

A-86

АРТАМОНОВ АЛЕКСЕЙ АНАТОЛЬЕВИЧ

**ТРЕХУРОВНЕВАЯ СИСТЕМА АГЕНТНОГО ПОИСКА И
ОБРАБОТКИ НАУЧНО-ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ**

Специальность 05.13.01 – Системный анализ, управление и обработка
информации (в информационных системах)

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Автор:



Москва – 2016

Работа выполнена в Национальном исследовательском ядерном университете «МИФИ»

Научный руководитель: **Оныкий Борис Николаевич**
доктор технических наук, профессор, заведующий кафедрой «Анализ конкурентных систем» НИЯУ МИФИ, г. Москва

Официальные оппоненты: **Сенаторов Михаил Юрьевич**
доктор технических наук, профессор, советник руководителя Московского филиала Gazprom EP International B.V., г. Москва


Крюков Александр Павлович
кандидат физико-математических наук, заведующий лабораторией аналитических вычислений в физике высоких энергий, Научно-исследовательский институт ядерной физики имени Д.В. Скобельцына Московского государственного университета имени М.В. Ломоносова (НИИЯФ МГУ), г. Москва

Ведущая организация: Объединенный институт ядерных исследований, г. Дубна, Московская область

Защита диссертации состоится «22» февраля 2017 г. в 15 час. 00 мин. На заседании диссертационного совета Д 212.130.03 на базе Национального исследовательского ядерного университета «МИФИ» по адресу 115409, Москва, Каширское шоссе, 31.

С диссертацией можно ознакомиться в библиотеке Национального исследовательского ядерного университета «МИФИ» и на сайте: <http://ods.mephi.ru/>.

Отзывы в двух экземплярах, заверенные печатью организации, просьба направлять по адресу 115409, Москва, Каширское шоссе, 31, диссертационные советы НИЯУ МИФИ (тел. +7(499)324-84-98).
Автореферат разослан: « 11 » января 2017 г.

Ученый секретарь
диссертационного совета Д 212.130.03  Н.М. Леонова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

В настоящее время быстро растет количество ресурсов, содержащих информацию научно-технической направленности, которая вместе с тем рассеивается по большому количеству источников. Несмотря на попытки стандартизации информационных ресурсов, количество различных систем доступа к этой информации постоянно увеличивается.

Проблемная ситуация состоит в том, что использование диалоговых систем поиска и структуризации этой информации становится практически невозможным в связи с большим количеством источников (порядка нескольких сотен (10^2 степени) по каждому тематическому направлению) и скоростью наращивания в них научно-технической информации (порядка 10^3 новостных сообщений по каждому тематическому направлению в неделю).

С 90-х годов XX века ведутся работы по исследованию возможностей агентных систем для решения задач обработки больших массивов информации (Big Data).

В 1996 году была создана международная организация по стандартизации агентных технологий – Foundation for Intelligent Physical Agents (FIPA Международное сообщество разработчиков интеллектуальных агентов), в которую вошли как университеты, так и крупные промышленные организации. Направленность работы этой организации состояла в разработке спецификаций и стандартов для программных агентов и агентных систем.

В 1996 году на конференции FIPA было принято следующее определение агента: «Агент – это сущность, которая находится в некоторой среде, от которой она получает данные, отражающие события, происходящие в среде, интерпретирует их и исполняет команды, которые воздействуют на среду».

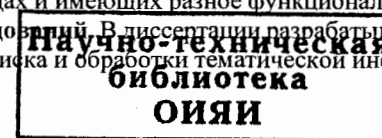
Автор идеи веб-технологий, профессор Тим Бернес Лее в 2001 году опубликовал в Scientific American статью «The Semantic web», в которой выдвигалась идея создания агентных технологий и показывалась широта их возможного полезного использования в веб системах. По его инициативе был создан Консорциум Всемирной паутины (World Wide Web Consortium, W3C), который на сегодняшний день объединяет более 400 организаций, среди которых находятся университеты и промышленные компании. Основная деятельность Консорциума состоит в создании веб-стандартов, в том числе и для агентных систем.

Консорциум Всемирной паутины дает следующее определение пользовательского агента.

Пользовательский Агент – это любое программное обеспечение, которое привлекает, обрабатывает и облегчает взаимодействие конечного пользователя с веб-контентом, или программное обеспечение, чей пользовательский интерфейс реализован с использованием веб-технологий.

Автор диссертации использует оба определения при создании агентов, работающих в разных средах и имеющих разное функциональное назначение.

Актуальность исследования. В диссертации разрабатывается трехуровневая модель агентного поиска и обработки тематической информации, а также



методы её программно-технической реализации. В диссертации обобщаются научные работы автора, выполненные в аспирантуре НИЯУ МИФИ на кафедре «Анализ конкурентных систем» за 2012 – 2016 годы (теоретические и экспериментальные исследования проводились на примере таких тематических направлений как «Физика плазмы», «Лазерные промышленные технологии», «Фотоника», «GRID системы»). Работы выполнялись в рамках Федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» по проекту «Мультиагентные информационно-аналитические системы по естественно-научным и технологическим направлениям» №16.740.11.0129 от 02 сентября 2010 года.

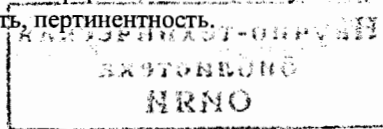
За последние пять лет актуальность работ по созданию агентных информационно-аналитических систем значительно выросла. Это связано с быстрым ростом в сети Интернет количества источников научно-технической информации, таких, например, как сайты университетов, научных центров, научных подразделений (кафедр и лабораторий) и, наконец, сайты известных специалистов и творческих групп.

Извлечение новостной информации из такого большого количества источников без использования агентных технологий стало практически невозможным. Поэтому тема диссертации, посвященной решению вопросов эффективного агентного поиска и обработки научно-технической информации, является актуальной.

