочетания и его перевода, наличие словарных соответствий между ними (или контекстами, в которых они были получены) и т.д. Данные ограничения будут подробно описаны далее.

В нашем распоряжении имеется корпус параллельных текстов. Для поиска словосочетаний используем следующую эвристику. Сопоставим деревья синтаксического разбора параллельных предложений и будем учитывать не только частоту появления той или иной комбинации слов, но то, как она переводится на другой язык. Предположение состоит в том, что словосочетание переводится на другой язык чаще всего одинаково (например, «гаечный ключ» — «wrench»).

Данная работа преследует следующие цели:

1. Разработка алгоритма поиска переводных словосочетаний (с учётом некоторых ограничений, описанных ниже).
2. Получение статистических данных (словосочетаний) для улучшения работы синтаксического анализатора (используемого в том числе в данной работе).
3. Расширение переводного словаря за счёт нахождения новых переводных статей.
4. Создание ТМ-базы (Translation memory) словосочетаний для использования лингвистами в качестве справочного материала.

1.2. Известные подходы к решению задачи

Существует ряд известных работ, посвящённых извлечению словосочетаний. В первую очередь следуют упомянуть работы F.A. Smadja [1, 2], считающиеся классическими в этой области. В их основе лежит статистический подход. Словосочетания порождаются для слов, часто встречающихся совместно и в определённых позициях друг относительно друга.

Высокая частота совместной встречаемости слов, как оказалось, не позволяет утверждать об устойчивости словосочетания (например, в силу специфики корпуса). Поэтому было разработано большое количество метрик, определяющих «меру ассоциации» (от англ. «association measures», или силу связанности) слов коллокации друг с другом. Описание наиболее известных мер есть, например, в [3]. В основном, эти меры учитывают частоту совместной встречаемости слов коллокации и частоту слов в отдельности по корпусу. Несколько примеров таких метрик: взаимная информация (Mutual Information, MI) [4], t-score, логарифмическая функция правдоподобия (log-likelihood) [5].

Использование различных чисто статистических подходов ([1, 2, 6]) можно в первую очередь объяснить их простотой и отсутствием общедоступных и, в то же время, достаточно надёжных синтаксических анализаторов. Описываемый в данной работе подход в свою очередь базируется на использовании синтаксического анализатора, разработанного в компании АВВУУ и имеющего достаточно хорошую точность. Это позволяет рассматривать предложение не как случайный поток слов, а как дерево, определяющее связи между словами предложения. В этом случае словосочетание представляет собой подграф, для которого мы знаем не только слова, входящие в словосочетание, но и зависимости между ними («меловой период» — главным является «период», подчинённым определением — «меловой»). Это позволяет нам задать лингвистические критерии фильтрации «шумовых» (неинтересных нам) словосочетаний (например, с сочинительной связью¹). Мы можем определить понятие вложенности словосочетаний в терминах теории графов, а не теории множеств («союз республик» является вложенным словосочетанием по отношению к «союз советских республик», но не является подмножеством словосочетания «республика в союзе»).

¹ Под сочинительной связью понимается такая связь, при которой отсутствует грамматическая зависимость одного компонента синтаксической конструкции от другого компонента [7]. Например, перечисления через союзы «и», «или».

Часто для уточнения семантической значимости получаемых словосочетаний применяют наложение синтаксических шаблонов (например, «существительное+прилагательное»). Словосочетания, не попадающие ни под один известный шаблон, считаются случайными и выбрасываются. Такой подход применяется, например, в уже упомянутых работах Frank A. Smadja [2]. В данной работе ищутся словосочетания различной длины, поэтому задача создания полной системы синтаксических шаблонов является довольно трудоёмкой. На данный момент от такого подхода решено отказаться.

Кроме того, существуют подходы, использующие синтаксическую информацию (см., например, [3]). Однако в них речь чаще всего идёт о словосочетаниях, состоящих всего из двух слов. Нас же интересуют многословные словосочетания, при поиске которых возникает проблема удаления «частичных» словосочетаний. Данная работа предлагает новый подход к её решению.

1.3. Используемые средства

В работе используются следующие алгоритмы и данные, разработанные ранее в компании АВВУУ:

1. Переводной словарь (русско-английский).
2. Синтаксический анализатор (парсер).
3. Алгоритм пословного сопоставления предложений.

В основе переводного словаря лежат семантические инварианты (межъязыковые статьи). Для каждого из языков описаны различные возможные реализации этих инвариантов — синонимы (например, «бегемот» и «гиппопотам» будут лежать в одном классе). В то же самое время, омонимы будут принадлежать сразу нескольким статьям (слово «bank» будет относиться и к финансовому учреждению, и к подводной мели). Задача разрешения омонимии производится на этапе анализа текста и выходит за рамки данной работы. Используемый словарь достаточно полный и состоит из более чем 60 000 семантических классов.

Синтаксический анализатор строит (как правило) несколько различных вариантов деревьев синтаксического разбора каждого предложения². Для наших целей извлечения словосочетаний используется лучшее дерево синтаксического разбора (на основе внутренних оценок качества деревьев, получающихся при разрешении омонимии). В вершинах этого дере-

ва расположены семантические инварианты, рёбра дерева — связи («главное-подчинённое»). Алгоритм может ошибаться и возвращать неправильное дерево (с неверно разрешённой омонимией). В этом случае мы полагаем, что либо словосочетания не будут порождены совсем (деревья на разных языках слишком сильно отличаются), либо будут порождены неправильные (а значит редкие) переводные словосочетания, которые будут удалены при фильтрации.

Пословное сопоставление (выравнивание) — алгоритм, ставящий в соответствие друг другу слова параллельных фрагментов текста. Данный алгоритм во время своей работы использует результаты синтаксического разбора предложения.

2. Описание применяемого подхода

Поиск словосочетаний можно разделить на следующие этапы :

1. Пословное сопоставление предложений.
2. Генерация одноязычных словосочетаний по деревьям синтаксического разбора.
3. Генерация переводных словосочетаний.
4. Фильтрация кандидатов с учётом частоты их появления в корпусе текстов.
5. Сортировка полученных результатов (словосочетания и новые переводы для словаря).

Ниже рассмотрим каждый этап подробнее.

2.1. Пословное выравнивание предложений

На данном этапе производится сопоставление слов двух параллельных предложений. Ищется наилучшее возможное паросочетание — каждому слову одного предложения ставится в соответствие не более одного слова параллельного предложения (соответствие может быть найдено не для всех слов). Для этой операции используется двуязычный словарь. Для лучшего сопоставления также учитываются зависимости в деревьях синтаксического разбора параллельных предложений.

2.2. Генерация одноязычных словосочетаний

Наложим следующие ограничения на словосочетания (и их переводы):

- ◆ Количество значимых (неграмматических) слов в пределах от одного до пяти ($l_{\max} = 5$).
- ◆ Слова образуют поддерево в дереве разбора предложения.

² Различные представления синтаксической структуры предложения описаны, например, в [8].

- ◆ Среди слов словосочетания нет местоимений.
- ◆ Вершиной синтаксического дерева словосочетания не может быть грамматическая часть речи (стоп-слово).
- ◆ Разница в количестве значимых (неграмматических) слов по сравнению с переводом — не больше одного.
- ◆ Не более одной «дырки»³ в словосочетании ограниченной длины.

Будем идти по дереву разбора и строить всевозможные сочетания слов, удовлетворяющие описанным выше критериям.

2.3. Генерация переводных словосочетаний

Воспользуемся результатами, полученными на двух предыдущих шагах — пословным выравниванием параллельных предложений и множеством вариантов словосочетаний, полученных по этим же предложениям. Найдём на основе этих данных всевозможные переводные словосочетания (точнее, кандидатов на роль переводных словосочетаний). Наложим на них следующие ограничения:

- ◆ Разница в длине словосочетания и его перевода (без учёта грамматических частей) должна быть не больше одного слова.
- ◆ Наличие пословных соответствий среди слов словосочетания (чем длиннее словосочетание, тем больше должно быть соответствий).
- ◆ Для коротких словосочетаний (1-2 слова) пословных соответствий может не быть, но тогда должны совпадать предок корня и все исходящие вершины в дереве синтаксического разбора.
- ◆ Если удалось сопоставить входящие или исходящие вершины для одного из словосочетаний, то эти связи должны соответствовать соответственно входящим или исходящим вершинам второго словосочетания.

На этом этапе порождаются всевозможные словосочетания. На корпусе текстов размером $\approx 4,2$ млн. фрагментов (параллельных предложений) получается порядка 107 млн. различных переводных словосочетаний, из которых только около 7 млн. встречаются 2 и более раза.

2.4. Фильтрация кандидатов

На предыдущем шаге было получено большое количество словосочетаний и их переводов, встре-

чающихся в корпусе параллельных текстов. Теперь возникает задача отобрать из них семантически значимые и устойчивые. Будем полагаться на следующий принцип: устойчивые словосочетания (почти) всегда переводятся одинаково.

Будем отбрасывать словосочетания, переводящиеся по-разному в разных документах, а также редкие (встретившиеся всего несколько раз в корпусе). Некоторые другие эвристики будут описаны ниже.

Выделим следующие этапы фильтрации полученных ранее словосочетаний:

1. Удаляем редкие словосочетания (предварительная фильтрация по частоте).
2. Удаляем словосочетания, содержащие стоп-слова.
3. Удаляем вложенные словосочетания (например, «объединённых наций» — часть словосочетания «организация объединённых наций»), если их частота не сильно превышает частоту объемлющего словосочетания.
4. Аналогично удаляем «объемлющие» словосочетания, если они встречаются значительно реже вложенных («глава организации объединённых наций»).
5. Из множества неоднозначных переводов выбираем наиболее вероятные. При этом проверяем, чтобы выбранный вариант был доминирующим (составлял не менее 70% от возможных переводов словосочетания).
6. Снова удаляем редкие словосочетания, но уже с большим порогом.
7. Удаляем известные пословные словарные переводы.
8. Сортируем полученные словосочетания.

Рассмотрим каждый из фильтров подробнее.

2.4.1. Удаление редких словосочетаний

На этапе генерации словосочетаний по деревьям синтаксического разбора мы получаем огромное количество вариантов. Большинство из них — случайные комбинации слов, не представляющие интереса. Исключим их, используя частотный фильтр. Установим порог f_{\min} , который определяет минимальную частоту словосочетания. Посредством анализа результатов при варьирования величины f_{\min} было найдено оптимальное значение $f_{\min}^* = 5$.

2.4.2. Удаление словосочетаний, содержащих стоп-слова

Стоп-словами называются слова, запрещённые в силу каких-либо обстоятельств. Например, в дан-

³ «Дыркой» назовём разрыв в линейном представлении словосочетания в предложении. Её размер измеряется в количестве слов.

ной работе к ним относятся числа, названия денежных единиц. Стоп-слова делятся на две категории:

1. Запрещённые в любой позиции словосочетания.
2. Запрещённые в качестве корня.

К первой категории относятся, например, числа («вторая конференция», «двухтысячный год», «два с половиной»). Ко второй — названия денежных единиц и некоторые другие классы слов.

Жёсткое удаление словосочетаний, в которых встретилось стоп-слово, может привести к потере интересных нам переводов. Поэтому фильтрация происходит только если стоп-слово встретилось в обеих частях словосочетания.

2.4.3. Удаление вложенных словосочетаний

Алгоритм генерации словосочетаний по предположению порождает все возможные их варианты. Например, в предложении «Десятое заседание Организации Объединённых наций», наряду с правильным словосочетанием «Организация Объединённых наций» мы так же получим варианты «организация наций» и «объединённых наций». Эти варианты корректны с точки зрения условий генерации словосочетаний, но не представляют интереса (в силу своей несамостоятельности).

В обратную сторону: среди переводов названия месяца «November» встречается словосочетание «ноябрь месяц», которое в общем случае не является правильным. Будем удалять и такие словосочетания.

В силу большого числа словосочетаний, попарное их сравнение для выяснения вложенности не представляется возможным. Основная идея данного этапа — введение вспомогательной структуры. Для каждого переводного словосочетания строится множество вложенных в него «подсловосочетаний». Например, из словосочетания «первый полёт на Луну» — «first moonflight» получатся «полёт на луну» — «moonflight», «первый полёт» — «first flight» и т.п. Каждому получившемуся вложенному словосочетанию приписывается частота его «родительского» словосочетания. Затем одинаковые «вложенные» словосочетания (от разных «родителей») объединяются в одно, а частотой получившегося словосочетания берётся максимальная из частот. Как результат, мы получаем новый список из словосочетаний, который можно легко сравнить с исходным. Для удобства назовём этот список массивом «кусочков».

Если для словосочетания удаётся найти пару (такое же) в массиве «кусочков», это означает, что есть (минимум одно) словосочетание большего размера, для которого данное является вложенным. Нам известна максимальная частота «большого» словосочетания (как атрибут «кусочка»). Если она достаточно велика (не сильно меньше частоты рассматриваемого словосочетания), то рассматриваемое словосочетание является неотъемлемой частью некоего большего по размеру и должно быть удалено. Таким образом, удаётся избавиться от «вложенных» словосочетаний.

Для решения обратной задачи (удаления «объемлющих» словосочетаний) необходимо модифицировать массив «кусочков», дописав в атрибуты так же ссылки на исходные словосочетания, из которых были получены кусочки, и их частоты. В этом случае можно так же оценить частоту «большого» словосочетания (соответствующего «кусочку»). Если она существенно ниже частоты рассматриваемого словосочетания, то оставить нужно именно «вложенное» словосочетание. А «внешнее» — удалить.

2.4.4. Разрешаем неоднозначности переводов

В нашей модели словосочетания — устойчивые выражения. Переводные словосочетания также обладают устойчивым переводным соответствием. Значит, для «правильного» словосочетания среди всевозможных вариантов его перевода должен быть один доминирующий.

Например, из-за неточностей перевода и ошибок разбора наряду с правильным переводом «левый» — «left» среди словосочетаний появляются соответствия «левый» — «right». Для их отсекаания в основном и предназначен рассматриваемый фильтр. На этом этапе накладываются достаточно жёсткие требования — доминирующий перевод должен составлять не менее 70% всех переводов.

2.4.5. Повторная фильтрация по частоте

Для того, чтобы позволить последующим фильтрам отработать правильно на словосочетаниях с близкой к пороговой частотой появлений, мы временно оставляем не очень редкие, но всё же «непроходные» по нашим порогам словосочетания. На данном этапе мы их окончательно удаляем.

2.4.6. Удаление пословных переводов

Для переводных словосочетаний, каждая из частей которых состоит из одного слова, проверим, не являются ли они переводами по словарю («левый» – «left»). Удаляем их в подобных случаях.

2.4.7. Сортировка полученных результатов

Полученные в результате фильтрации словосочетания можно условно разделить на две группы:

1. Несловарные переводы слов (один к одному), которых нет в нашем словаре.
2. Собственно, переводные словосочетания.

Первая группа состоит из переводных словосочетаний с одним значимым словом в каждой из частей (условия генерации словосочетаний их допускают). Полученные таким образом соответствия служат лингвистам справочным материалом для обнаружения недостающих связей в словаре, а так же для исправления описаний.

Вторая группа содержит искомые переводные словосочетания.

2.5. Порядок применения фильтров

Рассмотренные выше фильтры выполняются в том же порядке, в котором они описаны в данной работе. Рассмотрим, какими факторами это обусловлено.

Частотный фильтр удаляет словосочетания, которые априори не представляют интереса для дальнейшего рассмотрения. Мы считаем, что если словосочетание редкое, то дальнейшими статистическими проверками мы не сможем доказать его значимость. Фильтр стоп-слов так же удаляет априори неинтересные варианты, уменьшая необходимую работу на последующих этапах. Важным также является тот факт, что после применения этих фильтров количество словосочетаний уменьшается на 90-95%.

Фильтры «вложенных»/«внешних» словосочетаний, а также неоднозначных переводов применяются на одном множестве словосочетаний, так как каждый из них требует максимально полной информации о «близких» к ним словосочетаниях (отличающихся дополнительными словами или другими переводами).

Повторная фильтрация по частоте призвана удалить те редкие словосочетания, которые мы намеренно оставили для корректной работы других фильтров (например, фильтра неоднозначных

переводов). Наряду с фильтром пословных переводов они удаляют то, что не должно войти в результат по итогам фильтрации. Их порядок значения не имеет (суммарный результат остаётся одинаковым).

На этапе сортировки получаются окончательные результаты. Этот фильтр работает теми словосочетаниями, которые представляют практический интерес. Соответственно, он выполняется последним.

3. Количественные результаты

Эксперименты проводились на русско-английском корпусе, содержащем 4,2 млн. фрагментов (параллельных предложений).

Алгоритм порождения переводных словосочетаний даёт 62 млн. различных пар (всего порождается около 107 млн. словосочетаний). Из них подавляющее большинство (56 млн.) встречаются только один раз. Динамика количества остающихся словосочетаний на различных этапах фильтрации показана в табл. 1.

Таблица. 1.

Динамика фильтрации словосочетаний

Название фильтра	Словосочетаний на выходе
По частоте (предварительно)	2,5 млн.
По списку стоп-слов	1,1 млн.
Вложенные словосочетания	568 тыс.
Неоднозначные переводы	105 тыс.
Словарные переводы	66,5 тыс.
По частоте (окончательно)	42 636
Сортировка результатов	42 535

Таким образом, на экспериментальном корпусе удаётся получить $\approx 42,5$ тыс. переводных словосочетаний. Несколько примеров новых переводов (отсутствовали в используемом словаре), найденных на этом этапе приведены в табл. 2.

Таблица. 2.

Примеры новых словарных связей (отсутствовали в словаре)

Английский	Русский
Glib	Бойкий
Ledger	Гроссбух
Reciter	Чтец

Несколько примеров из найденных словосочетаний приведены в табл. 3.

Таблица 3.

Примеры найденных словосочетаний

Английский	Русский	Встретилось в корпусе
job time	срок задания	12
galaxy space	космическое пространство	13
other foreign object	иной посторонний объект	5
air transport field	область воздушного транспорта	30
to be beyond the scope of book	выходить за рамки книги	29
to establish under article	учреждать в соответствии со статьёй	75

Существуют две общепринятые меры измерения качества полученных результатов — точность и полнота. В данной работе основной упор на повышении качества получаемого результата. Это связано с тем, что практически невозможно вручную просмотреть несколько десятков тысяч словосочетаний для поиска в них ошибок. Полнота же может быть увеличена, например, за счёт увеличения текстовой базы. В отличие от статистических подходов, мы предпочитаем удалять редкие словосочетания, так как достаточно высока вероятность встретить среди них те, которые получены в результате ошибок разбора или пословного сопоставления.

Для оценки качества мы используем два подхода. Первый является косвенным — проводится синтаксический разбор тестовой выборки с и без полученных словосочетаний. Сравнение полученных результатов часто позволяет найти ошибки алгоритма фильтрации (те словосочетания, которые не должны были пройти этот этап).

Второй подход — это ручная проверка лингвистами случайной выборки. Пример такой оценки представлен в табл. 4. Словосочетания, забракованные экспертом, условно разделены на 3 категории: ошибки словаря (лингвистического описания), ошибки алгоритма (для которых известны пути исправления) и прочие ошибки. В частности, одна из основных проблем — отсутствие учёта дублирующихся фрагментов (одинаковые предложения из инструкций, приказов и подобных шаблонных документов, имеющихся в базе). С учётом добавления этого функционала можно достигнуть качества в 80% и более.

Таблица 4.

Экспертная оценка качества на выборке из 100 словосочетаний.

Качество	Процент
Хорошие словосочетания	67
Ошибки в описаниях	4
Недоработки алгоритма	16
Другое	12

4. Заключение

Результатом проведённого исследования является алгоритм, позволяющий находить переводные словосочетания по корпусу параллельных текстов.

Есть несколько направлений возможного дальнейшего его развития.

