



Экспериментальная система поиска и анализа патентной информации Exactus Patent

Илья Тихомиров
к.т.н., доцент
зав. лаб. 0-7

+7 (499) 135 04 63

117312, Москва,
пр-т 60-летия
Октября, 9





Что такое Exactus Patent?

Это исследовательский проект, направленный на изучение возможностей применения современных методов компьютерной лингвистики для решения задач патентного поиска и анализа

Exactus Patent построен на технологиях Exactus и коммерческом продукте TextAppliance

Все технологии являются полностью российскими

Почему Exactus Patent?

- В патентных поисковых машинах используется классический булев поиск со сложным языком запросов, слабо используются достижения современной компьютерной лингвистики
- До сих пор идут споры:
 - имеет ли смысл считать релевантность (как это делает Google)
 - искать по документу-образцу
 - использовать функцию поиска текстовых заимствований
 - и тд.
- Имеются существенные проблемы с поиском и анализом по российскому патентному массиву и поддержкой русского языка

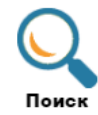


Отличительные особенности Exactus Patent

- Полный лингвистический анализ текстов патентов, включая морфологический и синтактико-семантический анализ
- Использование современных подходов и методов поиска и обработки больших массивов текстов
- Учёт релевантности при поиске и анализе
- Комплексное решение задач патентного поиска и анализа, расширенные аналитические функции
- Поддержка русского и английского языка
- Дружественный web-интерфейс, который не требует специальных знаний при поиске и анализе

Пример 1:

Семантический поиск текстовых
заимствований в российском патентном
массиве



Поиск



Аналитика

автоматизированная система идентификации доступных ресурсов

Поиск

[Дополнительные настройки](#)

Результаты поиска

Топ найденных документов:

1 2 3 4 5 6 7 8 9 10 > >>

1. АВТОМАТИЗИРОВАННАЯ СИСТЕМА ИДЕНТИФИКАЦИИ ДОСТУПНЫХ РЕСУРСОВ СЕТИ ИНТЕРНЕТ

(54) АВТОМАТИЗИРОВАННАЯ СИСТЕМА ИДЕНТИФИКАЦИИ ДОСТУПНЫХ РЕСУРСОВ СЕТИ ИНТЕРНЕТ. <...> Автоматизированная система идентификации доступных ресурсов сети Интернет, содержащая модуль селекции опорных адресов сайтов в базе данных сервера, информационный и синхронизирующий входы которого являются первыми информационным и синхронизирующим входами системы соответственно, при этом первый информационный вход системы предназначен для приема запросов электронных документов по их сетевым адресам, первый синхронизирующий вход системы предназначен для приема <...>

Авторы: Романов Анатолий Николаевич (RU)

Год публикации: 2008

Патентообладатель: Федеральное государственное унитарное предприятие "Научно-исследовательский институт "Восход" (RU)

Классы МПК: G06F12/00 (2006.01) G06F15/00 (2006.01)

http://www1.fips.ru/fips_servl/fips_servlet?DB=RUPM&DocNumber=76482&TypeFile=html

Заемствования и Похожие Резюме Ключевые слова





Заимствования и похожие



Патентный поиск и анализ^β

Поиск по запросу **Поиск по тексту**

Исходный документ:

АВТОМАТИЗИРОВАННАЯ СИСТЕМА ИДЕНТИФИКАЦИИ ДОСТУПНЫХ РЕСУРСОВ СЕТИ ИНТЕРНЕТ

Авторы: Романов Анатолий Николаевич (RU)

Год публикации: 2008

Патентообладатель: Федеральное государственное унитарное предприятие "Научно-исследовательский институт "Восход" (RU)

http://www1.fips.ru/fips_servl/fips_servlet?DB=RUPM&DocNumber=76482&TypeFile=html



Похожие



Заимствования



Аналитика

Поиск заимствований

Год публикации: 2008

Оценка оригинальности документа - 6.3%

Источники заимствования

Заимствования

1. СИСТЕМА ДИНАМИЧЕСКОЙ КОНТЕНТНОЙ ФИЛЬТРАЦИИ ЭЛЕКТРОННЫХ ДОКУМЕНТОВ

Авторы: Осипов Геннадий Семенович (RU), Тихомиров Илья Александрович (RU), Соченков Илья Владимирович (RU)

Патентообладатель: УЧРЕЖДЕНИЕ РОССИЙСКОЙ АКАДЕМИИ НАУК ИНСТИТУТ СИСТЕМНОГО АНАЛИЗА РАН (RU)

Год публикации: 2008

Тип публикации: ФИПС. Полезные модели

http://www1.fips.ru/fips_servl/fips_servlet?DB=RUPM&DocNumber=77465&TypeFile=html

93.7%

- Заимствованные фрагменты
- Резюме
- Ключевые слова



Заимствования и похожие

Проверяемый текст

. Известна система Microsoft Internet Security and Acceleration (ISA) Server [1], осуществляющая фильтрацию документов, передаваемых по высокоуровневым протоколам (в частности, HTTP).

. В качестве ресурсов сети Интернет выступают так называемые электронные документы, представляющие собой информационные объекты, доступные в сети передачи данных, которые могут быть получены путем информационного сетевого обмена оконечного оборудования пользователя с сервером по протоколу HTTP и далее воспроизведены с помощью оконечного оборудования пользователя в печатном или ином виде..

