

На правах рукописи

УДК 007.51::316.77::[001+378]

ГРЕБНЕВ Андрей Николаевич

**РАЗРАБОТКА МЕТОДОВ САМООРГАНИЗАЦИИ
КОММУНИКАЦИОННЫХ ИНФОРМАЦИОННЫХ СИСТЕМ
В НАУЧНО-ОБРАЗОВАТЕЛЬНОЙ СРЕДЕ**

Специальность 05.13.01 – «Системный анализ,
управление и обработка информации
(в машиностроении и вычислительной технике)»

АВТОРЕФЕРАТ
на соискание ученой степени
кандидата технических наук

Ижевск 2007

Работа выполнена в ГОУ ВПО «Удмуртский государственный университет» (УдГУ)

Научный руководитель: доктор физико-математических наук,
профессор
Бельтюков Анатолий Петрович

Официальные оппоненты: доктор технических наук,
профессор
Кучуганов Валерий Никанорович
г. Ижевск

доктор физико-математических наук,
профессор
Замов Наиль Калимович
г. Казань

Ведущая организация: **ГОУ ВПО «Нижегородский государственный университет» (ННГУ),**
г. Нижний Новгород

Защита состоится 27 марта 2007 г. в 14:00 часов на заседании Диссертационного совета К 212.065.01 при ГОУ ВПО «Ижевский государственный технический Университет» по адресу: 426069, г. Ижевск, ул. Студенческая, д. 7.

Отзыв на автореферат, заверенный гербовой печатью, просим высылать по указанному адресу на имя ученого секретаря диссертационного совета.

С диссертацией можно ознакомиться в библиотеке ГОУ ВПО «Ижевский государственный технический университет».

Автореферат разослан 20 февраля 2007 г.

Ученый секретарь диссертационного совета
кандидат технических наук, доцент _____ Сяктерев В.Н.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность. Три основания развития и процветания экономики страны – это образование, наука, бизнес. В свою очередь важным фактором их развития и интеграции будут социальные коммуникации. Автоматизация образовательных процессов, процессов выработки знаний, бизнес-процессов с помощью информационно-коммуникационных технологий (ИКТ) привела к появлению новых понятий: электронное образование (ЭО)¹, электронная наука (ЭН), электронный бизнес. Базой же для них являются электронные коммуникации (ЭК) и электронный документооборот (ЭДО).

Одним из институтов взаимодействия образования, науки и бизнеса является вуз. Вуз – это та среда, в которой молодежь получает профессиональное образование, приобщается к важным научным исследовательским понятиям (в том числе к конференции и публикации), готовится к вступлению в бизнес.

Научно-образовательные коммуникационные процессы в вузе

	Образовательная среда	Научная среда
Роли	Студент-студент, студент-преподаватель, преподаватель-преподаватель.	Автор-читатель, читатель-читатель, автор-автор; докладчик-слушатель, докладчик-докладчик, слушатель-слушатель.
Документы	Рефераты, курсовые и дипломные работы, конспекты лекций, программы курсов, методические пособия, учебники, учебные пособия.	Тезисы, статьи, доклады, авторефераты, монографии, диссертации.
События	Лекции, лабораторные, практические занятия, коллоквиумы, экзамены, защиты курсовых и дипломных работ.	Конференции, заседания, семинары, симпозиумы, защиты кандидатских и докторских диссертаций.

Рассмотрим коммуникационные процессы (см. табл.), протекающие в вузе в образовательной среде (образовательные коммуникации) и в научной среде (научные коммуникации). В таблице выделены три наиболее важных вида процессов, очевидно, что почти все они нуждаются в автоматизации с использованием ИКТ. Участники с указанными ролями должны иметь возможность взаимодействовать в синхронном (например, чаты) и асинхронном (например, форумы) режиме на удалении друг от друга. В случае документальной коммуникации, должны быть автоматизированы процессы, связанные со всеми этапами жизненного цикла (сбора и создания, управления и использования, хранения, сохранения, доставки и распространения) указанных документов. Автоматизация требуется и в коммуникативных

¹ Theory and Practice of Online Learning. Editors: Terry Anderson & Fathi Elloumi. Athabasca University, 2004. http://cde.athabascau.ca/online_book/

процессах, связанных с событиями, например: информационная поддержка события материалами, регистрация участников событий и т. п.

Таким образом, актуальность исследования обусловлена потребностью научно-образовательного сообщества в ускорении информационных социальных коммуникационных процессов в научно-образовательной среде, посредством автоматизации с помощью ИКТ.

Объект исследования – автоматизация информационных социальных коммуникационных процессов в научно-образовательной среде.

Предмет исследования – методология и методы автоматизации информационных социальных коммуникационных процессов в научно-образовательной среде.

Из объекта и предмета исследования видно, что работа выполнена на стыке социальных коммуникаций и информационных технологий с преобладанием последних.

Цель работы – разработать теоретическую и практическую основу для построения программного комплекса, автоматизирующего (и, как следствие, ускоряющего) информационные социальные коммуникационные процессы в научно-образовательной среде.

Для реализации поставленной цели необходимо решить следующие **задачи**:

- исследование и анализ существующих понятий и технологий социальных информационных коммуникаций (в том числе и электронных) в научно-образовательной среде;
- разработка новых методов автоматизации и ускорения научно-образовательных коммуникационных процессов;
- разработка новой единой концепции научно-образовательной системы и построение ее математической модели;
- разработка программной базы (основы) для построения программного комплекса, автоматизирующего научно-образовательные коммуникационные процессы.

Научная новизна исследования состоит в следующем:

- в результатах системного анализа новых возникающих понятий электронных коммуникаций (ЭК), электронного документооборота (ЭДО), электронного образования (ЭО), электронной науки (ЭН), выявлении их роли в научно-образовательных коммуникационных процессах;
- в адаптации современных направлений ЭК, ЭДО и методов веб 2.0 (таких как персонализация, коллаборация, децентрализация) к традиционным понятиям публикации и конференции;
- в предложенной трехуровневой модели эволюции понятий публикации и конференции;
- в разработанной математической модели метода взаимодействия «многие ко многим» через среду (стигмергия), в основе которого лежат человеко-базируемый генетический алгоритм и обучающийся веб (коллаборативная фильтрация и коллективный интеллект);
- в выработке концепции научно-образовательной самоорганизующейся метасистемы и требований к ее подсистемам (комплексам).

