

7. Фоменков С.А., Колесников С.Г. Информационное наполнение баз данных по физическим эффектам // Информационные технологии. 2004. № 6. С. 60-62.

8. Фоменков С.А., Колесников С.Г., Коробкин Д.М. Формирование и структура баз данных по физическим эффектам. // Известия Волгоградского государственного технического университета. 2013. Т. 18. № 22 (125). С. 153-157.

9. Фоменков С. А., Колесников С.Г., Дворянкин А.М. Использование структурированных физических знаний для прогнозирования новых нанотехнических систем // Известия Волгоградского государственного технического университета. 2012. Т. 4. № 13. С. 80-82.

Methods of updating physical effects database on example of nanotechnical systems

Sergey Alekseevich Fomenkov, Professor, Doctor of Technical Sciences

Sergey Grigoryevich Kolesnikov, Senior Research Associate

*Dmitriy Mikhaylovich Korobkin, Candidate of Technical Sciences, Associate Professor
Volgograd State Technical University*

The methods of updating the database of physical effects are presented in the article. The generalized procedures of the methods and the recommendation about their performance are described. The application methods are dwelled on.

Keywords: methods, database of physical effects, software complex, nanotechnical systems.

УДК 004.822

АВТОМАТИЧЕСКОЕ ПОЗИЦИОНИРОВАНИЕ ЗАЯВКИ НА ПОЛУЧЕНИЕ ПАТЕНТА В ГЛОБАЛЬНОМ ПАТЕНТНОМ ПРОСТРАНСТВЕ

Алла Григорьевна Кравец, д-р.техн.наук, проф. кафедры САПР и ПК

Тел.: (8442) 248 100, e-mail: agk@gde.ru

Дмитрий Михайлович Коробкин, канд.техн.наук, доц. кафедры САПР и ПК

Тел.: (8442) 248 100, e-mail: agk@gde.ru

Сергей Сергеевич Укустов, асп. кафедры САПР и ПК

Тел.: (8442) 248 100, e-mail: agk@gde.ru

Михаил Александрович Дыков, асп. кафедры САПР и ПК

Тел.: (8442) 248 100, e-mail: agk@gde.ru

Волгоградский государственный технический университет

www.vstu.ru

В статье рассматривается реализация пилотного проекта «E-patent Examiner», в рамках которого был создан макет системы и размещен на серверах «облачного» кластера. Для проведения экспериментов в базу знаний макета были загружена галерея патентов по предметным областям «электричество», «физика», «механика» из патентных массивов РФ и США.

Ключевые слова: «облачная» обработка естественного языка, семантический анализ, тематическое моделирование, статистический анализ текста, поиск по глобальному техническому уровню.

Система автоматического позиционирования заявочных материалов на получение патента на изобретение в глобальном патентном пространстве на основе статистико-семантических подходов E-patent Examiner (далее – Система) - это система для принятия решения экспертом при экспертизе заявки на получение патента.



А.Г. Кравец

Intellogist, PriorSmart, MaxVal, BizInt SmartCharts, Espacenet, AmberScope, Acclaim IP, Innography, IFI Claims, PatentInspiration.

Однако, все вышеперечисленные продукты осуществляют поиск релевантных заявок документов согласно запросу, сформулированному экспертом. Поэтому считать их прямыми аналогами разрабатываемой системы нельзя.

Основными целями создания Системы являются:

– обеспечение автоматического позиционирования заявочных материалов на получение патента на изобретение в глобальном патентном пространстве на основе статистико-семантических подходов;

– оптимизация технологических процессов обработки материалов связанных с проведением экспертизы изобретений на основе результатов автоматического позиционирования заявочных материалов на получение патента на изобретение в глобальном патентном пространстве на основе статистико-семантических подходов.

Задачи, которые необходимо решить для достижения целей:

1. Провести анализ предметной области и разработать структуру глобального патентного пространства.

2. Оценить варианты позиционирования Заявки в глобальном патентном пространстве.

3. На основе сравнительного анализа выявить требования к статистико-семантическим методам и моделям.

4. Разработать архитектуру и требования к системе «E-patent Examiner».



М.А. Дыков

5. Провести эксперименты для оценки эффективности разработанной концепции.