. Техническим результатом является повышение быстродействия системы путем локализации адресов поиска записей списков доступа базы данных сервера по идентификаторам электронных документов..

. Параллельно с этим, синхронизирующий импульс с входа 12 системы задерживается элементом 31 модуля 1 на время срабатывания регистра 25 и дешифратора 26, и далее опрашивает состояние элементов 28-30 И..

. Тогда $siteID=CRC32(http://www.site.com)$, $pathID=CRC32(foo/bar/doc.html)$,.. и идентификатор электронного документа - есть пара $(siteID, pathID)$, а идентификатор соответственного ресурса, к которому принадлежит документ - $(siteID, 0)$..

. КодКод Код признака доступасiteID pathID- либо разрешения доступа к электронному документу с данным адресом; -либо запрещения доступа к электронному документу с данным адресом Код pathID из первой считанной записи из базы данных с первого выхода регистра 45 поступает на один вход компаратора 46, на другой информационный вход которого с выхода 39 модуля 1 поступает код pathID запрашиваемого электронного документа..

Текст источника

Известна система Microsoft Internet Security and Acceleration (ISA) Server [1], осуществляющая фильтрацию документов, передаваемых по высокоуровневым протоколам (в частности, HTTP). <...>

Под электронным документом в сети Интернет подразумевается информационный объект, доступный в сети передачи данных, который может быть получен путем информационного сетевого обмена оконечного оборудования пользователя с сервером по протоколу HTTP и далее воспроизведен с помощью оконечного оборудования пользователя в печатном или ином виде.. <...>

Таким образом, введение новых модулей и новых конструктивных связей позволило существенно повысить быстродействие системы путем локализации адресов записей списка доступа базы данных сервера по идентификаторам электронных документов.. <...>

Параллельно с этим, синхронизирующий импульс с входа 12 системы задерживается элементом 31 модуля 1 на время срабатывания регистра 25 и дешифратора 26 и далее опрашивает состояние элементов 28-30 И.. <...>

Тогда $siteID=CRC32(http://www.site.com)$, $pathID=CRC32(foo/bar/doc.html)$,.. и идентификатор электронного документа - есть пара $(siteID, pathID)$, а идентификатор соответственного ресурса, к которому принадлежит документ - $(siteID, 0)$.. <...>

Код pathID из первой считанной записи из базы данных с первого выхода регистра 45 поступает на один вход компаратора 46, на другой информационный вход которого с выхода 39 модуля 1 поступает код pathID запрашиваемого электронного документа.. <...>

Пример 2:

анализ российского патентного массива с
использованием Exactus Patent и
Thomson Innovation по теме
«электронная книга»

Exactus Patent:

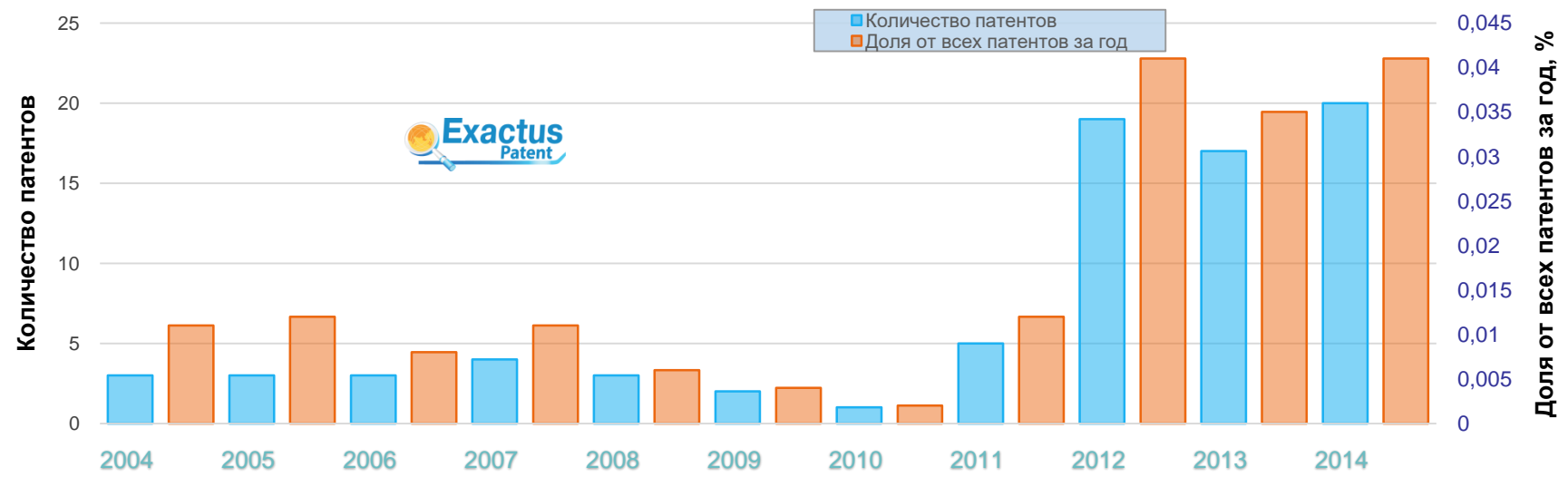
{электронная книга}

Thomson Innovation:

