

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
"Тверской государственный технический университет"
(ТвГТУ)

УДК 004.89:378

№ госрегистрации

Инв. №

ОТЧЕТ
О ПАТЕНТНЫХ ИССЛЕДОВАНИЯХ
Тема: "Интеллектуальная распределенная система информационной
поддержки инноваций в науке и образовании"

Проект РФФИ 13-07-00342\13

ТВЕРЬ 2013

Список исполнителей

Руководитель темы,

директор ЦНОЭР,

канд. техн. наук

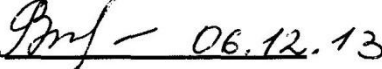
Исполнитель,

начальник отдела

информационных ресурсов

ЦНОЭР

 06.12.13 В.К. Иванов
подпись, дата

 06.12.13 Н.В. Виноградова
подпись, дата

Реферат

Отчет содержит 50 листов машинописного текста, 7 таблиц, 4 листа приложений.

Ключевые слова: базы данных, интеллектуальная обработка, интеллектуальный анализ данных, информационно-поисковые системы, матрица запросов, методика выполнения исследований, обработка результатов, параметры поиска, патентный поиск, поисковые запросы, результаты поиска, релевантность, фильтрация.

Цель патентного поиска – найти аналоги проектируемой интеллектуальной распределенной системы информационной поддержки инноваций в науке и образовании, установить степени патентной чистоты и научной новизны, выявить потенциальных конкурентов, минимизировать возможное дублирование уже существующих технических решений.

Структура и содержание отчета. Отчет включает:

- титульный лист;
- список исполнителей;
- реферат;
- содержание;
- перечень сокращений, условных обозначений, символов, единиц и терминов;
- введение;
- общие данные об объекте исследования;
- методику выполнения исследований;
- подготовку к выполнению поисковых операций;
- выполнение поиска в БД ФИПС;
- выполнение поиска в ИПС Google;
- результаты обработки данных, полученных после выполнения поисковых операций;
- заключение;
- приложения (задание на проведение патентных исследований, регламент поиска, отчет о поиске).

В первом разделе отчета приводятся общие сведения об объекте исследования.

Во втором разделе кратко представлены основные положения методики выполнения исследований.

В третьем разделе описана методика выполнения патентного (и не патентного) поиска: выбор баз данных и информационно-поисковых систем, порядок выполнения поиска по поисковым запросам и классификационным рубрикам, правила формирования основной группы результатов поиска и ее представление.

В четвертом разделе приведена методика обработки результатов поиска, включающая правила выполнения вычислений релевантности документов и различных показателей качества результатов выполненных поисковых операций.

В пятом разделе описывается подготовка к выполнению поисковых операций: создание матрицы запросов, выбор баз данных и информационно-поисковых систем, а также классификационных рубрик.

В шестом разделе представлены результаты поисковых операций, выполненных в базе данных ФГБУ ФИПС, и обработки полученных данных, сформулированы соответствующие выводы.

В седьмом разделе представлены результаты поисковых операций, выполненных в информационно-поисковой системе Google, и обработки полученных данных, сформулированы соответствующие выводы.

В заключении сделаны обобщающие выводы по результатам патентного исследования и даются рекомендации по совершенствованию проведения патентных исследований.

В приложения вынесены задание на проведение патентных исследований, регламент поиска, отчет о выполненных поисковых операциях.

Работы проводились при финансовой поддержке РФФИ (договор № НК13-07-00342\13).

Оглавление

Реферат.....	3
Оглавление.....	5
Нормативные ссылки	7
Перечень сокращений, условных обозначений, символов, единиц и терминов.....	8
Введение	9
1. Общие сведения об объекте исследования	10
2. Методика выполнения исследований	12
2.1. Общие положения	12
2.2. Временной интервал поиска.....	12
2.3. Критерии отбора.....	12
2.4. Классификационные рубрики	12
2.5. Разделы методики.....	13
3. Методика выполнения поиска.....	14
3.1. Порядок выполнения поисковых операций.....	14
3.2. Выбор множества баз данных и информационно-поисковых систем	14
3.3. Выполнение поиска по поисковым запросам	14
3.4. Выполнение поиска по классификационным рубрикам	16
3.5. Формирование основной общей группы результатов поиска	17
3.6. Представление результатов поиска	17
4. Методика обработки результатов поиска	18
4.1. Порядок обработки результатов поиска	18
4.2. Определение средней экспертной оценки релевантности документов	18
4.3. Формирование итоговой группы.....	18
4.4. Определение точности поиска.....	18
4.5. Определение полноты поиска	19
4.6. Определение сбалансированной F-меры	19
5. Подготовка к выполнению поисковых операций.....	21
5.1. Выбор баз данных и информационно-поисковых систем.....	21
5.2. Формирование множества ключевых слов	21
5.3. Создание матрицы запросов.....	21
5.4. Выбор классификационных рубрик.....	21
6. Выполнение поиска в БД ФИПС	22
6.1. Параметры поиска по поисковым запросам.....	22
6.2. Фильтрация результатов поиска по поисковым запросам	22
6.3. Формирование основной группы результатов поиска по поисковым запросам	22
6.4. Параметры поиска по классификационным рубрикам.....	22

6.5. Фильтрация результатов поиска по классификационным рубрикам	22
6.6. Формирование основной группы результатов поиска по классификационным рубрикам	23
6.7. Формирование основной общей группы результатов поиска	23
6.8. Итоговые результаты поиска.....	23
6.9. Обработка результатов поиска	24
6.10. Выводы по результатам поиска в БД ФИПС	26
7. Выполнение поиска в ИПС Google	26
7.1. Параметры поиска по поисковым запросам	26
7.2. Фильтрация результатов поиска	26
7.3. Формирование основной группы результатов поиска	27
7.4. Итоговые результаты поиска.....	27
7.5. Обработка результатов поиска	27
7.6. Выводы по результатам поиска в ИПС Google	29
Заключение	29
Список использованных источников	31
ПРИЛОЖЕНИЕ А	32
ПРИЛОЖЕНИЕ Б	34
ПРИЛОЖЕНИЕ В	36
ПРИЛОЖЕНИЕ Г	42

Нормативные ссылки

В настоящем отчете о патентных исследованиях использованы ссылки на следующие стандарты:

ГОСТ Р 15.011-96	Система разработки и постановки продукции на производство. Патентные исследования. Содержание и порядок проведения
ГОСТ 7.32-2001	Система стандартов по информации, библиотечному и издательскому делу. Отчет о научно-исследовательской работе. Структура и правила оформления

Перечень сокращений, условных обозначений, символов, единиц и терминов

В настоящем отчете использованы сокращения:

БД	База данных
ИПС	Информационно-поисковая система
МПК	Международная патентная классификация
РИ	Реферат изобретения
РФФИ	Российский фонд фундаментальных исследований
ТвГТУ	Тверской государственный технический университет
УДК	Универсальная десятичная классификация
ФГБУ ФИПС	Федеральное государственное бюджетное учреждение «Федеральный институт промышленной собственности»
ЦНОЭР	Центр научно-образовательных электронных ресурсов

В настоящем отчете использованы термины и определения:

База данных — представленная в объективной форме совокупность самостоятельных материалов, систематизированных таким образом, чтобы эти материалы могли быть найдены и обработаны с помощью электронной вычислительной машины (ЭВМ).

Информационно-поисковая система - комплекс программных и аппаратных средств, который обеспечивает отбор и представление электронных документов по заданным критериям.

Информационный поиск – отрасль знания, которая занимается представлением, хранением и доступом к информационным ресурсам.

Патентные исследования - исследования технического уровня и тенденций развития объектов хозяйственной деятельности, их патентоспособности, патентной чистоты, конкурентоспособности (эффективности использования по назначению) на основе патентной и другой информации

Поисковый запрос (запрос) - исходная информация для осуществления поиска с помощью поисковой системы. Формат поискового запроса зависит как от устройства поисковой системы, так и от типа информации для поиска.

Релевантность (лат. *relevare* — поднимать, облегчать) в информационном поиске — семантическое соответствие поискового запроса и поискового образа документа. В более общем смысле, одно из наиболее близких понятию качества «релевантности» — «адекватность», то есть не только оценка степени соответствия, но и степени практической применимости результата, а также степени социальной применимости варианта решения задачи.

Фильтрация документов — одна из задач информационного поиска, целью которого является разбиение множества документов на следующие категории: документы, удовлетворяющие заданному критерию, и не удовлетворяющие ему.

Введение

Патентные исследования проводились в соответствии с заданием, разработанным в рамках темы: «Интеллектуальная распределенная система информационной поддержки инноваций в науке и образовании» (проект РФФИ 13-07-00342\13) и требованиями ГОСТ Р15.011-96 «Система разработки и постановки продукции на производство. Патентные исследования. Содержание и порядок проведения».

Цель исследования: найти аналоги проектируемой интеллектуальной распределенной системы информационной поддержки инноваций в науке и образовании, установить степени патентной чистоты и научной новизны, выявить потенциальных конкурентов, минимизировать возможное дублирование уже существующих технических решений.

Объектом исследования является система интеллектуальной обработки результатов поиска информации, а предметом исследования – патентная (и не патентная) документация, характеризующая интеллектуальную обработку результатов поиска.

Исследования проводились с использованием ИПС ФГБУ «Федеральный институт промышленной собственности» и ИПС Google.

Содержание выполненных исследований:

- Общее описание объекта исследования.
- Методика выполнения поиска материалов для патентных исследований, включающая методики выполнения поиска, представления результатов поиска.
- Методика обработки результатов поиска, включающая правила выполнения вычислений релевантности документов и различных показателей качества результатов выполненных поисковых операций.
- Описание хода выполнения поиска с помощью информационно-поисковых систем ФИПС и Google, а также результатов выполненных поисковых операций.
- Результаты обработки данных, полученных после выполнения поисковых операций, с формулировкой соответствующих выводов.
- Выводы и рекомендации по результатам патентного исследования.

Работы проводились при финансовой поддержке РФФИ (договор № НК13-07-00342\13).

1. Общие сведения об объекте исследования

Объектом настоящих патентных исследований является технология интеллектуальной обработки результатов поиска информации. Ниже описываются основные особенности проекта, в рамках которого разрабатывается указанная технология и выполнены патентные исследования.

Создание новых эффективных методов поиска и синтеза научного знания (в частности, прорывных технологий и инновационных идей) является одной из наиболее приоритетных задач исследований и разработок, способствующих развитию как отдельно взятого научного направления – информационного поиска, так и современной науки в целом. Самые разнообразные исследования касаются особенностей поиска инноваций в экономике, науке, образовании. На решение этой задачи направлен и проект «Интеллектуальная распределенная система информационной поддержки инноваций в науке и образовании».

Основная идея проекта – исследование новых подходов к автоматизированной семантической обработке больших массивов научно-технической информации в дата-центре (ДЦ), позволяющей осуществлять поиск прорывных технологий и других инновационных идей. Важными особенностями, которые определили принятые подходы, являются:

- Искомые решения часто находятся на стыке смежных областей; отсюда сложности формулировки точного поискового запроса.
- Одновременно с информацией о собственно инновациях желательно получать сведения о применениях, рисках, особенностях, пользователях, авторах, производителях.
- Наличие альтернатив и необходимость одновременного использования различных критериев отбора наиболее эффективных решений.
- Разрозненность и неоднородность сведений об инновациях; их преимущественно внутриотраслевой характер.

Очевидно, что нужны эффективные методы создания и наполнения электронных коллекций новейших идей и технологий, содержащих не просто их описания, а специальным образом отобранные, классифицированные и ассоциированные данные. Эти данные могут быть использованы для анализа ретроспективы и перспективы конкретных инноваций, а также для оценки современных направлений развития.

В отличие от поиска определенных сведений (фактов), касающихся отдельных сторон искомой сущности, решить нетривиальную задачу формулировки поискового запроса непросто (иногда и невозможно). Например, требуется узнать экономические показатели шахты ОАО "Распадская" за первое полугодие 2013 года. Используя эту фразу как поисковый запрос, можно получить релевантный ответ в первом десятке результатов поиска Google. Но как подобрать материалы для анализа научно-технических, экономических и социальных факторов влияния на инновационные технические, технологические или финансовые механизмы угледобычи в восточных районах России?