6. Разработать макет системы с применением «облачных» технологий и методов параллельных вычислений.



Д.М. Коробкин



С.С. Укустов

Структура глобального патентного пространства (ГПП) состоит из трёх слоев (рис.1):

1. Фундаментальные исследования и знания.

2. Патенты и другие охраняемые документы, выдаваемые в различных странах. Этот слой пространства разделен на определенные классы с помощью действующих классификаций.

3. Открытые источники, то есть, источники, опубликованные в различных базах научного цитирования, в журналах, а также в сети Интернет.

Для определения вариантов позиционирования заявки на изобретение в ГПП необходимо выделить зону (1, Рис.1), которая зоной принятия решения о выдаче патента при экспертизе новой заявки. Именно в этой зоне появляются три основные пробле-

мы. Первая – это сверхбольшой объем неструктурированной информации, то есть, информации, хранящейся в виде текстов, рисунков в различных источниках, зачастую имеющих совершенно разную структуру. И вторая проблема также информационная – это неполнота информации, то есть отсутствие доступа к тем или иным патентным базам данных, открытым источникам, индексам цитирования, которые требуют дополнительных затрат на подключение, например. И третья проблема – это субъективизм эксперта, именно в этой зоне как раз он наиболее рискован.

Таким образом, поступившая заявка может находиться в 4 вариантах позиционирования относительно текущего уровня техники, представленного ГПП (рис.1).

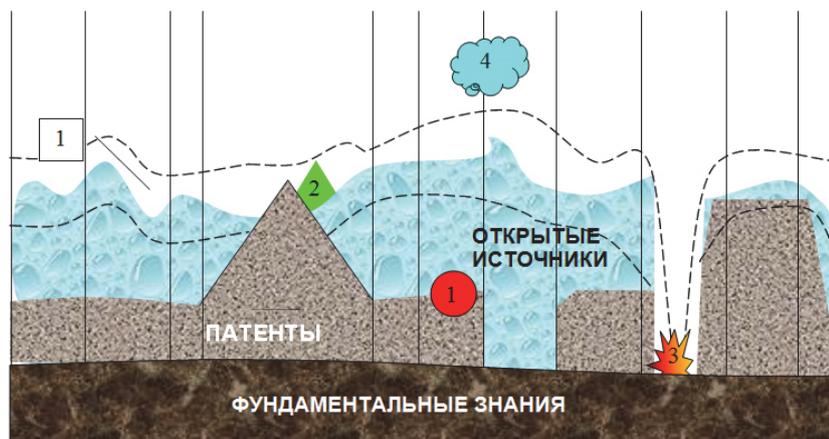


Рис. 1. Позиционирование поступившей заявки в ГПП

При экспертизе новой заявки может быть выявлено, что она соответствует условию промышленной применимости, но, при этом, она полностью покрывается текущим пространством, и, соответственно, она не соответствует условию новизны или изобретательского уровня, в зависимости от того, обнаружены ли все признаки формулы изобретения в одном источнике информации, или в нескольких, точнее не может быть выдан патент на представленные результаты каких-либо исследований и разработок.

Второе состояние заявки – это заявка, которая частично перекрывается патентным пространством и, соответственно, находится именно в зоне принятия решения о выдаче патента, то есть, какая-то часть этой заявки является новой и не покрывается техническим уровнем, существующим множеством патентов и открытых источников в ГПП. Именно такая заявка является патентноспособной, то есть на неё может быть выдан охраненный документ.

Третье состояние заявки – это заявка, которая претендует на пионерский патент и основана на серьёзных фундаментальных исследованиях, но, в данном случае, имеется разрыв в ГПП, который, позволяет предполагать наличие технически осуществимых решений в этой области, но, тем не менее, на данный момент их нет.

Четвёртое положение в патентном пространстве новой заявки – это заявка, которая находится в отрыве от зоны принятия решения, то есть от допустимой зоны технического уровня. Фактически, она, конечно, может иметь какие-то семантические связи с текущими открытыми источниками и даже патентами, уже не говоря о фундаментальных знаниях, но, кроме семантических связей, существует её достаточно сильная удалённость от текущего технического уровня. В связи с этим, происходит ослабление этой связи, она сильно удалена, что позволяет говорить о неприменимости предложенного решения и отсутствии возможности выдачи каких-либо охраненных документов.