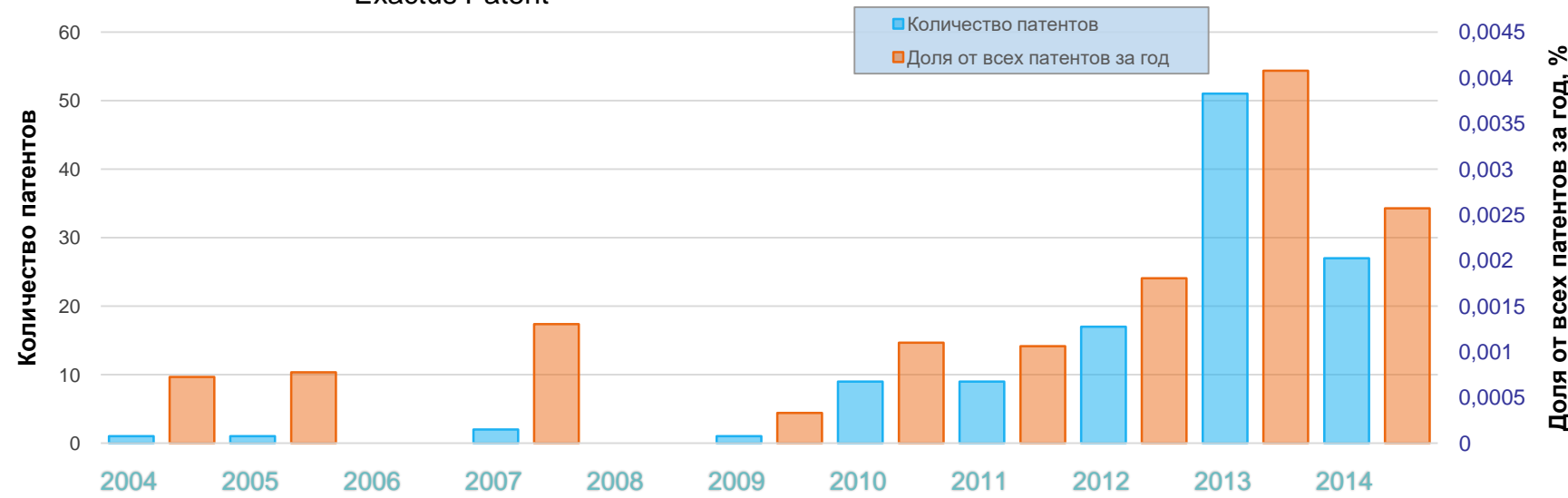
CC=RU AND CTB=(e-reader* OR eBook* OR e-book* OR eReader* OR Electron* ADJ (book* OR reader*)) AND (PY>=(2004) AND PY<=(2014)).