Цель диссертационной работы состоит в разработке и экспериментальном исследовании научно-методических и инструментальных средств агентных технологий, обеспечивающих масштабируемость системы и высококачественное информационное обслуживание коллективных и персональных пользователей по параметрам полноты, точности и pertinентности.

В работе решаются следующие основные задачи:

1. Разработка архитектуры трехуровневой агентной информационно-аналитической системы по естественно-научным и технологическим направлениям.
2. Выбор инструментально-технологической платформы для программно-технической реализации трехуровневой модели тематического обслуживания коллективных и персональных пользователей.
3. Реализация и экспериментальные исследования трехуровневой модели агентного поиска и обработки информации.
4. Разработка и экспериментальные исследования методов тематического и лингвистического масштабирования агентной системы.
5. Разработка и исследование технологий регулярного выпуска информационных продуктов, таких как дайджест, динамическое досье, семантическая сеть.
6. Постановка и решение задач по экспериментальной оценке значений параметров качества информационного обслуживания пользователей, таких как полнота, точность, pertinентность.



7. Разработка агентных и интерактивных методов решения типовых информационно-аналитических задач, таких как создание и ведение досье на объекты профессионального интереса; использование семантических сетей для выявления явных и неявных отношений между объектами; ситуационный экспресс анализ в заданных областях науки и техники.

8. Создание учебного курса по использованию агентных технологий при поиске и обработке данных для проведения информационно-аналитических исследований.

Объектом исследования являются агентные технологии поиска и обработки информации в информационно-аналитических системах.

Предметом исследования являются методы и средства агрегирования тематической информации, а также средства обеспечения тематической и лингвистической масштабируемости мультиагентной системы.

Методы исследования. В работе используются методы системного анализа, структурно-параметрического синтеза, системного проектирования, многокритериальные методы количественной экспертной оценки сложных объектов с оценкой согласованности полученных экспертных данных, экспериментальные методы оценки поведения трех типов агентов: «агент-почтальон», «агент-курьер», «агент-референт». Временные характеристики выполнения поисковых операций и обработки данных оцениваются методом тестирования. Pertinentность информации, направляемой пользователю оценивается высококвалифицированными специалистами в тематической области.

Научная новизна. Новые научные результаты, полученные лично автором, состоят в следующем:

1. Предложена трехуровневая модель агентного поиска информации для коллективных и персональных пользователей, обеспечивающая достижение высоких показателей полноты, тематической точности и pertinентности информации.
2. Разработаны и обоснованы функции агентов, работающих на каждом из трех уровней системы, и методы формирования поисковых предписаний для «агентов-почтальонов», «агентов-курьеров», «агентов-референтов».
3. Предложен и реализован метод агрегирования баз данных, управляющих агентным поиском, по каждому из тематических направлений.
4. Предложены и экспериментально проверены методы лингвистического масштабирования с использованием международного стандарта TМХ 1.4b Specification. Получены экспериментальные данные по агентному поиску научно-технической информации на русском, английском, китайском, испанском языках по различным тематическим направлениям («Физика плазмы», «Лазерные промышленные технологии», «Фотоника», «GRID – системы»).
5. Предложен метод агрегирования тезаурусов для поиска информации по пограничным тематическим областям для сокращения объемов выдачи.
6. Предложена методика проведения информационно-аналитических работ по ситуационному экспресс-анализу с помощью операционной карты в различных тематических областях.

7. Предложен новый подход к визуализации научно-технической информации, отражающий динамические свойства объектов для исследователей, принимающих решения.

Обоснованность и достоверность результатов работы подтверждается доказанностью выводов, полученных в результате экспериментов по агентным технологиям, апробацией основных результатов работ на российских и международных конференциях, публикациями, успешной реализацией предложенных методов в виде трёхуровневой модели поиска и обработки научно-технической информации.

Практическая значимость и реализация результатов работы

Теоретические и практические результаты диссертации использованы при создании Мультиагентной информационно-аналитической системы по естественно-научным и технологическим направлениям (МИАС), разработанной на кафедре «Анализ конкурентных систем» НИЯУ МИФИ.

Система внедрена в Научно-техническом институте межотраслевой информации (НТИМИ) для анализа развития отечественных и зарубежных технологий двойного назначения и в Центре плазменных и лазерных технологий НИЯУ МИФИ для информационно-аналитической поддержки научно-исследовательских и опытно-конструкторских работ (НИОКР), проводимых по данным тематическим направлениям.

Автором создан учебный курс «Информационно-аналитическая работа с использованием агентных технологий». Впервые курс прочитан в 2010/2011 годах как факультативный для студентов старших курсов, специализирующихся в области международного научно-технологического и промышленного сотрудничества (направление подготовки (специальность) – 41.03.05). В 2013/2014 учебном году курс дополнен обширным практикумом, реализованным в специально созданной учебно-научной лаборатории «Агентные информационно-аналитические системы». В 2014/2015 учебном году курс вошел в учебный план подготовки как обязательный. Практические занятия проводятся с использованием фактических данных, полученных при опытно-промышленной эксплуатации МИАС в упомянутых выше организациях.

Апробация результатов работы

Основные результаты исследований, проводимых в рамках данной работы, были представлены на следующих конференциях:

Международная конференция «Современные проблемы прикладной математики и информатики» (МРАМCS 2014), г. Дубна, август 2014 г.;

IV конференция «Информационные технологии на службе оборонно-промышленного комплекса», г. Иннополис, май 2015 г.;

Международный военно-технический форум АРМИЯ-2015, г. Кубинка июнь 2015 г.;

XXXIV Межотраслевая научно-методическая конференция «Актуальные вопросы исследования и прогнозирования военного потенциала зарубежных стран в интересах вскрытия угроз безопасности Российской Федерации», г. Москва, октябрь 2015 г.;

XXV Международный симпозиум по ядерной электронике и компьютерным технологиям (Symposium on Nuclear Electronics and Computing (NEC 2015), г. Будва, октябрь 2015 г.;

VI школа CERN/ОИЯИ/МИФИ «GRID и Административные информационные системы», г. Дубна, ноябрь 2015 г.