- ◆ Введение функции оценки качества полученных словосочетаний.
- ◆ Настройка пороговых значений для фильтров.
- ◆ Добавление новых фильтров.
- ◆ Учёт дублирования отдельных фрагментов. ■

Литература

1. Smadja F.A. Retrieving collocations from text: Xtract // Computational Linguistics. 1993 Vol. 19. No. 1. — Pp. 143-177.
2. Smadja F.A., McKeown K. Translating collocations for use in bilingual lexicons. — HLT, 1994.
3. Evert S. The Statistics of Word Cooccurrences Word Pairs and Collocations. PhD thesis / Universität Stuttgart. — Institut für Maschinelle Sprachverarbeitung (IMS), 2004.
4. Church K.W., Hanks P. Word association norms, mutual information, and lexicography // Computational Linguistics. 1990. Vol. 16. No. 1. — Pp. 22-29.
5. Dunning T. Accurate methods for the statistics of surprise and coincidence // Computational Linguistics. 1993. Vol. 19. No. 1. — Pp. 61-74.
6. Bouma G. Collocation extraction beyond the independence assumption // Proceedings of the ACL 2010 Conference Short Papers / ACLShort '10. — Stroudsburg, PA, USA: Association for Computational Linguistics, 2010. — Pp. 109-114.
7. Казарина В.И. Современный русский синтаксис: структурная организации простого предложения. — Елец: ЕГУ, 2007.
8. Bolshakov I.A., Gelbukh A.F. Computational Linguistics: Models, Resources, Applications. — IPN - UNAM - Fondo de Cultura Economica, 2004.

МОДЕЛЬ ДЛЯ ИДЕНТИФИКАЦИИ ЕСТЕСТВЕННОГО ЯЗЫКА ТЕКСТА

С.В. Гусев,
программист ЗАО «НОРСИ-ТРАНС»

А.М. Чеповский,
кандидат технических наук, профессор кафедры информационных систем
Московского государственного университета печати (МГУП);
доцент кафедры анализа данных и искусственного интеллекта НИУ ВШЭ

Адрес: г. Москва, ул. Прянишникова, д. 2а
E-mail: unk379@mail.ru, acher@adde.math.msu.su

В статье рассмотрена проблема автоматической идентификации естественного языка текста. Предлагается статистическая модель текстов на естественном языке. Рассматриваются алгоритмы определения естественного языка текста.

Ключевые слова: статистическая модель языка, идентификация естественного языка текста.

Введение

В информационных системах различного типа, предназначенных для обработки в автоматическом режиме больших объемов текстов на естественных языках, актуальны различные задачи распознавания текстовой информации. Требование автоматизации процессов обработки текстовой информации придает особую важность проблемам определения естественного языка, на котором написан текст, или часть текста.

Для информационных систем предприятий, систем документооборота особую важность при-

обретает задача определения в автоматическом режиме языка коротких текстовых сообщений размеров, характерных для сообщений электронной почты и аналогичных коммуникационных сервисов. В настоящее время известны достаточно точные методы распознавания языка для длинных текстов, содержащих десятки слов и предложений [1]. Модели, использующие частоты буквосочетаний, широко использовались для определения языка текста [2, 3]. В [3] отмечалось, что возможно использовать ранговые методы для идентификации языка текста,

но они не применимы для коротких текстов. В работе [3] делается вывод о том, что проблема определения языка коротких сегментов текста остается актуальной, а достижение более высокой точности осуществлялось за счет более крупных моделей и медленной скорости классификации.

В данной статье рассматривается статистическая модель строки текста на естественном языке и разрабатывается методика ее реализации с целью достижения эффективности использования статистики буквосочетаний символов языка для достижения высокой точности классификации текстов по используемым естественным языкам.

**Статистическая
модель текста
на естественном языке**

Задача определения языка текста является задачей распознавания образов и может решаться на базе вероятностной модели. Принцип Байесовского классификатора можно применить к строке символов, считая, что нам известны статистические характеристики для символов в текстах на конкретном естественном языке, или текстах, относящихся к заданному классу.

Рассмотрим строку s , состоящую из N символов c_n ($n = 1, \dots, N$), принадлежащих алфавиту Σ . Конкретное значение строки будем обозначать следующим образом: $s = \langle c_1 c_2 \dots c_N \rangle$, а значение отдельного символа строки на i -той позиции будем обозначать так: $s[i] = c_i$. Для решения задачи необходимо отнести данную строку к одному из классов Y_l ($l = 1, \dots, K$), где под Y_l ($l = 1, \dots, K$) понимается один из K языков.

Считаем, что каждый класс задает некоторое распределение вероятностей на множестве всех допустимых строк. В таком случае возможно применение статистического критерия максимального правдоподобия для определения класса, которому принадлежит классифицируемая строка.

Вероятность появления строки s в некотором языке равна произведению вероятностей появления каждого из её символов при условии появления всех символов, идущих перед ним:

$$\begin{aligned} P(s = c_1 \dots c_N) &= P(s[N] = c_N | s[1] = c_1, \dots, s[N-1] = c_{N-1}) * \\ &= P(s[N-1] = c_{N-1} | s[1] = c_1, \dots, s[N-2] = c_{N-2}) * \dots * P(s[1] = c_1) \end{aligned} \quad (1)$$

Предположим, что распределение вероятностей i -того символа, зависит не более чем от k предыдущих, тогда формула (1) примет вид:

$$\begin{aligned} P(s[i] = c_i | s[1] = c_1, \dots, s[i-1] = c_{i-1}) &= \\ = P(s[i] = c_i | s[i-k] = c_{i-k}, \dots, s[i-1] = c_{i-1}) \end{aligned} \quad (2)$$

По обучающей выборке производится оценка условных вероятностей. Для этого вычисляются частоты всех подстрок длины не более k , и значения условных вероятностей появления очередного символа принимаются равными отношению частот соответствующих подстрок:

$$\begin{aligned} P(Y_l, s[i] = c_i | s[i-m] = c_{i-m}, \dots, s[i-1] = c_{i-1}) &= \\ = \frac{f(c_{i-m} \dots c_i)}{f(c_{i-m} \dots c_{i-1})}, \forall m \leq k \end{aligned} \quad (3)$$

где $f(c_i)$ – частота встречаемости подстрок в обучающей выборке.

Для каждой классифицируемой строки вычисляется ее вероятностная оценка относительно каждого класса:

$$\begin{aligned} P(Y_l, s) &= P(Y_l, s[N] = c_N | s[N-k] = c_{N-k}, \dots, s[N-1] = c_{N-1}) * \\ &= P(Y_l, s[N-1] = c_{N-1} | s[N-k-1] = c_{N-k-1}, \dots, s[N-2] = c_{N-2}) * \dots * P(Y_l, s[1] = c_1) \end{aligned} \quad (4)$$

Классифицируемая строка относится к классу, относительно которого она имеет наибольшую оценку.

**Алгоритм реализации модели
для идентификации языка**

Каждый текст на естественном языке подвергается предварительной обработке, в результате которой из текста получается набор слов, состоящих только из символов алфавита соответствующего языка, приведённых к нижнему регистру. По полученному набору слов формируется частотный словарь буквосочетаний длиной от 1 до n с учётом количества вхождений слов в частотный словарь. Данная процедура выполняется при построении модели строки конкретного языка, которая строится на основе массива обучающих текстов и завершается заполнением базы данных профилей для каждого исследуемого языка.

К каждому слову текста прибавляется один пробел справа, и оно подаётся на вход автомата. Об-

работка слова начинается из начального состояния автомата, соответствующего буквосочетанию, состоящему из $(n - l)$ пробела. Для очередного символа вычисляется вероятность перехода по нему из текущего состояния в следующее состояние по формулам (1) – (4). Вероятностью появления данного слова считается произведение вероятностей всех переходов, произошедших во время обработки слова автоматом.

Для определения языка текста оценивается вероятность соответствия рассматриваемого текста моделям строк для каждого естественного языка. Выбирается максимальная вероятность, которая соответствует языку, на котором написан текст.

Предложенная методика и алгоритмы реализованы в виде библиотеки для платформы MS Windows x86.

Точность определения языка текста

Для фрагмента текста может быть вычислена числовая оценка того, насколько данный фрагмент соответствует модели строки текста на естественном языке. Пусть фрагмент текста s содержит N символов. Вероятность его появления в тексте на l -том языке может быть оценена в соответствии с формулой (4). Тогда оценку соответствия этого фрагмента l -тому языку будем определять следующим образом:

$$E_l(s) = \frac{\ln(P(Y_l, s))}{N} + const \quad (5)$$

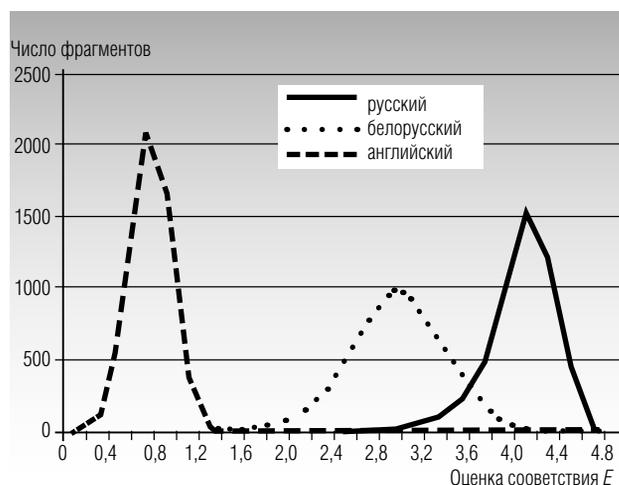


Рис. 1. Распределение оценки соответствия модели текста по количеству фрагментов длиной 20 символов для различных языков.

где $P(Y_l, s)$ – вероятность появления строки s в языке Y_l ;

N – длина строки s в символах;

$const$ – нормирующая постоянная величина.

Математическое ожидание такой оценки не зависит от длины фрагмента текста.

При вычислении оценки соответствия данному языку фрагментов текстов на других языках результат сильно зависит от используемого другими языками алфавита и степени похожести языков на данный язык. На рисунке 1 представлены распределения оценок (5) соответствия модели строки на русском языке для 5000 фрагментов длины 20 символов из текстов на русском, белорусском и английском языках.

Видно, что если алфавиты языков не пересекаются с русским алфавитом (английский), то тексты на этих языках будут отсекаются с высокой точностью. Если же алфавиты языков пересекаются с русским алфавитом (белорусский), то невозможно точно отделить тексты на другом языке при простейшем пороговом отсечении по максимальному значению вероятности.

При увеличении длины фрагмента шансы отделения неизвестного языка с помощью порогового значения увеличиваются. На рисунке 2 показано распределение оценок 2000 фрагментов длиной 200 символов для текстов на русском и белорусском языках. Видно, что для длины фрагмента текста в 200 символов уже можно отделить русский текст от белорусского с высокой степенью достоверности.

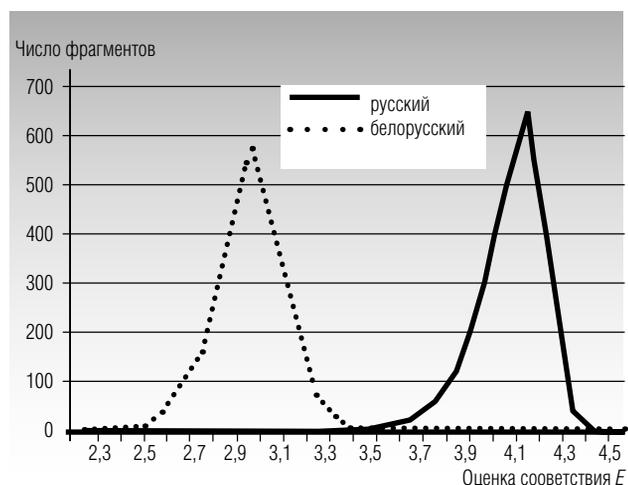


Рис. 2. Распределение оценки соответствия модели текста по количеству фрагментов длиной 200 символов для различных языков.

Приведенные результаты на рисунках 1 и 2 показывают существование серьезной проблемы при определении языков в условиях, когда входные тексты могут содержать языки, отсутствующие среди определяемых языков.

Для отсеечения неизвестных языков используется пороговое значение, которое вычисляется отдельно для каждого языка и каждой длины фрагмента:

$$T_l = M[E_l] - \gamma * \sigma[E_l], \quad (6)$$

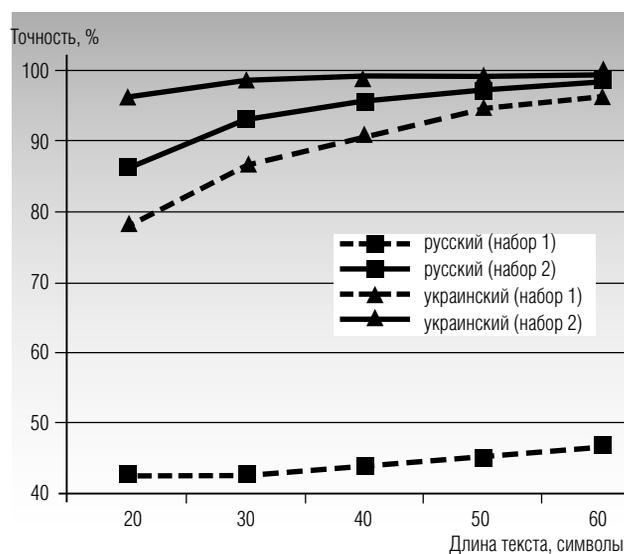


Рис. 3. Зависимость точности определения русского и украинского языков от длины текста для двух вариантов обучающих наборов.

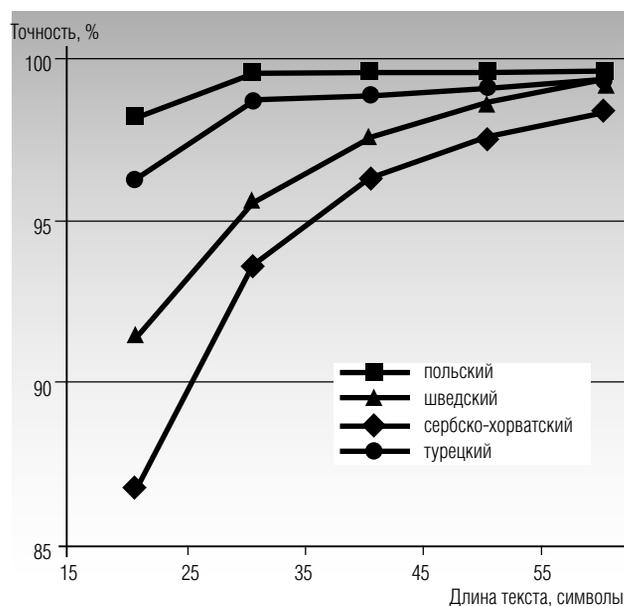


Рис. 4. Зависимость точности определения языка в зависимости от длины текста.

где $M[E_l]$ – математическое ожидание оценки соответствия для l -того языка;

$\sigma[E_l]$ – среднеквадратичное отклонение оценки соответствия для l -того языка;

γ – настраиваемый коэффициент порога отсеечения.

Для высокой точности определения языка в условиях, когда входные тексты могут содержать неизвестные языки, требуется уметь определять все языки, использующие такую же систему письменности (т.е. алфавиты должны существенно пересекаться). Тексты, написанные на языках, использующих другие системы письменности, можно отсечь с помощью порогового значения оценки даже в случае очень коротких фрагментов. Для увеличения качества распознавания языков необходимо использовать базы данных для наибольшего количества языков.

Экспериментальные исследования

Для оценки точности распознавания используются следующие характеристики. Коэффициент релевантности определяет, какова в результирующем наборе, принадлежащем определенному классу, доля файлов, которые действительно являются файлами на соответствующем естественном языке. Коэффициент полноты измеряет при тестовом эксперименте, какая доля файлов данного типа (на соответствующем естественном языке) из тестового набора правильно отнесена к данному типу файлов. В качестве главной характеристики качества распознавания используется усредненная точность, которая определяется как взвешенное гармоническое среднее коэффициента релевантности и коэффициента полноты.

Всего рассматривалось сорок пять естественных языков, включающих использующие кириллическую письменность языки (славянские и относящиеся к языкам тюркской и иранской групп), основные языки индоевропейской семьи, использующие латинскую письменность. Эксперименты по определению естественного языка текста проводились на тестовых массивах, суммарные размеры которых в зависимости от языка изменялись от 1Мб до 7Мб «чистого» текста. При этом 80% текстов использовалось для обучения, а 20% этих корпусов текстов рассматривались как тестовые наборы, язык которых распознавался.

Влияние обучающих выборок на качество распознавания языков изучалась на двух наборах естественных языков, использующих кириллическую письменность. Набор 1 состоял из шести славянских языков: Русский, Украинский, Белорусский, Болгарский, Сербскохорватский, Македонский. Набор 2 включал языки первого набора и десять других языков, использующих кириллическую письменность: Азербайджанский, Башкирский, Карачаево-балкарский, Казахский, Киргизский, Татарский, Узбекский, Осетинский, Таджикский, Монгольский. Идентификация языка текста проводилась для двух вариантов обучения, основанных на описанных выше двух наборах естественных языков.

На *рисунке 3* представлены зависимости усредненной точности определения русского и украинского языков в зависимости от размера анализируемых текстов, измеряемых в количестве алфавитных символов языка, исключая пробелы. Видно, что при увеличении обучающего набора с «Набор 1», включающего 6 языков, до «Набор 2», включающего 16 языков, точность определения языков существенно увеличивается.

Тестовые эксперименты по исследованию качества распознавания естественных языков проводились для 45 естественных языков. Для всего этого набора проводилось обучение системы распознавания языков, и идентифицировались языки для текстов на всех 45 языках. Типичные результаты представлены на *рисунке 4*, где приведены результаты

численных экспериментов на примерах распознавания польского, сербскохорватского, шведского и турецкого языков в зависимости от размеров анализируемых текстов. Достигнутые показатели качества аналогичны представленным в [3]. Видно, что для текстов даже небольших размеров достигается высокая (близкая к 100%) точность определения естественного языка.

9. Заключение

Таким образом, в статье предложена вероятностная модель строки текста на естественном языке и разработана методика автоматического определения языка текстового сообщения на естественном языке, включая небольшие по объему тексты.

Разработанная модель текста на естественном языке и методика ее реализации показала эффективность использования статистики буквосочетаний символов языка для решения задач распознавания текстов на естественном языке. Достигнута высокая точность распознавания естественного языка текста.

Методика эффективна при обработке больших объемов текстовой информации на различных естественных языках без использования словарей и грамматик этих языков. Предложенная методика может быть использована при решении практических задач автоматической обработки текстовой информации в информационных системах предприятий и организаций. ■

Литература

1. Paul B., McNamee. Language identification: a solved problem suitable for undergraduate instruction // Journal of Computing Sciences in Colleges, 2005, 20(3). — P. 94-101.
2. Cavnar W.B., Trenkle J.M. N-gram-based text categorization // Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval. — Las Vegas, US, 1994. — P. 161-175.
3. Vatanen T., Väyrynen J., Virpioja S. Language identification of short text segments with n-gram models // Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), 2010. — P. 3423-3430.

ПРОГРАММНЫЙ КОМПЛЕКС ДЛЯ ГЕНЕРАЦИИ СЕМЕЙСТВ ТРАНЗИТИВНЫХ ГРАФОВ СТЕПЕНИ 4

А.А. Незнанов,

кандидат технических наук, доцент кафедры анализа данных и искусственного интеллекта Национального исследовательского университета «Высшая школа экономики»

Ю.В. Старичкова,

ассистент кафедры прикладной математики Московского энергетического института (Технического университета)

Адрес: г. Москва, ул. Кирпичная, д. 33/5

E-mail: aneznanov@hse.ru, starichkovayv@mpei.ru

Описывается оригинальный программный комплекс для генерации бесконечных и конечных семейств связанных транзитивных графов степени 4, полностью покрывающих все известные графы до 30 вершин. Отличительной особенностью разработки является многокритериальная каталогизация семейств на основе характеристик симметрии, структурной сложности и визуализации симметричных диаграмм. Комплекс расширяет функциональные возможности АСНИ «Graph Model Workshop» и позволяет решать задачи, требующие синтеза транзитивных топологий.

Ключевые слова: структурный анализ, теория графов, теория групп, транзитивные графы, генерация, симметрия, структурная сложность, алгоритм, программная реализация.

1. Введение

Транзитивные графы (ТГ) – графы, все вершины которых эквивалентны по своим свойствам, то есть принадлежат одной орбите группы автоморфизмов. Такие графы играют особую роль как в теории, так и в многочисленных приложениях. Это связано с тем, что они являются моделями структур гомогенных систем. С теоретической точки зрения актуальны задачи конструк-

тивного перечисления и классификации ТГ, что необходимо для развития смежных разделов дискретной математики. Практически значимы задачи анализа и синтеза ТГ с необходимыми характеристиками, ставящиеся при исследовании систем управления, топологий вычислительных сетей, химических реакций, строения кристаллов, предельной связности, надёжности и инвариантных преобразований гомогенных систем, методов кодирования данных и т.п. Данные задачи удобнее всего

решать, выделив бесконечные семейства с требуемыми свойствами. Основой выделения семейств являются характеристики симметрии, структурной сложности и прорисовки диаграмм.