Полученные результаты дадут возможность дальнейшей разработки проблемы.

Практическая значимость работы обусловлена разработанным и удачно апробированным пилотным проектом AtLearn, который является базой для построения программного комплекса (подсистемы метасистемы), автоматизирующего социальные коммуникационные процессы в научно-образовательной среде. Разработанный каркас позволяет в значительной степени сократить временные и финансовые затраты на разработку программного комплекса.

Основными результатами, выносимыми на защиту, являются:

- результаты системного анализа новых возникающих понятий электронного образования (ЭО), электронной науки (ЭН), электронных коммуникаций (ЭК), электронного документооборота (ЭДО) и их технологий, выявления их роли в научно-образовательных коммуникационных процессах;
- адаптация современных направлений ЭК, ЭДО и методов веб 2.0 (таких как персонализация, коллаборация, децентрализация) к традиционным понятиям публикации и конференции;
- предложенная трехуровневая модель эволюции понятий публикации и конференции;
- разработанная математическая модель метода взаимодействия «многие ко многим» через среду (стигмергия), в основе которого лежат человеко-базируемый генетический алгоритм и обучающийся веб (коллаборативная фильтрация и коллективный интеллект);
- выработанная концепция научно-образовательной самоорганизующейся метасистемы и требования к ее подсистемам (комплексам);
- разработанный и успешно апробированный пилотный проект AtLearn, который является основой (каркасом) для построения программного комплекса (подсистемы метасистемы), автоматизирующего социальные коммуникационные процессы в научно-образовательной среде.

Методы исследования:

- методы системного анализа, теории множеств и теории графов;
- методы многоагентных систем, роевого и коллективного интеллекта, стигмергии, теории сложных самоорганизующихся систем и синергетики;
- CASE-технологии, методологии разработки программного обеспечения (RUP, Agile), объектно-ориентированные подходы в проектировании и программировании.

Апробация работы. Основные результаты работы докладывались и обсуждались на следующих конференциях, семинарах и заседаниях:

- XL Международной научной студенческой конференции «Студент и научно-технический прогресс», Новосибирск, 2002;
- 36-й Региональной молодежной конференции «Проблемы теоретической и прикладной механики», Екатеринбург, 2005;
- Третьей открытой Всероссийской конференции «Преподавание информационных технологий в РФ», Рыбинск, 2005;

- Конференции «Технологии Microsoft в теории и практике программирования», Нижний Новгород, 2006;
- XIII Всероссийской научно-методической конференции «Телематика-2006», Санкт-Петербург, 2006;
- на семинарах и заседаниях кафедр «Математического обеспечения ЭВМ» и «Теоретических основ информатики» ГОУ ВПО «Удмуртский государственный университет», Ижевск;
- на расширенном заседании кафедры «Информационные системы» ГОУ ВПО «Ижевский государственный технический Университет», Ижевск, 2006.

За теоретические разработки по данной тематике получены следующие дипломы:

- диплом второй степени за работу, представленную на Международной научной студенческой конференции «Студент и научно-технический прогресс», Сибирское отделение РАН (Новосибирск, 2002);
- диплом Министерства образования РФ по итогам открытого конкурса 2003 года на лучшую научную студенческую работу по естественным, техническим и гуманитарным наукам в вузах РФ (Москва, 2003).

За практическую реализацию проекта AtLeap получены следующие дипломы:

- диплом за второе место на конкурсе Java-проектов (<http://www.javakonkurs.ru>) в категории J2EE, проводимом корпорацией Sun Microsystems при официальной поддержке Министерства по информационным технологиям и связи РФ (Москва, 2005);
- диплом за второе место на Всероссийском конкурсе разработчиков программного обеспечения IntelCup (<http://www.intelcup.ru>), проводимом корпорацией Intel при поддержке Нижегородского государственного университета (Нижний Новгород, 2005);
- диплом за первое место за самое высокое качество реализации на Открытом дальневосточном конкурсе программных средств студентов, аспирантов и молодых специалистов «Программист-2006» (<http://www.contest.dvo.ru>) (Владивосток, 2006);
- диплом за третье место в номинации «Лучшая система управления сайтом» на открытом конкурсе молодежных Интернет-проектов, компьютерной графики и анимации «Золотая паутина-2006» (<http://www.goldweb.org>) (Омск, 2006);
- диплом за победу в номинации «Инструментарий» на конкурсе IBM WebSphere Community Edition 2006 (<http://linuxformat.ru/contest/was2006.phtml>) (Москва, 2006).

Публикации. Основное содержание диссертации опубликовано в 12 работах (из них 5 статей, из которых 1 в списке изданий, рекомендуемых ВАК).

Структура работы. Работа состоит из введения, трех глав, заключения и списка использованной литературы. Общий объем работы при сквозной нумерации составляет 142 страницы, рисунков 17, таблиц 4, библиография содержит 188 названий.

Достоверность и обоснованность полученных результатов исследования и вытекающих из них выводов обеспечивается использованием фундаментальных

положений теории сложных самоорганизующихся систем и искусственного интеллекта. Достоверность результатов практической реализации подтверждается наличием пяти дипломов за победы на конкурсах и достаточно большим количеством пользователей системы из многих стран мира.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обосновывается актуальность темы диссертационной работы, формулируются объект и предмет исследования, цель и задачи, научная новизна и практическая значимость полученных результатов, дается информация об апробации работы.

В первой главе был проведен системный анализ понятий электронных коммуникаций (ЭК), электронного документооборота (ЭДО), электронного образования (ЭО), электронной науки (ЭН), показана их роль и взаимосвязь.

Рассмотрена электронная социальная коммуникация, также называемая компьютерно-опосредованная коммуникация (Computer-Mediated Communication, СМС). Рассмотрен важный частный случай СМС – совместная деятельность, другие названия – совместная работа при поддержке компьютера (CSCW, Computer-Supported Cooperative Work) или коллаборация при поддержке компьютера (CSC, Computer-Supported Collaboration). Описаны и даны отличия между 3 терминами совместной деятельности: координация (coordination), кооперация (cooperation) и коллаборация (collaboration). Коллаборация – это сотрудничество более чем одной персоны для достижения общей цели при наивысшей степени индивидуальной свободы и в тоже время взаимозависимости между участниками. Коллаборация – это динамический процесс, в котором нет ясного определения ролей и механизмов решения проблемы, при этом часто используются эвристические методы. Дается обзор актуальных направлений электронных коммуникаций: коллаборативная фильтрация (collaborative filtering), коллаборативная маркировка (collaborative tagging) и социальные сети (social networks). Приведены примеры коммуникационного программного обеспечения (ПО). Описана новая эра развития всемирной паутины, называемая веб 2.0, заключающаяся в значительном усилении роли социальной составляющей в виде методов персонализации, коллаборации, децентрализации.