Результаты работ также докладывались на Научной сессии НИЯУ МИФИ в 2010-2016 гг.

Публикация результатов

Основные положения диссертации опубликованы в 22 печатных работах, в том числе: 4 статьи опубликованы в журналах из списка рекомендованных Высшей аттестационной комиссией, 4 работы опубликованы в изданиях, индексируемых базой Scopus. По научно-техническим разработкам в составе коллектива авторов получены 4 свидетельства о регистрации баз данных, как в Федеральной службе по интеллектуальной собственности, так и в библиотеке Конгресса США.

Основные положения, выносимые на защиту:

1. Трёхуровневая модель агентного поиска и обработки научно-технической информации в сети Интернет, обеспечивающая одновременно высокие показатели полноты, точности и пертинентности информации.

2. Система формирования поисковых предписаний для трех типов агентов, работающих в трехуровневой системе: «агентов-почтальонов», «агентов-курьеров», «агентов-референтов».

3. Метод агрегирования тематических баз данных, управляющих агентным поиском.

4. Метод лингвистического масштабирования агентной системы с использованием международного стандарта TMX 1.4b Specification.

5. Технология формирования персональных тезаурусов для пользователей.

6. Агентные и интерактивные методы решения типовых информационно-аналитических задач: создание и ведение досье на объекты профессионального интереса; использование семантических сетей для обнаружения явных и неявных отношений между объектами; ситуационный экспресс-анализ в заданных областях науки и техники.

7. Учебный курс по использованию агентных технологий при поиске и обработке данных для проведения информационно-аналитических исследований.

Личный вклад автора. Основные научные результаты, полученные автором лично, заключаются в разработке трехуровневой модели агентного поиска и обработки научно-технической информации в сети Интернет; разработке методов управления агентами на трех уровнях системы агентного поиска научно-технической информации; разработке методов и программных средств тематического и лингвистического масштабирования агентной системы; проведении экспериментальных исследований эффективности агентных технологий на трех уровнях агентного поиска и обработки данных. Автор

внес существенный вклад во внедрение разработанной трехуровневой системы в практику информационно-аналитической деятельности НТИМИ. Автором лично проведена настройка трехуровневой агентной системы на решение задач по информационному анализу технологий двойного назначения. Автором лично разработан и впервые прочитан студентам старших курсов НИЯУ МИФИ учебный курс «Информационно-аналитические исследования с использованием интеллектуальных агентных систем».

Структура и объем диссертационной работы. Диссертационная работа состоит из введения, четырех глав, заключения, списка литературы и 7 приложений. Общий объем работы составляет 115 страниц (без учета приложений). Работа содержит 24 иллюстрации и 27 таблиц. Список литературы состоит из 86 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обосновывается актуальность темы диссертационной работы, выделяются объект и предмет исследования, определяются цели и задачи работы, приводится список используемых методов исследования, даются сведения о структуре и объеме диссертационной работы.

В первой главе автором описываются объективные процессы увеличения объема и рассеивания профессионально значимой научно-технической информации. Существующие информационно-поисковые системы широкого использования (Яндекс, Google) не могут в полной мере удовлетворить пользователя-профессионала, так как они предоставляют справочную и энциклопедическую информацию по предмету, что не удовлетворяет критериям новизны и пертинентности. Сбор и структуризация традиционным диалоговым способом стала невозможной в связи с увеличением количества информационных источников и скорости их публикации. Решение данной проблемы стало возможным благодаря агентным технологиям, так как задачами поиска и обработки информации занимается не пользователь, а его представитель - агент.

Автором изучены понятия полноты (R) и точности (P) автоматизированных диалоговых информационно-поисковых систем, которые являются противоречивыми: полноту всегда можно повысить до единицы (при очень низкой точности), возвращая все документы на все запросы. При этом при увеличении количества найденных документов полнота не убывает, в то время как точность обычно снижается. Показатель, позволяющий найти баланс между точностью и полнотой поиска, называется F -мерой и представляет собой их среднее гармоническое взвешенное.

$$F = \frac{1}{\alpha \frac{1}{P} + (1 - \alpha) \frac{1}{R}} = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}, \quad (1)$$

где α — вес параметра P .

$$\beta^2 = \frac{1 - \alpha}{\alpha}, \alpha \in [0, 1], \text{ т. е. } \beta^2 \in [0, \infty]. \quad (2)$$

В агентную систему закладываются равнозначные показатели полноты и точности, исходя из интересов пользователя, что означает $\alpha = 1/2$, или $\beta = 1$, формула упрощается:

$$F = \frac{2PR}{P + R}. \quad (3)$$

P и R являются стохастическими параметрами. Вид F -меры с учетом статистического характера параметров P и R для системы определяется следующим образом:

$$F_{\text{сист}} = \frac{2(M[P] \pm K_P)(M[R] \pm K_R)}{(M[P] \pm K_P) + (M[R] \pm K_R)}. \quad (4)$$

Автором установлено целевое значение F -меры проектируемой системы на уровне 0,7.

Для достижения такого значения F -меры предложена трёхуровневая модель агентного поиска и обработки информации. Особенность данной модели состоит в том, что на каждом уровне возможна максимизация одного из параметров - полноты, точности, пертинентности. Основные принципы трехуровневой модели состоят в следующем:

1. Интеграция тематической информации из всех сетевых источников в единую полнотекстовую базу данных.
2. Тематическая рубрикация информации путем фильтрации полнотекстовой базы данных с использованием тематических тезаурусов.
3. Формирование пользовательских баз данных на основе персональных тезаурусов.

Автором вводится следующее понятие агента,

Агент — это поисковая программа (или поисковый робот), самостоятельно активизирующаяся в системе по заданному расписанию или запросу пользователя и выполняющая действия как в сети Интернет, так и внутри агентной системы по заранее заданным предписаниям.

Введены понятия трех типов агентов, работающих на каждом из уровней системы.