В работе рассматривается подкласс *связных транзитивных графов степени 4* (ТГС4). Он наиболее широко используется на практике, так как, например, предлагает удачный компромисс между уровнями структурной надёжности и структурной сложности (в первую очередь выражаемой числом рёбер). При этом, в отличие от транзитивных графов степени 3, ТГС4 исследованы недостаточно.

Целью классификации является выделение (по набору критериев) из рассматриваемого множества графов семейств подобных графов с растущим числом вершин. Исходными данными является полная база связных ТГС4 до 30 вершин включительно. Работа базируется на проведённой ранее авторами классификации [8], которая, в свою очередь, стала возможной благодаря предыдущей работе Незнанова А.А. и Кохова В.А. по вычислению характеристик и систематизации справочной информации о ТГС4 [3].

Результатом работы является программный комплекс «Полный генератор семейств транзитивных графов степени 4» синтеза семейств ТГС4, который расширяет функциональные возможности АСНИ «*Graph Model Workshop*» (авторы – Кохов В.А., Незнанов А.А., Ткаченко С.В.). Отметим, что АСНИ уже включает в качестве расширений теоретически-ориентированные генераторы различных классов и семейств, включая случайные структуры общего вида, деревья, последовательно-параллельные, планарные графы и т.д., а также практически-ориентированные генераторы молекулярных графов, регулярных топологий и др.

2. Основные определения

Базовые понятия теории графов и теории групп можно найти в [1-3]. Для понимания последующего изложения приведём следующие определения, связанные с исследуемым классом графовых моделей.

Пусть $G = (V, E)$ обозначает обыкновенный граф, где идентификаторами (номера) вершин и ребер выступают числа натурального ряда:

$$\begin{aligned} V &= \{v_1, v_2, \dots, v_p\} = \{1, \dots, p\}, \\ p &= |V| - \text{число вершин;} \\ E &= \{e_1, e_2, \dots, e_q\} = \{1, \dots, q\}, \\ q &= |E| - \text{число ребер.} \end{aligned}$$

Через $Aut(G)$ обозначим группу автоморфизмов графа G (ГАГ), а через $|Aut(G)|$ – порядок группы. Фиксатор (стабилизатор) вершины $v \in V$ – подгруппа $Aut(G, v)$, оставляющая неподвижной вершину v , то есть $Aut(G, v) = \{g \in Aut(G) : g(v) = v\}$. Фиксатор подмножества вершин $V^0 \subseteq V$ – подгруппа $Aut(G, V^0)$, оставляющая неподвижной каждую вершину множества V^0 , то есть

$$Aut(G, V^0) = \bigcap_{v \in V^0} Aut(G, v).$$

Стабилизатор подмножества вершин $V^1 \in V$ – подгруппа $Aut[G, V^1]$, оставляющая множество V^1 неподвижным, то есть

$$Aut[G, V^1] = \{g \in Aut(G) : \forall v \in V^1 [g(v) \in V^1]\}.$$

Орбита вершины $v \in V$ – подмножество $\Theta(Aut(G), v)$ вершин графа G , которые могут быть отображены на вершину v :

$$\Theta(Aut(G), v) = \{v' : [\exists g \in Aut(G) : g(v') = v]\}.$$

Граф называется *транзитивным*, если все его вершины принадлежат одной орбите.

Орбита вершины $v \in V$ относительно фиксатора $Aut(G, V^0)$ – подмножество $\Theta(Aut(G, V^0), v)$ вершин графа G , которые могут быть отображены на вершину v при условии фиксации подмножества $V^0 \subset V$:

$$\Theta(Aut(G, V^0), v) = \{v' : [\exists g \in Aut(G, V^0) : g(v') = v]\}.$$

Орбита вершины $v \in V$ относительно стабилизатора $Aut(G, V^1)$ – подмножество $\Theta(Aut(G, V^1), v)$ вершин графа G , которые могут быть отображены на вершину v при условии стабилизации подмножества $V^1 \subset V$:

$$\Theta(Aut(G, V^1), v) = \{v' : [\exists g \in Aut(G, V^1) : g(v') = v]\}.$$

Определим также следующие интегральные характеристики симметрии. Подмножество вершин $V^- \subset V$ называется *экстремальным подмножеством нетождественной стабильности* графа, если справедливо

$$(Aut(G, V^-) \neq E_p) \ \& \ (\forall v \in V \setminus V^- : Aut(G, V^- \cup \{v\}) \approx E_p).$$

Подмножество вершин $V^+ \subset V$ называется *экстремальным подмножеством тождественной стабильности* графа, если справедливо

$$(Aut(G, V^+) \approx E_p) \& (\forall v \in V^+ : Aut(G, V^+ \setminus \{v\}) \neq E_p).$$

Пусть Π^- и Π^+ обозначают соответственно множество всех подмножеств V^- и V^+ вершин графа G .

Тогда число тождественной стабильности (Ψ) и число нетождественной стабильности (χ) графа равны, соответственно:

$$\Psi = \min_{V^+ \in \Pi^+} |V^+|$$

$$\chi = \max_{V^- \in \Pi^-} |V^-|.$$

Числом тождественности $t(G)$ графа G называется минимальное число новых вершин, необходимых для построения тождественного надграфа OG графа G , то есть

$$t(G) = \min_{Aut(OG) = E_p} (|V_{OG}| - |V_G|).$$

Прорисовка диаграммы графа – получение одного из визуальных образов графа на плоскости. Мы будем рассматривать симметричные диаграммы ТГС4, то есть диаграммы, хотя бы частично отражающие симметрию вершин графа. *Вариантами прорисовки* ТГС4 назовём получение диаграмм, существенно отличающихся по расположению вершин и рёбер графа относительно осей симметрии. Наиболее популярными вариантами прорисовок транзитивных графов являются прорисовки с расположением вершин по нескольким окружностям с центром в одной и той же точке.

Семейство – последовательность (конечная или бесконечная) подобных по некоторым характеристикам графов с возрастающим числом вершин, для которых определена процедура построения следующего представителя последовательности по предыдущему.

3. История разработки и основные особенности комплекса

Основной задачей синтеза топологий ТГС4 является не только конструктивное перечисление всех структур на заданном числе вершин, но и получение ТГС4 с большим числом вершин и заданными характеристиками симметрии. Одним из подходов к синтезу ТГС4 с большим числом вершин и заданными характеристиками симметрии является классификация и генерация семейств с заданными характеристиками на основе уже перечисленных ТГС4.

Хотя серьёзное исследование класса транзитивных графов началось ещё в середине 20 века, полноценные результаты по конструктивному пере-

числению ТГ были получены уже в 1990-х годах и отражены в работах *H.P.Yap, B.D.McKay, G.F.Praeger* и *C.E.Royle* [4-6]. На данный момент в каталоге ТГС4 перечислены все структуры с числом вершин до 31 (297 графов).

В 1988 г. Коховым В.А. были представлены результаты оригинального исследования интегральных характеристик симметрии класса ТГС4 [7]. В дальнейшем, опираясь на ранее полученные Незнановым А.А. (2001-2004 гг.) и Киричек О.В. (2003-2004 гг.) результаты, начата разработка программного комплекса для синтеза данного класса графов. Следует отметить, что в работе, выполненной ранее Киричек О.В., присутствовал ряд проблем, не позволяющих говорить о полной завершенности созданного программного комплекса, в частности:

- ◆ неполнота классификации (не все графы, содержащиеся в базе, могли быть классифицированы, как представители одного из семейств, предложенных в работе);
- ◆ избыточная сложность классификации (подклассы, которые целесообразно объединять в один класс, были выведены в роли отдельных классов);
- ◆ некоторые изоморфные графы были отнесены в различные семейства, таким образом, формально отличаясь друг от друга.

Работа по улучшению классификации ТГ была начата в 2005 году. В ходе работы постоянно уточнялась классификация ТГ по критериям: порядок группы автоморфизмов; число тождественной стабильности; число нетождественной стабильности; структура орбит фиксатора вершины; внешний вид диаграммы графа (основной эстетический критерий выделения семейства). В 2008 году была опробована первая версия программного комплекса (ПК) «Полный генератор семейств транзитивных графов степени 4» (*TransGen*) и были приведены первые результаты полной классификации ТГС4 до 30 вершин включительно [8]. Однако было ясно, что результаты можно улучшить. В 2010 году была готова новая версия комплекса синтеза ТГС4, которая позволяет генерировать семейства:

- ◆ как конечные, так и бесконечные;
- ◆ графы которых либо имеют идентичные значения характеристик $|Aut(G)|$, Ψ , χ , $t(G)$, либо изменение этих значений при переходе к следующему графу семейства описывается одной и той же формулой;
- ◆ начиная как с первого графа семейства (выбранного так, чтобы обеспечить без избыточность

генерации), так и с первого *уникального представителя* семейства, для которого стабилизируются все характеристики семейства;

♦ задавая характеристики $|Aut(G)|$, ψ и χ , интересующие пользователя (с возможностью поиска в базе характеристик);

♦ выбирая одну из симметричных прорисовок диаграмм для семейства.

4. Функциональность программного комплекса

4.1. Основные возможности и объёмные характеристики

Основными возможностями программного комплекса *TransGen* является генерация выбранного пользователем подмножества 59 бесконечных и 73 конечных семейств ТГС4 с симметричной прорисовкой диаграмм синтезируемых графов.

Программный продукт *TransGen* является расширением АСНИ «*Graph Model Workshop*» (*GMW*), разработанной в научной группе Кохова В.А. *TransGen* создавался в среде *CodeGear RAD Studio 2007 (Delphi Personality)* в виде динамически компонуемой библиотеки, которая экспортирует функции, вызываемые АСНИ. В *табл. 1* содержится информация об общем объёме программных разработок.

Таблица 1.

Некоторые параметры разработанного программного комплекса

№	Параметр	Значение
1	Число строк авторского исходного кода	6280
2	Объём авторского исходного кода	164 КБ
3	Количество компилируемых строк кода	98742
4	Объём машинного кода	2,1 МБ

4.2. Архитектура комплекса и описание порождаемых семейств

На *рис. 1* представлена архитектура ПК *TransGen*. При реализации программного комплекса для построения семейств ТГС4 используются шаблоны расстановки вершин по окружности и шаблоны генерации рёбер. Шаблоны расстановки вершин по окружности определяются классификацией семейств, указание шаблона генерации рёбер позволяет получить подклассы семейств. В *табл. 2* приведены основные шаблоны расстановки вершин по

окружности и количество семейств, для построения которых они использованы (идентификаторы семейств имеют вполне определённый смысл, который, однако не важен прикладному пользователю).

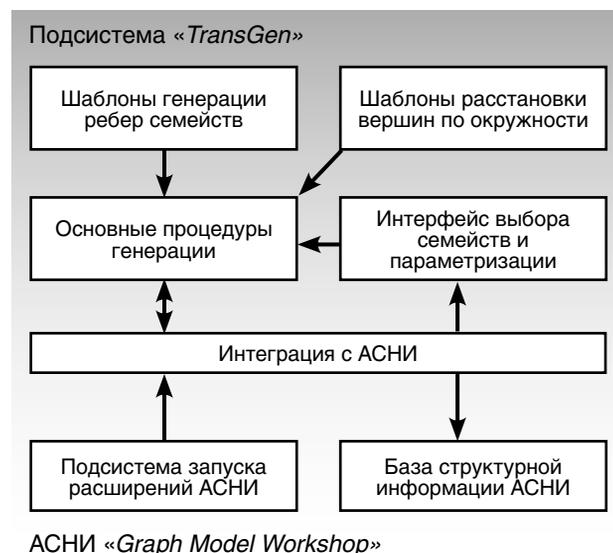


Рис. 1. Архитектура ПК *TransGen*.

Таблица 2.

Классы шаблонов расстановки вершин по окружности

№	Шаблоны расстановки вершин по окружности	Количество семейств
1	Вершины расположены на одной окружности, соседние вершины смежны. Подклассы определяются расположением остальных рёбер.	51 семейство. G_1, G_2, G_5, G_6, G_7, G_102, G_107, G_108, G_119, G_209, G_230, G_11, G_12, G_13, G_14, G_15, G_16, G_17, G_18, G_20, G_103, G_210, G_211, G_231, G_239, G_21, G_22, G_25, G_68, G_78, G_214, G_215, G_33, G_219, G_237, G_238, G_26, G_27, G_28, G_35, G_29, G_30, G_36, G_37, G_38, G_74, G_117, G_217, G_220, G_233, G_203
2	Вершины расположены на одной окружности и соседние несмежны.	4 семейства. G_42, G_125, G_221, G_222
3	Вершины расположены по нескольким окружностям.	1 семейство. G_56
4	Графы, имеющие прорисовку другого вида (тор).	3 семейства. G_10001 – 10004
5	Конечные семейства с уникальными шаблонами.	73 семейства. G_1001 – G_1004, G_1006 – G_1074
Итого:		59 бесконечных и 73 конечных семейства

Для дальнейшего построения представителей семейства после определения шаблона расстановки вершин по окружности применяется шаблон генерации ребер для расположения остальных ребер в графе. Пример схемы построения ребер для шаблона первого типа приведен на рис. 2, где:

- ◆ p – число вершин в графе;
- ◆ i – текущая обрабатываемая алгоритмом синтеза вершина (для графов с прорисовкой по одной окружности нумерация ведется последовательно по окружности по часовой стрелке);
- ◆ ребро, проходящее через центр графа – ребро, соединяющее вершины i и $(p/2-i)$;
- ◆ k – переменная, используемая для указания смежности вершин (задает номер вершины в соответствии с ее расположением на окружности).

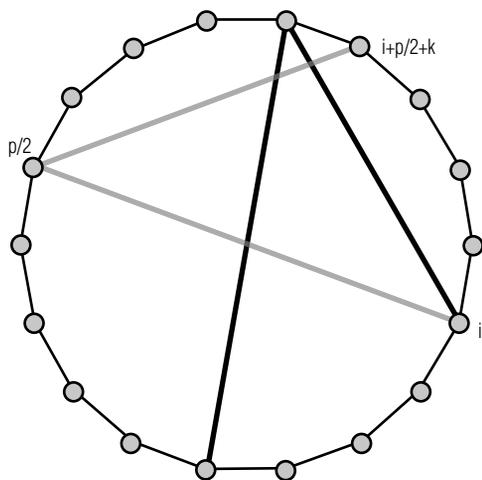


Рис. 2. Иллюстрация понятия шаблона генерации ребер.

Модуль параметризации шаблона хранит базу информации о семействах, в том числе:

- ◆ шаг SV – разность числа вершин между i и $(i-1)$ графами;
- ◆ первый граф семейства с числом вершин FV ;
- ◆ первый уникальный граф с числом вершин UV – представитель семейства, начиная с которого и для каждого следующего графа характеристики симметрии совпадают;
- ◆ число вершин FN последнего графа в семействе, если оно конечное;
- ◆ значения или формулы вычисления числа симметрии ($|Aut(G)|$), чисел тождественной (Ψ) и нетождественной (χ) стабильности;
- ◆ варианты прорисовок диаграмм для семейств, среди которых выделена каноническая.

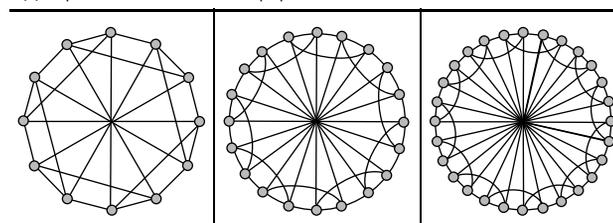
Пример информации о семействе G_{68} приведён в табл. 3, где n обозначает порядок группы предыдущего представителя семейства, а m – число нетождественной стабильности предыдущего представителя семейства.

Таблица 3.

Пример информации о семействе ТГС4

Название: G_68						
FV	UV	SV	Ψ	χ	$ Aut(G_i) $	$ Aut(G) $
12	12	8	2	$m+4$	48	$n+32$

Ранее конструктивно перечисленные представители (по каталогу [3]): 12-4-1, 20-4-16, 28-4-21
 Диаграммы канонической прорисовки:



На синтезированных семействах (в среднем до 150 вершин) проведены объёмные вычислительные эксперименты по уточнению строения групп автоморфизмов, анализу структурной сложности и структурной надёжности ТГС4. Всего потрачено более 30 часов машинного времени на первичные вычисления характеристик графов и более 8 суток на анализ результатов. Вычислительные эксперименты подтвердили *корректность* (принадлежность ТГС4), *безизбыточность* (отсутствие ТГС4, принадлежащих нескольким семействам) и *полноту* (покрытие всех известных ТГС4 до 30 вершин включительно) семейств ТГС4, порождаемых ПК *TransGen*.

4.3. Характеристики используемых алгоритмов

Методы конструктивного перечисления ТГ, предложенные в работах *Н.Р. Yap, G.F. Praeger, C.E. Royle, B.D. McKay* [4-6] и других исследователей, связаны с вычислением характеристик ГАГ. Возникающие при этом задачи для графов общего вида принадлежат либо к классу *NPI*, либо *NP*-полны. Например, задача определения орбит эквивалентна по сложности задаче распознавания изоморфизма графа и в общем виде (для произвольного графа) остаётся *NP*-сложной. Для графов с ограниченными степенями вершин Лакс Ю.М. ещё в 1985 году доказал возможность построения полиномиального алгоритма, но реализации практически полезного алгоритма с гарантированно полиномиальным временем работы авторам не из-

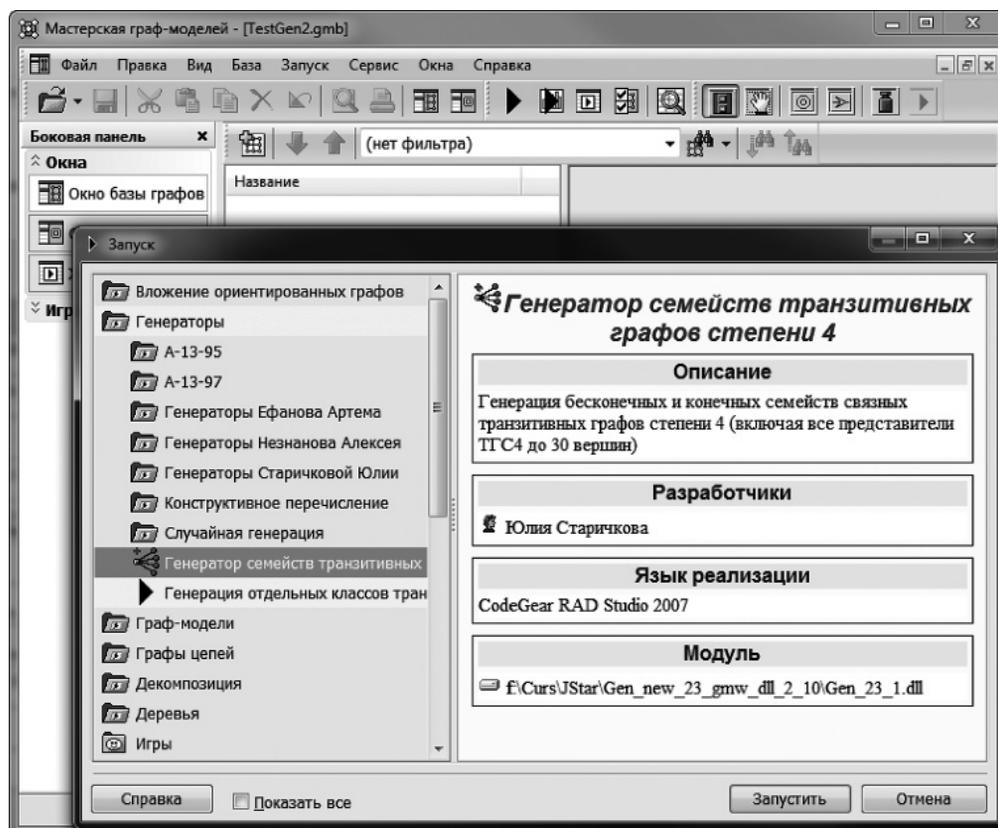


Рис. 3. Запуск TransGen в среде АСНИ Graph Model Workshop.

вестно. Однако для обработки ГАГ относительно небольших графов существуют хотя и переборные, но эффективные с практической точки зрения алгоритмы. Важнейшей задачей является задача построения порождающего множества (ПМ) ГАГ, поскольку, зная некоторое порождающее подмножество группы, можно определять характеристики этой группы, не перечисляя все её элементы. В.Д. McKay в 1981 году создал алгоритм построения минимального ПМ ГАГ, послуживший основой библиотеки *Nauty* [9], в которой реализованы алгоритмы построения ПМ ГАГ, поиска орбит, канонизации графа и решения смежных проблем. *Nauty* стала первым в целой серии реализаций, либо более специализированных, либо использующих дополнительные изощрённые техники для обработки «худших случаев».