Динамика патентной активности

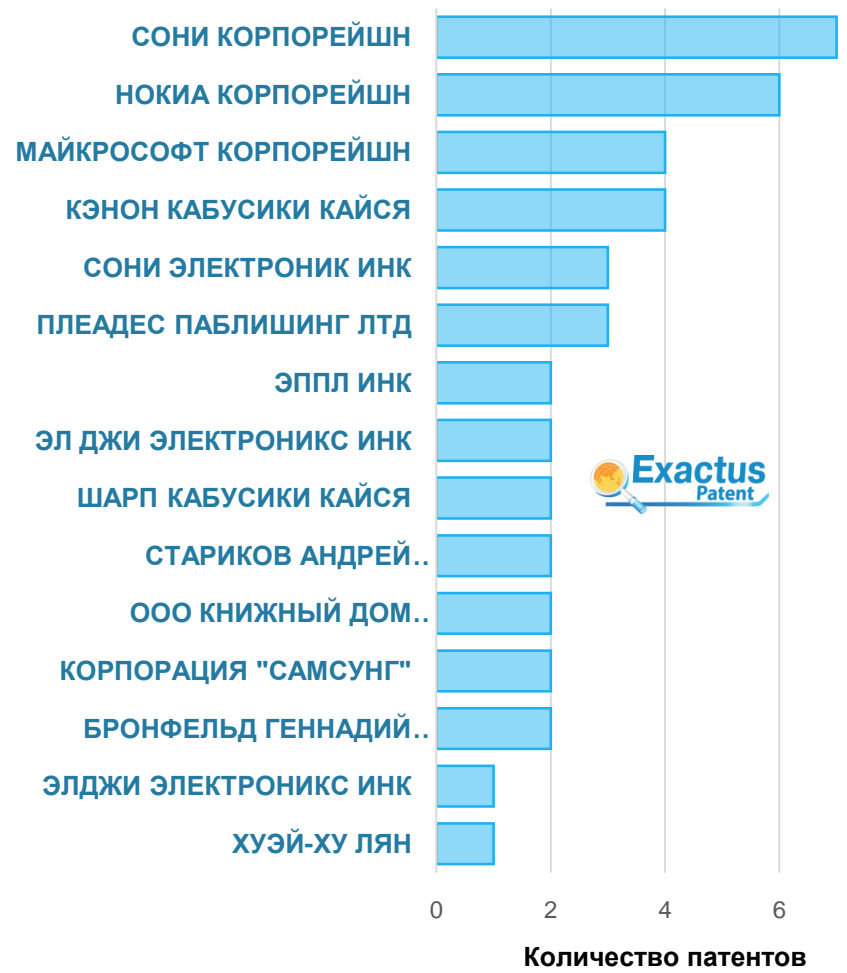


Exactus Patent

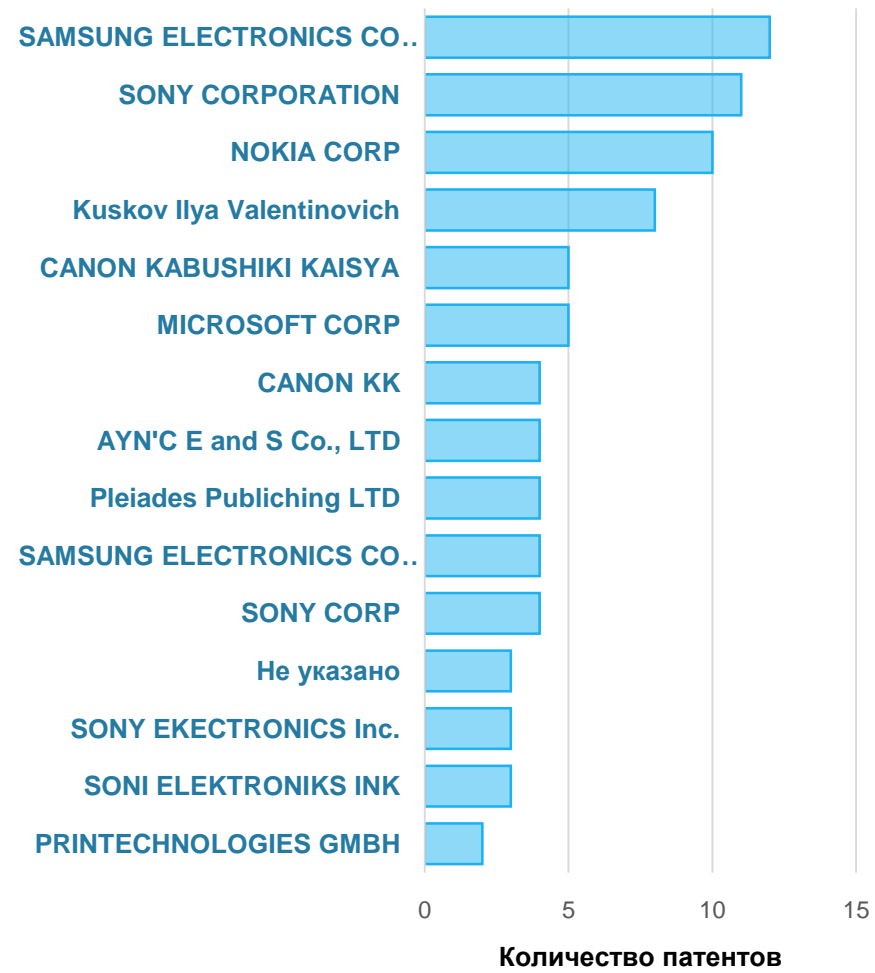


Thomson Innovation

Патентообладатели

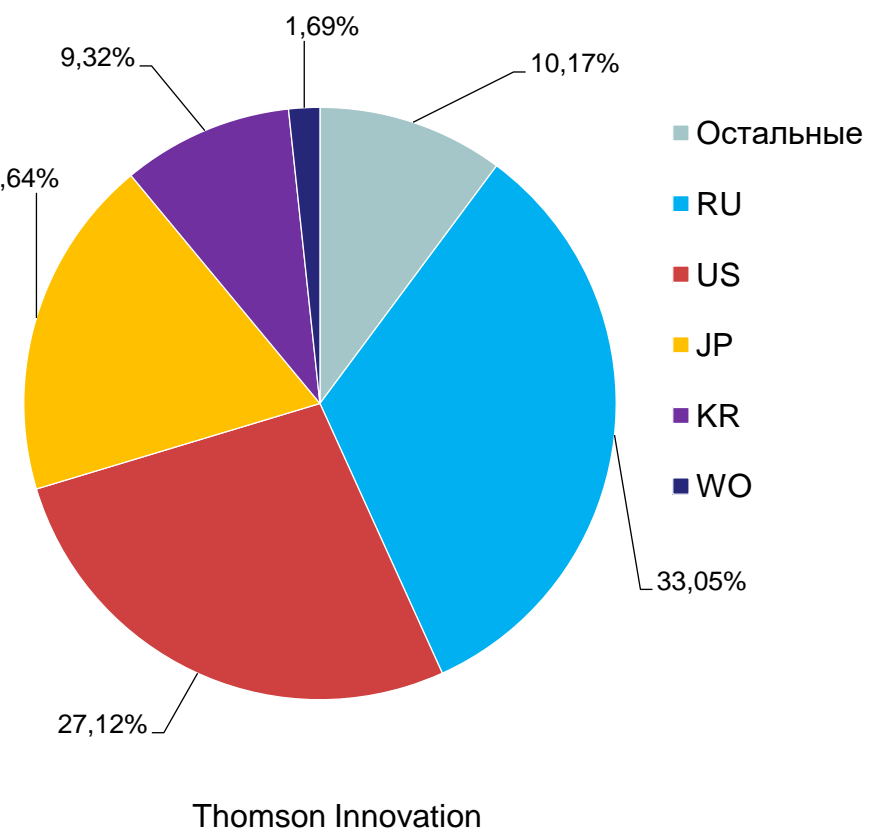
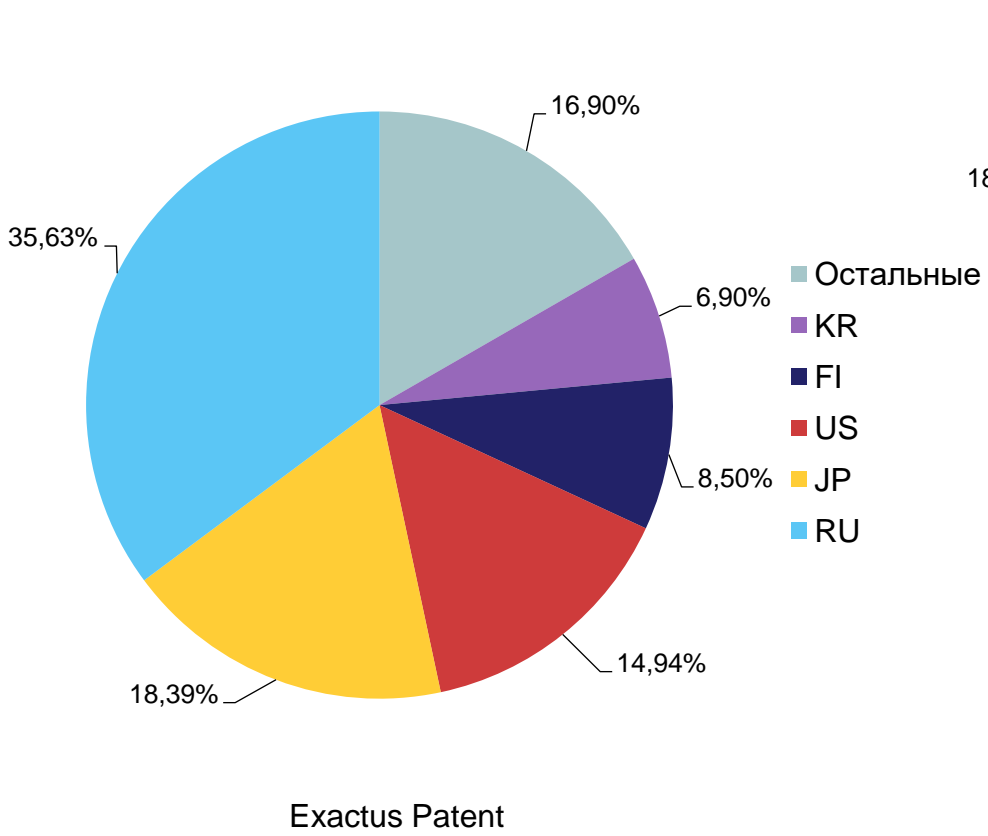