Рассмотрен частный случай ЭК – электронный документооборот, таким образом, рассмотрена информация с точки зрения ее жизненного цикла в электронной (цифровой) среде, т. е. сбора и создания, управления и использования, хранения, сохранения, доставки и распространения. Даны определения электронного документа (ЭД) и электронного издания (ЭИ). Важный упор был сделан на электронные библиотеки (ЭБ, Digital Library) как одно из основных направлений ЭДО. Даны примеры технологий, используемых при сохранении, систематизации, обмене, поиске и идентификации документов. Описаны стандарты метаданных и форматов документов. Приведены примеры реализации ЭБ.

Дано определение онлайн-образования и различия в его терминологии. Рассмотрена структура системы дистанционного образования. Описаны стандарты образовательных объектов. Приведены примеры ПО в виртуальной обучающей среде.

В разделе об электронной науке дано описание понятия колаборатории (col-laboratory), среды для решения проблем (СПП, Problem Solving Environment, PSE) и грид (GRID) среды.

Дается сравнение системы открытого образования Openet, единого научного информационного пространства РАН, коммуникационных пространств (Microsoft Spaces, Yahoo 360) и грид-технологий, как наиболее ярких примеров из ЭО, ЭД, ЭК, ЭН соответственно. Показаны их недостатки.

Показано, что на данный момент существуют лишь отдельные технологии и подходы, специфичные для каждого из направлений, не учитывающие неотрывность науки и образования и не использующие достижения современных методов в виде персонализации, коллаборации и децентрализации. Показана необходимость частичного перехода от процессов администрирования и управления, требующих огромных затрат, к модерированию и самоуправлению (самоорганизации).

Таким образом, обоснована необходимость выработки новой концепции автоматизации социальных коммуникационных процессов в научно-образовательной среде (НОС).

Во второй главе впервые предложена концепция метасистемы, автоматизирующей социальные коммуникационные процессы в НОС.

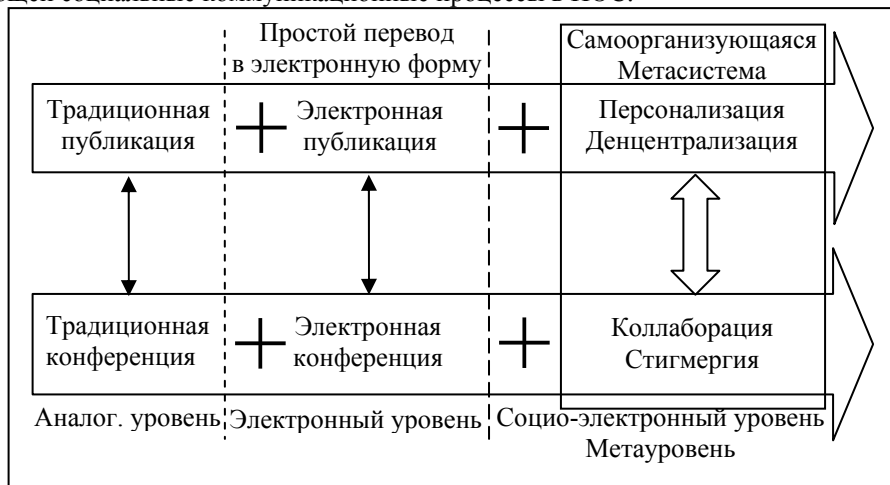


Рис. 1. Модель эволюции понятий публикации и конференции

В НОС два основных вида коммуникации представляются понятиями публикация (тезисы, статья и т. п.) и конференция (форум, симпозиум и т. п.). Развитие ИКТ позволяет перейти от классической публикации к электронной публикации, от классической конференции к электронной конференции. Никто не говорит о замене классических способов электронными – например, даже современные способы коммуникации все еще не могут заменить живое общение. Речь идет о дополнении классической формы коммуникации электронной формой (см. рис. 1).

Добавление электронной формы публикации и конференции привносит множество преимуществ. Однако в настоящее время этого недостаточно и существует возможность получить еще большую отдачу от ИКТ, примененных в НОС. В кон-

цепции современного развития всемирной паутины Тима Орейли (Tim O'Reilly) веб 2.0² показана всевозрастающая роль социальной составляющей в коммуникациях в виде персонализации, коллаборации, децентрализации. Под персонализацией понимается адаптация системы под желания и возможности конкретного пользователя, причем в большей степени на основании персональных данных, полученных от пользователя неявно. Технологии веб 2.0 могут и должны быть адаптированы к процессам электронной публикации и конференции.

Технологии персонализации и коллаборации уже присутствуют в разных областях. Существуют реализованные разработки ассоциативно-статистического подхода анализа текста, позволяющие проводить автоматическое реферирование, тематическую классификацию и кластеризацию текстов, смысловой поиск, автоматическое построение гипертекста и т. д. Для работы с цитатами и ссылками предназначено библиографическое ПО. Для персональной базы знаний хорошо подходит технология вики. Основу вики составляет граф, где узлы помечены ключевыми словами (терминами) и ассоциированными с ними словарными статьями. Вики подходит для представления научных знаний в виде системы понятий (терминов). Статьи другой современной технологии блог могут являться аналогией классического понятия тезисов (препринт). Эти и другие технологии образуют персональное АРМ (Автоматизированное рабочее место) исследователя, главным компонентом которого является персональная электронная библиотека. Становится заметен переход общих универсальных систем к персонализированным, переход от централизованных систем к децентрализованным.

С точки зрения коллаборации важными являются технологии: коллаборативной фильтрации, фолксономии, социальных сетей и т. п. Коллаборативная фильтрация, например, ссылок, книг, журналов, статей, персоналий позволит автоматически строить ассоциативные связи между этими важными сущностями. Фолксономия на основании выставленных пользователями ключевых слов позволит автоматически классифицировать опять же ссылки, книги, журналы, статьи, персоналии и т. п. Социальная сеть исследователей позволит видеть научные школы.