Для решения подобных задач пользователи вынуждены применять множество сочетаний ключевых понятий, уточняя их в ходе анализа промежуточных результатов поиска в ресурсах Интернет или в специализированных хранилищах, например, патентных базах данных (БД). Неочевидно, что при этом будет строго использована какая-либо обоснованная методика. В результате у пользователя в распоряжении будет большой объем данных, в той или иной степени релевантных запросам. При этом, как правило, нет возможности подробно рассмотреть все имеющиеся результаты. И возникают следующие вопросы:

- Как совместно оценить релевантность документов, найденных разными запросами? Правильно ли определялась релевантность документов?

- Является ли ранжирование результатов поисковой системы корректным с позиций ожиданий пользователя? Все ли результаты, доступные для непосредственной оценки, соответствуют ожиданиям пользователя?
- Все ли результаты, соответствующие ожиданиям пользователя, попали в число доступных для непосредственной оценки? Все ли искомые решения найдены вообще?
- Как отфильтровать документы, не относящиеся по сути к искомой тематике?
- Могут ли быть обнаружены эффективные решения, которые относятся к другим областям применения, но могут быть успешно использованы как инновации в данной области.
- Есть возможность визуально оценить множество найденных решений в совокупности со связанными объектами?

Невозможно однозначно ответить на эти вопросы в рамках тривиальных решений. Но в проекте «Интеллектуальная распределенная система информационной поддержки инноваций в науке и образовании» делается попытка предложить ряд нестандартных подходов.

2. Методика выполнения исследований

2.1. Общие положения

2.1.1. Объектом исследования является технология интеллектуальной обработки результатов поиска информации, а предметом исследования – патентная (и не патентная) документация, характеризующая интеллектуальную обработку результатов поиска.

2.1.2. При проведении поиска должны приниматься во внимание официальные публикации охранных документов, таких как патенты на изобретения, авторские свидетельства (для Российской Федерации и Советского Союза), свидетельства на полезные модели (для Российской Федерации) и патенты на полезные модели, а также национальные и международные заявки на патенты основных промышленно развитых государств, в частности, Великобритании, Германии, Японии, Кореи, США, КНР, государств СНГ.

2.1.3. При проведении поиска в БД ФГБУ ФИПС должны учитываться все патентные документы независимо от их правового статуса, в том числе патенты и заявки на регистрацию объектов промышленной собственности, а также международные заявки, сведения о которых были опубликованы в официальных бюллетенях патентных ведомств и в общедоступных базах данных патентной документации до даты завершения патентного поиска.

2.2. Временной интервал поиска

2.2.1. При поиске в ИПС Google должны учитываться все документы, которые соответствуют тематике поиска.

2.2.2. Временной интервал патентного поиска для целей настоящего патентного исследования не ограничивается. В целях исследования современного состояния и тенденций развития объекта исследований временной интервал поиска был установлен по 2013 год включительно, поскольку поданные патентные заявки уже опубликованы в базах данных патентных документов.

2.2.3. При выборе ретроспективы поиска по патентным документам учитывалось, что максимальный срок действия патента на изобретение составляет 20 лет, с возможностью его продления на три года. Кроме того, в соответствии с действующим патентным законодательством и международным правом в области патентования технических решений предусматривается включение в мировой технический уровень всех патентных документов, вне зависимости от даты приоритета, а также статуса документа, то есть в мировой технический уровень должны быть включены все опубликованные на дату проведения патентного поиска документы, вне зависимости от того, действуют они, или срок их действия уже истек.

2.3. Критерии отбора

2.3.1. Эвристический отбор охранных документов должен производиться, исходя из цели патентного поиска. Критерии поиска включают: функциональную принадлежность объекта патентного права; техническую сущность объекта патентного права; суть технического эффекта объекта патентного права; метод достижения технического эффекта. Могут быть использованы и другие критерии.

2.4. Классификационные рубрики

2.4.1. Выбор классификационных рубрик МПК должен производиться, исходя из необходимости охватить все технические решения, имеющие сходные функциональные возможности или принцип действия, без ограничения каким-либо конкретным назначением.

2.5. Разделы методики

2.5.1. Методика выполнения исследований обеспечивает надлежащий охват информации и позволяет выявить технические решения, относящиеся к области поиска, как уже охраняемые патентными документами, так и составляющие область потенциальной охраны в связи с установлением приоритета технического решения при подаче соответствующей патентной заявки.

2.5.2. Методика выполнения исследований включает:

- 1) Методику выполнения патентного (и не патентного) поиска.**
- 2) Методику обработки результатов поиска.**

3. Методика выполнения поиска

Методика выполнения поиска применима как к патентному, так и к не патентному поиску.

3.1. Порядок выполнения поисковых операций

- 1) Выбор множества баз данных (БД) и/или информационно-поисковых систем (ИПС).
- 2) Выполнение поиска по поисковым запросам.
- 3) Выполнение поиска по классификационным рубрикам.
- 4) Формирование основной группы результатов поиска.
- 5) Представление результатов.

3.2. Выбор множества баз данных и информационно-поисковых систем

Перед началом выполнения поиска необходимо выбрать БД, по которым будет производиться поиск, и ИПС, с помощью которых будет производиться поиск.

Ниже приведены примеры БД и ИПС:

1) Российские БД

- Роспатент (ФИПС) (платный и бесплатный доступ) <http://www.fips.ru/russite/default.htm>
- Международный центр научной и технической информации (МЦНТИ) (бесплатный доступ) <http://www.icsti.su>
- Всероссийский научно-технический информационный центр (ВНТИЦ) (бесплатный доступ) <http://www.rntd.citis.ru/> (Официальный сайт - www.vntic.org.ru)
- Государственная публичная научно-техническая библиотека (ГПНТБ) (бесплатный доступ) <http://www.gpntb.ru>
- Библиотека патентов и изобретений, зарегистрированных на территории РФ www.findpatent.ru

2) БД патентных ведомств мира (<http://www.sciteclibrary.ru>)

- Всемирная организация по интеллектуальной собственности <http://pctgazette.wipo.int>
- Американская (только американская) патентная база, (поиск бесплатно работает не глубже 1975 года, трудности со скачиванием рисунков) <http://www.uspto.gov>
- БД PAJ с 1993 года <http://www.jpo.go.jp>
- БД AIPN (патентные документы Японии с 1995 г., документы США (с 1987 г.), ЕПВ (с 1994 г.) и ВОИС (с 1994 г.)). <http://www.aipn.org>
- Сайт Европейской патентной организации <http://www.espacenet.com/access/index.en.html>
- Сайт Всемирной организации интеллектуальной собственности (WIPO) <http://www.wipo.int/ipdl/en/search/pct/search-adv.jsp>
- QUESTEL www.qpat.com

3) Поисковые системы Интернет

- Яндекс – <http://www.yandex.ru>
- Google – <https://www.google.ru>
- Bing – <http://www.bing.com>

4) Другие БД

eLibrary.ru - <http://elibrary.ru>

3.3. Выполнение поиска по поисковым запросам

3.3.1. Выполнения поиска по поисковым запросам включает:

- 1) Формирование множества ключевых слов
- 2) Создание матрицы запросов

- 3) Выполнение поисковых запросов
- 4) Фильтрация результатов поиска
- 5) Формирование основной группы результатов поиска

3.3.2. Формирование множества ключевых слов

3.3.3. При отборе терминов для списка ключевых слов следует:

- 1) Исключить общеупотребительные слова, к которым относятся предлоги, союзы, вспомогательные глаголы, частицы, местоимения и т. п.
- 2) Отбросить термины, которые слишком часто/редко встречаются в документах.
- 3) Ключевые слова поместить в таблицу:

№ п/п	Ключевые слова
1.	
2.	

3.3.4. Создание матрицы запросов

Матрица запросов создается для обеспечения полноты поиска. Элемент матрицы соответствует поисковому запросу, состоящему из комбинации двух ключевых слов. Например, число 12 в матрице показывает, что поисковый запрос состоит из первого и второго ключевого слова. Пример матрицы для пяти поисковых запросов:

Номер запроса	1	2	3	4	5
1	x	x	x	x	x
2	12	x	x	x	x
3	13	23	x	x	x
4	14	24	34	x	x
5	15	25	35	45	x

3.3.5. Выполнение поисковых запросов

Выполнение поисковых запросов следует производить для каждой выбранной БД и/или ИПС. Порядок действий следующий:

- 1) Осуществить поиск по каждому поисковому запросу из матрицы запросов. Получаем результаты поиска по каждому поисковому запросу, представляющие собой:
 - список адресов документов с фрагментами текста документа в ИПС Интернет (например, Google);
 - список названий и адресов документов в БД (например, в ФГБУ ФИПС).
- 2) Каждую страницу результатов поиска по каждому запросу сохранить как документ Microsoft Office Word.

3.3.6. Фильтрация результатов поиска

Фильтрация результатов поиска осуществляется в каждой группе результатов поиска по каждому запросу. Включает предварительную и основную фильтрацию.

3.3.7. Предварительная фильтрация документов происходит при первоначальном просмотре каждой страницы результатов поиска. При этом из результатов исключаются: документы, тематика которых явно не соответствует предмету поиска; несуществующие документы; ресурсы, зараженные вирусами.

3.3.8. При основной фильтрации документы открываются и просматриваются более детально. При этом исключаются документы, тематика которых формально является релевантной, но по некоторым причинам не должна быть предметом поиска.

Пример:

При предварительной фильтрации по запросу "классификация результатов поиска, отбор результатов поиска" документ "Система обработки информации и способ ввода для оптимизации оцифровки кода операций" был оставлен.

При основной фильтрации, после просмотра документа, выяснилось, что в нем представлен способ кодирования и ввода знаков, и вычислительное устройство для кодирования и ввода знаков, поэтому документ был удален, так как не соответствовал теме поиска.

3.3.9. Формирование основной группы результатов поиска

Для формирования основной группы результатов поиска следует:

- 1) Объединить все страницы результатов поиска в общий список – исходную группу.
- 2) Сформировать основную группу результатов поиска, удалив из исходной группы дублированные документы.

3.4. Выполнение поиска по классификационным рубрикам

3.4.1. Методика выполнения поиска по классификационным рубрикам МПК включает:

- 1) Выбор классификационных рубрик
- 2) Выполнение поисковых запросов
- 3) Фильтрацию результатов поиска
- 4) Формирование основной группы результатов поиска

3.4.2. Выбор классификационных рубрик

Выбор группы осуществляется в соответствии с МПК. При выборе классификационных рубрик следует:

- Выбрать одну или несколько рубрик в соответствии с МПК.
- Название и индекс рубрик записать в таблицу:

№ п/п	Индекс рубрики	Название рубрики

3.4.3. Выполнение поисковых запросов

Выполнение поисковых запросов следует производить для каждой выбранной БД. Порядок действий следующий:

- 1) Выполнить поисковый запрос по каждой классификационной рубрике МПК. В результате получатся результаты поиска по каждому поисковому запросу, представляющие собой список названий и адресов документов в БД (например, в ФГБУ ФИПС).
- 2) Каждую страницу результатов поиска по каждой классификационной рубрике сохранить как текстовый документ (например, документ Microsoft Office Word).

3.4.4. Фильтрация результатов поиска

Фильтрация результатов поиска осуществляется в каждой группе результатов поиска по каждой классификационной рубрике. Включает предварительную и основную фильтрацию:

- 1) Предварительная фильтрация документов происходит при первоначальном просмотре каждой страницы результатов поиска. Исключаются документы, тематика которых явно не соответствует предмету поиска.

- 2) При основной фильтрации документы открываются и просматриваются более детально. При этом исключаются документы, тематика которых формально является релевантной, но по некоторым причинам не должна быть предметом поиска.

3.4.5. Формирование основной группы результатов поиска по классификационным рубрикам

- 1) Объединить все результаты поиска по всем запросам в общий список - исходную группу результатов поиска.
- 2) Сформировать основную группу результатов поиска по классификационным рубрикам, удалив из исходной группы дублированные документы.

3.5. Формирование основной общей группы результатов поиска

3.5.1. Порядок действий следующий:

- 1) Соединить основные группы результатов поиска по поисковым запросам и классификационным рубрикам.
- 2) Удалить дублированные документы.
- 3) Сформировать основную общую группу результатов поиска.