На этапе уточнения классификации и верификации семейств, а также исследования дополнительных свойств семейств, использовались авторские алгоритмы анализа свойств ГАГ, в основу которых положен вариант построения ПМ ГАГ, предложенный в проекте *Bliss* [10], который, в свою очередь, основан на *Nauty*.

Однако, при синтезе используются специальные шаблоны генерации рёбер (вариантов топологий) и координат вершин (вариантов прорисовок), которые образуют каталог, обращение к которому с передачей параметров генерации приводит к построению ТГС4 с заданным числом вершин за время, *прямо пропорциональное числу генерируемых рёбер*. А поскольку число рёбер в регулярном графе степени 4 равно $2p$, где p – число вершин, то алгоритм синтеза представителя семейства имеет линейную асимптотическую оценку временной вычислительной сложности относительно его числа вершин.

Таким образом, почти *два года* машинного времени, затраченного при разработке ПК на проведение вычислительных экспериментов, привели к тому, что в результате *TransGen* работает настолько быстро, что любые дополнительные оптимизации процесса генерации бессмысленны.

4.4. Интерфейс с пользователем и проведение вычислительных экспериментов

Возможно несколько вариантов сценария использования ПК *TransGen*. Основной сценарий – генерация заданного числа ТГС4 выбранных

семейств, начиная с первого или первого уникального представителя. Выбор семейств производится как прямой пометкой семейств, так и применением фильтров по заданным характеристикам $|Aut(G)|$, ψ и χ (см. далее). Однако существуют и дополнительные сценарии использования: 1) конструктивное перечисление всех ТГС4 до 30 вершин включительно, что очень полезно при решении теоретических задач; 2) генерация графов в заданном диапазоне числа вершин, что обычно требуется на практике при синтезе топологий реальных систем.

При любом сценарии возможно задание дополнительных параметров.

1. Указание количества графов.
2. Пропуск представителей. Например, можно генерировать каждый 4 граф в семействе.
3. Выбор либо канонической, либо одной из других полностью симметричных прорисовок диаграмм для данного семейства (семейство может предлагать десятки вариантов прорисовки).

Запуск расширения осуществляется через меню «Запуск» АСНИ *GMW* или в диалоге интерактивного запуска расширений (рис. 3). Отметим, что генератор бессмысленно запускать в пакетном режиме, зато можно запускать, не открыв предварительно базу графов, так как генератор умеет создавать новые базы графов с выбранным стилем и параметрами.

Расширение обладает развитым интерфейсом с пользователем, который строится вокруг списка семейств. Главное окно генератора показано на рис. 4. Непосредственно в списке отображаются три первых представителя с выбранным вариантом прорисовки (по умолчанию используется каноническая прорисовка по каталогу [3]) и комментарии, которые можно изменять. Через контекстное меню

можно получить информацию о характеристиках семейств. Особый интерес представляет возможность фильтрации семейств. В текущей версии возможна фильтрация по характеристикам симметрии. Поиск семейств по характеристикам и применение фильтров производится в отдельном окне, которое показано на рис. 5. После применения фильтра изменяется состав помеченных семейств в списке.

Приведем сценарий использования ПК (как расширения в составе АСНИ *GMW*) для синтеза высоко-симметричных топологий. Допустим, мы хотим синтезировать топологию со сложным строением группы автоморфизмов максимального порядка. Задав фильтр на порядок группы, обнаруживаем, что искомым семейством будет семейство *G_25*. Применяв фильтр, зададим необходимые параметры генерации представителей семейства: 50 графов для генерации, начиная с первого представителя, с канонической прорисовкой и созданием новой базы со стилем визуализации, диктуемым предметной областью. Результатом работы ПК *TransGen* будет база структур с соответствующими параметрами (рис. 6).

5. Исследование характеристик семейств ТГС4

Приведем примеры применения разработанной системы. Первая из трудоемких задач – построение индексов структурной сложности для семейств ТГС4 и установление связи между характеристиками симметрии и сложности. Суть вычислительного эксперимента заключалась в определении корреляций значений индексов структурной спектральной сложности [11] в различных базах для сгенерированных представителей семейств ТГС4 с числом вершин до 120. В качестве базисов структурных

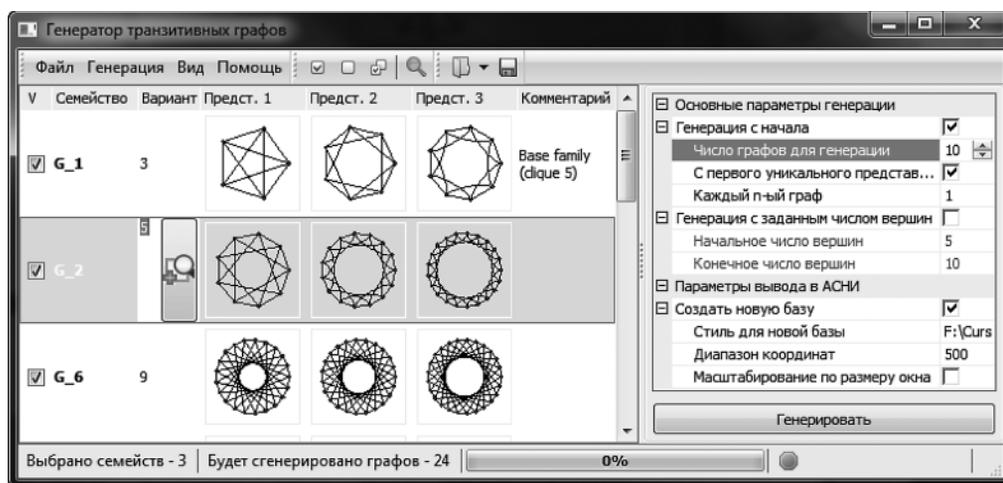


Рис. 4. Интерфейс подсистемы *TransGen*.

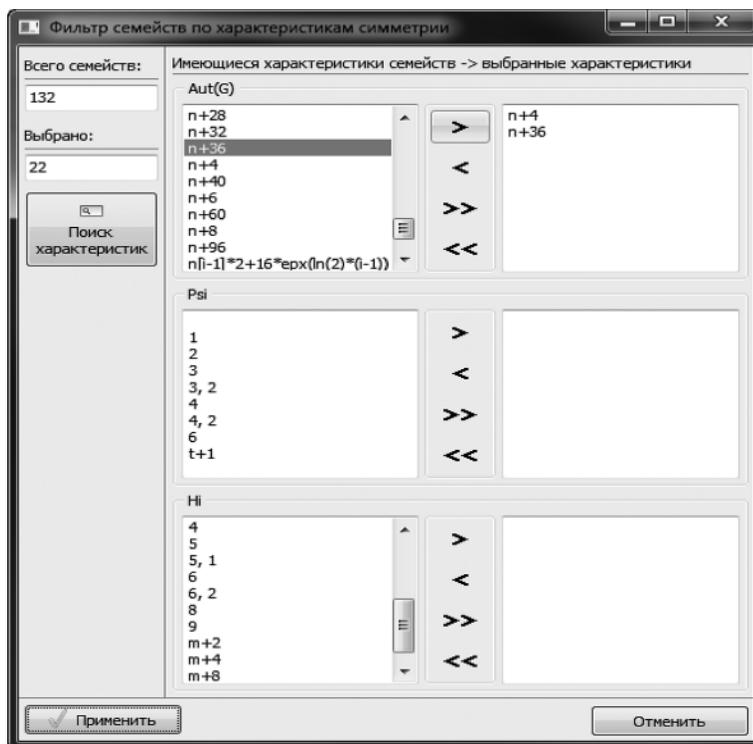


Рис. 5. Диалог выбора семейств по базовым характеристикам симметрии.

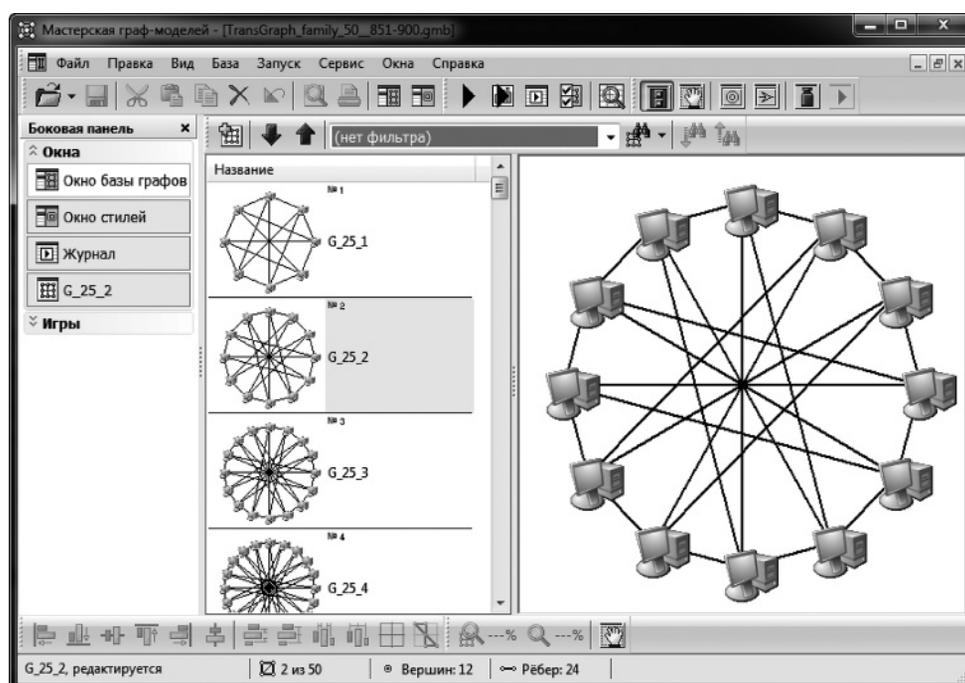


Рис. 6. Пример результата работы подсистемы TransGen (50 первых представителей семейства G_25).

дескрипторов использовались базы цепей, циклов и деревьев с ограничением на размер (число вершин). Таким образом, проведено по 20 вычислительных экспериментов для 59 бесконечных семейств ТГС4, всего обработано 2950 графов. В результате построены графики зависимости индексов сложности в различных базах от числа вершин представителя семейства ТГС4, выявлены семей-

ства граничные по значению индексов сложности. Пример сравнения индексов сложности в базе деревьев приведен на рис. 7. Проведено сравнение характеристик симметрии и значений индексов сложности для семейств ТГС4. Оказалось, что спектральная сложность семейств в разных базах совершенно по-разному коррелирует с интегральными характеристиками симметрии, что очень важно

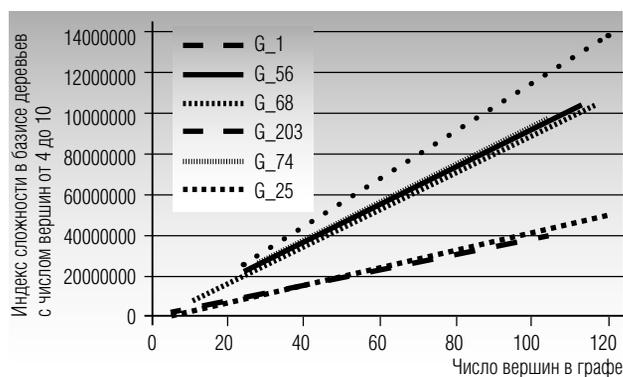


Рис. 7. Графики значений индекса спектральной сложности в базисе всех деревьев с числом вершин от 4 до 10 для нескольких показательных семейств.

при выборе топологий реальных систем с заданным уровнем надёжности и минимизацией сложности.

Второй из задач, для которой использовались порожденные семейства, являлась задача повышения эффективности базовых алгоритмов структурного анализа на высоко-симметричных графах [12].

6. Заключение

Описана цель создания, история развития и функциональность программного комплекса «Полный генератор семейств транзитивных графов степени 4». Он расширяет функциональность АСНИ «Graph

Model Workshop» и активно используется при проведении теоретических исследований в области теории графов и теории групп, а также при решении прикладных задач, связанных с синтезом топологий высокопроизводительных вычислительных систем и систем связи. Так, были синтезированы логические топологии для эффективной реализации параллельных алгоритмов на системах с различным числом узлов, причём топологии содержали заданные типы циклических фрагментов. Различные варианты симметричных топологий были применены при решении задач оптимизации системы управления и документооборота в комитете финансов Новгородской области и филиалах. Однако структуры степени 4 оказались избыточны и в итоге были внедрены системы с кубической топологией, показавшие лучшее соотношение стоимость/надёжность.

Уже начата работа по созданию подсистемы синтеза транзитивных топологий степени 5. Для ТГС4 по просьбам пользователей будет добавлена база характеристик структурной надёжности. Также планируется включение в генератор фильтрации по базовым моделям структурной сложности (как минимум – в базисах цепей и циклов), поскольку оказалось, что число небольших фрагментов заданных типов (а особенно – их отсутствие) является практически значимой характеристикой семейства. ■

Литература

1. Берж К. Теория графов и ее применение. – М.: Издательство иностранной литературы, 1962. – 320 с.
2. Харари Ф. Теория графов. – М.: Мир, 1973. – 303 с.
3. Незнанов А.А., Кохов В.А. Справочник по теории графов. Характеристики симметрии и сложности связанных транзитивных графов степени 4 с числом вершин до 30 включительно. – М. Деп. в ВИНТИ, №1094-В2004, 2004. – 418 с.
4. Yarp H.P. Point symmetric graphs with $p < 13$ points, Nanta Math. 6 (1973), PP. 8-20.
5. Royle G.F., Praeger C.E. Constructing the vertex-transitive graphs of order 24, Journal of Symbolic Computation, Volume 8, Issue 4, 1989, PP. 309-326.
6. McKay B.D., Royle G.F. The transitive graphs with at most 26 vertices, ArsCombinatoria, 30 (1990), PP. 161-176.
7. Кохов В.А. Диаграммы, числа стабильности и цикловые индексы групп автоморфизмов транзитивных графов // Исследования по прикладной теории графов. – Новосибирск: Наука. Сиб. отд., 1986. – С. 97-125.
8. Старичкова Ю.В., Незнанов А.А. Улучшенная классификация и генерация транзитивных графов степени 4. // Доклады международной конференции «Информационные средства и технологии». (МФИ-2008), Т.2, Москва, 2008. – С. 75-79.
9. McKay B.D. The nauty page (<http://cs.anu.edu.au/~bdm/>)
10. Junttila T., Kaski P., bliss: A Tool for Computing Automorphism Groups and Canonical Labelings of Graphs (<http://www.tcs.hut.fi/Software/bliss/>)
11. Кохов В.А. Концептуальные и математические модели сложности графов. – М.: Издательство МЭИ, 2002. – 157 с.
12. Незнанов А.А., Кохов В.А. О связи строения стационарных подгрупп группы графа и эффективности учёта симметрии при решении переборных задач структурного анализа // Труды Московского физико-технического института. 2009. Т. 1. № 2. – С. 77-83.

МЕТОД ОЦЕНКИ ЗРЕЛОСТИ ОРГАНИЗАЦИИ С ИСПОЛЬЗОВАНИЕМ МОДЕЛИ ВРММ И ЕГО РЕАЛИЗАЦИЯ

А.Н. Бирюков,

кандидат физико-математических наук, доцент кафедры
корпоративных информационных систем Национального исследовательского
университета «Высшая школа экономики»

Д.Ю. Клецких,

выпускник образовательной программы «Второе высшее образование»
Высшей школы бизнес-информатики Национального исследовательского
университета «Высшая школа экономики»

Адрес: г. Москва, ул. Кирпичная, д. 33/5

E-mail: abiryukov@hse.ru, oraworker@yandex.ru

В статье рассматриваются вопросы оценки зрелости бизнес-процессов организации. Описывается метод оценки зрелости организации с иерархической структурой, основанный на известной модели ВРММ. Разработанный метод, позволяет организации самостоятельно количественно оценить зрелость бизнес-процессов. Представлен подход к автоматизации метода.

Ключевые слова: управление бизнес-процессами, уровень зрелости организации, модель зрелости ВРММ, SLA, оценка зрелости бизнес-процессов.

Введение

Согласно существующим на данный момент подходам, наиболее адекватное понимание того, как работает и насколько эффективен бизнес, возникает в результате анализа его бизнес-процессов [1]. На сегодняшний день разработан обширный инструментарий для описания и анализа бизнес-процессов. Это и средства моделирования,

и методические справочники, и отраслевые эталонные модели, существующие, главным образом, в форме стандартов, разрабатываемых международной организацией по стандартизации (ISO) или профессиональными организациями. Теоретически понятно, что эти модели могут служить ориентирами при поиске ответа на вопрос, насколько управление организацией соответствует лучшим практикам. В реальности, однако, все гораздо сложнее.

Например, как оценить близость бизнес-процессов организации к эталону, если часть работ некоторого процесса выполняется, а часть – нет? Что вообще однозначно свидетельствует о выполнении работы: документ, устное заявление исполнителя, негласное соглашение между исполнителями или что-то еще? Как быть, если сам процесс иногда выполняется для решения определенной задачи, а иногда нет?

Деятельность по оценке близости бизнес-процессов организации к эталону называют оценкой зрелости организации, и для этой деятельности существует целый ряд методик. Обычно методика включает два компонента: эталонную модель процессов и метод оценки зрелости в соответствии с этой моделью. Наиболее известными примерами таких методик служат СММ (называемая также СММ for Software) [2], группа методик CMMI (CMMI for Acquisition, CMMI for Services и др.) [3], методику SPICE [4], превратившуюся впоследствии в стандарт ISO/IEC 15504 (русский аналог – ГОСТ Р ИСО/МЭК 15504 [9]), методику SCAMPI (Standard CMMI Appraisal Method for Process Improvement) [5], OPM3 (Organizational Project Management Maturity Model [6] – методика оценки зрелости процессов корпоративного управления проектами на основе РМВоК [7]). Общим для всех этих методик является то, что они предназначены для оценки только проектных организаций, т.е. организаций, бизнес которых по преимуществу имеет проектный характер.

Относительно недавно появилась эталонная модель, демонстрирующая попытку избавиться от ограничения, связанного с проектной природой бизнеса организации. Эта модель была предложена Object Management Group (OMG) и называется ВРММ (Business Process Maturity Model) [8]. Однако и она обладает целым рядом ограничений. Во-первых, ВРММ описывает только двухуровневые организации, т.е. организации, состоящие из структурных подразделений, находящихся на одном и том же уровне иерархии. Во-вторых, для нее отсутствует описание метода оценки зрелости. Таким образом, для практического применения ВРММ необходимо прежде всего разработать такой метод. Традиционно оценка зрелости – очень трудоемкая работа, выполняемая специально обученными аудиторами (см. напр. [10]). Это приводит к тому, что на практике точный метод оценки зрелости применяется крайне редко. Неформальные же оценки страдают неточностью и субъективностью. В то же время, потребность в таких оценках очень высока [10]. Это важно при выборе субподрядчиков

и поставщиков, анализе эффективности организации, оценке и реинжиниринге бизнес-процессов.

В настоящей статье предлагается формальный метод самостоятельной оценки зрелости процессов организацией в ходе своей повседневной деятельности без привлечения аудиторов и консультантов. Метод основан на модели процессов ВРММ. Описан подход к автоматизации метода.

Модель ВРММ и ее свойства

Полное описание ВРММ можно найти в [8]. Здесь будут изложены только те ее свойства, которые необходимы для построения метода оценки. Согласно ВРММ, в структурных подразделениях организации реализуются конкретные экземпляры исполнения процессов, причем происходит это на периодической основе. Процессы представлены так называемыми процессными областями, состоящими из отдельных работ, называемых специфическими практиками. Реальный бизнес-процесс состоит в выполнении некоторого набора практик. Ни конкретный набор процессов, ни точный порядок их выполнения (сценарий процесса), ни даже наборы практик, входящие в процесс, не определены. Это вполне соответствует сложившемуся подходу к описанию моделей процессов в методиках оценки зрелости.

Как уже говорилось, ВРММ исходит из упрощенного представления об организации как о двухуровневой структуре, включающей высшее руководство и структурные подразделения. Подразделение представляет собой неделимую сущность, структура у которой отсутствует. Модель устанавливает пять уровней зрелости процессов организации, различающихся присущими им наборами процессных областей (это похоже на группы ключевых процессов в СММ). У каждой процессной области есть свой набор специфических целей и множество составляющих её специфических практик.