В добавление к уже достаточно давно существующему понятию импакт-фактор (impact factor) предложено определять рейтинг информации не только на количестве обратных ссылок, но и на: количестве просмотров; количестве сообщений в форуме, ассоциированным с документом (автором, журналом); количестве участников в чате, форуме; рейтинге участников.

Итак, если на весь комплекс вышеизложенных механизмов посмотреть извне, то можно увидеть, что система самостоятельно способна: агрегировать знания (вики на примере Wikipedia); перерабатывать существующие и вырабатывать новые знания (обсуждение в блоге и форуме); классифицировать знания и строить взаимосвязи между понятиями (коллаборативная фильтрация, фолксономия, социальные се-

² O'Reilly Tim. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. O'Reilly. September 30, 2005. <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>

ти); выявлять актуальность знания с точки зрения сообществ (импакт-фактор). Ввиду этого видны признаки искусственного интеллекта (ИИ) и самоорганизации.

Далее в работе рассмотрены человеко-базированный генетический алгоритм (HBGA, Human-based genetic algorithm) как одно из направлений эволюционных вычислений в ИИ. Алгоритм HBGA – это генетический алгоритм, где генетические операторы (инициализация, мутация, кроссовер и селекция) поручены людям – при этом используется множество агентов (многоагентная система). Рассмотрен эволюционный подход к моделям передачи информации, называемый меметика и основанный на концепции мемов (meme), рассматривающей идеи как единицы культурной информации, распространяемые между людьми посредством имитации, научения и т. п. Определено понятие самоорганизации из теории сложных систем и его связь с кибернетикой и синергетикой. Описаны понятия коллективного интеллекта (collective intelligence) и мирового мозга (Global Brain). Описаны понятия роевого интеллекта (swarm intelligence) и стигмергии (stigmergy) на примере самоорганизации общественных насекомых (муравьев, термитов). Взаимодействия делят на прямые и косвенные. Прямые взаимодействия – это «очевидные» взаимодействия: сенсорное исследование, визуальный контакт, химический контакт и т. п. Две особи взаимодействуют косвенно, когда одна из них модифицирует окружающую среду, а вторая отвечает на изменения по прошествии некоторого времени. Такое взаимодействие есть пример стигмергии.

Приведены примеры стигмергии в веб 2.0: блог, вики, алгоритм PageRank поисковой системы Google. Обучающийся веб (Learning Web) приведен как пример со стыка коллаборативной фильтрации и коллективного интеллекта.

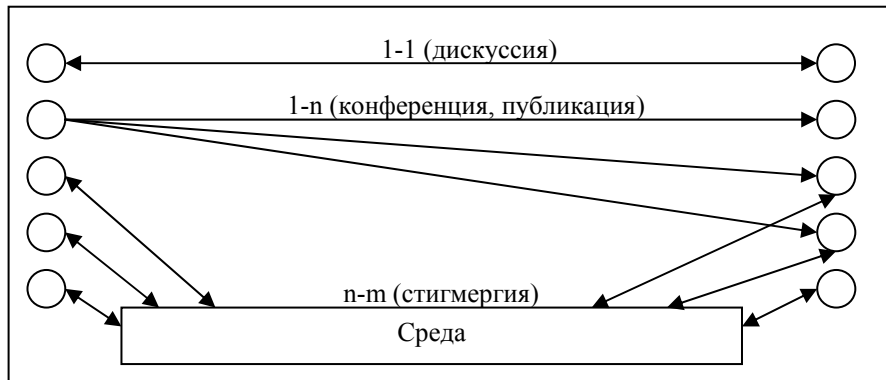


Рис. 2. Три метода взаимодействия исследователей

Современные темпы развития общества требуют все более тесного взаимодействия все большего количества исследователей. Классические формы взаимодействия, будь то «один к одному» или «один ко многим» в виде конференции или публикации, даже при переводе в электронную форму не могут покрыть эти потребности. Необходим механизм, который позволял бы параллельно взаимодействовать большому количеству исследователей в режиме «многие ко многим». Таким механизмом является взаимодействие через изменения в среде – стигмергия (см.

рис. 2). Стоит отметить, что при конференции и публикации значительный поток информации идет от докладчика к слушателям, от автора к читателям, обратная связь незначительна или отсутствует вовсе. В случае стигмергии все исследователи имеют равный доступ к среде как на чтение, так и, что особенно важно, на запись (публикацию) информации.

Предлагается многоагентная система, где агентами выступают сами исследователи (люди). Агенты взаимодействуют как с помощью уже распространенных методов коммуникации «один к одному» (например, дискуссия) или «многие ко многим» (например, конференция и публикация), так и с помощью стигмергии, т. е. взаимодействия «многие ко многим» через среду.

Рассмотрим множество людей-коммуникаторов, взаимодействующих с системой, т. е. множество агентов A . Рассмотрим отношение посылки информационного сообщения $send$. В случае коммуникации:

- многие ко многим – $send$ будет отношение на множестве A ,
- один ко многим – функция $A \rightarrow A$,
- один к одному – биекция $A \rightarrow A$.

Согласно утверждению Марка Эллиота³, в случае если $|A| > 25$ и отношение A симметрично, т. е. сообщения отсылаются в обе стороны, и каждый коммуникатор является как получателем, так и отправителем, то появляется необходимость в косвенном взаимодействии через среду – стигмергия.

Среда – это направленный граф $Env = G(R, L)$, где: R – это не пустое множество ресурсов (в частном случае это документ или даже веб-страница), являющихся узлами графа, $L \subset R \times R$ – это множество связей (ссылок), являющихся дугами графа.

Рассмотрим множество персональных страниц агентов $P = \{r \in R \mid \exists a \in A : pp(a) = r\}$, где $pp : A \rightarrow R$ – это функция отображения агентов к их персональным страницам, т. е. $P \subset R$. Тогда социальной сетью будет подграф $SN = G'(P, L')$, $L' \subset P \times P$.

Каждый ресурс $r = \langle V, cv, h \rangle$, где V – это множество версий (редакций) ресурса, $cv \in V$ – это текущая версия ресурса, $h \in H$ – это уникальный идентификатор ресурса, например URI. Такая структура позволяет хранить всю историю изменений ресурса, и в случае необходимости «откатиться» к одной из старых версий.

Версия $v \in V$, $v = \langle in, metain \rangle$, где $in \in IN$ – информация ресурса, а $metain \in METAIN$ – это метаинформация ресурса. Информация – это последова-

³ Elliott Mark. Stigmergic Collaboration: The Evolution of Group Work. M/C Journal. Volume 8, Issue 2, May 2006. <http://bitworking.org/news/Stigmergy>

тельность мемов (идей) $m \in M$, заданная функцией зависящей от агента, т. е. субъекта.