3.6. Представление результатов поиска

3.6.1. После выполнения поиска для каждой БД или ИПС отдельно и по каждому запросу или классификационной рубрике необходимо представить следующие результаты:

- 1) Число найденных документов.
- 2) Число отобранных документов.

3.6.2. Также для каждой БД или ИПС следует представить следующие суммарные результаты:

- 1) Общее число найденных документов.
- 2) Число документов в основной группе результатов поиска по поисковым запросам.
- 3) Число документов в основной группе результатов поиска по классификационным рубрикам.
- 4) Число документов в основной группе результатов поиска.

3.6.3. Результаты поиска по каждому запросу поместить в следующую таблицу:

Номер запроса в матрице	Число найденных документов	Число отобранных документов	Pi	Ri	Fi
1	3	4	5	6	7

Примечания:

- При заполнении графы 1 номер запроса в матрице см. в п. 3.3.4.
- Сумма значений графы 3 должна соответствовать общему числу найденных документов.
- Сумма значений графы 4 должна соответствовать общему числу документов в основной группе результатов поиска (см. п.п. 3.3.9 и 3.4.5).
- Методику расчетов значений в графах 5, 6 и 7 см. в разд. 4.

4. Методика обработки результатов поиска

4.1. Порядок обработки результатов поиска

- 1) Определение средней экспертной оценки релевантности документов.
- 2) Формирование итоговой группы.
- 3) Определение точности поиска.
- 4) Определение полноты поиска.
- 5) Вычисление сбалансированной F-меры.

4.2. Определение средней экспертной оценки релевантности документов

4.2.1. Средняя экспертная оценка характеризует степень соответствия информации в документе и реальной информационной потребности пользователя. Значение $grade \in [0,3]$, где 3 - «релевантный документ», 2 - «частично релевантный(+) документ», 1 - «частично релевантный(-) документ», 0 - «нерелевантный документ»

4.2.2. Средняя экспертная оценка релевантности документов $grade$ должна быть определена для каждого документа основной группы результатов поиска.

4.2.3. Результаты должны быть помещены в следующую таблицу:

№ п/п	Название документа	grade

4.3. Формирование итоговой группы

4.3.1. На основании средней экспертной оценки релевантности $grade$ из общей группы сформировать итоговую группу документов, в которую войдут документы, значения $grade$ которых равны 3 или 2.

Примечание: перечень патентов итоговой группы поместить в Приложение В отчета.

4.4. Определение точности поиска

4.4.1. Точность поиска (precision) в общем случае определяется как отношение количества отобранных документов к количеству найденных в БД или ИПС документов.

4.4.2. Точность поиска для запроса, которая определяется как отношение количества отобранных документов к количеству найденных документов по каждому поисковому запросу:

$$P_i = \frac{R_{ai}}{A_i} \quad (1)$$

где

P_i - точность поиска для запроса;

R_{ai} - количество отобранных документов по каждому поисковому запросу;

A_i - количество найденных документов на поисковый запрос.

4.4.3. Групповая точность поиска, которая определяется как отношение количества отобранных документов (итоговая группа) к количеству найденных документов (основная группа):

$$P = \frac{R_a}{A} \quad (2)$$

где

P - групповая точность поиска;

R_a – количество отобранных документов (итоговая группа);

A – количество найденных документов (основная группа).

4.5. Определение полноты поиска

4.5.1. Полнота поиска (recall) в общем случае определяется как отношение количества отобранных документов к сумме всех найденных.

4.5.2. Полнота поиска для запроса, которая определяется как отношение количества отобранных документов по каждому поисковому запросу к общему числу найденных документов по всем поисковым запросам:

$$R_i = \frac{R_{ai}}{D_i} \quad (3)$$

где

R_i - полнота поиска для запроса;

R_{ai} – количество отобранных документов по каждому поисковому запросу;

D_i – общее число найденных документов.

- Групповая полнота поиска, которая определяется как отношение количества отобранных документов (итоговая группа) к количеству всех найденных документов:

$$R = \frac{R_a}{D} \quad (4)$$

где

R - групповая полнота поиска;

R_a – количество отобранных документов (итоговая группа);

D – общее количество найденных документов.

4.6. Определение сбалансированной F-меры

4.6.1. Мерой для совместной оценки точности и полноты является F -мера, которая определяется как взвешенное гармоническое среднее точности P и полноты R . Если точности и полноте придается одинаковый вес, то такая F -мера называется сбалансированной.

4.6.2. Сбалансированная F -мера для запроса:

$$F_i = \frac{2P_iR_i}{P_i + R_i} \quad (5)$$

где

F_i - сбалансированная F -мера для запроса;

P_i - точность поиска для запроса;

R_i - полнота поиска для запроса.

4.6.3. Групповая сбалансированная F -мера:

$$F = \frac{2PR}{P + R} \quad (6)$$

где

F - групповая сбалансированная F -мера;

P - групповая точность поиска;

R - групповая полнота поиска;

Результаты всех вычислений следует поместить в таблицу (см. п. 3.6.3).

5. Подготовка к выполнению поисковых операций

При выполнении поисковых операций использовалась методика, изложенная в разд. 2 и 3.

5.1. Выбор баз данных и информационно-поисковых систем

5.1.1. Для исследований были выбраны БД ФГБУ ФИПС (патентный поиск) и ИПС Google (не патентный поиск).

5.1.2. Поиск в ИПС Google был произведен в связи с тем, что новизна изобретения, определяется не только зарегистрированными в России патентами, но также имеющейся во всем мире общедоступной информацией.

5.2. Формирование множества ключевых слов

5.2.1. Множество ключевых слов представлено в табл. 1.

Таблица 1

№ п/п	Ключевые слова
1.	уточнение поискового запроса
2.	классификация результатов поиска
3.	интеллектуальный анализ результатов поиска
4.	фильтрация результатов поиска
5.	семантика связей между объектами
6.	определение подмножества семантически связанных данных
7.	отбор результатов поиска
8.	контроль тематики результатов поиска

5.3. Создание матрицы запросов

5.3.1. Матрица запросов составлена в соответствии с таблицей 1. Каждый ненулевой элемент матрицы соответствует определенному запросу. Создано 28 комбинаций запросов, которые представлены в таблице 2.

Таблица 2

№ ключевого слова	1	2	3	4	5	6	7	8
1								
2	12							
3	13	23						
4	14	24	34					
5	15	25	35	45				
6	16	26	36	46	56			
7	17	27	37	47	57	67		
8	18	28	38	48	58	68	78	

5.4. Выбор классификационных рубрик

5.4.1. Содержание классификационных рубрик по МПК раскрывается в таблице 3.

Таблица 3

№ п/п	Индекс рубрики	Название рубрики
1.	G06F17/30	Информационный поиск; структуры баз данных для этой цели

6. Выполнение поиска в БД ФИПС

6.1. Параметры поиска по поисковым запросам

6.1.1. Патентный поиск по поисковым запросам осуществлялся по информационной базе ФГБУ ФИПС "Патентные документы РФ (рус.)". Для проведения экспертной оценки был выполнен отбор следующих источников патентной информации:

- Рефераты российских изобретений (РИ).
- Рефераты российских изобретений (РИ).
- Заявки на российские изобретения (ЗИ).
- Полные тексты российских изобретений из трех последних бюллетеней (НИ).
- Формулы российских полезных моделей (РПМ).
- Формулы российских полезных моделей из трех последних бюллетеней (НПМ).
- Перспективные российские изобретения (ПИ).
- Полные тексты российских изобретений из трех последних бюллетеней (НИ).

6.1.2. Ретроспективность поиска (по дате приоритета охранного документа) определена регламентом и составляла 20 лет.

6.1.3. Для ввода поисковых запросов использовалось поле "Основная область запросов".

6.2. Фильтрация результатов поиска по поисковым запросам

6.2.1. Предварительная фильтрация охранных документов для каждого поискового запроса включала удаление из списка результатов найденных документов, тематика которых явно не относится к данной области знаний. До фильтрации - 1145 документов, после фильтрации - 283 документов.

6.2.2. Основная фильтрация охранных документов осуществлялась для каждой страницы результатов поиска по каждому запросу детальным просмотром каждого документа. Исключались патенты, тематика которых формально является релевантной, но по некоторым причинам не должна быть предметом поиска. До фильтрации - 283 документов, после фильтрации - 161 документ.

6.3. Формирование основной группы результатов поиска по поисковым запросам

6.3.1. Основная группа результатов поиска образуется после объединения всех результатов запросов в общий список, его сортировки и удаления дублетных документов. До формирования основной группы - 161 документ, после - 50 документов.

6.4. Параметры поиска по классификационным рубрикам

6.4.1. Патентный поиск по классификационным рубрикам МПК осуществлялся по информационной базе ФГБУ ФИПС "Патентные документы РФ (рус.)". Для проведения экспертной оценки был выполнен отбор следующих источников патентной информации:

- Рефераты российских изобретений (РИ)
- Заявки на российские изобретения (ЗИ)
- Полные тексты российских изобретений из трех последних бюллетеней (НИ)

6.4.2. Для ввода индекса рубрики G06F17/30 использовалось поле "(51) МПК".

6.5. Фильтрация результатов поиска по классификационным рубрикам

Производим предварительную фильтрацию результатов поиска. Исключаем документы, тематика не относится к данной области знаний.

6.5.1. До предварительной фильтрации - 100 документов, после - 74 документа.

6.5.2. До основной фильтрации - 74 документа, после - 61 документ.

6.6. Формирование основной группы результатов поиска по классификационным рубрикам

6.6.1. Основная группа результатов поиска образуется после объединения всех результатов запросов в общий список, его сортировки и удаления дублированных документов. До формирования основной группы - 61 документ, после - 37 документов.

6.7. Формирование основной общей группы результатов поиска

6.7.1. Основная общая группа результатов поиска образуется после объединения результатов поиска по поисковым запросам и классификационным рубрикам и последующего удаления дублированных документов. До формирования основной общей группы - 87 документов, после - 60 документов.

6.8. Итоговые результаты поиска

6.8.1. Результаты поиска по каждому поисковому запросу представлены в табл. 4:

Таблица 4

№ запроса в матрице	Число найденных документов	Число отобранных документов	P	R	F1
12	97	15	0,15	0,01	0,01875
13	100	4	0,04	0,003	0,00558
14	90	2	0,02	0,002	0,00364
15	0	0	0	0	0
16	100	2	0,02	0,002	0,00364
17	100	5	0,05	0,004	0,00741
18	100	2	0,02	0,002	0,003636
23	4	1	0,25	0,00008	0,000160
24	25	6	0,24	0,005	0,009796
25	100	5	0,05	0,004	0,007407
26	2	0	0	0	0
27	8	3	0,37	0,003	0,005952
28	0	0	0	0	0
34	4	0	0	0	0
35	0	0	0	0	0
36	1	0	0	0	0
37	0	0	0	0	0
38	0	0	0	0	0
45	1	0	0	0	0
46	1	0	0	0	0
47	12	1	0,08		0
48	0	0	0	0	0
56	100	14	0,14	0,01	0,018667
57	100	15	0,15	0,01	0,01875
58	100	9	0,09	0,008	0,014694
67	0	0	0	0	0
68	0	0	0	0	0
78	100	7	0,07	0,006	0,011053

6.8.2. Общее число найденных охранных документов - 1145.

6.8.3. Число документов в основной общей группе результатов поиска - 60 патентов.