У каждой процессной области А имеется ровно одна институциональная цель и выглядит она всегда одинаково: «Специфические практики процессной области А внедрены (в терминологии ВРММ – институционализированы – Авт.)». Это означает, что все бизнес-процессы, которые составляются из специфических практик, должны удовлетворять пяти условиям:

1. описания бизнес-процессов процессной области [имя процессной области] документированы;
2. все работы бизнес-процессов процессной области [имя процессной области] спланированы;

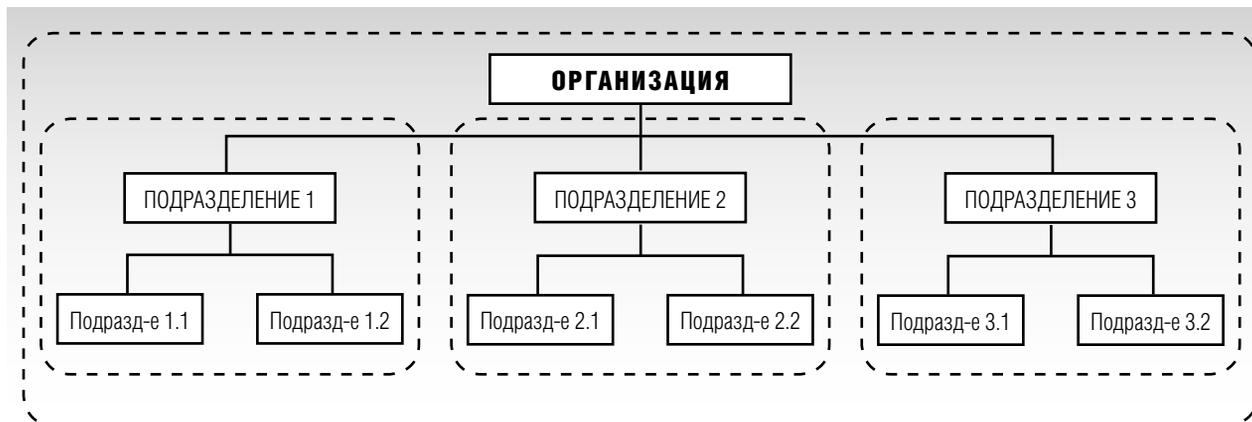


Рис. 1. Вложенные организации.

3. участники бизнес-процессов процессной области [имя процессной области] обладают необходимыми знаниями и навыками;

4. производительность, работы, текущее состояние и результаты бизнес-процессов процессной области [имя процессной области] контролируются и измеряются; при необходимости выполняются корректирующие действия;

5. специфические практики процессной области [имя процессной области] объективно проверяются на соответствие законам, требованиям регуляторов, стандартам, организационным политикам, бизнес-правилам, описаниям бизнес-процессов, рабочим процедурам; все несоответствия надлежащим образом обрабатываются.

Эти пять условий называются в ВРММ институциональными практиками. То, насколько реализована процессная область, определяется двумя факторами. Во-первых, тем, какие из составляющих ее специфических практик реально выполняются в организации (другими словами, насколько достигаются специфические цели области). Во-вторых, тем, в каком объеме выполняются для неё институциональные практики. Определить это – задача экспертов. Соответствующей методики, как уже говорилось, в модели ВРММ нет. После того, как будет установлено, насколько реализованы процессные области, можно будет, зная, какой набор областей достаточен для достижения определенного уровня зрелости, построить профиль организации или определить ее уровень зрелости.

Понятно, что профиль зрелости организации может выглядеть довольно сложно, если в ней одновременно (отчасти) внедрены специфические практики процессных областей, относящихся к разным уровням зрелости.

Расширим теперь представление об организации, лежащее в основе модели ВРММ. Для этого рассмотрим пример организации, показанный на рис. 1.

В показанном примере присутствует четыре «организации» – собственно организация и три вложенные структуры, также обладающие, с точки зрения ВРММ, всеми признаками организации. Действительно, у них есть структурные подразделения (пронумерованные с 1.1 до 3.2), своё руководство (руководство подразделений 1-3), свои бизнес-цели (декомпозированные из бизнес-целей организации) и т.п. Таким образом, имеет смысл говорить о зрелости бизнес-процессов вложенных организаций.

Возникает естественный вопрос: связана ли зрелость процессов организации верхнего уровня со зрелостью процессов вложенных организаций? Рассмотрим, например, частный случай, когда все вложенные организации (подразделения 1-3 на рис. 1) находятся на втором уровне зрелости. На каком уровне будет находиться организация верхнего уровня?

С точки зрения ВРММ тот факт, что Подразделение N находится на втором уровне зрелости означает, что в нём присутствуют и полностью внедрены специфические практики следующих процессных областей (в скобках указаны английские аббревиатуры названий).

- ◆ Ответственность руководства за улучшение процессов организации (OPL).
- ◆ Ответственность руководства за управление организацией и эффективность её работы (OBG).
- ◆ Управление требованиями к подразделениям (WURM).
- ◆ Планирование работ и распределение ответственности между подразделениями (WUPC).

- ◆ Контроль за работами в подразделениях (WUMC).
- ◆ Распределение и выполнение работ в подразделениях (WUP).
- ◆ Управление конфигурациями в подразделениях (WUCM).
- ◆ Управление сорсингом на уровне организации (SM).
- ◆ Соответствие процессов и продуктов внешним требованиям (PPA).

Чтобы вся организация находилась на втором уровне зрелости, необходимо, чтобы такие же специфические практики существовали и были полностью внедрены на уровне организации в целом.

Таблица 1.

Взаимное влияние специфических практик организации верхнего уровня и вложенной организации

Процессные области подразделения	Процессные области подразделения									
	OPL	OBG	WURM	WUPC	WUMC	WUP	WUCM	SM	PPA	
OPL	- >	-	-	-	-	-	-	-	-	-
OBG	-	-	-	-	-	-	-	-	-	-
WURM	-	- >	-	-	-	-	-	-	-	-
WUPC	-	-	-	- >	-	-	-	-	-	-
WUMC	-	-	-	-	- >	-	-	- >	-	-
WUP	-	-	- !	-	- !	- !	-	-	-	-
WUCM	-	-	-	-	-	-	-	-	-	-
SM	-	-	-	-	-	-	-	-	-	-
PPA	-	-	-	-	-	-	-	-	-	-

Зададимся вопросом: существует ли какая-либо корреляция между уровнями зрелости организации верхнего уровня и вложенных организаций? Если, например, вышестоящая организация находится на втором уровне зрелости, следует ли отсюда, что нижестоящие организации также находятся на втором уровне, т.е. соответствующие специфические практики для них присутствуют и внедрены? Верно ли обратное утверждение? Результаты сведены в *таблицу 1*. Клетки таблицы имеют следующий смысл:

◆ в клетке имеется две позиции – верхняя (влияние специфических практик процессной области вложенной организации на практики процессной

области организации верхнего уровня) и нижняя (обратное влияние);

◆ в каждой из позиций может стоять либо «-» - влияния нет, либо стрелка (указывающая направление влияния), либо «!» – соответствующие практики пересекаются (см. ниже).

Например, стрелка вниз в клетке WUMC-SM означает, что среди специфических практик процессной области WUMC организации верхнего уровня есть, как минимум, одна (это практика SP3 «Ресурсы, отличные от человеческих, и необходимые для работы подразделения приобретаются и разворачиваются в соответствии с планами и потребностями подразделения»), которая заставляет выполняться, как минимум, одну практику процессной области SM вложенного подразделения (SP1 «Описания потребностей подразделения в сорсинге устанавливаются и поддерживаются»). Другими словами, зрелость организации верхнего уровня в этом случае стимулирует появление соответствующих практик во вложенной организации.

Ситуация, которая привела к появлению восклицательного знака в клетке WUP-WURM, состоит в следующем. В процессной области WUP организации верхнего уровня существует практика SP2 «Определены и согласованы заинтересованными лицами зависимости и взаимодействия между индивидуумами и рабочими группами, необходимые для выполнения работы». В процессной области WURM вложенной организации существует практика SP5 «Утверждена, согласована с подразделением и сопровождается спецификация требований к работе подразделения». Поскольку для организации верхнего уровня «индивидуумы и рабочие группы» соответствуют подразделениям вложенной организации, указанные практики во многом пересекаются. Впрочем, из таблицы видно, что таких случаев, как и вообще зависимостей между уровнями зрелости вложенных организаций очень мало.

Аналогичный вывод имеет место и для других сочетаний уровней зрелости вложенных организаций. Строго говоря, проверка гипотезы о том, что «все вложенные организации полностью независимы в смысле их профилей зрелости» - отдельная тема, выходящая за рамки настоящей статьи. Однако, как видно из *табл. 1*, такое допущение достаточно реалистично. Мы дальше будем считать, что корреляция по зрелости между организацией верхнего уровня и вложенными организациями отсутствует. Это означает, что при оценке зрелости все вложенные организации можно оценивать независимо.

Услуги подразделений

Для того чтобы построить метод оценки зрелости иерархической организации, будем считать, что её подразделения оказывают друг другу некоторые услуги, т.е., по существу, являются сервисными организациями. Наиболее известная реализация такого подхода описана в библиотеке ITIL (IT Infrastructure Library) [4], где в качестве сервисного подразделения рассматривается ИТ-служба. Опыт внедрения процессов ITIL показал, однако, что сервисная парадигма оказывается полезной при анализе и реализации не только ИТ-услуг, но и услуг энерго- и теплоснабжения, автотранспортных услуг и ряда других, включая кадровые и финансовые услуги.

На рис. 2 показан пример организации, где подразделения оказывают услуги друг другу. Это распространённая практика: служба HR обеспечивает приём на работу и другие кадровые услуги для всех подразделений организации, финансовая служба участвует в подготовке бюджетов других подразделений, наконец, ИТ-служба оказывает услуги по автоматизации бизнес-процессов подразделений. Образуется сложная сеть взаимного влияния подразделений друг на друга. Взгляд на организацию как на совокупность (вложенных) подразделений, оказывающих услуги друг другу, позволяет учесть в оценке зрелости процессов такой важный фактор, как качество этих услуг. Естественно предположить, что чем выше качество услуги, тем на более высоком уровне зрелости находится оказывающее её подразделение и наоборот.

Стрелками на рис. 2 показаны услуги подразделений. В терминологии модели VRMM продукты и услуги производит только организация в целом, а подразделения лишь генерируют результаты своей работы. В нашем случае, поскольку каждое подраз-

деление можно рассматривать, как самостоятельную организацию, будем говорить о продуктах и услугах применительно также и к подразделению.

В качестве примера рассмотрим оказание услуг ИТ-службой. В соответствии с современными взглядами будем считать, что услуга ИТ-службы – автоматизированное или автоматическое выполнение бизнес-процесса или его части, включая отдельные работы. Такие услуги, как обеспечение доступа к ИТ-ресурсам, например, предоставление выхода в интернет или обеспечение защиты персональных данных пользователей с этой точки зрения представляют собой лишь автоматизацию решения отдельных задач в ходе выполнения сложного бизнес-процесса.

В терминологии VRMM услуги ИТ-службы сводятся к:

1. (частичной) автоматизации и сопровождению выполнения специфических практик;
2. созданию новых специфических практик, отсутствовавших ранее (например, практики по контролю и измерению эффективности работ и соответствия их планам);
3. институционализации процессной области за счёт:
 - ◆ создания и сопровождения достоверного и актуального электронного описания бизнес-процессов в информационной системе;
 - ◆ (частичной) автоматизации выполнения бизнес-процессов, т.е. обеспечению того, что выполнение процессов (частично) спланировано, а их эффективность измеряется и контролируется.

Более простые составляющие услуги оказываются подразделениями ИТ-службы, например, подразделение, занимающееся сопровождением, может оказывать услуги по:

- ◆ сопровождению выполнения специфических практик;

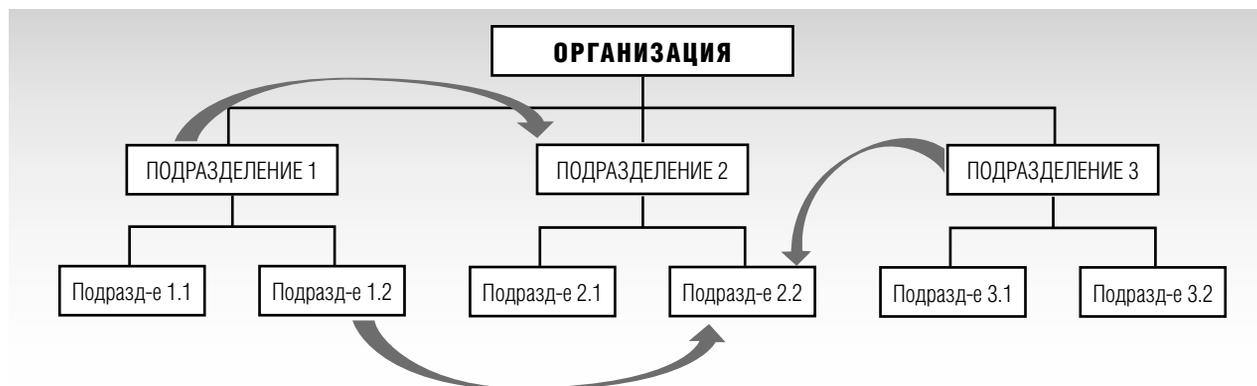


Рис. 2. Услуги подразделений друг другу.

✧ сопровождению и актуализации описаний бизнес-процессов;

✧ сопровождению ИС, обеспечивающих измерение и контроль эффективности бизнес-процессов.

Очевидно, состав услуг, оказываемых ИТ-службой и её подразделениями, будет разным для разных подразделений-клиентов. Другими словами, должны существовать свои SLA (Service Level Agreements) [4, 10] для каждой пары «подразделение-провайдер услуг и подразделение-клиент».

Услуги ИТ-службы и оценка зрелости подразделений-клиентов

Из сказанного следует, что услуги ИТ-службы не только характеризуют её зрелость, но и способствуют повышению зрелости подразделений-клиентов. Действительно, автоматизация выполнения специфических практик есть не что иное, как способ повышения эффективности бизнес-процессов процессной области подразделения (институциональная практика InP4), а автоматизация бизнес-процессов стабилизирует их выполнение и гарантирует повторяемость. Строго говоря, между оценками вклада ИТ-службы в выполнение специфических практик и в институционализацию про-

цессной области нет прямой связи. Например, все практики могут быть максимально автоматизированными, но описание бизнес-процесса может отсутствовать (островная автоматизация). Более того, сам бизнес-процесс может быть неспланированным (не выполняется InP2), а содержание отдельных специфических практик может противоречить внешним требованиям (не выполняется InP5).

Это означает, что комплексная оценка вклада ИТ-службы в зрелость подразделения-клиента — сложная и трудноформализуемая задача. Для решения её необходимо привлекать экспертов из числа участников и владельцев бизнес-процессов.

Эксперты должны ответить на следующие вопросы:

- как оценить вклад ИТ в выполнение конкретной специфической практики?
- как оценить вклад услуги ИТ-службы в институционализацию процессной области?

Пример инструмента, который позволяет ответить на первый вопрос, приведен в *таблице 2*, определяющей максимальные вклады услуг ИТ-службы в выполнение специфических практик некоторой области (выбрана область WUMC) некоторого подразделения. Реальные вклады не могут превышать указанных границ.

Таблица 2.

Пример оценок максимального вклада ИТ-службы в выполнение специфических практик

Практика	Макс вклад ИТ в выполнение практик
SP 1 Запросы на работу, получаемые подразделением, анализируются, получают приоритеты и включаются в план работы подразделения	0.2
SP 2 Люди необходимые для выполнения работы подразделения, привлекаются, готовятся и получают задание в соответствии с планами и потребностями	0.2
SP 3 Ресурсы, отличные от человеческих, необходимые для работы подразделения приобретаются и разворачиваются с планами и потребностями подразделения	0.5
SP 4 Устанавливается и поддерживается распределение работ по исполнителям и рабочим группам подразделения	0.2
SP 5 Поток работ в подразделении контролируется на соответствие его описанию	1.0
SP 6 Работа, порученная исполнителям и рабочим группам подразделения регулярно контролируется и корректируется	0.2
SP 7 Для управления работой используются показатели, определенные планом	0.2
SP 8 Производительность и состояние работ, результатов и услуг, необходимых подразделению, регулярно сравниваются с требованиями, планами и обязательствами	0.5
SP 9 Критические данные и информация, необходимая для подразделения, собирается, запоминается и используется соответственно планам управления данными	1.0
SP 10 Осуществляется управление рисками для подразделения	0.1
SP 11 Значительные отклонения от требований подразделения, его оценок, планов и обязательств выявляются и обрабатываются	0.1
SP 12 Вероятные причины перечисленных в SP 11 отклонений выявляются и обрабатываются	0
SP 13 Состояние подразделения, его достижения, проблемы и риски анализируются заинтересованными лицами	0.1
SP 14 Планы подразделения пересматриваются так, чтобы отражать его состояние, изменения рисков, обязательств и параметров планирования	0
SP 15 Опыт, накопленный в ходе выполнения работ в подразделении, фиксируется и используется при разработке требований, оценок, планов и обязательств	0.3

Приведем пример расчёта оценки существования некоторой специфической практики. Рассмотрим практику SP8 «Производительность и состояние работ, результатов и услуг, необходимых подразделению, регулярно сравниваются с требованиями, планами и обязательствами». Это практика процессной области WUMC некоторого подразделения. Потенциальный вклад ИТ-службы в выполнение практики (согласованная услуга ИТ-службы) состоит в том, что с помощью некоторой информационной системы регулярно готовятся все необходимые отчёты. Если это так, то в соответствии с таблицей, выполнение практики обеспечено на 50%. На оставшиеся 50% выполнение практики от ИТ-службы не зависит, например, отчёты могут вообще не использоваться для сравнения актуальных и плановых показателей. Допустим, эксперты приходят к выводу, что в реальности анализ производительности и состояния работ в среднем регулярно сравниваются только с половиной (50%) планов, а ИТ-служба реализовала лишь 30% задуманных отчётов. Значит, общая оценка выполнимости практики – 15%. Это среднее число случаев, когда сравнение опирается на достоверную информацию. В остальных случаях, либо сравнение вообще не выполняется, либо оно производится «на глаз», без использования отчётной информации.

Конкретная согласованная методика оценки вклада ИТ-службы должна быть частью соответствующего SLA. Таким образом, можно построить профиль выполнимости практик каждой процессной области подразделения.

Для того, чтобы полностью ответить на вопрос о степени реализации процессной области, нужно оценить вклад ИТ-службы в стабильность, контролируемость и измеримость бизнес-процессов. Методика расчёта степени реализации процессной области также определяется в SLA. Например, каждый владелец бизнес-процесса может оценивать степень документированности плана процесса по тому, существует ли электронный или бумажный план процесса, насколько он актуален, как часто он обновляется и т.п., после чего оценки разных процессов усредняются и дают оценку для процессной области в целом. Ответ на вопрос, строить ли интегральную оценку, включающую профиль процессной области и оценки зрелости всех процессов, зависит от поставленных руководством целей.

Услуги подразделений и оценка зрелости ИТ-службы

Очевидно, способность ИТ-службы оказывать услуги зависит от её зрелости, и, в частности, от того, насколько повлияли на её зрелость услуги, оказываемые ей другими подразделениями.

Можно определить такие услуги подразделений ИТ-службе:

1. (частичное) выполнение специфических практик силами другого подразделения;
2. институционализация процессной области за счёт:
 - ◆ выкраивания бизнес-процессов ИТ-службы из стандартных бизнес-процессов другого подразделения;
 - ◆ (частичного) контроля за выполнением бизнес-процессов, измерения их эффективности и соответствия результатов внешним требованиям.

Примером такого рода услуг могут служить услуги HR-службы, которая берет на себя выполнение специфических работ по обучению персонала (часть выполнения практики SP2 процессной области WUMC). Особенно заметную роль такие услуги будут играть для процессной области «Организационное развитие компетенций», приписанной в ВРММ к третьему уровню зрелости. Подобно тому, как ИТ-служба своими ресурсами способствует созданию стабильных и эффективных бизнес-процессов в других подразделениях, эти подразделения, в свою очередь, используют свои ресурсы для повышения эффективности и стабильности бизнес-процессов ИТ-службы. Таким образом, возникают SLA между бизнес-подразделениями и ИТ-службой.

Реализация и внедрение метода

Суммируя вышесказанное, приходим к модели взаимодействия подразделений, показанной на *рис. 3*.