Для реализации статистико-семантических методов и моделей (ССМ) были разработаны алгоритмы: первичной обработки ГПП для формирования статистико-семантических профилей существующих патентов (рис. 2); обработки поступившей заявки для построения её статистико-семантического профиля (рис.3).

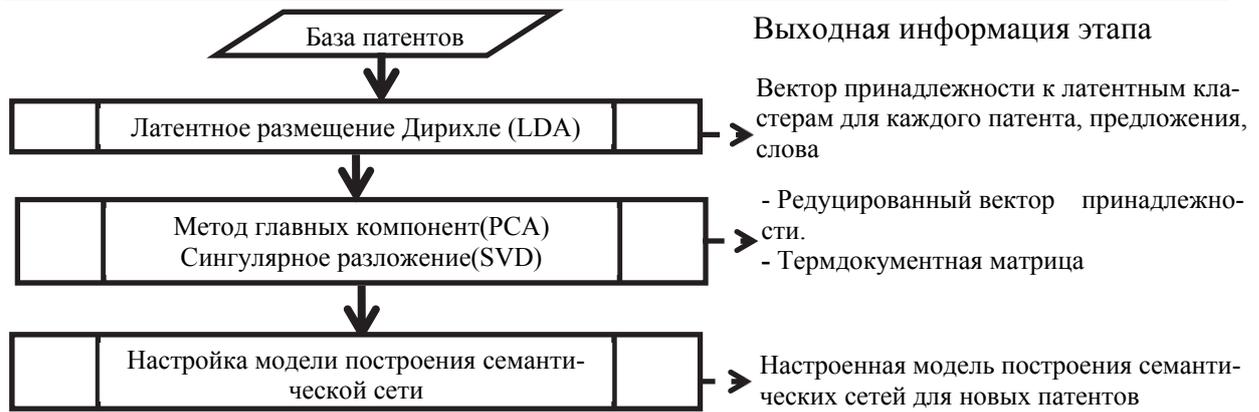


Рис. 2. Алгоритм обработки существующей базы патентов

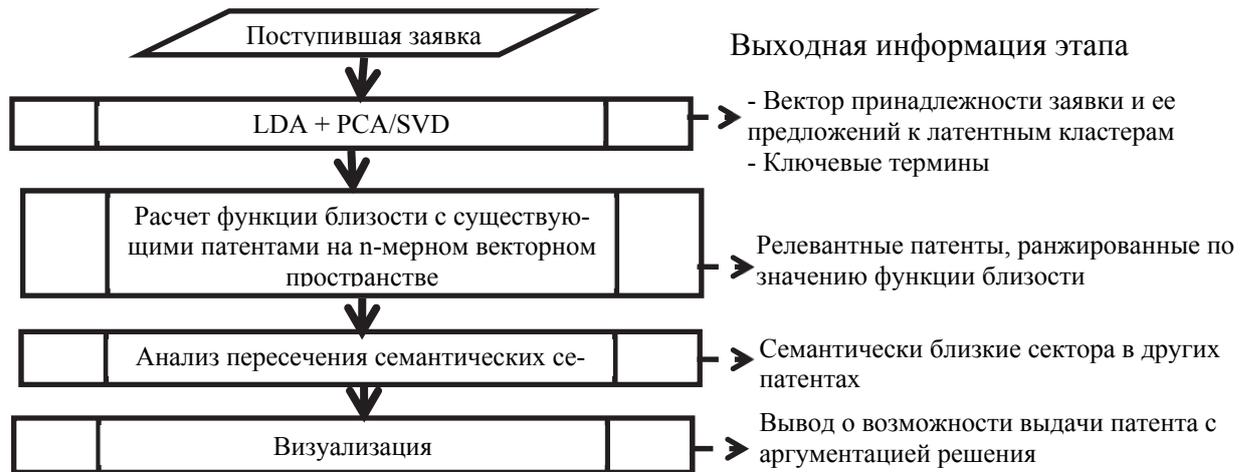


Рис. 3. Алгоритм обработки поступившей заявки

Рассмотрим фрагмент трёхслойной структуры ГПП, который представляет собой многомерный массив данных, с разбиением согласно действующей классификации (МПК) (рис.4). Ядро ГПП представляет собой сложную сетевую структуру, также многомерную и, кроме этого, разделённую, вне зависимости от классификации, на тематические зоны. Именно такой подход позволит реализовать экспертизу новых заявок и прогнозирование возможных новых технических решений в смежных областях, вне зависимости от классификаций. Это многомерное пространство может быть расширено, например, за счёт вектора географической привязки патентов и охранных документов, календарной привязки, то есть, к годам выдачи патента, и прочей аналитической информации.