6.9. Обработка результатов поиска

6.9.1. Средняя экспертная оценка каждого отобранного охранного документа представлена в табл. 5:

Таблица 5

№ п/п	Название документа	grade
1.	Способ классификации веб-страниц и организации соответствующего информационного наполнения	3
2.	Объединение результатов поиска	2
3.	Способ построения семантической модели документа	3
4.	Способ поиска информации в политематических массивах неструктурированных текстов	3
5.	Способ формирования и регулярной доставки пользователям информационных изданий и система для осуществления способа	2
6.	Способ кластеризации и сокращения размерности многомерных данных для их индексации и поиска	3
7.	Способ организации данных в Интернет	1
8.	Способ поиска и выборки информации с повышенной релевантностью	1
9.	Способ автоматизированной обработки информационных текстовых материалов	2
10.	Способ передачи классифицированной информации с заданной приоритетностью	0
11.	Способ синтеза самообучающейся системы извлечения знаний из текстовых документов для поисковых систем	3
12.	Поиск произвольного текста и поиск по атрибутам в данных электронного руководства по программам	2
13.	Архитектура многоуровневого брандмауэра	0
14.	Система и способ для клиент-обоснованного поиска веб-агентом	1
15.	Оснащение пользовательского интерфейса расширением поисковых запросов	1
16.	Предложение родственных терминов для многосмыслового запроса	3
17.	Проверка релевантности между ключевыми словами и содержанием веб-сайта	2
18.	Модель данных для объектно-реляционных данных	1
19.	Способ поиска и выборки информации из различных баз данных	2
20.	Система автоматизированного упорядочения неструктурированного информационного потока входных данных	0
21.	Система и способ предоставления предпочтительного языка упорядочивания результатов поиска	1
22.	Информационно-поисковая система предоставления данных об объектах	1
23.	Способ, система и читаемый компьютером носитель информации для синхронизации поддающихся изменению документов для множества клиентов	0
24.	Способ предоставления данных об объектах и система для его реализации	1
25.	Динамический опыт пользователя посредством семантически богатых объектов	0
26.	Системы и способы для поиска с использованием запросов, написанных на языке и/или наборе символов, отличном от такового, для целевых страниц	1
27.	Устройство обработки информации, способ обработки информации и программа обработки информации	0
28.	Способ и система для вычисления значения важности блока в дисплейной странице	0

№ п/п	Название документа	grade
29.	Способ автоматизированной обработки текста на естественном языке путем его семантической индексации, способ автоматизированной обработки коллекции текстов на естественном языке путем их семантической индексации и машиночитаемые носители	2
30.	Аннотация посредством поиска	1
31.	Предложение родственных терминов для многосмыслового запроса	2
32.	Функции ранжирования, использующие модифицированный наивный байесовский классификатор запросов с инкрементным обновлением	2
33.	Способ и система для индексирования и поиска в базах данных	1
34.	Семантическая навигация по веб-контенту и коллекциям документов	1
35.	Прикладной программный интерфейс для извлечения и поиска текста	1
36.	Относительные результаты поиска на основе пользовательского взаимодействия	1
37.	Функции ранжирования, использующие смещенное расстояние, измеряемое количеством последовательных переходов, до документа в сети	1
38.	Система интерактивного поиска и отображения информации	1
39.	Информационно-поисковая система	1
40.	Устройство и способ автоматического построения приложений из спецификаций и из имеющихся в наличии компонентов, выбранных с помощью семантического анализа	0
41.	Диаграмма ранжирования	3
42.	Эффективная навигация результатов поиска	2
43.	Монетизация и определение приоритета результатов распределенного поиска	1
44.	Способ семантической обработки естественного языка с использованием графического языка-посредника	1
45.	Устройство обработки информации, способ обработки информации и программа обработки информации	1
46.	Индексация и поиск динамически изменяющихся поисковых корпусов	1
47.	Способ информационного поиска (варианты) и компьютерная система для его осуществления	3
48.	Система и способ агрегации данных с множества веб-сайтов	3
49.	Визуализация структуры сайта и разрешение переходов сайта для результата поиска или связанной страницы	2
50.	Устройство обработки информации, способ обработки информации и компьютерный носитель данных	0
51.	Способ поиска информационных ресурсов с использованием переадресаций	2
52.	Генерация запроса с использованием конфигурации среды	3
53.	Способ создания и использования рекурсивного индекса поисковых машин	3
54.	Предварительный просмотр результатов поиска для предложенных терминов уточнения и вертикальных поисков	2
55.	Способ и устройство обработки ресурсов страниц	0
56.	Способ и система семантического поиска электронных документов	2
57.	Способ обнаружения идентификационных признаков для различных буквенно-знаковых систем письменности	2
58.	Способ анализа и прогноза развития сложно построенной системы и устройство для его осуществления	3
59.	Фильтр по образцу	3
60.	Способ повышения скорости поиска данных с использованием адаптивных носителей данных	1

6.9.2. На основании средней экспертной оценки релевантности сформирована итоговая группа результатов поиска (см. подразд. 4.3), общее количество документов которой составляет 28 патентов. Перечень патентов итоговой группы помещен в Приложение В «Отчет о поиске». Реферативный список отобранных патентов находится в Приложении Г.

6.9.3. Точность поиска для каждого поискового запроса вычислена по формуле (1) и помещена в табл. 4.

6.9.4. Групповая точность поиска вычислена по формуле (2). $P=28/60=0,467$.

6.9.5. Полнота поиска для каждого поискового запроса вычислена по формуле (3) и помещена в табл. 4.

6.9.6. Групповая полнота поиска вычислена по формуле (4). $R=28/1145=0,024$.

6.9.7. Сбалансированная F-мера для каждого поискового запроса вычислена по формуле (5) и помещена в табл. 4.

6.9.8. Групповая сбалансированная F-мера вычислена по формуле (6). $F=0,046$.

6.10. Выводы по результатам поиска в БД ФИПС

6.10.1. Поисковые запросы, в которые входят ключевые понятия «семантика связей между объектами», «определение подмножества семантически связанных данных», "уточнение поискового запроса", "контроль тематики результатов поиска" имеют наименьшие точность, полноту поиска, а также сбалансированную F-меру (табл. 4). Это свидетельствует о том, что эти вопросы недостаточно изучены.

6.10.2. Наибольшую точность поиска имеют поисковые запросы, включающие ключевые понятия «классификация результатов поиска», "интеллектуальный анализ результатов поиска" (табл. 4), что показывает актуальность данной тематики.

6.10.3. Групповая сбалансированная F-мера характеризует состояние анализируемой темы в целом и равна $F=0,046$, т.е. она, практически, близка к нулю. Это говорит о том, что данная тема изучена еще недостаточно.

6.10.4. Анализ по дате подачи заявки показывает, что наибольшее количество документов, отобранных для рассмотрения (60 док.), приходится на 2004-2006 гг. Для сравнения в период 2010-2011 гг. подано 7 заявок или 11,7% от общего количества рассматриваемых заявок.

7. Выполнение поиска в ИПС Google

7.1. Параметры поиска по поисковым запросам

7.1.1. Поиск осуществлялся с использованием матрицы запросов.

7.1.2. Число документов, просматриваемых на странице результатов поиска – 50.

7.2. Фильтрация результатов поиска

7.2.1. Предварительная фильтрация результатов поиска для каждого поискового запроса включала удаление следующих документов: словари, глоссарии, справочники; учебные пособия; контрольные работы и рефераты студентов; учебные программы, лекции, тесты, билеты, шпаргалки для студентов и школьников; материалы для раскрутки сайтов; сайты компаний, сайты для закупок, социальные сайты; блоги; рекламные объявления. До фильтрации - 1800 документов, после фильтрации - 623 документа.

7.2.2. Основная фильтрация документов осуществлялась для каждой страницы результатов поиска по каждому запросу детальным просмотром каждого документа. Исключались документы, тематика которых формально является релевантной, но по некоторым причинам

не должна быть предметом поиска. До фильтрации - 623 документа, после фильтрации - 279 документов.

7.3. Формирование основной группы результатов поиска

7.3.1. Основная группа результатов поиска образуется после объединения всех результатов запросов в общий список, его сортировки и удаления дублированных документов. До формирования основной группы - 279 документов, после - 166 документов.

7.4. Итоговые результаты поиска

7.4.1. Итоговые результаты поиска по каждому поисковому запросу представлены в табл. 6:

Таблица 6

№ запроса в матрице	Число найденных документов	Число отобранных документов	P	R	F1
12	50	5	0,1	0,003	0,00583
13	50	12	0,24	0,007	0,01360
14	50	3	0,06	0,002	0,00387
15	50	2	0,04	0,001	0,00195
16	50	25	0,5	0,01	0,01961
17	50	3	0,06	0,002	0,00387
18	50	3	0,06	0,002	0,00387
23	50	37	0,74	0,02	0,03895
24	50	20	0,4	0,01	0,01951
25	50	35	0,7	0,02	0,03889
26	50	37	0,74	0,02	0,03895
27	50	23	0,46	0,01	0,01957
28	50	21	0,42	0,01	0,01954
34	50	20	0,4	0,01	0,01951
35	50	43	0,86	0,02	0,03909
36	50	33	0,66	0,02	0,03882
37	50	13	0,26	0,007	0,01363
38	50	11	0,22	0,006	0,01168
45	50	27	0,54	0,015	0,02919
46	50	30	0,6	0,02	0,03871
47	50	6	0,12	0,003	0,00585
48	50	5	0,1	0,003	0,00583
56	50	34	0,68	0,02	0,03886
57	50	27	0,54	0,015	0,02919
58	50	20	0,4	0,01	0,01951
67	50	25	0,5	0,01	0,019608
68	50	25	0,5	0,01	0,01961
78	50	8	0,16	0,004	0,00781

7.4.2. Общее число документов, найденных в ИПС Google - 1800.

7.4.3. Общее число документов в основной группе результатов поиска - 166.

7.5. Обработка результатов поиска

7.5.1. Средняя экспертная оценка каждого отобранного документа представлена в табл. 7:

Таблица 7

№ доку- мента	grade
1	1
10	0
100	1
101	2
102	3
103	1
104	3
105	2
106	1
107	1
108	1
109	0
11	2
110	2
111	3
112	0
113	2
114	1
115	2
116	1
117	0
118	0
119	1
12	0
120	0
121	1
122	1
123	0
124	3
125	2
126	1
127	3
128	3
129	1
13	0

№ доку- мента	grade
130	0
131	0
132	1
133	3
134	2
135	1
136	1
137	2
138	0
139	1
14	1
140	1
141	1
142	2
143	3
144	3
145	2
146	3
147	2
148	1
149	1
15	2
150	0
151	1
152	1
153	1
154	0
155	1
156	1
157	0
158	0
159	0
16	1
160	2
161	0

№ доку- мента	grade
162	0
163	1
164	0
165	1
166	2
17	2
18	1
19	1
2	2
20	0
21	1
22	1
23	2
24	0
25	0
26	0
27	0
28	3
29	1
3	0
30	2
31	1
32	2
33	1
34	1
35	3
36	2
37	1
38	1
39	2
4	1
40	1
41	3
42	1
43	3

№ доку- мента	grade
44	3
45	2
46	1
47	2
48	0
49	0
5	3
50	0
51	3
52	1
53	1
54	0
55	1
56	1
57	1
58	3
59	2
6	1
60	1
61	0
62	1
63	1
64	2
65	3
66	3
67	3
68	1
69	2
7	3
70	0
71	2
72	1
73	1
74	1
75	1

№ доку- мента	grade
76	2
77	1
78	3
79	3
8	2
80	1
81	2
82	2
83	3
84	1
85	1
86	0
87	1
88	3
89	2
9	0
90	3
91	1
92	1
93	1
94	1
95	1
96	1
97	2
98	1
99	1

7.5.2. На основании средней экспертной оценки релевантности (см. табл. 7) сформирована итоговая группа результатов поиска (см. подразд. 4.3), общее количество документов в которой составляет 61.

7.5.3. Точность поиска для каждого поискового запроса вычислена по формуле (1) и помещена в табл. 6.

7.5.4. Групповая точность поиска вычислена по формуле (2). $P=61/166=0,37$

7.5.5. Полнота поиска для каждого поискового запроса вычислена по формуле (3) и помещена в табл. 6.

7.5.6. Групповая полнота поиска вычислена по формуле (4). $R=29/1800=0,02$.

7.5.7. Сбалансированная F-мера для каждого поискового запроса вычислена по формуле (5) и помещена в табл. 6.

7.5.8. Групповая сбалансированная F-мера вычислена по формуле (6). $F=0,038$.

7.6. Выводы по результатам поиска в ИПС Google

7.6.1. Поисковые запросы, в которые входит ключевое понятие "уточнение поискового запроса" имеют наименьшие точность, полноту поиска, а также сбалансированную F-меру (табл. 6). Это свидетельствует о том, что вопрос по уточнению поискового запроса изучен недостаточно.

7.6.2. Наибольшую точность поиска имеют поисковые запросы, включающие ключевые понятия «классификация результатов поиска», "интеллектуальный анализ результатов поиска", "семантика связей между объектами", "определение подмножества семантически связанных" (табл. 6), что показывает актуальность данной тематики. (табл. 6).