1. *Агент-почтальон (P)* — агент, взаимодействующий с информационными ресурсами сети Интернет по заданному расписанию и агрегирующий полученную информацию в текстовую базу данных.
2. *Агент курьер (K)* — агент, взаимодействующий с агентом-почтальоном и текстовой базой данных и систематизирующий информацию по соответствующим тематическим направлениям.

3. *Агент-референт (R)* – агент, взаимодействующий с пользователем и агентом-курьером в целях доставки pertinentной информации из баз данных системы пользователю.

На рисунках 1-3 представлены схемы уровней агентной системы в соответствии с их назначением.

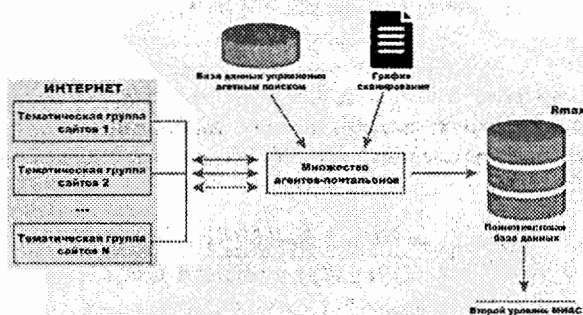


Рис.1. Схема первого уровня «Сбор информации с внешних сетевых источников»

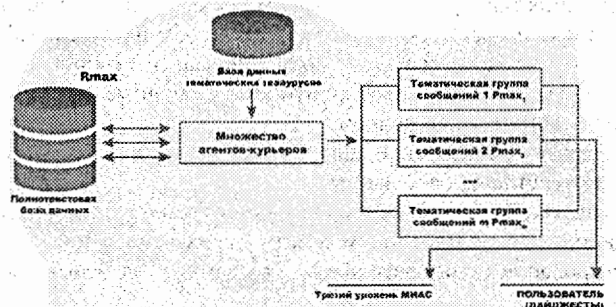


Рис.2. Схема второго уровня «Фильтрация и рубрикация сообщений»

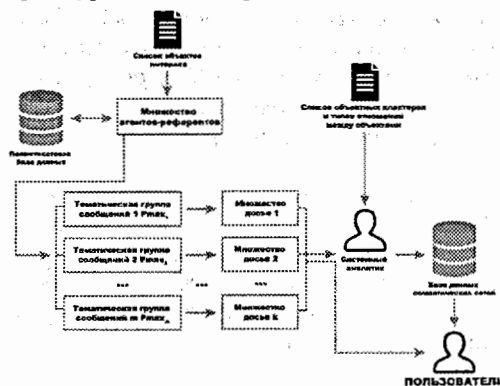


Рис. 3. Схема третьего уровня «Персональное обслуживание»

Каждый из перечисленных выше типов агентов подразделяется на категории: функциональный и целевой. Функциональный агент (индекс F , например, функциональный агент-курьер - K_F) выполняет стандартизированную операцию по заданному расписанию, целевой агент (индекс C , например, целевой агент-курьер - K_C) выполняет операцию по настроенным пользователем параметрам.

Введены две группы пользователей – коллективные и персональные, - обоснованы различные требования этих групп. В первом случае должен быть максимизирован параметр полноты информации, во втором - точность, причем с возможностью предоставления информации с любым пересечением тематических направлений.

В результате работы система формирует для пользователя следующие типы отчетов: дайджест, досье на объект интереса, семантическая сеть.

Также в первом разделе обоснован выбор программной платформы для реализации агентных технологий.

Вторая глава посвящена экспериментальному исследованию тематических групп сайтов в вебе. Информационное сопровождение тематического направления «Лазерные промышленные технологии» выявило, что с 2012 по 2015 годы количество обнаруженных источников выросло в 12 раз - с 22 до 278. Общее количество сообщений в среднем за один месяц составляет 3277 сообщений.

Введено деление информационных источников на три типа – организация, издание, конференция.

Время, затрачиваемое на сканирование соответствующей тематической группы в диалоговом режиме можно определить следующим выражением:

$$T_{\text{общ}} = N(t_{\text{сайта}} + M \cdot t_{\text{оценки}}), \quad (5)$$

где $T_{\text{общ}}$ – время, затраченное на анализ сайтов по одной тематической области за один сеанс сканирования, N – количество источников тематической области, $t_{\text{сайта}}$ – время, затраченное на работу с одним сайтом, M – количество новых новостных сообщений на сайте, $t_{\text{оценки}}$ – время оценки pertinentности нового сообщения.

Время, затрачиваемое на один сеанс сканирования тематической группы сайтов, составляет порядка 2 224 минут или 37 часов, что составит 4,7 человеко-дней при 8-часовом рабочем дне. Анализ информационных сообщений на релевантность и pertinentность был проведен совместно со специалистами «Лазерного промышленного центра» НИЯУ МИФИ. Было выявлено, что 1072 (33%) сообщений являются релевантными (встречается хотя бы одно слово из набора ключевых слов), а 462 (14 %) сообщений являются pertinentными тематике «Лазерные промышленные технологии». Проверка на pertinentность (соответствие результатов поиска информационной потребности (ожиданиям) пользователя) проводилась посредством прочтения всех сообщений.

Трехуровневая модель позволяет существенно автоматизировать работу по поиску и структуризации информации, а именно:

На первом уровне создается маршрутная база данных, состоящая из информационных источников, причем все информационные источники маркируются в соответствии с уровнем доступа к ресурсу.

На втором уровне формируются тематические тезаурусы. Построенный тезаурус в форме базы данных, содержащей как термины, так и их определения, позволяет, в частности, определять ценность терминов для рубрикации агентами-курьерами. Введем понятие «индекс общности термина» (D_n^N) – количество терминов, в названии или в определении которых используется данный термин с порядковым номером n из словаря общим объемом N .