Таким образом, ГПП представляет собой многомерную сеть, узлом которой является портрет патента, представляющий из себя совокупность статистического профиля документа, семантического профиля документа и метаданных источника, то есть, метаданные того текста документа, который хранится в той или иной базе данных:

узел = статистический профиль ⊕ семантический профиль ⊕ метаданные источника

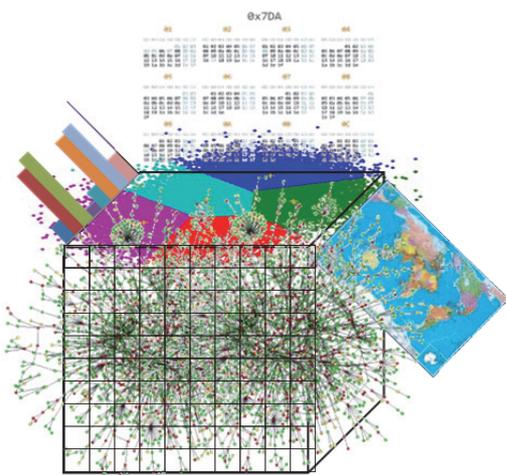


Рис. 4. Многомерная структура ГПП

Связями между отдельными профилями являются статистическая мера близости и доля пересечения семантических сетей разных документов:

связь = статистическая мера близости \oplus доля пересечения семантических сетей

Чем больше связей существует у каждого узла, тем более значимым он является в целом для сети.

Архитектура распределённой базы знаний системы «E-patent Examiner» представлена на рис. 5.

Предлагаемый подход позволяет избежать переформатирования сверхбольших объёмов баз данных патентов и позволяет получить доступ к этим патентам для экспертизы, а также поиска новых технических решений в тех областях, где патентов либо нет, либо они находятся на пионерском уровне, например, и т.д.

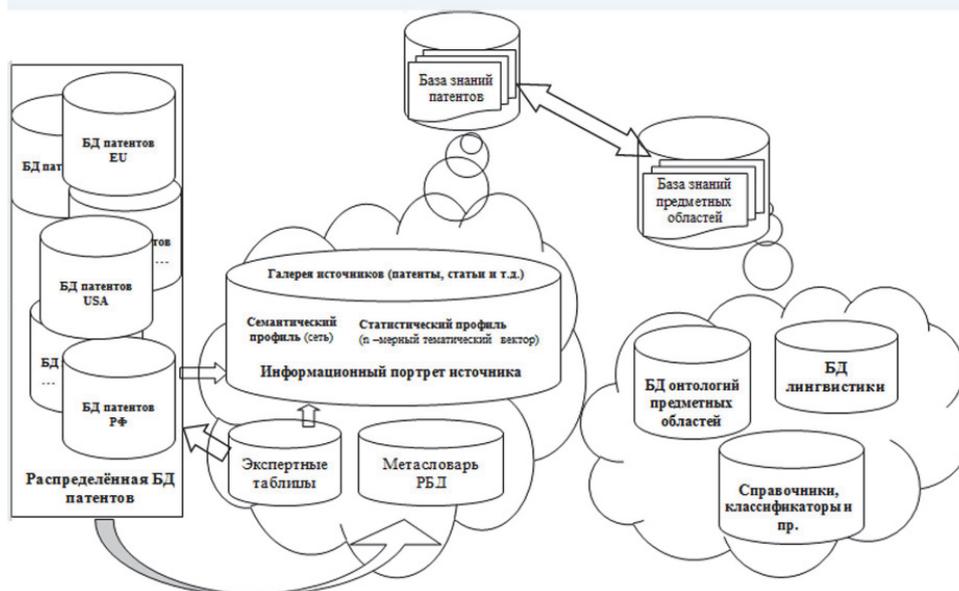


Рис.5. Архитектура распределенной базы знаний

То есть, эти пробелы в патентном пространстве можно заполнять с помощью анализа базы знаний, которые представлены в виде информационных портретов источников.