7.6.3. Групповая сбалансированная F-мера характеризует состояние анализируемой темы в целом и равна $F=0,038$, т.е. она, практически, близка к нулю. Это говорит о том, что данная тема изучена еще недостаточно.

Заключение

Выполненное патентное исследование свидетельствует, что в настоящее время вопрос интеллектуальной обработки результатов поиска информации изучен недостаточно. Так, всего 4% охраняемых документов было отобрано для дальнейшей обработки.

Однако на основании достаточного количества имеющихся результатов можно сделать обобщающие вывод, что на данный момент нет никаких серьезных препятствий для разработки новых методов интеллектуальной обработки результатов поиска информации, так как прямых аналогов проектируемой интеллектуальной распределенной системы информационной поддержки инноваций в науке и образовании найдено не было. То есть, степень патентной чистоты и научной новизны удовлетворительна, вероятность дублирования уже существующих технических решений минимальна, а потенциальные конкуренты не выявлены.

Тем не менее отметим, что поиск был произведен частично - ограничивающим фактором являлось наличие в тексте охранного документа только выбранных ключевых слов (понятий).

Текущие рекомендации следующие:

- 1) Необходимо проведение дополнительных патентных исследований для выявления новых патентов и не патентных работ в рассматриваемой области - интеллектуальная обработка результатов поиска.
- 2) Считать приоритетным направлением поиск охраняемых и не охраняемых документов, которые мешали бы на данном этапе проводить разработку тематики проекта.

- 3) Расширить перечень используемых БД и ИПС: включить в их состав зарубежные патентные базы данных с открытым доступом и другие популярные поисковые системы Интернет.
- 4) Расширить перечень ключевых слов (понятий) используемых для выполнения поисковых операций.
- 5) Продолжить использование разработанной методики проведения патентных исследований, уточнив методы математической обработки результатов поиска.
- 6) Ввести в состав используемой методики проведения патентных исследований правила интерпретации полученных и обработанных результатов поиска.

Список использованных источников

- 1) Иванов, В.К. Эвристический алгоритм фильтрации и семантического ранжирования результатов поиска документов [Электронный ресурс] : статья / В.К. Иванов, Н.В. Виноградова // Вестник Тверского государственного университета : научный журнал : Серия "Прикладная математика" / Твер. гос. ун-т. - 2013.
- 2) Ишков, А.Д. Проведение патентных исследований : справочное пособие / А.Д. Ишков, А.В. Степанов ; под ред. А.Д. Ишкова ; М-во образования и науки РФ, Моск. гос. строит. ун-т. — М. : МГСУ, 2012. — 132 с. - (Библиотека научно-педагогического работника).
- 3) Система разработки и постановки продукции на производство. Патентные исследования. Содержание и порядок проведения : ГОСТ Р 15.011-96. – Введ. 1996-01-30. – М. : ГОССТАНДАРТ РОССИИ, 1996.
- 4) Система стандартов по информации, библиотечному и издательскому делу. Отчет о научно-исследовательской работе. Структура и правила оформления [Электронный ресурс] : ГОСТ 7.32-2001. - Введ. 2002-07-01. - Минск : Межгосударственный совет по стандартизации, метрологии и сертификации, 2003.

ПРИЛОЖЕНИЕ А

УТВЕРЖДАЮ

Руководитель работ

канд. техн. наук. В. К. Иванов

«28»мая 2013 г.

ЗАДАНИЕ № 2013-05-22-01

на проведение патентных исследований

Наименование работы (темы): Интеллектуальная распределенная система информационной поддержки инноваций в науке и образовании

Шифр работы (темы): 13-07-00342\13

Этап работы: определение степени патентной чистоты и научной новизны.

Сроки выполнения: 24.05.2013 г. – 19.12.2013 г.

Задачи патентных исследований: найти аналоги проектируемой интеллектуальной распределенной системы информационной поддержки инноваций в науке и образовании, установить степени патентной чистоты и научной новизны, выявить потенциальных конкурентов, минимизировать возможное дублирование уже существующих технических решений.

КАЛЕНДАРНЫЙ ПЛАН

Виды работ по патентным исследованиям	Подразделения-исполнители (соисполнители)	Ответственные исполнители (Ф.И.О.)	Сроки выполнения патентных исследований. Начало. Окончание	Отчетные документы
1. Определение задач патентных исследований, видов и методов их проведения, разработка задания	ЦНОЭР ТвГТУ	Иванов В.К., Виноградова Н.В.	24.05.2013 - 27.05.2013	Форма приложения А по ГОСТ Р 15.011-96
2. Разработка регламента поиска, определение требований к поиску патентной информации	ЦНОЭР ТвГТУ	Иванов В.К., Виноградова Н.В.	28.05.2013- 29.05.2013	Форма приложения Б по ГОСТ Р 15.011-96

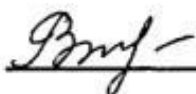
Виды работ по патентным исследованиям	Подразделения-исполнители (соисполнители)	Ответственные исполнители (Ф.И.О.)	Сроки выполнения патентных исследований. Начало. Окончание	Отчетные документы
3. Поиск, подбор, систематизация и анализ патентной информации	ЦНОЭР ТвГТУ	Иванов В.К., Виноградова Н.В.	30.05.2013- 19.11.2013	Форма приложения В по ГОСТ Р 15.011-96
4. Оформление результатов исследований в виде отчета	ЦНОЭР ТвГТУ	Иванов В.К., Виноградова Н.В.	20.11.2013- 19.12.2013	Отчет о патентных исследованиях

Руководитель проекта,
директор Центра научно-образовательных
электронных ресурсов


В.К. Иванов

28.05.13
дата

Начальник отдела
информационных
ресурсов


Н.В. Виноградова

28.05.13
дата

ПРИЛОЖЕНИЕ Б

УТВЕРЖДАЮ

Руководитель работ

канд. техн. наук. В. К. Иванов

«29»мая 2013 г.

Регламент поиска № 2013-05-24

Наименование работы (темы): Методы интеллектуальной обработки результатов поиска информации

Шифр работы (темы): 13-07-00342\13

Этап работы: определение степени патентной чистоты и научной новизны.

Сроки выполнения: 24.05.2013 г. – 19.12.2013 г.

Номер и дата утверждения задания: 2013-05-27-01, 27.05.2013 г.

Цель поиска информации: создание перечня документов, которые потенциально могут нарушить патентоспособность разрабатываемой технологии и препятствовать их применению на территории Российской Федерации.

Обоснование регламента поиска:

Начало поиска: 30 мая 2013 г.

Окончание поиска: 19 ноября 2013 г.

Предмет поиска (объект исследования, его составные части, товар)	Страна поиска	Источники информации, по которым будет проводиться поиск								Ретроспективность	Наименование информационной базы (фонда)
		патентные		НТИ*		конъюнктурные		другие			
		Наименование	Классификационные рубрики: МПК (МКИ)*, МКПО*, НКИ* и другие	Наименование	Рубрики УДК* и другие	Наименование	Код товара: ГС*, СМТК*, БТН*	Наименование	Классификационные индексы		
1	2	3	4	5	6	7	8	9	10	11	12
Методы интеллектуальной обработки результатов поиска информации	Россия	База данных Федеральной службы по интеллектуальной собственности, патентам и товарным знакам Российской Федерации	G06F17/30	-	-	-	-	-	-	20 лет	Роспатент, www.fips.ru

Руководитель проекта, директор Центра научно-образовательных электронных ресурсов



В.К. Иванов

29.05.13

дата

Начальник отдела информационных ресурсов



Н.В. Виноградова

29.05.13

дата

*МПК (МКИ)
НКИ
МКПО
НТИ
ГС
СМТК
БТН
УДК

- международная патентная классификация (международная классификация изобретений);
- национальная классификация изобретений;
- международная классификация промышленных образцов;
- научно-техническая информация;
- гармонизированная система (гармонизированная товарная номенклатура);
- стандартная международная торговая классификация ООН;
- Брюссельская таможенная номенклатура;
- универсальная десятичная классификация.

ПРИЛОЖЕНИЕ В

УТВЕРЖДАЮ

Руководитель работ

канд. техн. наук. В. К. Иванов

«06» декабря 2013 г.

ОТЧЕТ О ПОИСКЕ

Поиск проведен в соответствии с заданием на проведение патентных исследований № 2013.05.27-01 от 27 мая 2013 г. и регламентом поиска 2013.05.27-01 от 29 мая 2013 г.

Начало поиска – 30 мая 2013 г. Окончание поиска – 19 ноября 2013 г.

Сведения о выполнении регламента поиска: поиск выполнен частично. Документов, которые могут препятствовать применению разрабатываемой технологии на территории Российской Федерации, не найдено.

На дальнейших этапах работ по указанному контракту необходимо проведение дополнительных патентных исследований для выявления новых патентов и не патентных работ в рассматриваемой области.

Материалы, отобранные для последующего анализа:

Таблица В.6.1 - Патентная документация

№ п/п	Страна выдачи, вид и номер охранного документа. Классификационный индекс*	Заявитель (патентообладатель), страна. Номер заявки, дата приоритета, конвенционный приоритет, дата публикации*	Название изобретения (полной модели, образца)	Сведения о действии охранного документа или причина его аннулирования (только для анализа патентной чистоты)
1.	Россия, патент RU 2487404 МПК G06F17/30	Патентообладатель: МО/ДТЭК ОНТВЕРПЕН Б.В. (NL) Заявка: 2009127889/08, 19.12.2007 Конвенционный приоритет: 19.12.2006 IT MI2006A002436 Дата публикации: 27.01.2011	Способ классификации веб-страниц и организации соответствующего информационного наполнения	действует
2.	Россия, RU 2011148231 МПК G06F17/30	Заявитель(и): МАЙКРОСОФТ КОРПОРЕЙШН (US) Заявка: 2011148231/08, 27.05.2010 Дата публикации заявки: 10.07.2013	Объединение результатов поиска	экспертиза по существу

№ п/п	Страна выдачи, вид и номер охранного документа. Классификационный индекс*	Заявитель (патентообладатель), страна. Номер заявки, дата приоритета, конвенционный приоритет, дата публикации*	Название изобретения (полной модели, образца)	Сведения о действии охранного документа или причина его аннулирования (только для анализа патентной чистоты)
3.	Россия, патент RU 2487403 МПК G06F17/27	Патентообладатель(и): Федеральное государственное бюджетное учреждение науки Институт системного программирования Российской академии наук (RU) Заявка: 2011148742/08, 30.11.2011 Дата начала отсчета срока действия патента: 30.11.2011 Приоритет(ы): Дата подачи заявки: 30.11.2011 Опубликовано: 10.07.2013	Способ построения семантической модели документа	действует
4.	Россия, патент RU 2409849 МПК G06F17/30	Патентообладатель(и): Закрытое акционерное общество "ТЕЛЕФОРМ" (RU) Заявка: 2008130492/08, 24.07.2008 Дата начала отсчета срока действия патента: 24.07.2008 Приоритет(ы): Дата подачи заявки: 24.07.2008 Дата публикации заявки: 27.01.2010 Опубликовано: 20.01.2011	Способ поиска информации в политематических массивах неструктурированных текстов	прекратил действие, но может быть восстановлен
5.	Россия, патент RU 2177638 МПК G06F17/30	Патентообладатель(и): Зоммерс Олег Харисович Заявка: 99119985/09, 22.09.1999 Дата начала отсчета срока действия патента: 22.09.1999 Опубликовано: 27.12.2001	Способ формирования и регулярной доставки пользователям информационных изданий и система для осуществления способа	действует
6.	Россия, патент RU 2000112647 МПК G06F17/30	Заявитель(и): ИНТЕРНЭШНЛ БИЗНЕС МАШИНЗ КОРПОРЕЙШН (US) Заявка: 2000112647/09, 27.10.1998 Дата публикации заявки: 10.04.2002	Способ кластеризации и сокращения размерности многомерных данных для их индексации и поиска	нет данных
7.	Россия, патент RU 2003104608 МПК G06F17/27	Заявитель(и): ОНТОС АГ (CH) Заявка: 2003104608/09, 18.02.2003 (43) Дата публикации заявки: 20.09.2004	Способ автоматизированной обработки информационных текстовых материалов	Нет данных
8.	Россия, патент RU 2273879 МПК G06F17/30, G09B19/00	Патентообладатель(и): Насыпный Владимир Владимирович (RU), Насыпная Галина Анатольевна (RU) Заявка: 2004131643/09, 28.05.2002 (24) Дата начала отсчета срока действия патента: 28.05.2002 (43) Дата публикации заявки: 20.04.2005 (45) Опубликовано: 10.04.2006	Способ синтеза самообучающейся системы извлечения знаний из текстовых документов для поисковых систем	действует