Exactus Patent



Thomson Innovation

Государственная принадлежность патентообладателей

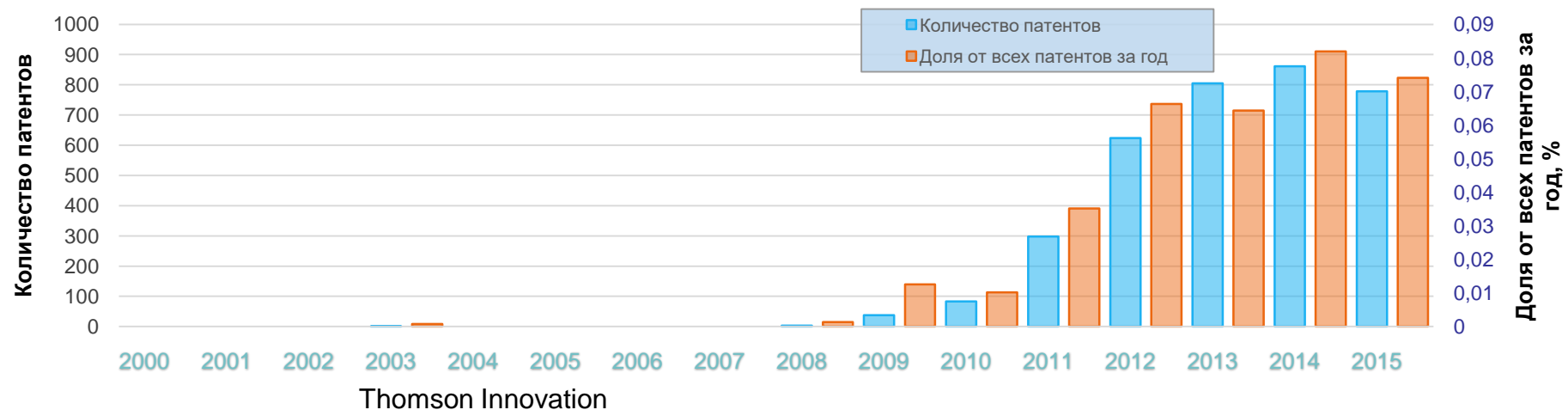
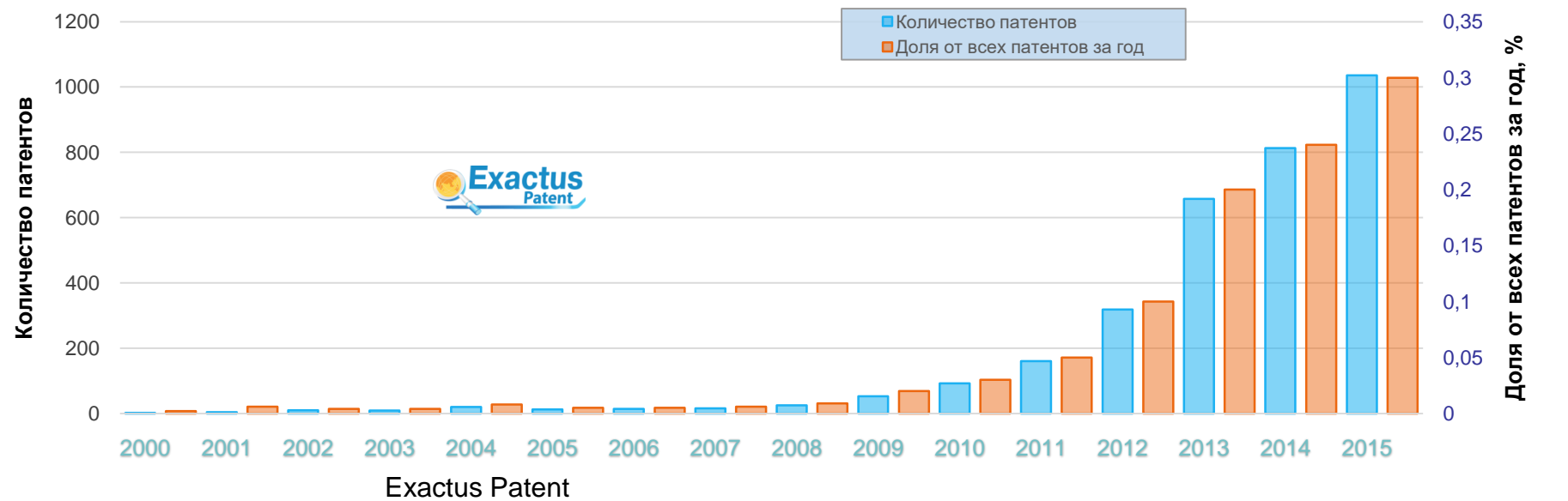