На основе индекса общности (от максимального значения к минимальному) происходит ранжирование терминов тематического направления, что приводит к экономии времени при работе агентов-курьеров. В таблице приведено распределение индексов общности по двум тематическим направлениям: «Физика плазмы» и «Лазерные промышленные технологии».

Таблица 1 Распределения индексов общности по направлениям «Физика плазмы» и «Лазерные промышленные технологии»

	Наименование направления	Количество терминов с $D_n^N \geq 5$	Количество терминов с $0 < D_n^N < 5$	Количество терминов с $D_n^N = 0$
1.	«Лазерные промышленные технологии»	13	69	315
2.	«Физика плазмы»	26	32	218

На третьем уровне формируются персональные базы данных пользователей. Для специалиста отсутствует единый критерий точности информации, и ее тематическая направленность определяется лично пользователем и может содержать различные рубрики. Также проводится рубрикация информации по стадиям жизненного цикла новых изделий, по областям приложения знаний и технологий и т.д.

Также у персонального пользователя существует необходимость наблюдения за определенными объектами интереса в календарном режиме. Под объектом понимается любая сущность, например, экспериментальная установка, технология, проект, организация и т.д. Это определяет необходимость создания и поддержания персональных списков объектов интереса пользователей.

Персональные словари могут быть нескольких видов:

1. Словари, описывающие тематическую область, интересующую определенного пользователя.
2. Словари, описывающие объект.
3. Словари, описывающие стадии (например, жизненный цикл) объекта.

В связи с увеличением количества тематических направлений в маршрутной базе данных стали появляться дублирующие источники, в связи с чем агенты-почтальоны по несколько раз собирали информацию с одного и того же ресурса.

При работе с сайтами на различных языках выяснилось, что если основной сайт публикует информацию на национальном языке, то переводное сообщение существенно запаздывает по времени. Особенно ярко выявленное свойство заметно при работе с сайтами на иероглифических языках (китайский, японский и т.д.). Переводная версия материалов сайта с китайского языка на английский имеет значительные задержки, а иногда вообще не переводится. Для решения этих задач автором была разработана единая база данных, хранящая как маршрутную информацию по всем тематическим направлениям, так и тезаурусы.

В третьей главе приводятся эксперименты по количественной оценке полноты и точности агентной системы для коллективного и персонального пользователя. Например, по «Лазерным промышленным технологиям» в маршрутной базе содержится около 1000 маршрутных адресов для агентов – почтальонов (P_F), тематический тезаурус состоит из 70 терминов на русском, английском и испанском языках. Были получены следующие показатели полноты, точности при формировании ежемесячных дайджестов на втором и третьем уровне (таблица 2).

Таблица 2 Результаты экспериментов по оценке полноты и точности на третьем уровне МИАС

	Наименование показателя	Значение показателя (второй уровень)	Значение показателя (третий уровень)
1.	Исходное множество, обрабатываемых документов (N)	2311	53
2.	Выданная информация, pertinentная запросу (A)	32	32
3.	Не выданная информация, pertinentная запросу пользователя (B)	0	0
4.	Выданная информация, не pertinentная запросу пользователя (C)	21	0
5.	Не выданная информация, не pertinentная запросу пользователя (D)	2258	21
6.	Полнота информационного поиска (R)	1	1
7.	Точность информационного поиска (P)	0,6	1
8.	Коэффициент информационного шума (K)	0,4	0

По итогам оценки информационной выдачи по параметрам релевантности и pertinentности было выявлено, что на втором уровне достигается релевантность информационной выдачи для заданного тематического направления, на третьем уровне за счет использования персональных словарей достигается максимальное значение параметра pertinentности для профессионального пользователя.

Формирование дайджеста с использованием регулярных агентных технологий занимает 15 минут. Пользователь достигает полной информированности

без затрат времени на поисковые операции, тратит время только на анализ содержательной части заинтересовавших его сообщений.

Выполнение регулярных агентных технологий по формированию досье на динамический объект было построено на примере объекта – программа «Электронные масштабируемые нейроморфные адаптивные-пластиковые системы» (Systems of Neuromorphic Adaptive Plastic Scalable Electronics (SyNAPSE)) - отдела оборонных исследований агентства DARPA, США. Агентами-почтальонами была получена информация с 2009 по 2015 год, связанная с финансированием, основными результатами проекта и т.д. Полученная информация была структурирована по годам и получен информационный документ, где на одном листе отображены все основные разработки по программе в каждом году, а также приведены статистические данные по всему проекту.

Третьим типом документа, получаемым в результате выполнения регулярных агентных технологий, является Семантическая сеть. Особенностью документа является такая структуризация информации, которая позволяет отразить на одном листе как один или несколько объектов интереса, так и различные типы отношений между объектами, а также явные и неявные связи.

Была построена семантическая сеть по тематике «Лазерные промышленные технологии» по международной организации Laserlab-Europe, которая объединяет множество европейских научных и коммерческих организаций, работающих в сфере лазерных технологий (рис. 4).

Для решения этой задачи был проведен анализ информационного ресурса – Laserlab Europe. Данный сайт состоит из более чем 10 000 страниц. Информация, представляющая интерес, находится на 171 информационной странице. Для выполнения поставленной задачи необходимо было агрегировать информацию со 171 страницы и структурировать ее соответствующим образом, а именно - выделить все объекты и отношения между ними.

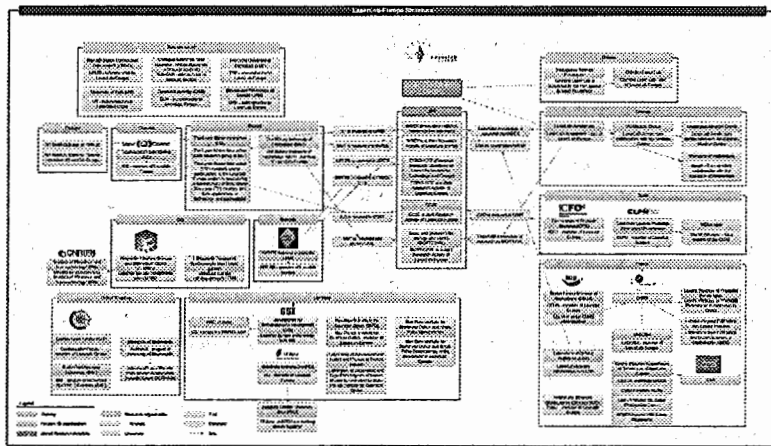


Рис.4. Семантическая сеть Laserlab-Europe

Построение таких сетей в агентной системе автоматизировано. Агенты-куреры (K_F, K_C) выделяют факты из поступающих материалов и создают связи между соответствующими сущностями в базе данных. После чего автоматизировано строится граф между сущностями, которые выбрал пользователь.