Авторы считают, что в данной работе получены новые положения и результаты, которые могут найти применение в следующих областях:

1. Поиск новых научно-технических решений: технология и автоматизированная система, в качестве поисково-аналитического инструментария.
2. Экспертиза заявочных материалов на получение патента на изобретение: технология и автоматизированная система, в качестве системы поддержки принятия решений эксперта.
3. Подготовка заявочных материалов на получение патента на изобретение: автоматизированная система, в качестве инструментария для поиска аналогов и анализа текста Заявки.

Литература

1. М.А. Дыков, А.Г. Кравец, Д.М. Коробкин, С.С. Укустов, М.Ю. Сальников. Представление документа в виде вектора ключевых фраз для решения задачи поиска по уровню техники в описаниях патентов // Управление и высокие технологии: прикаспийский журнал. 2014. № 1: С. 148-154.
2. Дыков М.А., Кравец А.Г., Коробкин Д.М., Укустов С.С. Автоматизированная система принятия решений при патентной экспертизе // Известия ВолгГТУ. 2014. Выпуск 16. № 8. С. 85-91.
3. Dykov M.A., Kravets A.G., Korobkin D.M., Ukustos S.S, Strelkov O.I. Combining statistical and semantic features for patents prior-art retrieval. Materials of the international scientific practical conference «Innovative information technologies». 2013. P. 2. pp. 120-129.

Automatic positioning of application for patent in global patent space

Kravets A.G., Doctor of Technical Sciences, Professor of SAPR and Personal Computer chair

Dmitriy Mikhaylovich Korobkin, Candidate of Technical Sciences, Associate Professor of SAPR and Personal Computer chair

Dykov D.A., graduate student of SAPR and personal computer chair

Ukustov A.S., graduate student of SAPR and personal computer chair
Volgograd State Technical University

The authors describe the result of the pilot project «E-patent Examiner» within the frame of the system prototype which was created and it was hosted in the «cloud» of the server cluster. For the experiments in the prototype knowledge base a gallery of patents from subject areas such as «electricity», «physics», «mechanics» of the Russian Federation and the United States patent arrays was loaded.

Keywords: Natural Language «Cloud» Processing, Semantic analysis, Topic modeling, statistic analysis, global prior-art search.

УДК 303.447

О ЗАВИСИМОСТИ РАСПРЕДЕЛЕНИЯ ГОЛОСОВ ОТ ЯВКИ

*Андрей Юрьевич Бузин, канд. физ.-мат. наук, канд. юрид. наук,
доц. кафедры нелинейного анализа и оптимизации факультета
физико-математических и естественных наук
Тел.: +7(916) 115-75-15, e-mail: abuzin@votas.ru
Университет дружбы народов РУДН
<http://www.rudn.ru>*

Зависимость распределения голосов от явки является предметом острых дискуссий среди исследователей электоральной статистики. В статье сравниваются некоторые показатели такой зависимости. Особое внимание уделено показателю Собянина-Суховольского, применяемому во многих исследованиях для оценки уровня фальсификаций. На примере конкретных выборов проиллюстрировано, что сам этот показатель варьируется в зависимости от выборки избирательных комиссий.

Ключевые слова: выборы, явка, критерий Собянина-Суховольского elections, turnout, Sobyenin-Suhovolskiy criterium

Суть обсуждаемой проблемы

Изучение качества голосования и подсчета голосов часто заключается в исследовании распределения на декартовой плоскости с координатами (τ, u) точек, соответствующих участковым избирательным комиссиям (УИК), где τ – значение явки в УИК, а u – доля голосов (от общего числа избирателей УИК), полученных определенным претендентом в этой УИК [Собянин, 1995; Муягков, 2009; Кобак, 2012; Klimek, 2012; Бузин, 2008]. Так, например, А. Собянин и В. Суховольский, а вслед за ними и многие другие исследователи, изучают наклон a регрессионной прямой, проведенный через эти точки. Если наклон a существенно отличается от доли голосов (от числа проголосовавших избирателей) v , набранных претендентом, то обычно делается вывод о фальси-