Как уже было отмечено, агенты взаимодействуют со средой, как на чтение, так и на запись. Запись (публикация) информации в среду осуществляется в соответствии с человеко-базируемым генетическим алгоритмом, где операторы определяются следующим образом:

- Инициализация – это создание нового мема и/или создание ресурса.
- Мутация – модификация мема.
- Кроссовер – обмен мемами между ресурсами.
- Селекция – выбор одной из старых версий ресурса и/или удаление ненужного мема.

Дуги (связи) графа G помечены весами связности: $s : L \rightarrow W$. В соответствии с концепцией обучающегося веба существуют три правила:

- если $(r_i \rightarrow r_j)$, то $s(l_{ij}) + C_1$, т. е. при прохождении агента по ссылке вес ссылки увеличивается;
- если $(r_i \rightarrow r_j) \& (r_j \rightarrow r_k)$, то $s(l_{ik}) + C_2$, т. е. при прохождении агента по ссылкам либо создается новая транзитивная связь, либо увеличивается вес уже существующей связи;
- если $(r_i \rightarrow r_j)$, то $s(l_{ji}) + C_3$, т. е. при прохождении агента по ссылке либо создается новая симметричная связь, либо увеличивается вес уже существующей связи.

Таким образом, получаем самоорганизующуюся среду накопления, выработки, классификации и проверки актуальности знаний.

Турчин В.Ф.⁴ говорит об эволюционной кибернетике, эволюции информационных систем и о метасистемном переходе, когда некоторое число систем интегрируются в единое целое с возникновением нового уровня управления, – таким образом, новая система есть метасистема по отношению к старым. В случае публикации и конференции (см. рис. 1) явно показана эволюция этих понятий.

Предложена концепция единой территориально распределенной самоорганизующейся научно-образовательной метасистемы, состоящей из типовых независимых программно-информационных подсистем (комплексов). Программные комплексы взаимосвязаны по иерархическому принципу, т. е. субъект (отдельный пользователь, подразделение, институт и т. д.) либо устанавливает и поддерживает такую систему самостоятельно, либо использует сервисы программного комплекса вышележащих уровней иерархии.

Целью метасистемы является ускорение социальных информационных коммуникационных процессов в НОС посредством их автоматизации с помощью ИКТ. Для достижения поставленной цели комплекс должен решать следующие задачи:

⁴ Турчин В. Ф. Феномен науки: Кибернетический подход к эволюции. – Изд. 2-е. – М.: ЭТС, 2000. – 368 с.

автоматизация процессов обмена информационными сообщениями между участниками, разнесенными во времени и/или пространстве; автоматизация всех этапов жизненного цикла документов; автоматизация процессов, связанных с событиями.

Концепция метасистемы строится на принципах эволюции двух основных понятий: публикации и конференции. Переход на социо-электронный метуровень эволюции сопровождается появлением социальных свойств: персонализации и коллаборации, самоорганизации и стигмергии.

Итак, формальное определение научно-образовательной метасистемы, являющееся сутью концепции данной работы, будет выглядеть следующим образом. Научно-образовательная метасистема – это метасистема, состоящая из управляемых ею многих однородных подсистем A_1, A_2, \dots, A_n , организованных в виде корневого дерева, где каждая A_i – это тройка:

- модель информационных каналов Шеннона и Уивера⁵ для коммуникации «один к одному»;
- модель электронной библиотеки Маркоса Андре Гонсалеса, Эдварда Фокса и др.⁶ как пример коммуникации «один ко многим»;
- предложенная автором модель среды (см. выше Env) для взаимодействия «многие ко многим» (стигмергия).

На основании предложенной концепции выработаны и обоснованы общие требования к типовой подсистеме (программному комплексу) метасистемы: мобильность (переносимость, portability) программного обеспечения, масштабируемость (scalability) программного обеспечения, интероперабельность (interoperability), мобильность пользователей (users mobility), открытые стандарты (open standards), открытое программное обеспечение (open software), интернационализация (internationalization), расширяемость (extensibility), простота использования.

Обозначено, что в настоящий момент на рынке ПО не существует решения, удовлетворяющего всем перечисленным требованиям. Следовательно, необходимо начать разработку программного комплекса, отвечающего всем выдвинутым требованиям. Разработку следует начать с разработки фундамента (каркаса) будущего программного комплекса.

В третьей главе рассмотрен реализованный и успешно апробированный пилотный проект AtLeap, который является базой (каркасом) для построения типового программного комплекса (подсистемы метасистемы). Система AtLeap удовлетворяет большинству требований, обозначенных в главе 2. Обоснован выбор платформы J2EE (Java 2 Enterprise Edition) в сравнении с .NET.

Проект AtLeap был инициирован автором в середине 2004 года. AtLeap – 100% Java-каркас (framework) с открытым исходным кодом (под лицензией Apache

⁵ Shannon Claude E. and Weaver Warren: The Mathematical Theory of Communication. The University of Illinois Press, Urbana, Illinois, 1949. ISBN 0-252-72548-4.

⁶ Gonçalves Marcos André, Fox Edward. A., Watsom Layne T., Kipp Neill A. Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries. Technical Report TR-01-12, Virginia Tech, Blacksburg, VA, 2001. <http://www.dlib.vt.edu/projects/5S-Model/5s6.pdf>

2.0) для быстрого создания веб-приложений на его основе. Проект AtLeap (<https://atleap.dev.java.net>) разрабатывается на сайте java.net сообщества (в своем роде аналог freshmeat.net или sourceforge.net). Проект AtLeap – это веб-приложение, работающее на основе некоторых стандартов J2EE 1.4 платформы, включая: JSR 152, JSR 154. Каркас AtLeap, с одной стороны, содержит в себе большой набор уже интегрированных друг в друга библиотек, с другой стороны, – многоязычную систему управления содержимым (СУС) с полнотекстовым поиском. При разработке проекта использовались CASE-технологии, методологии разработки ПО (RUP, Agile), объектно-ориентированные подходы в проектировании и программировании.