Пример 3:

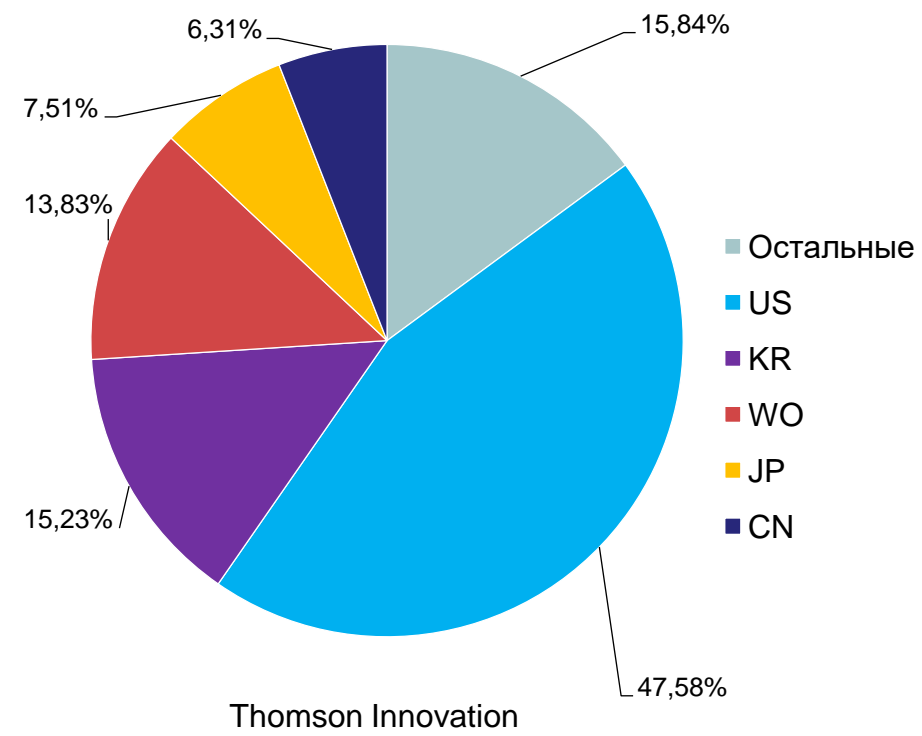
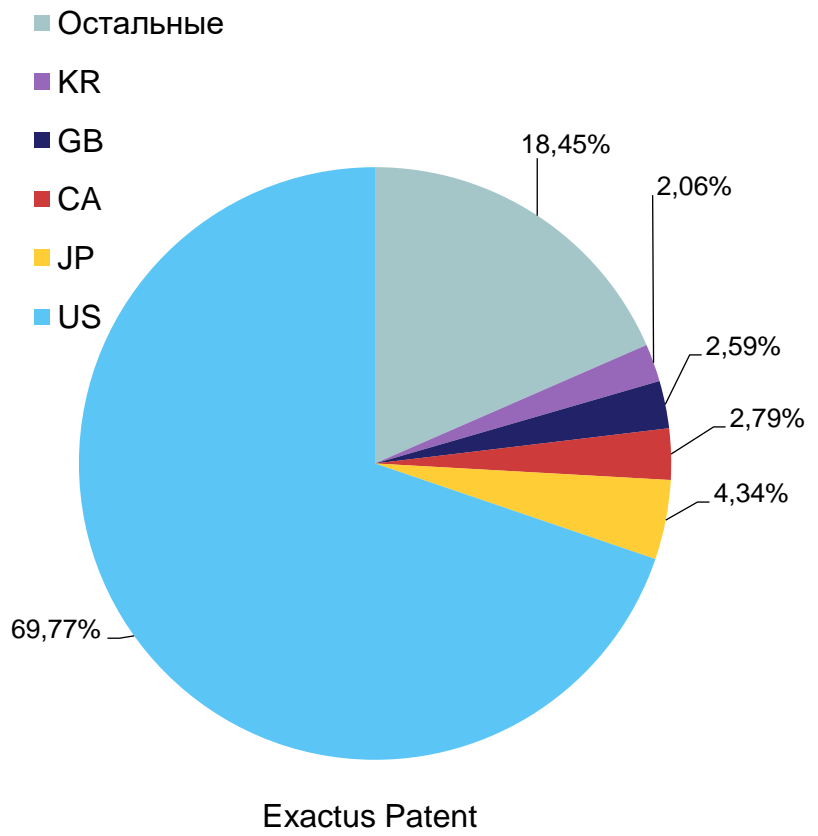
анализ массивов USPTO и WIPO с
использованием Exactus Patent и
Thomson Innovation по теме
«Умные энергосети» (Smart grid)