№ п/п	Страна выдачи, вид и номер охранного документа. Классификационный индекс*	Заявитель (патентообладатель), страна. Номер заявки, дата приоритета, конвенционный приоритет, дата публикации*	Название изобретения (полной модели, образца)	Сведения о действии охранного документа или причина его аннулирования (только для анализа патентной чистоты)
9.	Россия, патент RU 2365984 МПК G06F17/30, G06F12/00	Патентообладатель(и): МАЙКРОСОФТ КОРПОРЕЙШН (US) Заявка: 2004108667/09, 23.03.2004 Дата начала отсчета срока действия патента: 23.03.2004 Конвенционный приоритет: 24.03.2003 US 10/395,679 Дата публикации заявки: 27.09.2005 Опубликовано: 27.08.2009	Поиск произвольного текста и поиск по атрибутам в данных электронного руководства по программам	действует
10.	Россия, патент RU 2393533 МПК G06F17/00	Патентообладатель(и): МАЙКРОСОФТ КОРПОРЕЙШН (US) Заявка: 2005111000/09, 14.04.2005 Дата начала отсчета срока действия патента: 14.04.2005 Конвенционный приоритет: 15.04.2004 US 10/825,894 Дата публикации заявки: 20.10.2006 Опубликовано: 27.06.2010	Предложение родственных терминов для многосмыслового запроса	действует
11.	Россия, патент RU 2375747 МПК G06F17/30	Патентообладатель(и): МАЙКРОСОФТ КОРПОРЕЙШН (US) Заявка: 2005111001/09, 14.04.2005 Дата начала отсчета срока действия патента: 14.04.2005 Конвенционный приоритет: 15.04.2004 US 10/826,162 Дата публикации заявки: 20.10.2006 Опубликовано: 10.12.2009	Проверка релевантности между ключевыми словами и содержанием веб-сайта	действует
12.	Россия, патент RU 2305314 МПК G06F17/30	Патентообладатель(и): Общество с ограниченной ответственностью "Центр Компьютерного моделирования" (RU) Заявка: 2006122998/09, 28.06.2006 Дата начала отсчета срока действия патента: 28.06.2006 Опубликовано: 27.08.2007	Способ поиска и выборки информации из различных баз данных	действует
13.	Россия, патент RU 2399959 МПК G09B19/00, G06F17/27	Патентообладатель(и): Закрытое акционерное общество "Авикомп Сервисез" (RU) Заявка: 2008142648/12, 29.10.2008 Дата начала отсчета срока действия патента: 29.10.2008 Дата публикации заявки: 10.05.2010 Опубликовано: 20.09.2010	Способ автоматизированной обработки текста на естественном языке путем его семантической индексации, способ автоматизированной обработки коллекции текстов на естественном языке путем их семантической индексации и машиночитаемые носители	прекратил действие, но может быть восстановлен

№ п/п	Страна выдачи, вид и номер охранного документа. Классификационный индекс*	Заявитель (патентообладатель), страна. Номер заявки, дата приоритета, конвенционный приоритет, дата публикации*	Название изобретения (полной модели, образца)	Сведения о действии охранного документа или причина его аннулирования (только для анализа патентной чистоты)
14.	Россия, патент RU 2393533 МПК G06F17/00	Патентообладатель(и): МАЙКРОСОФТ КОРПОРЕЙШН (US) Заявка: 2005111000/09, 14.04.2005 Дата начала отсчета срока действия патента: 14.04.2005 Конвенционный приоритет: 15.04.2004 US 10/825,894 Дата публикации заявки: 20.10.2006 Опубликовано: 27.06.2010	Предложение родственных терминов для многозначного запроса	действует
15.	Россия, патент RU 2443015 МПК G06F17/30	Патентообладатель(и): МАЙКРОСОФТ КОРПОРЕЙШН (US) Заявка: 2009103140/08, 20.07.2007 Дата начала отсчета срока действия патента: 20.07.2007 Приоритет(ы): Конвенционный приоритет: 31.07.2006 US 11/496,373 Дата публикации заявки: 10.08.2010 Опубликовано: 20.02.2012	Функции ранжирования, использующие модифицированный наивный байесовский классификатор запросов с инкрементным обновлением	действует
16.	Россия, патент RU 2449357 МПК G06F17/30, G06F3/048	Патентообладатель(и): МАЙКРОСОФТ КОРПОРЕЙШН (US) Заявка: 2009120517/08, 03.11.2007 Дата начала отсчета срока действия патента: 03.11.2007 Конвенционный приоритет: 30.11.2006 US 11/565,224 Дата публикации заявки: 10.12.2010 Опубликовано: 27.04.2012	Диаграмма ранжирования	действует
17.	Россия, патент RU 2456661 МПК G06F3/048, G09G5/32	Патентообладатель(и): МАЙКРОСОФТ КОРПОРЕЙШН (US) Заявка: 2009109199/08, 20.08.2007 Дата начала отсчета срока действия патента: 20.08.2007 Конвенционный приоритет: 15.09.2006 US 11/532,369 Дата публикации заявки: 20.09.2010 Опубликовано: 20.07.2012	Эффективная навигация результатов поиска	действует
18.	Россия, патент RU 2011132437 МПК G06F17/30	Заявитель(и): Общество с ограниченной ответственностью "Алгоритм" (RU) Заявка: 2011132437/08, 01.08.2011 Приоритет(ы): Дата подачи заявки: 01.08.2011 Дата публикации заявки: 10.02.2013	Способ информационного поиска (варианты) и компьютерная система для его осуществления	экспертиза по существу
19.	Россия, патент RU 2011130218 МПК G06F17/30	Заявитель(и): Форнова Лтд (IL) Заявка: 2011130218/08, 27.12.2009 Конвенционный приоритет: 31.12.2008 US 61/193,862 27.09.2009 US 12/567,773 Дата публикации заявки: 10.02.2013	Система и способ агрегации данных с множества веб-сайтов	экспертиза по существу

№ п/п	Страна выдачи, вид и номер охранного документа. Классификационный индекс*	Заявитель (патентообладатель), страна. Номер заявки, дата приоритета, конвенционный приоритет, дата публикации*	Название изобретения (полной модели, образца)	Сведения о действии охранного документа или причина его аннулирования (только для анализа патентной чистоты)
20.	Россия, патент RU 2011134300 МПК G06F17/30	Заявитель(и): ГУГЛ ИНК. (US) Заявка: 2011134300/08, 16.01.2009 Приоритет(ы): Дата подачи заявки: 16.01.2009 Дата публикации заявки: 27.02.2013	Визуализация структуры сайта и разрешение переходов сайта для результата поиска или связанной страницы	экспертиза завершена
21.	Россия, патент RU 2453916 МПК G06F17/30	Патентообладатель(и): Лебедев Игорь Викторович (RU) Заявка: 2011118330/08, 05.05.2011 Дата начала отсчета срока действия патента: 05.05.2011 Приоритет(ы): Дата подачи заявки: 05.05.2011 Опубликовано: 20.06.2012	Способ поиска информационных ресурсов с использованием переадресаций	может прекратить свое действие
22.	Россия, патент RU 2454712 МПК G06F17/30	Патентообладатель(и): МАЙКРОСОФТ КОРПОРЕЙШН (US) Заявка: 2009136173/08, 28.02.2008 Дата начала отсчета срока действия патента: 28.02.2008 Конвенционный приоритет: 30.03.2007 US 11/731,619 Дата публикации заявки: 10.04.2011 Опубликовано: 27.06.2012	Генерация запроса с использованием конфигурации среды	действует
23.	Россия, патент RU 2459242 МПК G06F17/30	Патентообладатель(и): Серебренников Олег Александрович (RU) Заявка: 2011133310/08, 09.08.2011 Дата начала отсчета срока действия патента: 09.08.2011 Приоритет(ы): Дата подачи заявки: 09.08.2011 Опубликовано: 20.08.2012	Способ создания и использования рекурсивного индекса поисковых машин	действует
24.	Россия, патент RU 2011122663 МПК G06F17/30	МАЙКРОСОФТ КОРПОРЕЙШН (US) Заявка: 2011122663/08, 12.11.2009 Приоритет(ы): Конвенционный приоритет: 04.12.2008 US 12/328,570 Дата публикации заявки: 10.12.2012	Предварительный просмотр результатов поиска для предложенных терминов уточнения и вертикальных поисков	экспертиза завершена
25.	Россия, патент RU 2473119 МПК G06F17/30	Патентообладатель(и): Учреждение Российской академии наук Институт Системного Анализа РАН (ИСА РАН) (RU) Заявка: 2011132873/08, 05.08.2011 Дата начала отсчета срока действия патента: 05.08.2011 Приоритет(ы): Дата подачи заявки: 05.08.2011 Опубликовано: 20.01.2013	Способ и система семантического поиска электронных документов	действует

№ п/п	Страна выдачи, вид и номер охранного документа. Классификационный индекс*	Заявитель (патентообладатель), страна. Номер заявки, дата приоритета, конвенционный приоритет, дата публикации*	Название изобретения (полной модели, образца)	Сведения о действии охранного документа или причина его аннулирования (только для анализа патентной чистоты)
26.	Россия, патент RU 2473964 МПК G06F17/30	Патентообладатель(и): Государственное казенное образовательное учреждение высшего профессионального образования Академия Федеральной службы охраны Российской Федерации (Академия ФСО России) (RU) Заявка: 2011151684/08, 16.12.2011 Дата начала отсчета срока действия патента: 16.12.2011 Приоритет(ы): Дата подачи заявки: 16.12.2011 Опубликовано: 27.01.2013	Способ обнаружения идентификационных признаков для различных буквенно-знаковых систем письменности	действует
27.	Россия, патент RU 2474873 МПК G06F17/30, G06F19/00	Патентообладатель(и): Кашик Алексей Сергеевич (RU) Заявка: 2010153971/08, 29.12.2010 Дата начала отсчета срока действия патента: 29.12.2010 Приоритет(ы): Дата подачи заявки: 29.12.2010 Дата публикации заявки: 10.07.2012 Опубликовано: 10.02.2013	Способ анализа и прогноза развития сложно построенной системы и устройство для его осуществления	действует
28.	Россия, патент RU 2475833 МПК G06F17/30	Патентообладатель(и): КОНИНКЛЕЙКЕ ФИЛИПС ЭЛЕКТРОНИКС Н.В. (NL) Заявка: 2009124452/08, 26.11.2007 Дата начала отсчета срока действия патента: 26.11.2007 Конвенционный приоритет: 29.11.2006 EP 06124988.4 Дата публикации заявки: 10.01.2011 Опубликовано: 20.02.2013	Фильтр по образцу	действует

* Заполняется при необходимости.

Таблица В.6.2 - Научно-техническая, конъюнктурная, нормативная документация и материалы государственной регистрации (отчеты о научно-исследовательских работах)

Предмет поиска	Наименование источника информации с указанием страницы источника	Автор, фирма (держатель) технической документации	Год, место и орган издания (утверждения, депонирования источника)
1	2	3	4

* ПИ - патентные исследования; ТУ - технические условия; ПФ - патентный формуляр.

ПРИЛОЖЕНИЕ Г

УТВЕРЖДАЮ

Руководитель работ

канд. техн. наук. В. К. Иванов

«06» декабря 2013 г.