Каркас AtLeap инкапсулирует в себе более 50 различных библиотек – это Hibernate – объектный язык запросов (HQL) и персистентное хранение объектной модели приложения с ее отображением в таблицы реляционной СУБД; Spring – реализация паттерна проектирования Inversion of Control (IoC) и аспектно-ориентированного стиля программирования (AOP); Ant – система сборки приложений на XML; Struts – реализация слоя представления на основе второй модели паттерна проектирования MVC (Model View Controller); Lucene – поддержка полнотекстового поиска; XDoclet – реализация атрибутно-ориентированного программирования и т. д.

Система AtLeap обеспечивает независимость:

- от аппаратной платформы и операционной системы, за счет использования Java-машины, – она протестирована на 3-х JVM (Sun, Bea, IBM), не учитывая версии;
- от СУБД, за счет использования персистентного слоя, – протестирована на 7-и СУБД (MySQL, PostgreSQL, Oracle, DB2, MSSQL, Hypersonic, Derby), не учи-

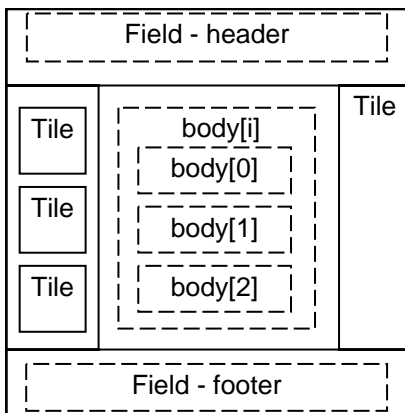


Рис. 3. Пример размещения плиток и полей на странице

тывая версии;

- от servlet-контейнера, за счет использования J2EE стандартов, – протестирована и имеет скрипты для интеграции с 11-ю серверами приложений (Apache Tomcat, Caucho Resin, Bea Weblogic, IBM WebSphere AS, IBM WebSphere AS Community

Edition, JBoss, ObjectWeb JOnAS, Geronimo, Jetty, Oracle AS, Sun Java System AS), не учитывая версии;

- от браузера – протестирована на 6-и браузерах (Internet Explorer 5.5+, Mozilla 1.3+, FireFox 1.0+, Netscape 7+, Opera 9+, Safari 2+).

В СУС AtLeap для генерации веб-страниц используется библиотека шаблонов Tiles входящая в поставку Struts. Основной сущностью Tiles является Tile (плитка), «строительный» блок, из плиток складывается результирующая страница (см. рис. 3). Плитки могут вкладываться друг в друга и наследоваться, переопределяя вложенные плитки. Плитки описываются в XML-файле, каждой плитке может соответствовать шаблон (template) – JSP-файл, описывающий размещение вложенных плиток с помощью тегов. Каждая страница СУС генерируется на основе макета (Layout), разрабатываемого дизайнером и верстальщиком (см. рис. 4). Макет соответствует одному определению (definition) плитки (tile), таким образом, макет может собираться из нескольких шаблонов (JSP-страниц) путем композиции и наследования плиток. Редакторам сайта вмешиваться в дизайн непозволительно, разработчики же, по причинам гибкости и скорости работы, предпочтут визуальному редактору непосредственное редактирование кода (HTML или CSS). Следовательно, через HTML-редактор необходимо позволять редактировать только отдельные части страницы. Эти места в макете называются полями (Field) – они определяются разработчиками и помечаются уникальными в рамках макета идентификаторами. Поля бывают 3 типов: однострочный текст; многострочный текст; HTML, редактируемый через визуальный WYSIWYG-редактор.

Поля содержат значения (Value) в кодировке UTF-8 для каждого языка, зарегистрированного в системе. Отдельно выделяется особый вид поля – индексированное. Оно позволяет динамически повторять блоки, достаточно сложные для редактирования в визуальном редакторе. Как было отмечено, плитки могут наследоваться, а значит, могут наследоваться и макеты, переопределяя значения полей с одинаковыми идентификаторами или добавляя новые поля.

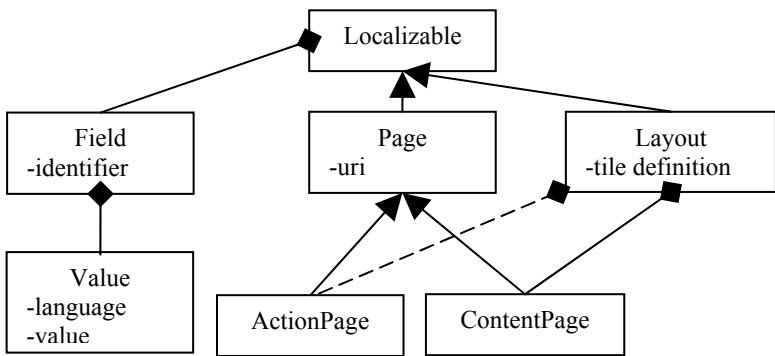


Рис. 4. Часть объектной модели идеологии СУС в AtLeap

В AtLeap существуют два понятия страниц: статическая страница (Content Page) и динамическая страница (Action Page). Статическая страница основывается

на макете – она полностью виртуальная, хранится в БД. Статическая страница может переопределять значения полей своего макета. Динамическая страница – это результат выполнения Struts действия (action). Это позволяет редактировать содержимое даже на динамически генерируемых (например, на основе данных из БД) страницах. Результатом может быть одна из нескольких страниц, поэтому макет, на котором основывается динамическая страница, определяется только во время выполнения. Динамическая страница может переопределять значения полей макета или/и добавлять новые поля.

Большое значение уделено простоте использования. Пользователь может перемещаться по сайту и с помощью двойного клика мыши или контекстного меню выбирать область для редактирования (в том числе и на страницах с динамическими данными из БД). Система AtLear поддерживает богатые средства форматирования текста (изменение шрифтов и цветов, вставка списков, таблиц, изображений, ссылок, а также импорт текста из Microsoft Word) за счет использования одного из двух WYSIWYG (What You See Is What You Get) DHTML-редакторов: FCKEditor или TinyMCE.

Каркас AtLear позволяет помещать ресурсы (изображения, документы и вообще файлы) в систему, с сохранением их в БД. Система AtLear имеет возможности построения меню, независимого от структуры страниц. Начальная структура меню определяется в конфигурационном XML-файле. Элементы меню могут быть переопределены (или добавлены) на одном из макетов или странице.

Проект AtLear имеет возможность хранить содержимое на нескольких языках одновременно, при этом 100% данных контента хранятся в БД. Текстовые сообщения пользовательского интерфейса административной консоли приложения переведены на 6 языков: английский, русский, китайский, немецкий, итальянский и испанский. Система AtLear генерирует дружелюбные для внешних поисковых систем (Google, Yandex и т.д) идентификаторы ресурсов (URL).