Эта модель определяет структуру данных, которую можно применить при реализации изложенного подхода. Следует добавить только, что наряду с оценками следует сохранять и момент их расчёта, чтобы обеспечить возможность временного анализа зрелости. Отдельный вопрос, требующий обсуждения, касается метода построения интегральных оценок процессной области. По существу, такие интегральные оценки носят характер упрощения при проведении исследования, поэтому вопрос их построения – это вопрос интерпретации результатов и полностью зависит от исследователя.

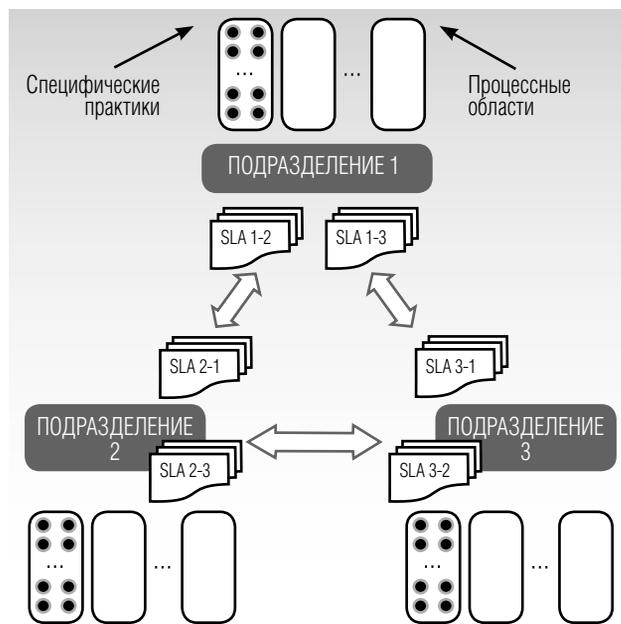


Рис. 3. Модель взаимодействия подразделений.

Изложенный метод, позволяет оценивать уровень зрелости бизнес-процессов в организации, а получаемый профиль развитости процессов управления в организации позволяет оценить сильные и слабые стороны управления.

Экспериментальные исследования показали, что взаимодействие подразделений с помощью SLA встречается на практике. Были изучены две компании – крупный ритейлер и розничный оператор связи. В ходе изучения опыта работы этих компаний оказалось возможным обобщить практику взаимодействий подразделений с помощью SLA с целью построения модели зрелости и практической апробации изложенного метода оценки зрелости бизнес-процессов.

Таблица 3.

Объект	Атрибуты
Структурное подразделение	Название, место в иерархии, процессные области
Направленная связь объект-объект	SLA
Процессная область	Специфические практики, максимальная оценка вклада всех подразделений, реальная оценка вклада всех подразделений, итоговая оценка области
Специфическая практика	Название, максимальная оценка вклада всех подразделений, реальная оценка вклада всех подразделений, итоговая оценка выполнимости
Макс оценка вклада подразделения	Число
Реальная оценка вклада	Число, время
Итоговая оценка	Число, время
SLA	Перечень услуг, методика расчёта реальных оценок

В табл. 3 определены следующие объекты и связи, необходимые для реализации модели в программной среде:

В основе архитектуры соответствующей информационной системы лежит хранилище данных, куда загружаются экспертные оценки, относящиеся к выполнению всей совокупности SLA, связывающих подразделения. Оценки в хранилище включают и время их возникновения. Аналитические средства, реализованные для хранилища, должны давать ответы на вопросы о том, какое подразделение в данный момент является наиболее зрелым, как оцениваются его услуги, каков вклад в его зрелость других подразделений и т.п. Хранилище данных может также предоставлять информацию для выявления узких мест бизнес-процессов и анализа необходимости и целесообразности их автоматизации. ■

Литература

1. Бирюков А.Н. Лекции о процессах управления информационными технологиями: учебное пособие. – М.: Интернет-университет информационных технологий: БИНОМ. Лаборатория знаний, 2010. – 215 с.
2. Paulk M.C., Curtis B. et al. Capability Maturity Model for Software (SW-CMM), version 1.1, CMU Software Engineering Inst., 1993.
3. Capability Maturity Model Integration (CMMI), Version 1.1, CMU Software Engineering Inst., CMU/SEI-2002-TR-011 ESC-TR-2002-011.
4. SPICE Consolidated Product, parts 1-9, ISO/IEC Software Process Assessment, 1995, <http://www-sqi.cit.gu.edu.au/spice>
5. Iqbal M., Nieves M. ITIL Service Strategy. The Stationery Office. – 2007.
6. Organizational Project Management Maturity Model (OPM3), 2nd Ed., PMI.
7. A Guide to the Project Management Body of Knowledge (PMBOK Guide), 4th Ed, PMI.
8. Business Process Maturity Model (BPMM), version 1.0, <http://www.omg.org/spec/BPMM/1.0/pdf>
9. ГОСТ Р ИСО/МЭК 15504. – М.: Стандартинформ, 2010.
10. Брукс П. Метрики для управления ИТ-услугами / Пер. с англ. – М.: Альпина Бизнес Букс, 2008. – 283 с.

ТЕХНОЛОГИЯ КОНТРОЛЯ КАЧЕСТВА ОБСЛУЖИВАНИЯ ТРЕБОВАНИЙ В ОРГАНИЗАЦИОННЫХ СТРУКТУРАХ, ПРЕДОСТАВЛЯЮЩИХ УСЛУГИ МАССОВОГО ХАРАКТЕРА

А.П. Шабанов,

кандидат технических наук, главный эксперт, ООО «ИБС»

М.А. Аракелян,

заместитель директора департамента, ООО «ИБС»

Адрес: 127434, г. Москва, Дмитровское шоссе, д. 9Б

E-mail: ashabanov@ibs.ru, marakelian@ibs.ru

Рассматривается технология контроля качества обслуживания в информационных бизнес системах требований субъектов организационной структуры массового обслуживания. Технология охватывает стадии проектирования и применения информационных бизнес-систем. Областью применения технологии являются организационные структуры, в основе деятельности которых лежит предоставление повторяющихся услуг их потребителям.

Ключевые слова: организационная структура, информационная бизнес система, контроль, качество обслуживания.

1. Введение

Современный этап развития информационных технологий характеризуется широким внедрением на предприятиях процессов контроля качества обслуживания и средств их автоматизации [1-6]. Рассматриваемый в настоящей

работе автоматизированный процесс контроля обслуживания строится на основе методического аппарата, описанного в [7]. Наиболее близким, к описываемому здесь процессу, является учётно-контрольный процесс, в котором функции контроля реализуются параллельно с реализацией процесса обслуживания¹ [8, 9]. Такой подход по-

¹ Организационная структура массового обслуживания это организационная структура (предприятие, организация, подразделение, служба), в основе деятельности которой лежит предоставление повторяющихся услуг их потребителям.

зволяет получать статистические значения показателей контроля в реальном масштабе времени. В настоящей работе рассматривается технология контроля над качеством обслуживания в организационной структуре массового обслуживания (далее по тексту, организационной структуре), оснащённой прикладной информационной системой (далее по тексту, информационной бизнес-системой)².

Характерной чертой данной технологии является охват ею всего жизненного цикла организационной структуры. На *рис. 1* приведена схема автоматизированного процесса контроля обслуживания, которая в общем виде отображает технологию контроля качества услуг, оказываемых работниками организационной структуры их потребителям.

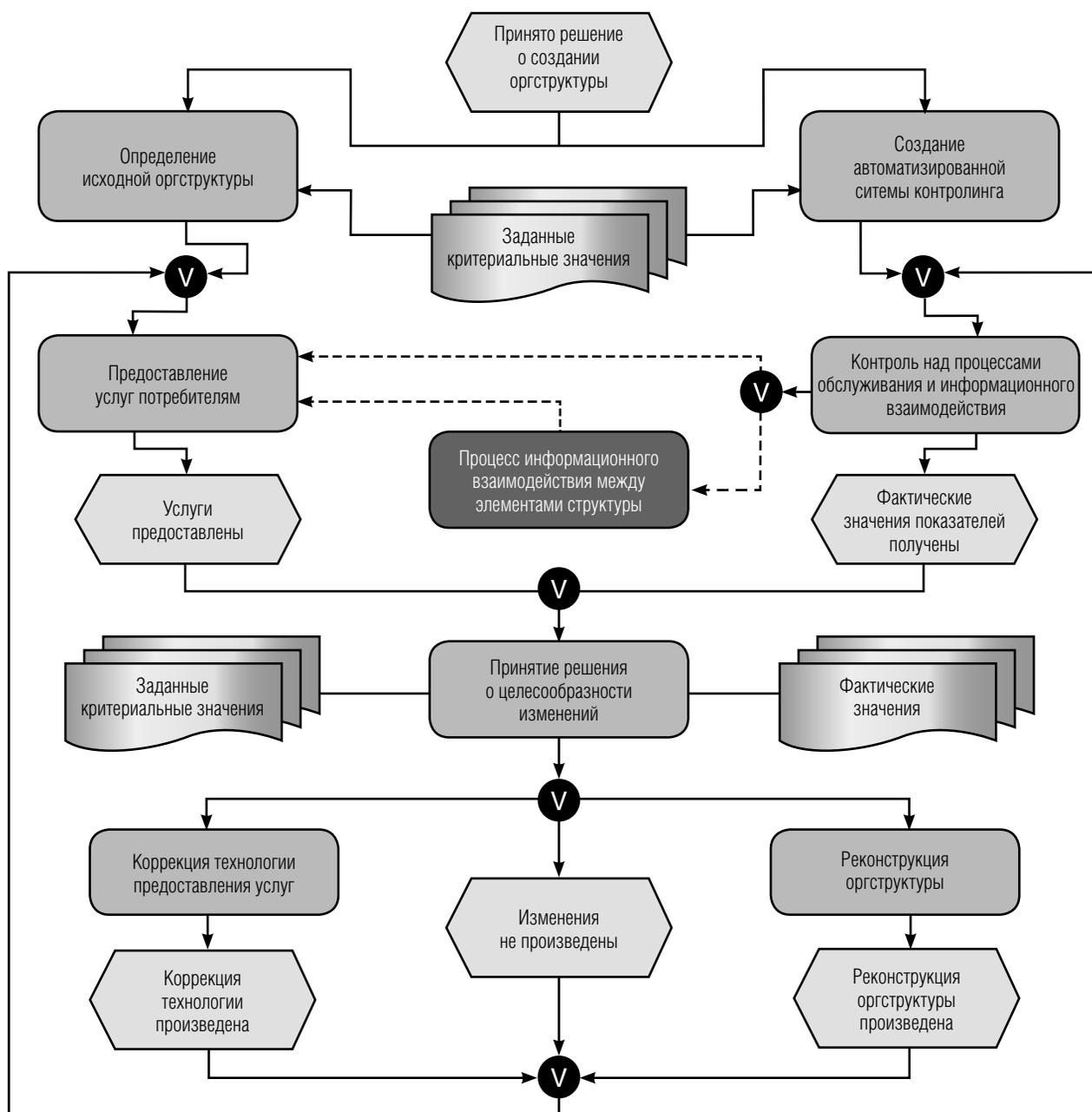


Рис. 1. Схема автоматизированного процесса контроля качества обслуживания.

² Прикладная информационная система (информационная бизнес-система) – составная часть информационной системы предприятия (министерства, ведомства, организации, подразделения, службы), предназначенная для поддержки одного или нескольких смежных направлений деятельности (направлений бизнеса); реализуется на основе специализированного прикладного программного обеспечения.

Целью контроля является обеспечение анализа и оценки соответствия численности и уровня автоматизации организационной структуры, выработки, при необходимости, рекомендаций по совершенствованию обслуживания.

2. Описание технологии контроля качества обслуживания

Контроль над качеством обслуживания требований субъектов организационной структуры осуществляется с помощью следующих критериев оценки: заданного времени $T_{зад}$ ожидания требованием обслуживания и заданной вероятности $P_{зад}$ ($\leq T_{зад}$) его непревышения. При этом в ходе анализа производится сравнение фактических значений с этими критериями. Результатом анализа является прогноз о наступлении фазы цикла обслуживания, в которой фактические значения не удовлетворяют заданным критериям. Такой прогноз является основанием для принятия указанных выше мер. Технология рекомендована к использованию в автоматизированной системе, сведения о которой приводятся в [10], и включает в себя следующие виды деятельности:

2. 1. Синтез исходной структуры

Производится в стадии проектирования в соответствии с методологией, изложенной в [7], базирующейся на моделях интервалов занятости системы массового обслуживания с ожиданием и позволяющей сформировать организационную структуру, как



Рис. 2. Пример 1 расчёта численности организационной структуры.



Рис. 3. Пример 2 расчёта численности организационной структуры.



Рис. 4. Пример 3 расчёта численности организационной структуры.

структуру вырожденного, линейно-конвейерного или конвейерно-древовидного типа. На рис. 2, рис. 3, ..., рис. 7 приведены примеры расчёта требуемой численности работников для организационных структур вырожденного, линейно-конвейерного и конвейерно-древовидного типа при разных значениях поправочного коэффициента k и в зависимости от максимальной длительности интервала занятости организационной структуры. Поправочный коэффициент k используется тогда, когда обслуживание осуществляется специалистом широкого профиля, k определяется исходя из сравнения со временем выполнения этой же функции специалистом узкого профиля³. Из примеров видно, что в зависимости от значения $N_{макс}$ максимально-допустимой длительности интервала занятости организационной структуры, удовлетворяющего заданным критериям $T_{зад}$ и $P_{зад}$ ($\leq T_{зад}$), может быть выбран тот или иной тип структуры.

³ Специалист узкого профиля выполняет функции, относящиеся к одному виду деятельности. Специалист широкого профиля выполняет функции, относящиеся к нескольким видам деятельности.

2.2 Создание автоматизированной системы контроля обслуживания

Осуществляется в соответствии с установленными стандартами автоматизации [11-13], стандарта качества [14, 15] и с учётом особенностей организационных структур.

Численность организационной структуры ($k=1,6$)

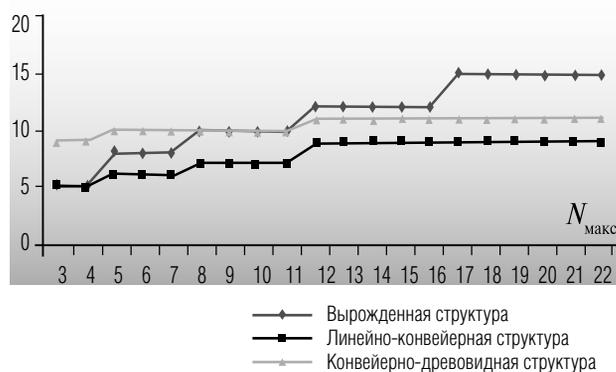


Рис. 6. Пример 5 численности организационной структуры.

Численность организационной структуры ($k=1,8$)

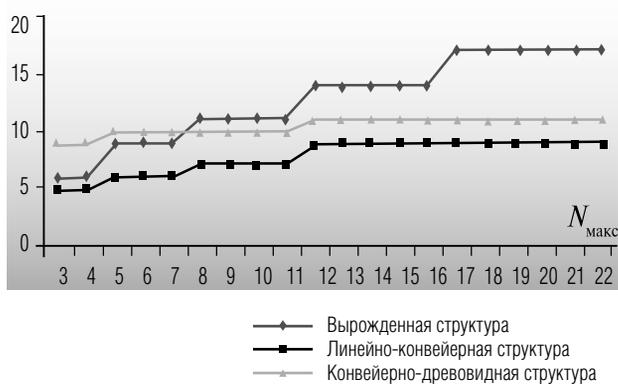


Рис. 7. Пример 6 расчёта численности организационной структуры.

Численность организационной структуры ($k=2$)

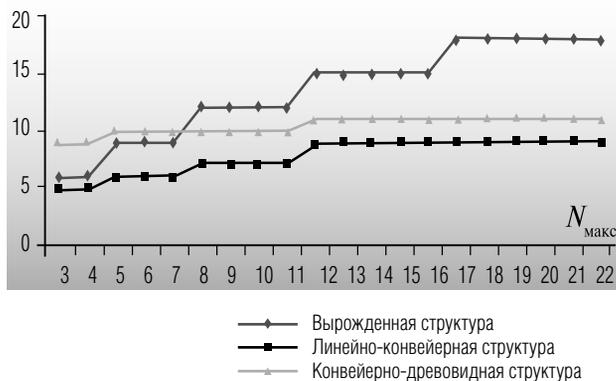


Рис. 7. Пример 7 расчёта численности организационной структуры.

2.3. Сбор и анализ статистической информации

◆ Производится в отношении каждого вида обслуживания, – каждого типа услуг поддерживаемых организационной структурой. В основе сбора статистической информации лежит модель контрольно-учётного процесса, реализуемого параллельно с реализацией процессов обслуживания [16]. В состав контрольно-учётного процесса входят типовые функции: «Классификация требования», «Формулирование проблемы», «Определение задач» и «Сопровождение задачи»; реализация процесса позволяет получить статистические значения показателей контроля в реальном масштабе времени. Сбор информации осуществляется путём накопления ретроспективной информации на этапах обслуживания. Например, осуществляется накопление информации в отношении всех действий, производимых при реализации функций «Классификация требования», «Формулирование проблемы», «Определение задач» и «Сопровождение задачи», а также в отношении результатов этих действий. На основе статистических выборок производится анализ качества обслуживания.

2.4. Разработка рекомендаций

Производится по следующим направлениям:

- ◆ изменения в организационной структуре, – численности работников, их квалификации, типе структуры;
- ◆ изменения в уровне автоматизации организационной структуры, – составе автоматизируемых функций, их уровне автоматизации.

3. Определение границ вероятности превышения нормативного времени ожидания требованиями обслуживания

Определение границ вероятности превышения нормативного времени ожидания требованиями обслуживания проводится в рамках методологии синтеза организационной структуры [7], основанной на оценке организационной структуры вырожденного типа при заданных критериях $T_{\text{зад}}$ и $P_{\text{зад}} (\leq T_{\text{зад}})$ и определения граничных значений G стабильного функционирования организационной структуры. Суть данного подхода заключается в следующем:

Граничному значению G сопоставляется соответствующее ему значение $N_{\text{макс}}$ максимально-допустимой длительности интервала занятости организационной структуры, выраженное числом требований, обслуженных в этом интервале. Физически, граничное значение G представляет собой время ожидания любым требованием, находящемся в этом интервале, обслуживания и удовлетворяющему критериям $T_{\text{зад}}$ и $P_{\text{зад}}(\leq T_{\text{зад}})$. Граничное значение G так же, как и значение $N_{\text{макс}}$, выражается числом требований. Для расчёта используются следующие формулы, приведённые в работе [17] для описания вероятностей ожидания требованиями обслуживания в интервале занятости, в зависимости от их порядкового места в этом интервале и при учёте того, что время обслуживания одного требования равно одному интервалу обслуживания:

$$P_N^k(j) = \frac{(N-1)!}{N^{N-2}} \left\{ F1 \times F2 + \sum_{x=1}^{j-1} [F3 \times (F4 - F5)] \right\} \quad (1)$$

$$F1 = \frac{(k-j)^{k-j-2}}{(k-j-1)!} \quad (2)$$

$$F2 = \left[\frac{(N-k+j+1)^{N-k+j-1}}{(N-k+j)!} - \sum_{m=1}^{j-1} \frac{(N-k+j)^{N-k+j-m-1}}{(m-1)!(N-k+j-m)!} \right] \quad (3)$$

$$F3 = \sum_{y=0}^x \left[\frac{(-1)^y (x-y+1)^y}{y!(k-j+x-y-1)!} \times \frac{1}{\times (k-j+x-y)^{k-j+x-y-2}} \right] \quad (4)$$

$$F4 = \left[\frac{(x+1)(N-k+j+1)^{N-k+j-x-1}}{(N-k+j-x)!} \right] \quad (5)$$

$$P_N(j) = \frac{1}{N} \sum_{k=j+1}^N P_N^k(j) \quad (6)$$

$$F5 = \left[\sum_{z=0}^{j-x-1} \frac{(x+z)(N-k+j)^{N-k+j-x-z-1}}{z!(N-k+j-x-z)!} \right] \quad (7)$$

$$N = 3, 4, \dots; k = 3, \dots, N; j = 1, \dots, k-1;$$

где: $P_N^k(j)$ – вероятность ожидания обслуживания длительностью в j временных интервалов для k -го требования, по порядку его обслуживания в интервале занятости длительностью в N временных интервалов [7];

$P_N(j)$ – вероятность того, что время ожидания требованием обслуживания в интервале занятости составляет в среднем j интервалов обслуживания.