В первой главе было приведено целевое значение F -меры для мультиагентной системы равное 0,7. Проведенные эксперименты по формированию дайджестов позволили получить репрезентативную выборку по параметрам полноты и точности. На основе данных по 21 эксперименту были вычислены значения математического ожидания (M), дисперсии (D) и стандартного отклонения (K) для полноты, точности и коэффициента информационного шума (Таблица 3).

Таблица 3. Значения математического ожидания (M), дисперсии (D) и стандартного отклонения (K) для полноты, точности и коэффициента информационного шума.

$M[R]$	$M[P]$	$M[K]$
0,966802	0,938948	0,061052
$D[R]$	$D[P]$	$D[K]$
0,032096	0,057325	0,057325
$K[R]$	$K[P]$	$K[K]$
0,179153	0,239426	0,239426

На основе полученных данных вычислим максимальное и минимальное значение F -меры для системы:

$$F_{\text{сист}}^{\text{max}} = 1,162$$

$$F_{\text{сист}}^{\text{min}} = 0,741$$

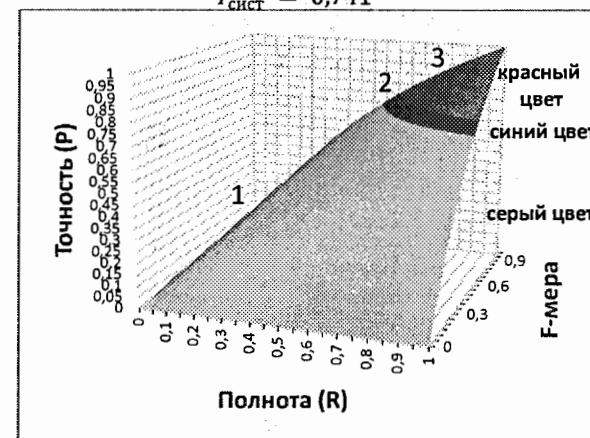


Рис.5. Полученные и целевые значения F -меры

На рис. 5 показана поверхность значений F -меры (Серый цвет, 1), область целевых значений (синий цвет, 2) и область полученных значений (красный

цвет, 3). Минимальное значение для F -меры больше, чем заданное значение в разделе 1, что отвечает заданным требованиям к значениям параметров системы.

В четвертой главе рассматриваются частные специальные случаи выполнения агентных технологий для получения различных информационно-аналитических документов. Приведен пример выполнения работы по подготовке к конференции «Нелинейная Фотоника: теория, материалы, приложения». Особенностью выполнения работы являлось то, что отсутствовала маршрутная база данных и тематический тезаурус, а заданный короткий срок не позволял настроить их. По персональному словарю, предоставленному сотрудниками лаборатории, состоящему из 3 терминов, и уже существующим источникам, к которым добавился профильный сайт организации, были проведены процедуры сбора и структуризации информации. Трехуровневая модель позволила существенно отсеять информационный шум и предоставить наиболее pertinentную информацию для пользователя.

Одна из задач информационно-аналитических исследований заключается в проведении ситуационного экспресс-анализа, который может быть представлен последовательностью операций. Особенность решаемой задачи состоит в том, что на технологической карте необходимо представить следующую информацию: информационные ресурсы, которые будут использованы в описании и анализе ситуации; инструментальные средства, необходимые для работы с выбранными информационными ресурсами; специалистов, которые отвечают за выполнение аналитической работы; логическую последовательность и интерпретацию результатов анализа. Графический документ, отражающий описанную выше информацию, назовем операционной картой (рис.6).

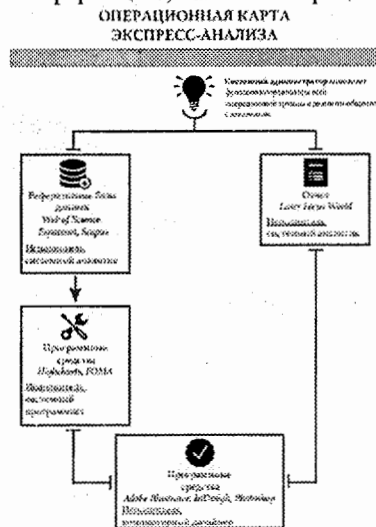


Рис. 6. Операционная карта экспресс-анализа

Основываясь на трехуровневой модели структуризации научно-технической информации, были определены следующие источники: материалы конференций, периодические научные издания, специализированные сайты, патенты.

В систему МИАС была загружена репрезентативная выборка с 42 крупнейших конференций в области «Лазерные технологии» за 2015 год. Источник «конференция» использовался в связи с тем, что на нем можно найти последнюю информацию о новейших разработках. Анализ частоты употребления ключевых слов из полученной выборки в 9 000 слов позволил получить следующее распределение (рис.7). На основе данного распределения можно сделать вывод о существующих трендах в области «Лазерные технологии».

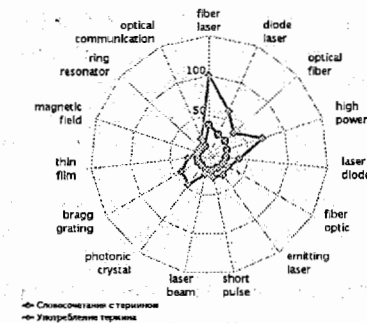


Рис. 7. Результаты анализа частоты употребления ключевых слов

В качестве источника данных по промышленным приложениям был взят результат мониторинга отчетов специализированных организаций, занимающихся публикацией данных об исследованиях и текущем состоянии лазерных технологий.