Встроенная функция полнотекстового поиска позволяет искать на 13-и языках (с учетом окончаний и стоп-слов). Система может индексировать документы в форматах: RTF, PDF, Word, Excel, PowerPoint, HTML, XHTML и т. д.

Каркас имеет собственную богатую библиотеку тегов (более 60), которая реализует возможности списков сущностей (grid), меню, контекстное меню и др.

Высокие требования предъявляются к стабильности. Используются транзакции БД (откаты и пессимистичные блокировки) и оптимистичный контроль параллелизма (транзакции приложения). Большие тексты будут сохранены, даже если закончилась HTTP-сессия. В случае одновременного редактирования больших текстов у пользователя имеется возможность слияния версий (merge). Корректность данных форм проверяется на сервере и на клиенте. Поддерживаются двунаправленные связи между ссылками и ресурсами для предотвращения появления «сломанных ссылок».

С использованием библиотеки DBUnit реализуется возможность резервного копирования БД, в независимый от СУБД, XML-формат непосредственно через Административную консоль. Поддерживается настройка автоматического создания резервной копии.

Для обеспечения быстродействия применяется система 3-х уровневой кэширования на сервере и серверная поддержка кэширования в браузере (expiration and validation).

В системе реализованы тернарные связи между сущностями «Пользователь», «Роль» и «Группа». Почти каждому действию в системе присвоена своя роль (более 130), что обеспечивает возможность гибкой настройки прав доступа. Из возможностей обеспечения безопасности следует отметить наличие поддержки HTTPS и функции «Запомнить меня».

Проект AtLeap имеет достаточно хорошую документацию – более 150 страниц в Руководстве администратора и Руководстве разработчика. Документация генерируется из формата DocBook в форматы PDF и HTML. Проект AtLeap содержит более 5-и мегабайт исходного кода. Все классы и методы снабжены внутренней документацией (javadoc). Качество исходного кода контролируется утилитами CheckStyle и PMD.

На примере сущности «Пользователь» реализовано автоматическое тестирование на всех слоях приложения, в том числе и тестирование веб-интерфейса с помощью JUnit, JMock, DBUnit, HTMLUnit.

В системе реализован сбор статистики посещений сайта пользователями. Она позволяет собирать и анализировать: типы браузеров и операционных систем посетителей, их разрешение и глубину цвета их мониторов; страницы входа и выхода посетителей, пути следования по сайту; внешние страницы, с которых пришли посетители; поисковые системы и ключевые слова поиска, которые использовали посетители, чтобы попасть на сайт, и др.

AtLeap позволяет автоматически строить карту сайта на основании иерархии созданных страниц и ресурсов.

В системе реализованы следующие модули: «Новости», «Рекомендации», «Свяжитесь с нами», «Форум», «Анкетирование».

Приведены результаты тестирования производительности на конфигурации: процессор: AMD Athlon XP 1.1 GHz; оперативная память: 1 Gb; ОС: Windows XP sp2; JDK: Sun JDK 1.5.0_05; Servlet-контейнер: Apache Tomcat 5.5.15; СУБД: MySQL 5.0.15; AtLeap 0.51. Объектом тестирования была статическая страница (content page) About на 4 языках, т. е. 4 различные страницы, размером приблизительно 32 килобайта каждая, содержащие следующие компоненты: три поля, одно индексированное поле, меню, компонент выбора языков, компонент поиска по сайту. В качестве инструмента тестирования использовался Apache JMeter. Все три продукта – JMeter, Tomcat, MySQL – запускались на одной и той же машине, что положительно сказалось на уменьшении задержки при передаче пакетов по сети и отрицательно на объеме доступной оперативной памяти и вычислительных ресурсов процессора. Последнее особенно важно, т. к. JMeter является тоже Java-приложением. JMeter запускался со следующими параметрами: количество потоков 700, время запуска потоков 120 секунд, гауссовый таймер случайных чисел (от 10 до 20 сек.). Было проведено два различных теста. Первый тест осуществлялся с включенным серверным кэшированием всех компонентов (плиток) страницы, т. е. обращения к СУБД не происходило. Однако сборка страницы осуществлялась динамически, что остав-

ляет возможность включать бизнес-логику в страницу. Серверная поддержка кэширования страниц браузером была выключена. Во время выполнения теста нагрузка на процессор в среднем распределялась как 50% на Tomcat, 25% на JMeter и 25% на остальные приложения, запущенные в системе, т. е. процессор был загружен на 100%. Второй тест проводился с включенным кэшированием всей страницы целиком. Кэширование осуществлялось с использованием фильтра OSCache. Серверная поддержка кэширования страниц браузером была выключена. Во время выполнения теста нагрузка на процессор в среднем распределялась как 20% на Tomcat, 10% на JMeter и 10% на остальные приложения запущенные в системе, т. е. общая нагрузка процессора была приблизительно 40%. Тесты динамических страниц не применялись, по причине того, что их время выполнения сильно зависит от логики выполняемой задачи. Медиана первого теста (2,625 секунды) является вполне приемлемой для динамически собираемой страницы при семистах одновременно работающих пользователей. Второй тест охватывает наиболее часто используемый способ подачи страниц. Из его результатов видно, что среднее время выполнения запроса в 15 миллисекунд является отличным результатом.

Стоимость разработки проекта AtLeap была оценена независимыми компаниями www.koders.com и www.ohloh.net в 0,6 миллиона долларов США. В качестве примера использования можно привести сайт крупной китайской компании www.chinistar.cn или сайт испанской ИТ-компании www.eusa.net, которая обладает ISO 9001:2000, CMMI Maturity Level 4 Production. К проекту проявляют интерес пользователи и разработчики со всего мира, среднее количество скачиваний дистрибутива в неделю составляет 150. Демонстрационная версия системы доступна на бразильском хостинге, спонсируемом корпорацией Sun Microsystems, по адресу: <http://javanet1-ws.locaweb.com:8080/atleap/>.

Утверждается, что объектная модель, разработанная автором и примененная в AtLeap, позволила обеспечить как гибкость и расширяемость приложения, так и простоту редактирования содержимого. На момент написания данной работы, бесплатных решений на платформе J2EE с открытым кодом подобного рода нет, т. е. проект AtLeap является уникальным.