С помощью формул (1) – (7) определяется значение $N_{\text{макс}}$, при котором ещё соблюдаются заданные критерии $T_{\text{зад}}$ и $P_{\text{зад}}(\leq T_{\text{зад}})$. Для определения $N_{\text{макс}}$ и соответствующего ему граничного значения G используется ряд значений $P_N(j)$.

Ниже в табл. 1 приведена выборка из этих значений.

Таблица 1.

Выборка значений $P_N(j)$

	$N=10$	$N=11$	$N=12$	$N=13$	$N=14$	$N=15$
$j=0$	0,1	0,09091	0,08333	0,07682	0,07143	0,06667
$j=1$	0,42872	0,39786	0,37108	0,34763	0,32693	0,30854
$j=2$	0,76379	0,72922	0,69692	0,66684	0,63887	0,61288
$j=3$	0,93814	0,91896	0,8989	0,87839	0,85776	0,83725
$j=4$	0,99016	0,98431	0,97711	0,96872	0,95929	0,949
$j=5$	0,99912	0,99813	0,99661	0,99448	0,9917	0,98827
$j=6$	0,99996	0,99988	0,99969	0,99936	0,99883	0,99806
$j=7$	1	1	0,99998	0,99995	0,99989	0,99978
$j=8$	1	1	1	1	0,99999	0,99998
$j=9$	1	1	1	1	1	1

Каждое из значений $P_N(j)$ сопоставляется с каждым нормированным значением $P_{\text{зад}}(\leq T_{\text{зад}})$, при этом параметр N принимается равным $N_{\text{макс}}$. При достижении наименьшего значения $P_N(j)$, удовлетворяющего неравенству:

$$|P_N(j)| > |P_{\text{зад}}(\leq T_{\text{зад}})|, \quad (8)$$

значению $N_{\text{макс}}$ максимально-допустимой длительности интервала занятости присваивается значение параметра N , удовлетворяющему неравенство (8);

граничному значению G присваивается значение параметра j , удовлетворяющему неравенство (8).

Например (табл. 1), если $P_{\text{зад}}(\leq T_{\text{зад}}) = 0,94$, то этому значению соответствуют следующие искомые граничные значения: $N_{\text{макс}} = 15$ и $G = 4$. Данному примеру соответствует следующая реальность:

- ◆ значение $G = 4$ сопоставляется с числом работников организационной структуры;

- ◆ значение $N_{\text{макс}} = 15$ является эталонным значением для проведения оценки качества обслуживания.

Сбор статистических данных осуществляется с использованием известных положений теории динамических систем [18], при этом в отношении каждого интервала занятости производятся следующие действия:

а) измеряется его длительность $T_{\text{из}}$;

б) значение $T_{\text{из}}$ делится на G и на длительность одного интервала обслуживания (например, это может быть нормированное значение длительности выполнения одной работы), в результате определяется фактическое значение $N_{\text{факт}}$;

в) производится сравнение $N_{\text{факт}}$ с эталонным значением $N_{\text{макс}}$.

Анализ зависимости границы вероятности превышения нормативного времени ожидания требования от характеристик предметной области и численности организационной структуры показывает, что характер изменений значений параметра N при постоянном значении параметра j показывает нелинейное изменение этих значений в сторону их уменьшения. Исходя из этого, можно сделать вывод о том, что с уменьшением значения времени обслуживания одного требования, например, в силу повышения степени автоматизации организационной структуры, число задач, решаемых этой структурой, может быть увеличено, без изменения числа работников.

4. Контроль над процессом информационного взаимодействия

С целью контроля над информационным обменом в автоматизированный процесс контроля качества обслуживания (рис. 1) введена функция контроля над процессом информационного взаимодействия. Контроль осуществляется следующим образом:

- ◆ структура технической системы, используемой для информационного обмена между работниками,

представляет собой совокупность групповых трактов между отдельными элементами или/и уровнями организационной структуры;

- ◆ каждый групповой тракт описывается моделью системы массового обслуживания; для каждого группового тракта задаются свои критические значения времени $T_{\text{зад}}$ ожидания требованием обслуживания и вероятности $P_{\text{зад}}(\leq T_{\text{зад}})$ его не превышения;

- ◆ над каждым групповым трактом ведётся контроль, по аналогии с контролем над процессом обслуживания, описанным выше.

В результате анализа полученных значений показателей качества информационного обмена, может быть принято одно из следующих решений:

- ◆ существующая структура локальных и магистральных сетей передачи данных, а также используемые в них аппаратные и программные средства, удовлетворяют заданным требованиям к обмену информацией между вычислительными ресурсами информационной системы;

- ◆ требуется разработать рекомендации по реорганизации структуры информационного обмена или изменению состава используемых аппаратных и/или программных средств.

5. Заключение

В настоящей работе рассмотрена технология контроля качества обслуживания в организационных структурах, основанная на динамическом анализе результатов сравнения фактического значения длительности интервала занятости организационной структуры с заданным критерием — эталонным значением.

При этом:

- ◆ в качестве эталонного, принимается максимально-допустимое значение длительности интервала занятости, которое ещё удовлетворяет заданным критериям оценки времени ожидания требованием обслуживания и вероятности его не превышения;

- ◆ результатом анализа является прогноз наступления фазы в цикле производства услуг, на которой фактическое значение превышает заданный критерий;

- ◆ такой прогноз является основанием для принятия мер по реорганизации организационной структуры или изменениям в степени автоматизации выполняемых работ. ■

Литература

1. Попов В.М. Глобальный бизнес и информационные технологии. Современная практика. – М.: Финансы и статистика, 2001. – 272 с.
2. Карминский А.М. Контроллинг в бизнесе. Методологические и практические основы построения контроллинга в организациях, 2-е изд. – М.: Финансы и статистика, 2003. – 256 с.
3. Москвин В.А. Управление качеством в бизнесе: Рекомендации для руководителей предприятий, банков и риск-менеджеров. – М.: Финансы и статистика, 2006. – 384 с.
4. Калянов Г.Н. Моделирование, анализ, реорганизация и автоматизация бизнес-процессов. – М.: Финансы и статистика, 2006. – 240 с.
5. Мельников В.В. Безопасность информации в автоматизированных системах. – М.: Финансы и статистика, 2003. – 368 с.
6. Володин К.И. и др. Автоматизированная система научно-техн. информации – разработка и эксплуатация. – М.: Финансы и статистика, 2004. – 192 с.
7. Шабанов А.П., Беляков А.Г. Организационные структуры массового обслуживания: научное издание. – М.: Институт проблем управления им. Трапезникова РАН, 2007. – С. 23-28, 36-40, 50-54, 57-60, 65-71, 25.
8. Шабанов А.П. Подход к автоматизации деятельности организационных структур // Труды V Всероссийской научно-практической конференции: Системы автоматизации в образовании, науке и производстве. – Новокузнецк: СибГИУ, 2005. – С. 124-128.
9. Шабанов А.П. Подход к выбору направления автоматизации деятельности // Труды VI Всероссийской научно-практической конференции: Системы автоматизации в образовании, науке и производстве. – Новокузнецк: СибГИУ, 2007. – С 81-85.
10. О выборе стратегии реализации системы учёта и контроля информационно-телекоммуникационных услуг / Труды XLVIII научной конференции: Современные проблемы фундаментальных и прикладных наук. – М.: МФТИ, 2005. – Часть III, С. 90-94.
11. Государственный стандарт Российской Федерации. ГОСТ 34.003-90. Информационная технология. Автоматизированные системы. Термины и определения.
12. Государственный стандарт Российской Федерации. ГОСТ 34.602-89. Информационная технология. Комплекс стандартов на автоматизированные системы. Техническое задание на создание автоматизированной системы.
13. Государственный стандарт Российской Федерации. ГОСТ 34.601-90. Информационная технология. Комплекс стандартов на автоматизированные системы. Стадии создания.
14. Государственный стандарт Российской Федерации. ГОСТ Р ИСО 9000-2001. Системы менеджмента качества. Основные положения и словарь. Дата введения 31.08.2001.
15. Государственный стандарт Российской Федерации. ГОСТ Р ИСО 9001-2001. Системы менеджмента качества. Требования.
16. Шабанов А.П. Подход к формализации учётно-контрольного процесса при решении проблемы минимизации ресурсов организационной структуры // Сборник трудов научно-практической конференции: Современные сложные системы управления. – Воронеж: ВГАСУ, 2005. – С. 186-190.
17. Шабанов А.П. Исследование условий стабильности информационных систем / Бизнес-Информатика. – 2010. – № 2(12). – С. 24-36.
18. Евланов Л.Г. Контроль динамических систем, 2-е изд., перераб. и доп. / – М.: Наука. Главная редакция физико-математической литературы, 1979. – 432 с.

О НЕОБХОДИМОСТИ ПРАВОВОЙ КЛАССИФИКАЦИИ ОПЕРАТОРОВ СЕТИ ИНТЕРНЕТ

А.К. Жарова,

кандидат юридических наук,

доцент кафедры инноваций и бизнеса в сфере информационных технологий

Национального исследовательского университета «Высшая школа экономики»

Адрес: г. Москва, ул. Кирпичная, д. 33/5

E-mail: ajarova@hse.ru

В отношениях, связанных с использованием цифрового произведения в среде Интернет, участвуют различные субъекты. В статье предлагается правовая классификация таких информационных субъектов, как сервис-провайдеры. Проводится сравнительно-правовой анализ информационного законодательства России, законов США и положений Директивы об электронной торговле Европейского союза 2000/31/ЕС. Указывается на необходимость введения дополнительных критериев в существующую правовую классификацию лиц, оказывающих услуги в сети Интернет, связанных с предоставляемыми услугами и выполняемыми функциями.

Ключевые слова: интернет, интернет-провайдер, оператор, ответственность лиц, международная классификация провайдеров, законодательство США, ЕС.

В настоящее время в обществе все чаще говорят о необходимости защиты от распространения вредной информации в глобальной сети Интернет, и связано это с тем, что развитие данной технологии создает большие возможности для совершения противоправных действий различными субъектами. Так, в Декларации ОЭСР о перспективах Интернет-экономики [2], заявляется о необходимости создания условий стандартизации в государствах — участниках применяемых цифровых технологий и протоколов, с целью развития инвестиций в развитие высоких технологий, позволяющих использовать возможности Интернет и Интернет-услуг, а также увеличить показатели Интернет-экономики. В упомянутой Декларации

определяется необходимость более эффективного использования спектра радиочастот, благодаря которым государства-участники представляют новые виды услуг своим гражданам, а также указывается на необходимость развивать применение протокола IPv6, благодаря которому субъектам государства будет облегчен доступ к формирующимся государственным услугам. Для этого Декларация ОЭСР призывает объединить усилия государств с целью формирования единого информационного пространства.

При преобразовании в цифровую форму произведение, вне зависимости от его первоначального замысла, становится просто данными, которые

могут перемещаться, копироваться, преобразовываться и вновь переводиться в форму, доступную человеческому пониманию. Из-за того, что цифровая среда вводит дополнительную прослойку между человеком и произведением, и эта новая прослойка контролируется только с помощью технических устройств, использование такого произведения становится более опосредованным. Развитие сети Интернет способствовало появлению новых информационных субъектов и новых информационных отношений, в правовом регулировании которых не все ясно. Это, в свою очередь, заставляет пересмотреть параметры классификации субъектов таких отношений

Новые технологии используются не только как возможности нового порядка, позволяющие ускорить и облегчить позитивное взаимодействие субъектов, но и как возможности создания конфликтных отношений. В глобальной сети действия субъектов носят как позитивный, так и негативный характер, некоторые субъекты используют информацию, размещенную в сети Интернет, считая, что она размещена там с согласия правообладателя, хотя, зачастую, это не так.

В информационных правонарушениях, субъектный состав участников очень разнообразен и от способности лица осознавать противоправность своих действий, возможности наступления вредоносных последствий (интеллектуальный аспект) и желания наступления этих последствий (волевой аспект), зависят определение вины и наступление ответственности за совершенные действия.

Главными правовыми проблемами определения ответственности за действия в сети Интернет являются определение места подсудности и определение субъектного состава возникающих отношений.

Следует также обратить внимание на то исключительное обстоятельство, что в общем случае при использовании Интернета большинство пользователей не идентифицируются, что значительно затрудняет возможность определения, к какой категории лиц они относятся и к территории какого государства имеют привязку. Это значит, что в случае судебных разбирательств, в первую очередь, могут возникнуть споры с определением подсудности.

Именно по этой причине, говоря о защите от распространения вредной информации в сети Интернет, в первую очередь, необходимо определиться с субъектным составом таких отношений.

В сфере Интернет можно выделить следующих субъектов – участников правоотношений:

1. Государство – отдельный субъект правоотношений.
2. Юридические лица:
 - 2.1. операторы связи (технологий);
 - 2.2. операторы информации;
 - 2.3. получатели услуг.
3. Физические лица:
 - 3.1. автор;
 - 3.2. обладатель информации или информационного ресурса;
 - 3.3. поставщики содержания информации;
 - 3.4. конечные пользователи – граждане.

Возможно также определение субъектного состава в среде Интернет через взаимодействие лиц, например, объединения в сети физических или юридических лиц. Кроме того, можно построить, четырехугольник телекоммуникационных отношений, вершинами которого являются следующие субъекты: поставщики информации, пользователи информации и информационных ресурсов, операторы связи (технологий), операторы информации.

За неправомерные действия предусмотрена различная юридическая ответственность, определяемая ГК РФ, УК РФ, КОАП РФ. Одним из видов ответственности субъектов является различного рода компенсация. Так в ст. 1250 ГК РФ указывается, что отсутствие вины нарушителя не освобождает его от обязанности прекратить нарушение интеллектуальных прав, а также не исключает применение в отношении нарушителя мер, направленных на защиту таких прав.

Кроме того, необходимо учитывать и то, что гражданско-правовая ответственность, в отличие от уголовной и административной, базируется на основе презумпции виновности, так п.2. ст. 401 ГК РФ, определяет, что «отсутствие вины доказывается лицом, нарушившим обязательство». Именно по этой причине Верховный суд РФ указал, что «игнорировать наличие или отсутствие вины возможно только в случае, если взыскание компенсации вообще не является мерой ответственности или же она применяется в предпринимательских отношениях п.3 ст.401 ГК РФ. Поскольку ответчики обычно предприниматели, то нет необходимости отыскивать вину. В случаях с гражданами, не являющимися предпринимателями, вину требуется установить» [5]. Таким образом, можно определить

правовую схему, по которой все юридические и физические лица, занимающиеся предпринимательской деятельностью, попадают под презумпцию виновности.

В соответствии с законами информационного законодательства, такими как ФЗ РФ от 27 июля 2006 г. №149-ФЗ «Об информации, информационных технологиях и о защите информации» и ФЗ РФ «О связи» от 07.07.2003 «№126-ФЗ (далее ФЗ «О связи»)), классификация субъектов, определяется через понятие оператора. Например, ФЗ «Об информации, информационных технологиях и о защите информации» (далее ФЗ «Об информации») п.12. ст.2, определяет оператора информационной системы, как «гражданина или юридическое лицо, осуществляющие деятельность по эксплуатации информационной системы, в том числе по обработке информации, содержащейся в ее базах данных». Под данное определение оператора информационной системы попадают все субъекты, которые связаны, так или иначе, с работой на компьютере и обработкой информации, содержащейся в базах данных компьютера.

Помимо данного определения, ФЗ «Об информации» в п.3. ст. 17 содержит также и термин «лицо, оказывающее услуги», к которому, по-видимому, должны относиться юридические лица или индивидуальные предприниматели. Однако, к сожалению, ФЗ РФ «Об информации» не дает расшифровки данного термина.

Следующий ФЗ «О связи», в соответствии со ст. 2 операторов классифицирует, как:

1. оператор, занимающий существенное положение в сети связи общего пользования;
2. оператор связи;
3. оператор универсального обслуживания;
4. оператор обязательных общедоступных телеканалов и (или) радиоканалов.

Оператор связи определяется как юридическое лицо или индивидуальный предприниматель, оказывающие услуги связи на основании соответствующей лицензии. Данный закон направлен на обеспечение технологических процедур взаимодействия операторов связи, с созданием и эксплуатацией всех сетей связи и сооружений связи, использованием радиочастотного спектра, оказанием услуг электросвязи и почтовой связи на территории Российской Федерации и на находящихся под юрисдикцией Российской Федерации территориях.

ФЗ «О связи» в ст. 68 указывает, что «в случаях и в порядке, которые установлены законодательством Российской Федерации, лица, нарушившие законодательство Российской Федерации в области связи, несут уголовную, административную и гражданско-правовую ответственность».

Однако ФЗ «О связи» не раскрывает вопрос, несет ли оператор связи или другой вид оператора по данному закону ответственность за передачу информации, предоставленной другим лицом и обеспечению доступа к ней. Исходя из цели ФЗ «О связи», необходимо полагать, что операторы связи, оказывающие технологическую поддержку в коммуникации субъектов, не должны нести ответственность за вышеуказанные действия, но в тоже время п. 3. ст. 17 ФЗ «Об информации» раскрывает исключение из гражданско-правовой «ответственности лиц, оказывающих услуги в случае, если лицо:

1. оказывает услуги по передаче информации, предоставленной другим лицом, при условии ее передачи без изменений и исправлений;
2. оказывает услуги по хранению информации и обеспечению доступа к ней при условии, что это лицо не могло знать о незаконности распространения информации».

В свою очередь по указанной классификации ФЗ «Об информации» операторы связи попадают под часть 1. п.3. ст.17, как лица, оказывающие услуги по передаче информации, предоставленной другим лицом.

Для устранения определенной неоднозначности лиц, оказывающих услуги по предоставляемым услугам и выполняемым функциям, необходимо классифицировать по двум большим уровням:

1. операторы работы с технологиями;
2. операторы работы с содержанием – информацией.

Так в соответствии с положениями Директивы об электронной торговле Европейского союза 2000/31/ЕС [3] принято выделять три категории провайдеров – провайдеры содержания (контента), хост-провайдеры и провайдеры доступа. Эти категории определяются выполняемыми ими функциями. Это означает, что отдельный сервис-провайдер, исполняющий несколько функций, может относиться к нескольким категориям, например, если он обеспечивает доступ к данным в Интернете и в то же время предлагает свое содержание с собственного сервера.

Директива об электронной торговле ЕС 2000/31/ЕС функции таких лиц определяет следующим образом:

Провайдеры содержания (контент-провайдеры) предоставляют собственное содержание и обеспечивают его доступность. По функциям, которые на них возложены, такие провайдеры создают содержание и, соответственно, берут на себя ответственность за содержательную часть собственной информации. Кроме того, они сохраняют контент на своей технической базе или под своим контролем, а значит, могут контролировать доступ к нему. К этой категории относится всякое лицо, которое создает веб-сайт на собственном сервере или на арендованной веб-площадке хост-сервера. На контент-провайдеров возлагается полная ответственность перед законом в соответствии с общими законоположениями. Здесь нет оснований для освобождения от ответственности, поскольку каждый отвечает за то содержание, которое он может полностью контролировать. Не может служить оправданием даже большой объем информации. Организация может установить соответствующую систему контроля за содержанием или не размещать информацию в сети, в противном случае, ей придется смириться с риском привлечения к ответственности за содержание.

Хост-провайдеры предоставляют содержание третьих сторон и обеспечивают его доступность. Они сохраняют чужое содержание (обычно оно принадлежит их клиенту) на собственной или арендованной технической базе. Типичные хост-провайдеры – это, к примеру, операторы сервера, сдающие в аренду объем памяти операторам веб-сайтов, или операторы телеконференции без ведущего, принимающие статьи от пользователей Интернета. Хост-провайдер, как правило, не может контролировать содержание на своем сервере: в большинстве случаев объем записанных в память данных и частая замена поступающих данных не позволяют установить механизм ручного контроля, особенно если хост-провайдер не загружает содержание на сервер, а обеспечивает своим клиентам прямой доступ к нему. Автоматический контроль, например, с помощью фильтрующего программного обеспечения, неспособен выполнить непростую задачу отделения законного материала от незаконного.

Данные в нестандартных форматах, в сжатой форме или в зашифрованном виде могут даже быть нечитабельными для провайдера. Кроме того, частный оперативный контроль не является целью пу-

бличной политики: огромную часть общения в сети составляет частная или другая секретная информация, которую не следует раскрывать провайдеру.

Наконец, даже если провайдер может выявить контрафактный материал, то его удаление без блокировки доступа ко всем услугам может оказаться слишком обременительной или невозможной задачей.