Благодаря быстрой и эффективной работе поисковых агентов, грамотно отобранным источникам и инструментам поиска стало возможно создание качественного, информативного отчета за максимально сжатые сроки (5 дней). Сочетание точности и новизны данных, систематизации и визуализации делает этот документ универсальным для принятия управленческих решений руководителем любого уровня.

Для масштабирования использования агентных систем автором был разработан учебный курс по теме «Информационно-аналитические исследования с использованием интеллектуальных агентных систем». Помимо технологических процедур агентного поиска тематической информации учебный курс содержит учебные материалы для освоения высокоавтоматизированных методов формирования таких выходных документов, как многоязычные дайджесты, семантические сети, динамические досье, аналитические материалы. Дисциплина развивает у студентов навыки информационно-аналитической работы, проведения анализа информационных источников на достоверность и актуальность, а также дает возможность получения углубленных знаний по подготовке аналитических материалов, необходимых для принятия решений,

направленных на повышение эффективности деятельности организаций в сфере международного научно-технологического и промышленного сотрудничества.

В заключении отражены основные результаты, полученные в данной работе.

Приложение содержит акты о внедрении результатов диссертационной работы; результаты итоговых информационно-аналитических документов, полученных на основе трехуровневой системы агентного поиска и обработки научно-технической информации; результаты экспериментов по оценке полноты и точности агентной системы, а также полученные свидетельства о регистрации баз данных.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

В диссертации решена актуальная научно-техническая задача разработки мультиагентной информационно-аналитической системы по естественно-научным и технологическим направлениям с тремя уровнями агентного поиска и обработки научно-технической информации, обладающая расширенными функциональными возможностями по регулярному и запросно-ответному обслуживанию профессиональных, коллективных и индивидуальных пользователей, а также высокими значениями параметров полноты, точности и пертинентности.

Основные результаты диссертационной работы заключаются в следующем:

1. Предложена, обоснована и реализована трёхуровневая архитектура агентной системы, обеспечивающая одновременно высокие значения параметров полноты, точности и пертинентности информации.

2. Разработаны и обоснованы функции агентов, для каждого из трех уровней системы, и методы формирования поисковых предписаний для «агентов-почтальонов», «агентов-курьеров», «агентов-референтов».

3. Поставлена и решена задача многокритериальной экспертной оценки и выбора программно-технической платформы для реализации агентных технологий.

4. Исследованы отношения между терминами экспериментально построенных тематических тезаурусов, на их основе определена количественная характеристика иерархического типа отношений – индекс общности.

5. Разработана методика формирования персональных словарей пользователя для управления «агентом-курьером» в целях повышения пертинентности информационной выдачи мультиагентной системы

6. Разработаны и экспериментально исследованы процедуры регулярного автоматизированного выпуска важных информационных продуктов: дайджестов, досье динамических объектов, семантических сетей.

7. Разработана методика построения семантических сетей для решения задач по идентификации прямых и косвенных отношений между объектами.

8. С использованием *F-меры* проведен анализ соответствия экспериментально полученных значений параметров полноты и точности заданным значениям.

9. Проведены исследования решения задач обработки больших объемов данных из разных источников с использованием агентных технологий и интерактивных процедур поиска и обработки данных.

10. Предложена методика проведения информационно-аналитических работ по ситуационному экспресс-анализу с помощью операционной карты в различных тематических областях.

11. Результаты диссертационного исследования внедрены и используются в практической деятельности Научно-технического института межотраслевой информации для анализа развития отечественных и зарубежных технологий и в Центре плазменных и лазерных технологий НИЯУ МИФИ для информационно-аналитической поддержки научно-исследовательских и опытно-конструкторских работ (НИОКР), а также в рамках учебного процесса НИЯУ МИФИ, что подтверждается актами о внедрении результатов диссертационного исследования и свидетельствует об обеспечении высококачественного информационного обслуживания пользователей.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

I. Публикации, представленные в международной базе цитирования Scopus

1. Artamonov A.A. 'Algoritmization of search operations in multiagent information-analytical systems / A.G. Ananieva, A.A. Artamonov, I.U. Galin, E.S. Tretyakov, D.O. Kshnyakov // Journal of Theoretical and Applied Information Technology, Nov 2015, Vol. 81, No. 1. - pp. 11-17.

2. Артамонов А.А. Визуализация семантических отношений в мультиагентных системах / А.А. Артамонов, Д.В. Леонов, В.С. Николаев, Б.Н. Оныкий, Л.В. Проничева, К.А. Соколова, И.А. Ушмаров // Научная визуализация, Т. 6, № 3, 2014. - С. 68-76.

3. Artamonov A.A. The express case analysis results visualization / A.A. Artamonov, A.G. Ananieva, E.S. Tretyakov, B.N. Onykiy, L.V. Pronicheva, K.V. Ionkina, A.S. Suslina // Scientific Visualization, 2016, Vol. 8, No.3. - pp 25-34.

4. Artamonov A.A. A three-tier model for structuring of scientific and technical information / A.A. Artamonov, A.G. Ananieva, E.S. Tretyakov, D.O. Kshnyakov, B.N. Onykiy, L.V. Pronicheva // Journal of Digital Information Management, June 2016. Vol.14. No.3. - pp 184-193.

II. Публикации в журналах, включенных в перечень периодических изданий ВАК Российской Федерации

5. Артамонов А.А. Мультиагентная информационно-аналитическая система по естественно-научным и технологическим направлениям / А.А. Артамонов, Д.В. Леонов, Б.Н. Оныкий, Л.В. Проничева // Системы высокой доступности, Т. 10, № 2, 2014, - С. 45-49.