Динамика патентной активности



Государственная принадлежность патентообладателей

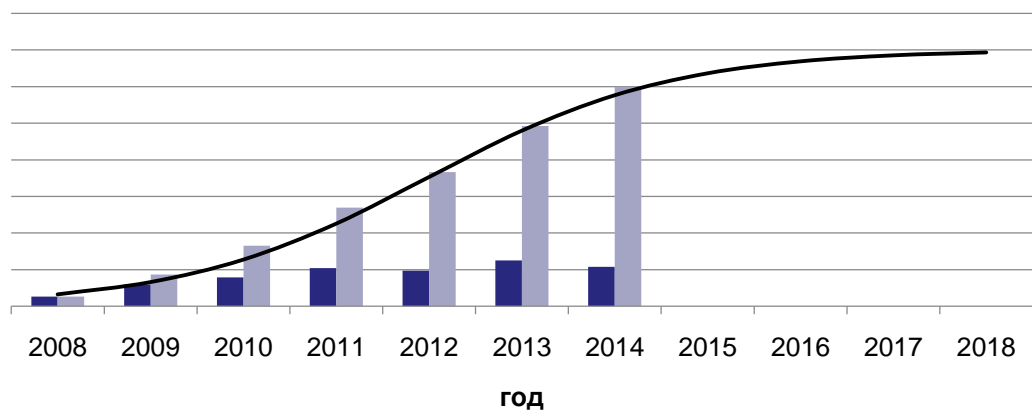


Пример 4:

Развивающиеся и стагнирующие
направления в российском патентном
массиве

Методика оценки направлений

- Для выявления стагнирующих/растущих направлений используется аппроксимация кумулятивной патентной массы логистической кривой



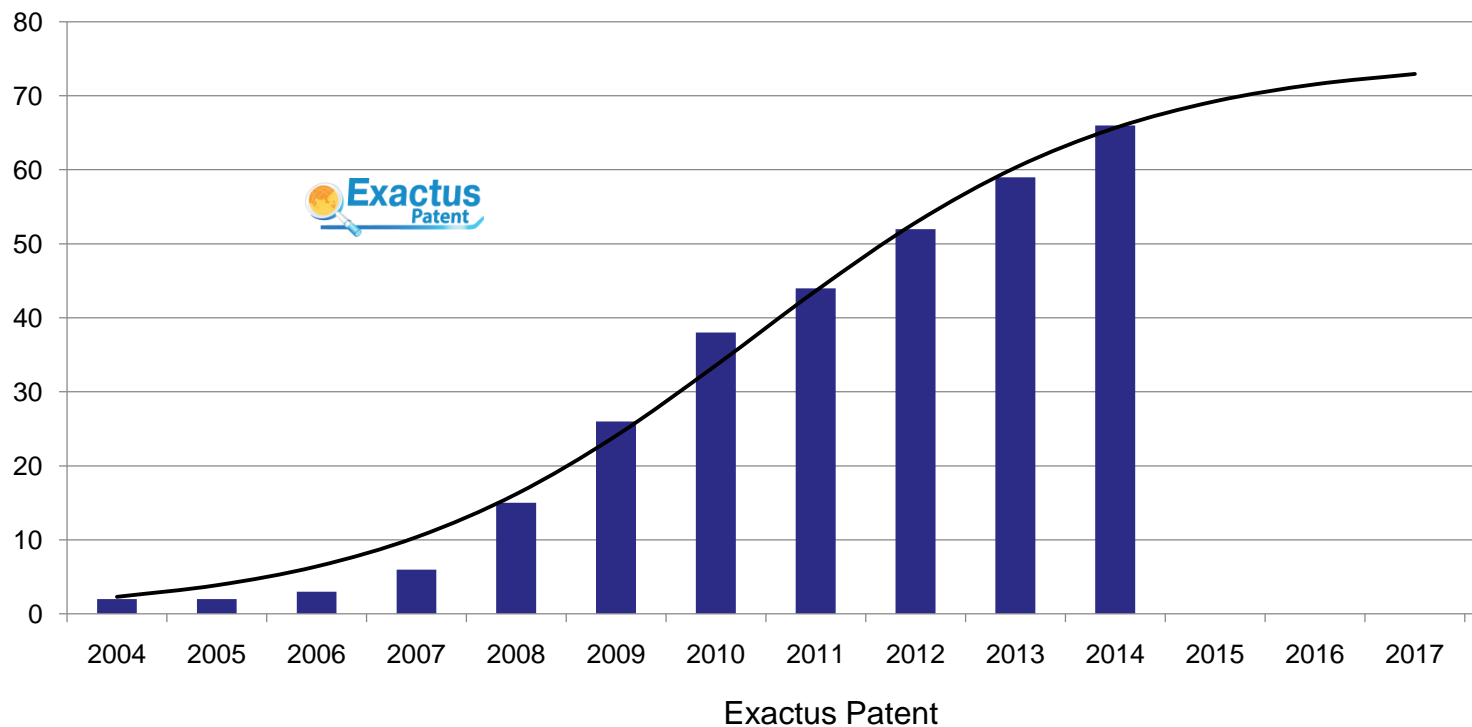
Morrison, Jeffrey S. "Life-cycle approach to new product forecasting." *The Journal of Business Forecasting* 14.2 (1995): 3.