Реферативный список патентов

Способ классификации веб-страниц и организации соответствующего информационного наполнения

(RU 2487404; Статус: действует)

Изобретение относится к области классификации веб-страниц и организации соответствующего информационного наполнения. Техническим результатом является повышение производительности поисков, выполняемых по веб-страницам. Раскрывается способ классификации веб-страниц и организации соответствующего информационного наполнения, включающий этап записи Интернет-адресов, в свою очередь включающий выполнение процессов автоматической записи указанных Интернет-адресов, и этап выбора, на котором назначают указанным Интернет-адресам соответствующий коэффициент релевантности. На этапе выбора указанный коэффициент релевантности назначают пропорционально его записи на этапе записи и выбирают Интернет-адреса, коэффициент релевантности которых выше заданного порогового значения. Способ также содержит этап сокращения, на котором удаляют Интернет-адреса, коэффициент релевантности которых выше порогового значения, не удовлетворяющие по меньшей мере одному критерию существенности, и этап проверки, на котором проверяют ряд указанных Интернет-адресов, удовлетворяющих указанным критериям существенности, причем этап проверки содержит действия, выполняемые человеком. 2 н. и 16 з.п. ф-лы, 6 ил.

Объединение результатов поиска

(RU 2011148231; Статус: Экспертиза по существу)

Реферат: Нет данных

Способ построения семантической модели документа

(RU 2487403; Статус: действует)

Изобретение относится к области обработки данных при семантическом анализе текстовых данных и построении семантической модели документов. Техническим результатом является обеспечение возможности обогащения документов метаданными, позволяющими улучшить и увеличить скорость восприятия основной информации, а также обеспечение возможности определять и подсвечивать ключевые термины текста, что позволяет ускорить его чтение и улучшить понимание. Способ построения семантической модели документа состоит из двух основных шагов. На первом из внешних информационных ресурсов, которые содержат описания отдельных объектов предметной области, извлекают онтологию. На втором - связывают текстовую информацию документа с концепциями онтологии и строят семантическую модель документа. В качестве информационных источников используют электронные ресурсы как связанные, так и не связанные структурой гипертекстовых ссылок. Технический результат, в частности, достигается за счет того, что сначала выделяют все термины документа и связывают их с концепциями онтологии таким

образом, чтобы каждому термину соответствовала единственная концепция, являющаяся его значением, а затем значения терминов ранжируются по важности к документу. 14 з.п. ф-лы, 6 ил.

Способ поиска информации в политематических массивах неструктурированных текстов

(RU 2409849, Статус: прекратил действие, но может быть восстановлен)

Изобретение относится к области информационных технологий, в частности к способам поиска информации в больших документальных базах данных (БД). Техническим результатом является сокращение времени поиска. В способе формируют базу терминов по определенной тематике, запрос и при поиске по запросу и нахождении каждого документа заносят в память компьютера его заголовок и адрес, пословесно сравнивают заголовки с терминами из базы терминов по определенной тематике и заносят в память количество совпавших терминов для каждого документа и его адреса, проверяют наличие документов, для которых количество совпавших терминов равно нулю, и удаляют их заголовок и адрес из памяти компьютера, сортируют заголовки и адреса оставшихся документов по количеству совпавших терминов, осуществляют проверку выданного документа на соответствие первоначальному запросу, если выданный документ не соответствует первоначальному запросу, то всю информацию о нем удаляют и осуществляют выдачу на экран монитора следующего по критерию документа. 4 з.п. ф-лы, 2 ил.

Способ формирования и регулярной доставки пользователям информационных изданий и система для осуществления способа

(RU 2177638; Статус: действует)

Изобретение относится к области информационного обслуживания с использованием компьютерных сетей. Технический результат заключается в повышении избирательности и уровня персонализации информационных изданий, обеспечения непрерывной адаптации предоставляемых материалов к изменяющимся потребностям пользователей, расширения круга потенциальных распространителей издания и повышения оперативности его доставки; оптимизации сетевой инфраструктуры, обеспечивающей возможность формирования практически неограниченного множества тематических изданий при сохранении единого центра, осуществляющего поиск, отбор и систематизацию материалов. Подготовка персонализированного информационного издания осуществляется на основании обработки данных в востребованных ранее материалах и анализе полученных откликов об их качестве и соответствии области интересов. Основную роль в подготовке и распространении персонализированного информационного издания играет издатель, который выбирает информационные материалы из доступных источников и помещает их в очередной выпуск издания. Каждый выпуск готовится персонально для пользователя или группы пользователей, подписавшихся на информационное издание, исходя из области их интересов и предпочтений. Распространение издания может осуществляться непосредственно издателем через собственный информационный сервер и через сеть независимых распространителей, получивших от издателя права на регистрацию пользователей и распространение издания через свои серверы. 5 с. и 74 з.п.ф-лы, 1 ил.

Способ кластеризации и сокращения размерности многомерных данных для их индексации и поиска

(RU 2000112647; Статус: Нет данных)

Реферат: Нет данных

Способ автоматизированной обработки информационных текстовых материалов

(RU 2003104608; Статус: Нет данных)

Изобретение относится к обработке информационных естественно-языковых текстовых материалов. Его использование позволяет получить технический результат в виде

автоматизированной обработки информационных текстовых материалов на основе выделения смысловых категорий в текстах на естественном языке с графическим представлением смысла всего текста в виде когнитивной карты, а также в виде повышения эффективности подготовки обрабатываемых информационных текстовых материалов к аналитической их обработке в условиях персонализированного информационного обеспечения. Этот технический результат достигается тем, что предварительно информационные потребности пользователя по определенной тематике формируют в виде ориентированного графа, а обработку текста ведут в фазе препроцессии, включающей этап морфологического анализа текста, этап поиска устойчивых словосочетаний и этап сегментации текста на предложения, в фазе процессии, включающей этап выделения примеров типовых отношений и этап выделения примеров типов объектов, и в фазе постпроцессии, состоящей из этапа формирования элементарных графов и этапа слияния элементарных графов в когнитивную карту. 28 з.п.ф-лы, 8 ил.

Способ синтеза самообучающейся системы извлечения знаний из текстовых документов для поисковых систем

(RU 2273879; Статус: действует)

Изобретение относится к вычислительной технике, информационно-поисковым и интеллектуальным системам. Его использование обеспечивает возможность автоматического формирования знаний путем извлечения их из текстовых документов, представленных на различных языках в электронном виде, и интеллектуальную обработку текстовой информации и запросов пользователей с целью извлечения знаний на любом иностранном языке. Этот результат достигается благодаря обеспечению механизма самообучения в виде стохастически индексированной системы искусственного интеллекта; производят морфологический и синтаксический анализ, а также стохастическое индексирование текстовых документов по заданной теме для формирования баз знаний семантического анализа; преобразуют запрос пользователя в стохастически индексированном виде во множество новых запросов, эквивалентных исходному запросу, и осуществляют выбор стохастически индексированных фрагментов текстовых документов со всеми словосочетаниями преобразованного запроса, из которых формируют стохастически индексированную семантическую структуру, формируют краткий ответ системы на основе этой структуры и проверяют релевантность краткого ответа системы запросу путем их сравнения. 2 н. и 18 з.п. ф-лы, 1 ил., 3 табл.

Поиск произвольного текста и поиск по атрибутам в данных электронного руководства по программам

(RU 236598; Статус: действует)

Изобретение относится к поиску произвольного текста и атрибутов данных электронного руководства по программам. Изобретение позволяет упростить поиск в электронном руководстве по программам. Объект изобретения включает в себя машину поиска данных электронного руководства по программам (ЭРП) и связанные с ним способы. Текстовая поисковая цепочка может быть нормализована в доступные для поиска элементы и элементы, интерпретируемые или как текстовые элементы поиска, или как элементы поиска по атрибутам. Создают один или более запросов, имеющие условия поиска различной степени сложности, в соответствии с интерпретацией элементов поисковой цепочки. Одну или более операций поиска в базах данных ЭРП и/или ресурсах сети выполняют на основании интерпретации элементов текста и атрибутов и результатам присваивают значение ранжирования результатов согласно интерпретации. Объединенные результаты поиска могут быть сгруппированы, ранжированы и фильтрованы для отображения пользователю. Результаты могут также быть отображены прогрессивно, по мере того как каждый символ поисковой цепочки вводится пользователем. 5 н. и 30 з.п. ф-лы, 7 ил.

Предложение родственных терминов для многозначного запроса

(RU 2393533; Статус: действует)

Изобретение относится к системе и способу для предложения родственных терминов. Техническим результатом является расширение функциональных возможностей за счет улучшения идентификации ключевых слов, релевантных для содержания веб-сайта. В способе формируются кластеры терминов в зависимости от вычисленного подобия векторов терминов. Вектор каждого термина создается из результатов поиска, связанных с группой предыдущих запросов с высокой частотой появления (ЧП), поданных ранее в поисковую машину. В ответ на получение термина/фразы от объекта этот термин/фраза оценивается, принимая во внимание термины/фразы в кластерах терминов, чтобы идентифицировать одно или более предложений родственных терминов. Система реализует указанный способ. 4 н. и 42 з.п. ф-лы, 4 ил., 1 табл.

Проверка релевантности между ключевыми словами и содержанием веб-сайта

(RU 2375747; Статус: действует)

Изобретение относится к информационному анализу данных и, в частности, к проверке релевантности между терминами и содержанием веб-сайта. Изобретение позволяет находить ключевые слова, в большей степени релевантные содержанию веб-сайта. Находится содержание сайта из URL предложения. Вычисляется(ются) расширенный(ые) термин(ы), семантически и/или контекстуально связанный(ые) с термином(ами) предложения. Вычисляются измерения подобия содержания и расширенного подобия из соответствующих комбинаций термина(ов) предложения, содержания сайта и расширенных терминов. Определяются измерения подобия категории между расширенными терминами и содержанием сайта с учетом обучаемого классификатора подобия. Обучается обучаемый классификатор подобия из проанализированного содержания сайта, связанного с данными каталога. Определяется значение достоверности, обеспечивающее объективную меру релевантности между термином(ами) предложения и содержанием сайта, из измерений подобия содержания, расширенного подобия и подобия категории, оценивающих множество оценок подобия с учетом обучаемой модели классификатора релевантности. 4 н. и 37 з.п. ф-лы, 4 ил., 1 табл.

Способ поиска и выборки информации из различных баз данных

(RU 2305314; Статус: действует)

Заявленное изобретение относится к способу поиска и идентификации документов по их описаниям, находящимся в различных базах данных и информационных ресурсах с различными стандартами формирования документов. Технический результат заключается в повышении точности поиска и проведения анализа полученной информации. Технический результат достигается за счет того, что сформированные пользователем поисковые запросы передаются в поисковую систему сервера, которая производит обработку упомянутых запросов путем выбора документов из различных баз данных, поисковая система объединяет все выбранные документы в единый список, сортирует упомянутые выбранные документы по тематикам, формирует папки, которые содержат упомянутые документы, одной тематики, снова сортируют упомянутые отсортированные документы с учетом окончательного рейтинга. После чего на основе пользовательского запроса определяют разделы будущего отчета, с помощью поисковой системы определяют текстовые признаки начала и завершения разделов, проводят разметку текста выбранных с наибольшими показателями окончательного рейтинга документов, внутри каждого раздела выделяют сегменты текста, проводят сортировку сегментов в соответствии с датой публикации, подготавливают итоговый отчет, в котором сегменты текста, отсортированные в соответствии с датой публикации оригинального документа, объединены в единый текстовый массив, после чего передают на пользовательский терминал через телекоммуникационные средства связи итоговый отчет. 4 ил., 1 табл.

Способ автоматизированной обработки текста на естественном языке путем его семантической индексации, способ автоматизированной обработки коллекции текстов на естественном языке путем их семантической индексации и машиночитаемые носители

(RU 2399959 Статус: прекратил действие, но может быть восстановлен)

Изобретение относится к области информационных технологий. Текст сегментируют в электронной форме на элементарные единицы. Выявляют устойчивые словосочетания, формируют предложения. Выявляют семантически значимые объекты и семантически значимые отношения между ними. Формируют для каждого семантически значимого отношения множество триад, в которых единственная триада первого типа соответствует связи, устанавливаемой семантически значимым отношением между двумя семантически значимыми объектами. Каждая из триад второго типа соответствует значению конкретного атрибута одного из этих семантически значимых объектов. Каждая из триад третьего типа соответствует значению конкретного атрибута самого семантически значимого отношения. Индексируют на множестве сформированных триад все связанные семантически значимыми отношениями семантически значимые объекты по отдельности. Запоминают в базе данных сформированные триады и полученные индексы вместе со ссылкой на исходный текст, из которого сформированы эти триады. Техническим результатом изобретения является повышение точности и скорости поиска релевантных фактов и документов. 4 н. и 8 з.п. ф-лы, 16 табл., 7 ил.