6. Артамонов А.А. Тематические тезаурусы в агентных технологиях поиска научно-технической информации в интернет (на примере тезауруса по теме "Физика плазмы") / А.А. Артамонов, И.Ю. Галин, К.В. Ионкина, В.А. Курнаев, К.А. Соколова, А.И. Черкасский // Математическое моделирование, Т. 27, № 7, 2015. - С. 4-9.

7. Артамонов А.А. Понсковые агентные технологии с многоязычным тезаурусом / А.А. Артамонов, И.Ю. Галин, Д.В. Леонов, Е.К. Михина, Б.Н. Оныкий, К.А. Соколова // Вестник Национального исследовательского ядерного университета "МИФИ", Т. 4, № 4, 2015. - С. 369-376.

8. Артамонов А.А. Анализ деятельности и информационного ресурса Агентства по перспективным оборонным научно-исследовательским разработкам США (DARPA) / А.А. Артамонов, А.Г. Ананьева, Н.А. Александрова, Д.О. Кшняков, Е.С. Третьяков // Наноинженерия, № 8, 2015. - С. 43-48.

III. Публикации в других изданиях и материалы конференций

9. Артамонов А.А. Современные проблемы прикладной математики и информатики (MPAMCS'2014) / А.А. Артамонов, К.А. Соколова, Д.В. Леонов, А.И. Черкасский // Формирование тематических кластеров для агентного поиска научно-технической информации в интернет (на примере тематического направления "Физика плазмы"). Дубна. 2014. - С. 41

10. Артамонов А.А. Современные проблемы прикладной математики и информатики (MPAMCS'2014) / А.А. Артамонов, И.Ю. Галин, К.А. Соколова, В.С. Николаев, А.И. Черкасский // Тематические тезаурусы в агентных технологиях поиска научно-технической информации в Интернет (на примере тезауруса по "Физике плазмы"). Дубна. 2014. - С. 160-161.

11. Артамонов А.А. III Всероссийская научно-практическая конференция «Информационные технологии в образовании XXI века» / А.А. Артамонов, В.С. Николаев, А.И. Черкасский, Д.В. Леонов // Применение мультиагентных информационно-аналитических систем в учебно-научной деятельности. Москва. 2013. - С. 69-73.

12. Артамонов А.А. Применение семантических сетей в аналитических исследованиях в области международного научно-технического и международного промышленного сотрудничества / А.А. Артамонов, В.С. Николаев, А.И. Черкасский // Научная сессия НИЯУ МИФИ – 2011, Москва. 2011. Т. 3 – С. 245.

13. Артамонов А.А. Аналитические исследования с применением интеллектуальных агентных систем / А.А. Артамонов, А.И. Черкасский, В.С. Николаев // Научная сессия НИЯУ МИФИ – 2013, Москва. 2013. Т. 3. - С. 50.

14. Артамонов А.А. Современные системы искусственного интеллекта и их приложения в науке. Всероссийская научная Интернет-конференция с международным участием / А.А. Артамонов, А.И. Черкасский, К.А. Соколова, А.Г. Ананьева // Экспериментальные исследования эффективности тематического агентного поиска. Казань. 2013. - С. 8-11.

15. Артамонов А.А. Современные системы искусственного интеллекта и их приложения в науке. Всероссийская научная Интернет-конференция с международным участием / А.А. Артамонов, Б.Н. Оныкий, А.И. Черкасский, Д.В.

Леонов // Учебный курс по агентным информационным технологиям в аналитической деятельности. Казань. 2013. - С. 58-62.

16. Артамонов А.А. Автоматизированное создание новостных изданий (дайджестов) по мировым источникам информации / А.А. Артамонов, В.С. Николаев, А.И. Черкасский // Научная сессия НИЯУ МИФИ – 2012, Москва. 2012. Т. 3. - С. 48.

17. Артамонов А.А. Современные системы искусственного интеллекта и их приложения в науке. Всероссийская научная Интернет-конференция с международным участием / К.А. Соколова, В.А. Курнаев, А.А. Артамонов, А.И. Черкасский // Интеллектуальная агентная система по физике плазмы. Казань. 2013. - С. 68-73.

18. Артамонов А.А. Анализ и стендовые экспериментальные исследования на рынке агентных информационно-аналитических систем / А.А. Артамонов, В.С. Николаев, А.И. Черкасский // Научная сессия НИЯУ МИФИ – 2012, Москва. 2012. Т. 3 – С. 46.

IV. Свидетельства о государственной регистрации баз данных

19. Артамонов А.А., Ионкина К.В., Оныкий Б.Н., Ананьева А.Г., Проничева Л.В., Галин И.Ю., Ушмаров И.А., Суслина И.В., Петровский В.Н., Быковский Д.П., Соколова К.А., Горяинова А.Е. Управляющая база данных для агентного поиска в Интернет новостной информации по тематическому направлению "Лазерные технологии", Свидетельство о государственной регистрации базы данных 2016620164, февраль 04, 2016.

20. Артамонов А.А., Соколова К.А., Баламутенко А.В., Николаев В.С., Леонов Д.В., Суслина И.В., Ананьева А.Г., Проничева Л.В., Ушмаров И.А., Черкасский А.И. Мировые научно-исследовательские и технологические организации по физике плазмы, Свидетельство о регистрации базы данных 2014620346, Февраль 26, 2014.

21. Артамонов А.А., Курнаев В.А., Оныкий Б.Н., Галин И.Ю., Соколова К.А., Курнаев А.А., Николаев В.С., Баламутенко А.В., Леонов Д.В., Проничева Л.В., Фомина Ю.Е. Тезаурус по физике плазмы в международном стандарте TMX 1.4b Specification, Свидетельство о регистрации базы данных 2015620043, Январь 12, 2015.

22. Artamonov A., Balamutenko A., Nikolayev V., Sokolina K., Leonov D., Suslina I., Ananieva A., Pronicheva L., Ushmarov I., Cherkasskij A. World Plasma Physics Research Organisations Database, TXU 1-904-126, February 13, 2014.