- Выход объема патентной массы на плато логистической кривой говорит о стагнации направления

Пример стагнирующего направления

- Структурная диагностика наноматериалов (G01N + B82B)

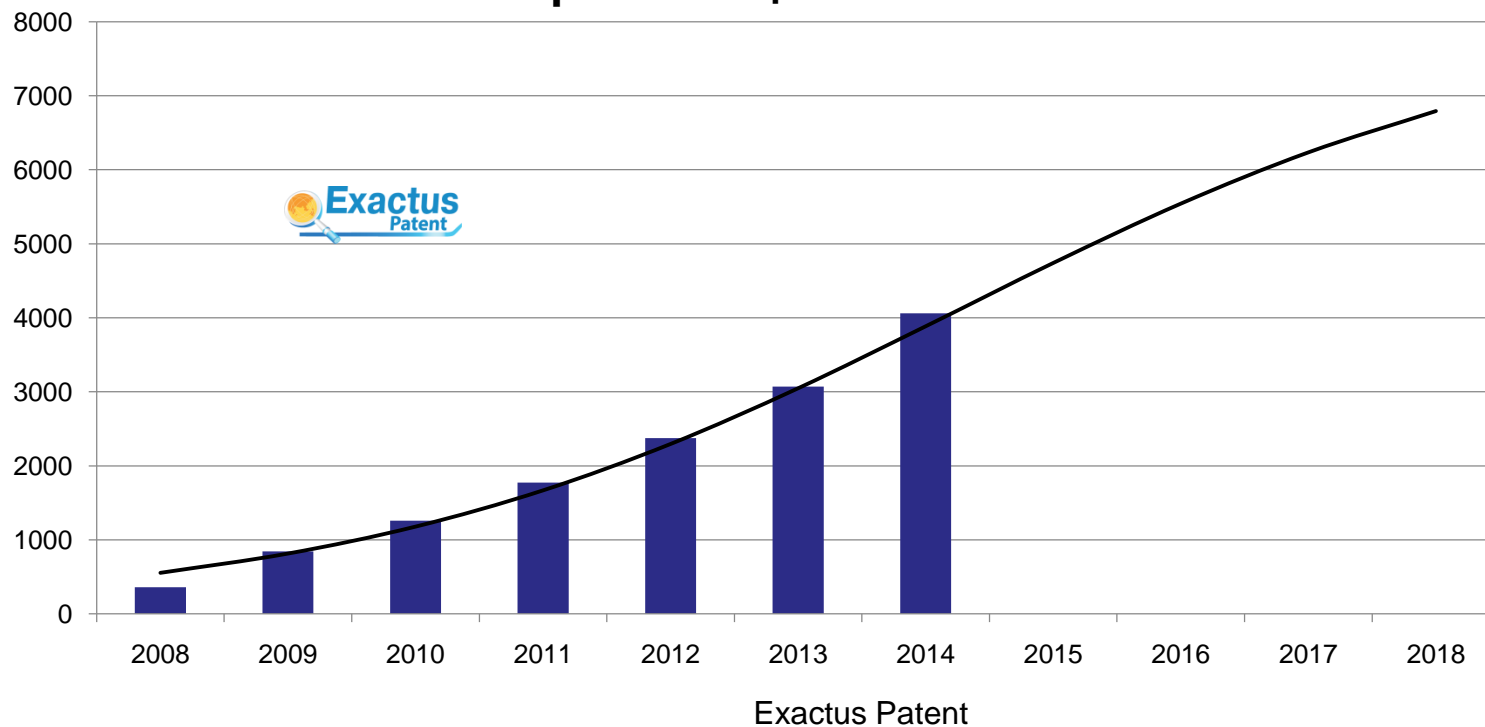
Кумулятивное количество патентов и его аппроксимация



Пример растущего направления

- Машины, двигатели вообще (F01).

Кумулятивное количество патентов и его аппроксимация



Находится в стадии бета-тестирования и практически готов к использованию

- Кластер под управлением Debian GNU/Linux
- Распределенные вычисления и балансировка нагрузки
- Проиндексированные базы:
 - ФИПС
 - WIPO (2000-2016)
 - USPTO (2002-2016)



Заключение

Exactus Patent – российская разработка для патентного поиска и анализа информации, которая обеспечивает:

- качественную поддержку русского языка
- широкий набор поисковых и аналитических функций
- результаты, которые как минимум не хуже аналогичных платных зарубежных разработок

patent.exactus.ru

Илья Тихомиров
ФИЦ ИУ РАН

117312, Москва, пр-т. 60-летия Октября, 9

Телефон/факс: +7 (499) 135-04-63

e-mail: tih@isa.ru