Предложение родственных терминов для многосмыслового запроса

(RU 2393533; Статус: действует)

Изобретение относится к системе и способу для предложения родственных терминов. Техническим результатом является расширение функциональных возможностей за счет улучшения идентификации ключевых слов, релевантных для содержания веб-сайта. В способе формируются кластеры терминов в зависимости от вычисленного подобия векторов терминов. Вектор каждого термина создается из результатов поиска, связанных с группой предыдущих запросов с высокой частотой появления (ЧП), поданных ранее в поисковую машину. В ответ на получение термина/фразы от объекта этот термин/фраза оценивается, принимая во внимание термины/фразы в кластерах терминов, чтобы идентифицировать одно или более предложений родственных терминов. Система реализует указанный способ. 4 н. и 42 з.п. ф-лы, 4 ил., 1 табл.

Функции ранжирования, использующие модифицированный наивный байесовский классификатор запросов с инкрементным обновлением

(RU 2443015; Статус: действует)

Изобретение относится к средствам ранжирования документов в сети, использующим систему с инкрементным обновлением. Техническим результатом является сокращение времени поиска документов в сети с повышением соответствия результатов поиска поисковому запросу. В способе ранжирования документов в сети в ответ на введенный пользователем поисковый запрос, содержащий один или несколько термов поискового запроса, используют классификатор запроса с инкрементным обновлением на основе данных использования с отображением ранжированных документов с вычисленными значениями параметров просмотра, сохраненными для каждого из термов поискового запроса, при этом обновляют значения подсчета с использованием инкрементного обновления. 4 н. и 13 з.п. ф-лы, 6 ил.

Диаграмма ранжирования

(RU 2449357; Статус: действует)

Изобретение относится к вычислительной технике. Технический результат заключается в расширении функциональных возможностей за счет фильтрации релевантности результатов

поиска на основе визуальной информации. Система для передачи усовершенствованных результатов запроса, содержащая процессор, который исполняет следующие исполняемые компьютером компоненты, сохраненные на компьютерном носителе хранения: компонент интерфейса, принимающий запрос, соответствующий данным, и преобразующий по меньшей мере одно из аудио или видео элемента данных для определения местоположения одного или более поисковых терминов запроса в пределах элемента данных, когда запрашиваемые данные содержат по меньшей мере одно из аудио или видео элемента данных; и визуальный компонент, передающий инструкции для отображения результата запроса по меньшей мере частично на основе запроса, при этом результат запроса идентифицирует один или более элементов данных и включает график местоположений поискового термина для каждого из одного или более элементов данных, причем график местоположений поискового термина идентифицирует местоположение вхождения в пределах элемента данных по меньшей мере одного поискового термина запроса. 3 н. и 16 з.п. ф-лы, 12 ил.

Эффективная навигация результатов поиска

(RU 2456661; Статус: действует)

Настоящее изобретение относится к системе и способу для навигации по результатам поиска. Техническим результатом является ускорение доступа к релевантной информации. В ответ на прием поискового запроса по меньшей мере один отдельный результат поиска представляется в перекрывающем окне, причем перекрывающее окно конфигурируется, чтобы перекрывать по меньшей мере часть окна просмотра документов. Навигация между перекрывающим окном и окном просмотра документов может выполняться в ответ на прием ввода по меньшей мере одной заранее определенной команды. Содержимое, связанное с выбранным отдельным результатом поиска (или другим идентификатором документа), может быть представлено в окне просмотра документов, и по существу одновременно перекрывающее окно может быть скрыто из виду. Последовательно, в ответ на прием по меньшей мере одной заранее определенной команды, перекрывающее окно может быть повторно представлено так, чтобы оно снова перекрывало по меньшей мере часть окна просмотра документов. 3 н. и 10 з.п. ф-лы, 4 ил.

Способ информационного поиска (варианты) и компьютерная система для его осуществления

(RU 2011132437; Статус: Экспертиза по существу)

Реферат: нет данных

Система и способ агрегации данных с множества веб-сайтов

(RU 2011130218; Экспертиза по существу)

Реферат: нет данных

Визуализация структуры сайта и разрешение переходов сайта для результата поиска или связанной страницы

(RU 2011134300; Статус: Экспертиза завершена)

Реферат: нет данных

Способ поиска информационных ресурсов с использованием переадресаций

(RU 2453916; Статус: может прекратить свое действие)

Изобретение относится к способам поиска информационных ресурсов с использованием переадресаций с исходных адресов информационных ресурсов на результирующие адреса информационных ресурсов. Техническим результатом является расширение функциональных возможностей доступа к ресурсам в сети за счет организации способа

переадресации. Сущность заявляемого технического решения заключается в том, что в клиентское приложение введен клиентский модуль и при определении факта отсутствия доступа к запрашиваемому информационному ресурсу клиентское приложение осуществляет обработку исходного адреса информационного ресурса, в результате которой при интерпретации исходного адреса информационного ресурса как псевдонима или адреса доступного информационного ресурса клиентское приложение производит доступ к информационному ресурсу по результирующим адресам, соответствующим исходному адресу информационного ресурса, иначе если исходный адрес не интерпретирован как псевдоним или адрес доступного информационного ресурса, то клиентское приложение формирует и передает запрос, содержащий исходный адрес информационного ресурса, к БД клиентского модуля и серверу информационной системы, которая осуществляет обработку исходного адреса информационного ресурса с определением и/или поиском результирующих адресов информационного ресурса, соответствующих исходному адресу информационного ресурса. 74 з.п. ф-лы, 1 ил.

Генерация запроса с использованием конфигурации среды

(RU 2454712; Статус: действует)

Изобретение относится к способам поиска информации. Техническим результатом является расширение функциональных возможностей поиска информации. Запрос для системы помощи включает в себя данные о системе пользователя и задаче, которую предпринимает пользователь. Запрос может быть использован процессором поиска, чтобы генерировать релевантные результаты, чтобы помочь пользователю. Данные системы пользователя могут включать в себя данные конфигурации об аппаратном обеспечении и программном обеспечении. Данные задачи могут быть получены из текущего состояния устройства или из операционной предыстории, которая может быть создана от одного пользователя или группы пользователей. Запрос может иметь механизм, чтобы взвешивать различные ключевые слова или компоненты запроса, а система обратной связи может регулировать веса для будущих запросов. 5 н. и 15 з.п. ф-лы, 4 ил.

Способ создания и использования рекурсивного индекса поисковых машин

(RU 2459242; Статус: действует)

Изобретение относится к способу построения и использования индекса поисковых машин и относится к области компьютерной техники и способам обработки информации. Техническим результатом является расширение функциональных возможностей поиска информации. Настоящее изобретение позволяет реализовать функциональность «вытягивания текста» из индекса поисковой машины с произвольного места в тексте, а также существенно снизить трудоемкость анализа совместной встречаемости слов в тексте, что, в свою очередь, позволяет делать предположения о неявных (причинно-следственных, ассоциативных и прочих) связях между объектами в их последовательностях. Изобретение расширяет возможность применения поисковых машин в области индексирования и поиска информации, представленной последовательностью объектов, отличных от объектов текстовой информации. Настоящее изобретение предлагает в дополнение информации Базового Хита размещать в Расширенном Хите по меньшей мере wordID для двух слов: «предыдущего слова», позиция которого в документе на единицу меньше и равна (position-1), и «следующего слова», позиция которого в документе на единицу больше и равна (position+1). 2 н. и 33 з.п. ф-лы, 19 ил.

Предварительный просмотр результатов поиска для предложенных терминов уточнения и вертикальных поисков

(RU 2011122663; Статус: Экспертиза завершена)

Реферат: нет данных

Способ и система семантического поиска электронных документов

(RU 2473119; Статус: действует)

Предложенная группа изобретений относится к средствам автоматизированного информационного поиска и обеспечивает высокую полноту и точность информационного поиска. Техническим результатом является повышение полноты и точности поиска электронных документов. Предложенный способ семантического поиска электронных документов включает дополнение индексной структуры электронного документа маркерами, занимающими одинаковый и фиксированный размер в компьютерной памяти, хранящими в зависимости от своего типа как метаинформацию электронных документов, так и информацию о вхождениях слов естественного языка в электронные документы, полученную с помощью лингвистического анализа. Предложенная система семантического поиска электронных документов включает соответствующие модули: модуль формирования коллекции и выделения метаинформации электронных документов, модуль хранения индексных структур электронных документов, модуль формирования и выдачи пользователю поисковой системы перечня электронных документов, ранжированных по релевантности, модуль формирования маркеров, хранящих метаинформацию электронных документов, модуль лингвистического анализа текста на естественном языке и формирования маркеров, хранящих информацию о вхождениях слов естественного языка. 2 н. и 2 з.п. ф-лы, 3 ил.

Способ обнаружения идентификационных признаков для различных буквенно-знаковых систем письменности

(RU 2473964; Статус: действует)

Изобретение относится к области информатики и вычислительной техники и может использоваться для обработки информационных потоков и обнаружения в них заданных эталонных признаков, представленных в различных буквенно-знаковых системах письменности. Техническим результатом является расширение функциональных возможностей в применении способа для обнаружения идентификационных признаков, представленных в различных буквенно-знаковых системах письменности, а также повышение эффективности использования памяти для хранения эталонных информационных признаков. Способ заключается в том, что осуществляют расчет уникального трехбайтного адреса базы эталонных информационных признаков (БЭИП) и записывают по данному адресу значение уровня логической единицы. Представление эталонного информационного признака в виде трехбайтного адреса позволяет при необходимости дополнять БЭИП без перерасчета значений адресов и морфологических коэффициентов информационных признаков. Трехбайтные адреса для информационных признаков не зависят от символьной длины словарных признаков. 2 ил., 7 табл.

Способ анализа и прогноза развития сложно построенной системы и устройство для его осуществления

(RU 2474873; Статус: действует)

Изобретение относится к области информационных технологий и относится к визуальному анализу информации, характеризующей состояние и прогноз развития различных областей человеческой деятельности. Техническим результатом является расширение функциональных возможностей визуализации данных. При проведении анализа и прогноза развития сложно построенной системы проводят формирование базы данных объектов системы в трехмерном пространстве и исследование базы данных слайсированием. Базу данных формируют в виде многоступенчатой многоуровневой иерархической базы данных объектов системы, начиная с объектов низшего уровня иерархии. Базу данных формируют из параметров объектов низшего уровня в трехмерном пространстве с осями: объект нижнего уровня - параметры нижнего уровня - время. Объекты второго уровня формируют математической обработкой параметров объектов первого уровня. Совокупность объектов

второго уровня формируют в трехмерном пространстве с осями: ось объектов второго уровня - параметры объектов второго уровня - время. Объекты третьего уровня формируют из объектов второго уровня аналогично получению объектов второго уровня из объектов первого и т.д. Слайсирование полученной таким образом иерархической базы данных проводят поуровненно с выявлением закономерностей состояния и развития сложно построенной системы. 2 н. и 5 з.п. ф-лы, 9 ил.

Фильтр по образцу

(RU 2475833; Статус: действует)

Изобретение относится к идентификации определенного объекта данных из набора объектов данных. Техническим результатом является расширение функциональных возможностей системы за счет возможности получать и использовать пользовательскую обратную связь, содержащую причину, указывающую, почему найденные данные релевантны и/или не релевантны запросу. Система содержит: блок (110) составления для составления запроса для идентификации определенного объекта данных; блок (120) идентификации для идентификации возможного объекта данных из набора объектов данных на основе запроса; блок (130) представления для представления описания возможного объекта данных пользователю; блок (140) обратной связи для приема пользовательской обратной связи по отличительной характеристике из множества отличительных характеристик, отличительная характеристика описывает возможный объект данных; и блок (150) обновления для обновления запроса для идентификации определенного объекта данных на основе пользовательской обратной связи по отличительной характеристике, описывающей возможный объект данных. Таким образом, пользователь может оценить отличительную характеристику возможного объекта данных на основе информации, полученной из блока представления и предоставить пользовательскую обратную связь, содержащую причину, указывающую, почему найденные данные являются релевантными и/или нерелевантными для запроса. 5 н. и 6 з.п. ф-лы, 6 ил.