ОСНОВНЫЕ ВЫВОДЫ И РЕЗУЛЬТАТЫ РАБОТЫ

1. Проведен системный анализ новых возникающих понятий: электронное образование (ЭО), электронная наука (ЭН), электронные коммуникации (ЭК), электронный документооборот (ЭДО) и их технологий. Выявлены их роли в научно-образовательных коммуникационных процессах. Приведены примеры бесплатных и коммерческих конкретных программных продуктов и/или сервисов из всех четырех направлений. Проведенный анализ позволил выявить недостаточность простого перевода в электронную форму, без усиления социальной составляющей.
2. Адаптированы современные направления ЭК, ЭДО и методы веб 2.0 (такие как персонализация, коллаборация, децентрализация) к традиционным понятиям публикации и конференции. Разработана трехуровневая модель эволюции этих

- понятий: аналоговый уровень, электронный уровень, социо-электронный уровень (метауровень).
3. Разработана математическая модель метода взаимодействия «многие ко многим» через среду (стигмергия), в основе которого лежат человеко-базируемый генетический алгоритм и обучающийся веб (коллаборативная фильтрация и коллективный интеллект). Модель позволяет организовывать параллельное взаимодействие больших групп людей и усилить обратную связь.
 4. Выработана концепция научно-образовательной самоорганизующейся метасистемы, где каждая подсистема определяется как тройка методов взаимодействий: «один к одному», «один ко многим» и «многие ко многим» (стигмергия). Такой подход позволяет ускорить коммуникационные процессы посредством усиления социальной составляющей. Определены и обоснованы требования к подсистемам метасистемы (программным комплексам).
 5. Разработан и успешно апробирован пилотный проект AtLeap, который является основой (каркасом) для построения программного комплекса (подсистемы метасистемы), автоматизирующего социальные коммуникационные процессы в научно-образовательной среде. Разработанный каркас позволит в значительной степени сократить временные и финансовые затраты на разработку программного комплекса.
 6. Обобщенным результатом данной работы является, с одной стороны, теоретическая база (в виде концепции метасистемы), а с другой стороны, практическая база (в виде проекта AtLeap), которые в совокупности образуют основу для построения информационной системы, автоматизирующей социальные коммуникационные процессы в научно-образовательной среде.
 7. В результате проведенного исследования можно сделать следующие выводы. В публикации и конференции перехода на электронный уровень недостаточно, нужен переход на социо-электронный уровень, в комплексе с персонализацией, коллаборацией и децентрализацией. При большом количестве агентов методов взаимодействия «один к одному» (дискуссия) и «один ко многим» (публикация, конференция) недостаточно, необходим третий метод взаимодействия – «многие ко многим» через среду (стигмергия). При автоматизации социальных коммуникационных процессов необходима самоорганизующаяся метасистема.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Гребнев А. Н. Система электронных публикаций научных статей // Материалы XL Международной научной студенческой конференции «Студент и научно-технический прогресс» : Информационные технологии / Новосиб. гос. ун-т. – Новосибирск, 2002. – С. 32-33.
2. Гребнев А. Н. Научные информационные системы // Вестник УдГУ : Математика. – Ижевск : Изд-во УдГУ, 2003. – С. 99-106.
3. Гребнев А. Н. Информационная система научных коммуникаций // Материалы XLI Международной научной студенческой конференции «Студент и научно-технический прогресс» : Информационные технологии / Новосиб. гос. ун-т. – Новосибирск, 2003. – С. 31-32.

4. Гребнев А. Н. Жизненный цикл электронной научной публикации // Тезисы докладов 31-й итоговой студенческой научной конференции / УдГУ ; отв. ред. А. А. Грызлов. – Ижевск, 2003. – С. 37-39.
5. Гребнев А. Н. Коммуникационная научно-образовательная среда // Вестник УдГУ : Математика. – Ижевск : Изд-во УдГУ, 2004. С. 29-38
6. Гребнев А. Н. Научное общение через электронный форум и электронные публикации // Проблемы теоретической и прикладной математики. Труды 36-й региональной молодежной конференции. – Екатеринбург : УрО РАН, 2005. – С. 327-331.
7. Гребнев А. Н. Автоматизация научной коммуникации в образовательных учреждениях [Электронный ресурс] / Материалы третьей открытой Всероссийской конференции «Преподавание информационных технологий в РФ». – Электрон. статья. – М. : Ассоциация предприятий компьютерных и информационных технологий, 2005. – Режим доступа : <http://www.it-education.ru/2005/reports/Grebnev.htm>, свободный. – Загл. с экрана.
8. Гребнев А. Н. AtLear – Java-каркас с открытым исходным кодом для Web-приложений // Вестник ИжГТУ : период. науч.-теор. журн. – Ижевск : Изд-во ИжГТУ, 2006. – №1. – 116 с. – С. 64-68.
9. Гребнев А. Н. AtLear – база для ИС научных коммуникаций в высшей школе [Электронный ресурс] / Материалы конференции "Свободное программное обеспечение в высшей школе". – Электрон. статья. – Переславль-Залесский: Ин-т программных систем, Ун-т г. Переславля им. А. К. Айламазяна, 2006. – Режим доступа : <http://heap.altlinux.ru/pereslav12006/grebnev/abstract.html>, свободный. – Загл. с экрана.
10. Гребнев А. Н. AtLear.NET – бесплатный .NET каркас для быстрой разработки Web-приложений // Материалы конференции "Технологии Microsoft в теории и практике программирования" / под ред. проф. Р. Г. Стронгина. – Н. Новгород : Изд-во Нижегородского госуниверситета, 2006. – 338с. – С. 75-76.
11. Гребнев А. Н. Методология коммуникационной научно-образовательной среды // Труды XIII Всероссийской научно-методической конференции "Телематика-2006". Т.1. – СПб, 2006. – С. 220-222.
12. Гребнев А. Н. AtLear – Java-каркас для быстрой разработки веб-приложений // Открытый дальневосточный конкурс программных средств студентов, аспирантов и молодых специалистов "Программист-2006" : сборник докладов. – Владивосток : ИАПУ ДВО РАН, 2006. – 91 с. – С. 40-43.

Подписано в печать 15.02.2007. Формат 60x84 1/16
Печать офсетная. Усл. печ. л. 1,0. Тираж 100 экз.
Заказ № 195.

Типография Удмуртского государственного университета
426034, Ижевск, ул. Университетская, 1, корп. 4.