По анализу функций, возложенных на провайдера, и технологическим характеристикам аппаратно-программных средств провайдера и сети Интернет можно утверждать, что провайдер несет полную ответственность перед законом за причиненный ущерб с момента обнаружения им контрафактного материала или получения уведомления о нем от других лиц, если он не удалит этот материал сразу, когда технически это возможно сделать.

Провайдеры доступа предоставляют доступ к содержанию третьих лиц. Их услуга заключается главным образом в перемещении данных без их постоянного хранения. Типичный провайдер доступа подключает к Интернету конечного пользователя, связывая его из локального места присутствия с базовой сетью Интернет. Подобно хост-провайдерам, провайдеры доступа не могут и не должны отслеживать любые данные, передаваемые по их соединениям. Но даже если бы они и знали, что по их линиям регулярно пересылаются незаконные данные, они не могли бы отделить эти данные от остального массива. Они, в отличие от контент-провайдеров, не могут контролировать данные, поскольку не хранят их у себя постоянно. По этой причине провайдер доступа освобождается от всякой ответственности, даже если он знал о незаконности содержания.

Провайдер доступа не имеет возможности проконтролировать и заблокировать исходящую информацию, у него есть возможность проследить пришедшую информацию, которая хранится у него до тех пор, пока адресант не скопировал или «скачал» ее на свой компьютер. Тем более, что Директива об электронной торговле Европейского союза (Directive on Certain Legal Aspects of Information Society Services, in particular Electronic Commerce, in the Internal Market, Jul 17, 2000, 2000/31/EC) в Статье 15 прямо запрещает государствам-членам налагать на любых провайдеров общую обязанность отслеживать информацию или активно высискивать факты, указывающие на незаконную деятельность.

Сравнивая существующие положения нормативных актов РФ и положения Директивы

об электронной торговле Европейского союза 2000/31/ЕС, можно произвести следующую классификацию. Провайдеров доступа Директивы ЕС 2000/31/ЕС отнести под сферу действия ФЗ «О связи», приравняв их к операторам связи. Провайдеров содержания и хост-провайдеров Директивы ЕС 2000/31/ЕС можно отнести к ФЗ «Об информации» 2006 г., но однозначно классифицировать данных субъектов по ФЗ «Об информации» не удастся, поскольку ФЗ «Об информации» содержит как понятие оператора информационной системы (куда входят, в том числе и физические лица, не занимающиеся предпринимательской деятельностью), так и понятие лица, оказывающего услуги либо по передаче информации, предоставленной другим лицом, либо по хранению информации и обеспечению доступа к ней.

Ответственность провайдера (или, по российскому законодательству, оператора) либо лица, оказывающего услуги, может различаться. Соответственно, одно и то же содержание может создавать для разных сервис-провайдеров разные типы правовой ответственности: информация может предоставляться контент-провайдером в распоряжение пользователей и передаваться по сети провайдерами доступа.

В феврале 2009 г. эту позицию закрепил Президиум Высшего Арбитражного Суда РФ, разъяснивший, что провайдер (по смыслу судебного решения, речь идет о хост-провайдере) не несет ответственности за передаваемую информацию, если он не иницирует ее передачу, не выбирает получателя информации и не влияет на целостность передаваемых данных. Тем не менее, ВАС РФ рекомендовал другим судам при рассмотрении дел учитывать действия самого провайдера, т.е. принимал ли провайдер превентивные меры по пресечению возможных нарушений своими клиентами с использованием его услуг или нет.

Кроме того, бремя доказательства фактов несанкционированного размещения провайдером в сети Интернет защищенных авторским правом произведений лежит на правообладателе. Если не установлено, что провайдер знал или мог знать о незаконном распространении произведений, он не должен доказывать отсутствие факта использования им этих произведений [7].

В настоящее время вынесены на рассмотрение и обсуждение проект изменения четвертой части ГК РФ [6], где в ст. 1253' включены других субъекты – Интернет-провайдеры, осуществляющие

передачу материала в сети Интернет, и Интернет-провайдеры, оказывающие услуги по размещению материалов в сети Интернет. При этом не принято во внимание уже существующее разнообразие субъектов. Данные категории лиц не попадают ни под международную классификацию, ни под существующую российскую. Необходимо учесть, что законодательство РФ оперирует понятием оператора или лица, оказывающего услуги.

Другим примером правовой классификации лиц, оказывающих услуги в сети Интернет, является законодательство США. Закон «Об авторском праве» США (Copyright Law of the United States of America) 1976 года устанавливает, что при обнаружении контрафактного экземпляра произведения и направлении провайдеру соответствующего письма он должен немедленно прекратить доступ к такому материалу. Так, согласно §512 «Ограничения на ответственность, касающиеся материала онлайн» (Limitations on liability relating to material online) Закона об авторском праве США 1976 г. [1] определено, что поставщик услуг освобождается от ответственности в случае, если он осуществляет промежуточное или переходное хранение материала в ходе такой передачи, а также, если:

- 1) передача материала была начата не по инициативе поставщика услуг;
- 2) передача или хранение выполнены через автоматический технический процесс без выбора материала поставщиком услуг;
- 3) поставщик услуг не выбирал получателей материала;
- 4) никакая копия материала, сделанного поставщиком услуг в ходе такого промежуточного или переходного хранения, не хранится на сервере провайдера больше периода, требуемого для необходимой передачи;
- 5) материал передан через систему или сеть без модификации ее содержания. Сравнивая законодательство РФ и США можно провести аналогию со ст. 17 Закона «Об информации» 2006 г., но в нашем законе не учитываются, как процедуры передачи или хранения, выполненные через автоматический технический процесс без выбора материала поставщиком услуг (т.е. речь идет о кэш-памяти), так и период, в течение которого копия материала, сделанного поставщиком услуг в ходе промежуточного или переходного хранения, хранится на сервере провайдера.

Кроме того, закон США «Об авторском праве» (Copyright Law of the United States of America) 1976 г. содержит норму, которая определяет, что при обнаружении контрафактного экземпляра произведения субъектом и направлении провайдеру соответствующего письма последний должен немедленно прекратить доступ к такому материалу. Механизм применения данной нормы раскрывается в другом законодательном акте США – Федеральном законе «Об ограничении ответственности за онлайн-нарушения авторского права» [4], который детально определяет регламент подачи правообладателем Интернет-провайдеру заявки о блокировании сайта, нарушающего авторские права, и о действиях Интернет-провайдера в таких случаях. Данный Федеральный закон Соединенных Штатов определяет условия, при которых поставщики Интернет-услуг и другие Интернет-посредники освобождаются от ответственности за действия третьих лиц.

К сожалению, в российском законодательстве отсутствует норма подобного рода. Это делает возможным отсутствие реакции со стороны лица, предоставляющего услуги в сети Интернет, на осуществляемые неправомерные действия.

Кроме того, существует и обратная сторона данных отношений: в случае, если лицо, предоставляющее услуги, вовремя отреагирует и по требованию заинтересованной стороны заблокирует доступ к потенциально вредному источнику информации, то он не застрахован от ситуации, когда распространение информации происходило с разрешения правообладателя. В таком случае лицо, предоставляющее услуги, само окажется правонарушителем, поскольку оно нарушило право правообладателя на доведение до всеобщего сведения.

Таким образом, для российского законодательства назрела необходимость правовой классификации по предоставляемым услугам и выполняемым функциям лиц, предоставляющих Интернет-услуги, которая позволит таким лицам четко понимать ответственность и регламент отношений в случае возникновения конфликтной ситуации. Один из вариантов решения существующих проблем – принятие ФЗ «Об Операторах сети Интернет (провайдерах)», который позволит однозначно определить функции, виды таких субъектов, ответственность этих субъектов за совершаемые действия, а также регламент отношений между провайдерами, правонарушителями, правообладателями. ■

Литература

1. Copyright Law of the United States of America // <http://www.copyrighter.ru/full/index.html?uscopyrightact.htm>
2. Declaration for the Future of the Internet Economy (The Seoul Declaration) 18 June 2008 - C(2008)99 // <http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=113&Lang=en&Book=False>
3. Directive on Certain Legal Aspects of Information Society Services, in particular Electronic Commerce, in the Internal Market, Jul 17, 2000, 2000/31/EC // http://www.efama.org/index2.php?option=com_docman&task=doc_view&gid=148&Itemid=35
4. Online Copyright Infringement Liability Limitation Act (OCILLA) является частью Цифрового Закона об авторском праве Тысячелетия Digital Millennium Copyright Act (DMCA) (DMCA) // http://dictionnaire.sensagent.com/Online_Copyright_Infringement_Liability_Limitation_Act/en-en/
5. Документы Верховного суда РФ юридическая пресса. Вопросы практики применения правил о компенсации в связи с нарушением исключительных прав // <http://www.advocate-realty.ru/press/unitpress/?id=422616>
6. Концепция совершенствования Раздела VII Гражданского кодекса Российской Федерации «Права на результаты интеллектуальной деятельности и средства индивидуализации». Проект рекомендован Президиумом Совета к опубликованию в целях обсуждения (протокол от 13 мая 2009 г.) // http://www.consultant.ru/obj/file/doc/pr_t7.rtf
7. Сайт ВАС РФ. Определение Президиума ВАС РФ от 23.12.2008 № 10962/08. Дело о защите исключительных прав на музыкальные произведения от их незаконного воспроизведения и доведения до всеобщего сведения путем помещения в сети Интернет передано на новое рассмотрение в суд первой инстанции, так как нарушителем авторских прав необоснованно признано лицо, которое само не осуществляло действий по использованию объектов авторского права. // http://www.arbitr.ru/?id_sec=353&id_doc=19657



EDUCATIONAL ONLINE COMMUNITIES IN MULTICULTURAL CONTEXT: OPEN CONTENT

J. Taratukhina, O. Chamina

Annotation

In the paper the general trends in the organization of educational communications related with open content development in educational network communities in different cultural environments are discussed, taking into consideration the socio-cultural characteristics of educational systems.

Keywords: online communities, open content, information environment, communicative and pedagogical innovations, educational environment, cross-cultural didactics, multimedia didactics.



METHOD OF ASSESSMENT OF COMPETITIVENESS OF APPLICATION SOFTWARE

Y. Ehlakov, E. Yanchenko

Annotation

The paper deals with the competitiveness of an application software product. Algorithm for estimating competitiveness of an application software product based on analysis of existing estimation techniques products. The algorithm includes a set of criteria to assess the procedure for calculating and analyzing the integral index of competitiveness the integral index of competitiveness.

Key words: competitiveness, application software, integrated indicator of competitiveness.



A METHOD OF SEMANTIC TRANSFORMATION OF THE GENERALIZED SEARCH REQUESTS WITH THE HELP OF A GOAL BASE

A. Kirillov

Annotation

The paper describes a new method of constructing semantic expansions of search requests (of generalized character) for improving the results of Web search. This method is based on the theory of K-representations - a new theory of designing semantic-syntactic analyzers of natural language texts with the broad use of formal means for representing input, intermediary, and output data. The stated approach is implemented with the help of the programming language «Java»: an experimental search system AOS (Aspect Oriented Search) has been developed.

Key words: semantic search, semantic transformation, generalized request, conceptual basis, SK-language, conceptual goal base, transformation pattern.



AUTOMATIC RETRIEVAL OF PARALLEL COLLOCATIONS

V. Novitskiy

Annotation

The article describes an approach to automatic retrieval of parallel collocations (that means collocations and their translations to another language). The method is based on the comparison of syntactic trees of two parallel sentences. The key feature of the described method is set of filters that help to get more precise results.

Keywords: NLP, parallel collocations.



THE MODEL FOR IDENTIFICATION OF A NATURAL LANGUAGE OF THE TEXT

S. Gusev, A. Chepovskiy

Annotation

The article is viewing problem of automatic identification of a natural language of the text. The statistical model of texts in natural language are suggested. The algorithms of definition of a natural language of the text is offered

Key words: statistical model of language, identification of a natural language of the text.



SOFTWARE PACKAGE FOR GENERATING FAMILIES OF TRANSITIVE GRAPHS WITH DEGREE FOUR

A. Neznanov, Y. Starichkova

Annotation

The software package for generating infinite and finite families of connected transitive graphs with degree four is described. Those families cover all known graphs up to 30 vertices. Main feature of software is the catalog of families based on the characteristics of symmetry, structural complexity, and visualization of symmetric diagrams of graphs. The package extends the functionality of integrated research environment «Graph Model Workshop» and is used to solve problems that require synthesis of transitive topologies of systems.

Keywords: structural analysis, graph theory, group theory, transitive graphs, generating, symmetry, structural complexity, algorithm, software, implementation.



ORGANISATIONAL MATURITY APPRAISAL METHOD AND ITS IMPLEMENTATION

A. Biryukov, D. Kletskikh

Annotation

The article deals with organizational maturity evaluation. A method of organizational maturity evaluation based on the Business Process Maturity Model (BPMM) is proposed. The method allows organizations to assess maturity of their business processes. An approach to the implementation of the method is also presented.

Key words: business process management, level of organization maturity, capability maturity model BPMM, SLA, business processes maturity evaluation.



TECHNOLOGY OF REQUIREMENTS MAINTENANCE QUALITY CONTROL IN ORGANIZATIONAL STRUCTURES – MASS SERVICE PROVIDERS

A. Shabanov, M. Arakelian

Annotation

In the paper a technology of maintenance requirements quality control for organizational structures – mass service providers is examined. The technology covers the stages of design and application of information business systems. Such technology is applicable for organizational structures that perform activities based on repetitive services providing.

Keywords: organizational structure, business information system, control, quality of service.



LEGAL CLASSIFICATION OF INTERNET SERVICE PROVIDERS

A. Zharova

Annotation

In the relations associated with application of a digital product in Internet environment various subjects participate. In the article a legal classification of such information subjects as service providers is offered. A comparative legal analysis of the information legislation of Russia, laws of the USA and positions of the Instruction on electronic trade of the European union 2000/31/EC is carried out. It is underlined that introduction of additional criteria for existing legal classification of the persons rendering Internet services, associated with particular services and functions, is necessary.

Keywords: Internet, Internet provider, operator, responsibility of persons, international classification of providers, legislation of the USA, legislation of the European Union.

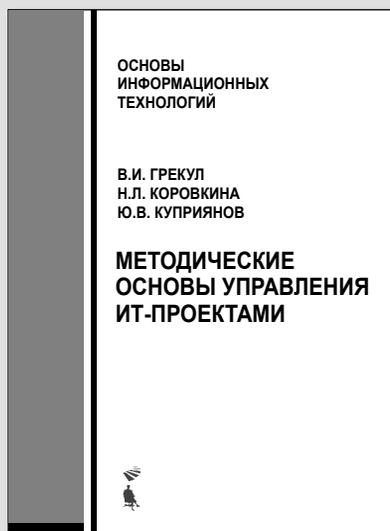
**МЕТОДИЧЕСКИЕ ОСНОВЫ УПРАВЛЕНИЯ
ИТ-ПРОЕКТАМИ**

Учебник

В.И. Грекул, Н.Л. Коровкина, Ю.В. Куприянов

Интернет-университет информационных технологий: Бином.

Лаборатория знаний, 2011.



При создании ИТ-решений перед всеми сторонами, вовлеченными в жизненный цикл проекта, возникает целый ряд вопросов, связанных с определением и детальным структурированием необходимых работ, с распределением прав и обязанностей, с управлением и контролем за исполняемыми работами. Одним из действенных инструментов для решения данных вопросов является использование унифицированных подходов, закрепленных в современных международных и российских стандартах и методологиях управления проектами. Представленный учебник содержит детальное описание процедур управления проектами внедрения информационных технологий. Отличительной особенностью данной книги является изложение материала с привязкой к этапам жизненного цикла создаваемого продукта, а не к фазам некоторого абстрактного проекта. Это позволяет читателю сформировать целостное представление о необходимых в ИТ-проекте управленческих процедурах, а также использовать материал последовательно во времени, по мере перехода от одного этапа технологического цикла создания продукта к другому.

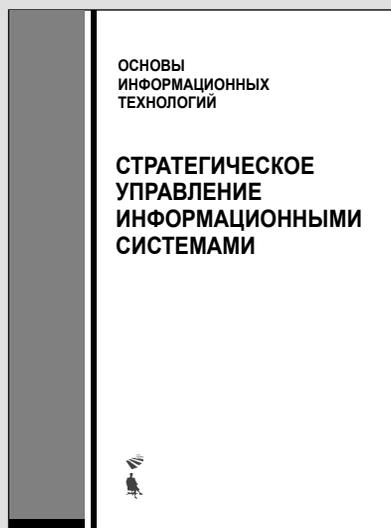
**СТРАТЕГИЧЕСКОЕ УПРАВЛЕНИЕ
ИНФОРМАЦИОННЫМИ СИСТЕМАМИ**

Учебник.

**Р.Б. Васильев, Г.Н. Калянов,
Г.А. Левочкина, О.В. Лукинова**

Под ред. Г.Н. Калянова

*Москва: Интернет-университет
Информационных технологий: Бином.
Лаборатория знаний, 2010.*



Учебник дает читателю представление о методологической базе и современных подходах и методах стратегического управления информационными системами, обеспечивающего целостный, процессно-ориентированный подход к принятию управленческих решений, направленных на повышение эффективности владения и развития информационных систем для достижения бизнес-целей предприятий и создания новых конкурентных преимуществ.

**АВТОМАТИЗАЦИЯ ДЕЯТЕЛЬНОСТИ
ПРЕДПРИЯТИЯ РОЗНИЧНОЙ ТОРГОВЛИ
С ИСПОЛЬЗОВАНИЕМ ИНФОРМАЦИОННОЙ
СИСТЕМЫ MICROSOFT DYNAMICS NAV**

**В.И. Грекул, Н.Л. Коровкина,
Д.А. Богословцев, Н.Н. Синайская**

Москва: Бином. Лаборатория знаний, 2009.

В книге рассмотрены процедуры настройки информационной системы Microsoft Dynamics Navision 4.0 и работа в системе при ее использовании в качестве инструмента автоматизации управления деятельностью торговой компании. Книга предназначена для подготовки пользователей системы, а также может использоваться в качестве пособия при проектировании информационных систем автоматизации торговли. В отличие от традиционных инструкций по применению программных продуктов, материал излагается в контексте выполнения задач производственной деятельности, который задается моделями типичных для отрасли бизнес-процессов.

ОСНОВЫ
ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ

В.И. ГРЕКУЛ
Н.Л. КОРОВКИНА
Д.А. БОГОСЛОВЦЕВ
Н.Н. СИНАЙСКАЯ

**АВТОМАТИЗАЦИЯ
ДЕЯТЕЛЬНОСТИ
ПРЕДПРИЯТИЯ
РОЗНИЧНОЙ ТОРГОВЛИ
С ИСПОЛЬЗОВАНИЕМ
ИНФОРМАЦИОННОЙ
СИСТЕМЫ MICROSOFT
DYNAMICS NAV**



**ПРОЕКТИРОВАНИЕ
ИНФОРМАЦИОННЫХ СИСТЕМ**

В.И. Грекул, Н.Л. Коровкина, Г.Н. Денищенко

Москва: Бином. Лаборатория знаний, 2008.

Курс направлен на изучение совместных методов и средств проектирования информационных систем в сфере экономики. Предусматривается изучение CASE-средств как программного инструмента поддержки проектирования информационных систем. Рекомендовано для студентов высших учебных заведений, обучающихся по специальностям в области информационных технологий.

ОСНОВЫ
ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ

В.И. ГРЕКУЛ
Н.Л. КОРОВКИНА
Г.Н. ДЕНИЩЕНКО

**ПРОЕКТИРОВАНИЕ
ИНФОРМАЦИОННЫХ
СИСТЕМ**



**ЛЕКЦИИ О ПРОЦЕССАХ УПРАВЛЕНИЯ
ИНФОРМАЦИОННЫМИ ТЕХНОЛОГИЯМИ**

Учебное пособие

А.Н. Бирюков

Москва: Интуит.РУ, БИНОМ.ЛЗ, 2010.

Рассматриваются основные процессные модели и методики, связанные с управлением ИТ, появившиеся в последние годы. Основное внимание уделяется анализу их взаимосвязей и выявлению общих концепций и подходов. Изложение в большой степени базируется на оригинальных материалах, не переведенных на русский язык.

ОСНОВЫ
ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ

А.Н. БИРЮКОВ

**ЛЕКЦИИ О ПРОЦЕССАХ
УПРАВЛЕНИЯ
ИНФОРМАЦИОННЫМИ
ТЕХНОЛОГИЯМИ**



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

Расценки:

Обложка: 2, 3, 4 страница обложки, полноцветная печать, полоса 210×290 мм (A4) – 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.

№	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 объем 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 – сверху 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 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, полученных в 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 обзора 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.