

ИНТЕГРАЦИЯ ЛИНГВИСТИЧЕСКИХ И СТАТИСТИЧЕСКИХ МЕТОДОВ ПОИСКА В ПОИСКОВОЙ МАШИНЕ «EXACTUS»

к.т.н. Тихомиров Илья Александрович

14-я международная конференция
Диалог-2008

Состояние дел в области поисковых алгоритмов

- Статистические алгоритмы поиска (с учетом морфологии).
- Лингвистические алгоритмы поиска (без учета статистики).

Цель разработчиков Exactus: объединение статистических и лингвистических методов поиска.

Особенности алгоритма поиска Exactus (1)

Учет статистических характеристик текста:

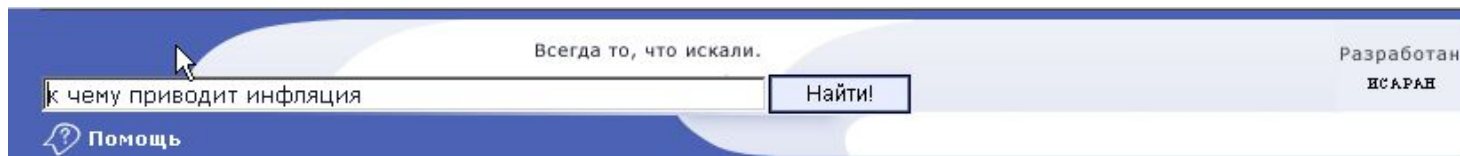
- TF*IDF веса термов (с некоторой модификацией).
- Значимость фрагментов текстов (заголовки, разметка, удаленность от начала документа и т.д.).

Особенности алгоритма поиска Exactus (2)

Учет лингвистических характеристик текста:

- Значения синтаксем (Золотова Г.А.).
- Семантические связи (Осипов Г.С.).





Особенности алгоритма поиска Exactus (3)



Вы искали: "к чему приводит инфляция". Найдено документов: 59. - Затраченное время: 2.68656 с.

Страницы: [1], 2, 3, 4, 5, 6.

- 1. Газета.Ru | Жуков и Кудрин считают, что рост пособий не приведет к инфляции**
...Галопирующая инфляция: вызывает беспокойство у домохозяйств и фирм; приводит к стремлению избавиться от денег покупкой. ...Галопирующая инфляция инфляция, нарастающая скачкообразными темпами. ...газета.ru | Жуков и Кудрин считают, что рост пособий не приведет к инфляции.
http://www.gazeta.ru/money/2006/05/10_n_635416.shtml
Найдено с помощью **Яндекс**
Релевантность: 1.000000
- 2. Словарь - Галопирующая инфляция - инфляция , ...**
...На первый взгляд, инфляция приводит к тому перераспределению дохода, когда вследствие роста цен увеличивается прибыль. ...Таким образом, рост спроса приводит к росту общего уровня цен и, следовательно, к развитию инфляции.
...Словарь Галопирующая инфляция инфляция,..
<http://www.finam.ru/dictionary/wordf00E64/default.asp?n=1>
Найдено с помощью **Яндекс**
Релевантность: 0.700000
- 3. Инфляция**
...К каким последствиям приводит инфляция? ...Инфляция. ...К социально экономическим последствиям инфляции относятся: . Снижение реальных доходов населения и, как следствие, .
<http://www.bishelp.ru/ekonomika/makro/inflyacia.php>
Найдено с помощью **Google**
Релевантность: 0.700000
- 4. www.365.az - Строительство непроизводственных мегаобъектов ...**
...Строительство непроизводственных мегаобъектов приводит к росту инфляции. и других материалов приводит к появлению на рынке новых партий личности.
<http://www.365.az/content/view/9754/28/>
Найдено с помощью **Google**
Релевантность: 0.630000

-  1. [... Изъятие из оборота мелких монет не **приведет к инфляции**, считает Кудрин](#)
Нет, это абсолютно не вызовет никого увеличения **инфляции**", - сказал он журналистам, отвечая на вопрос, **приведет ли к инфляции** изъятие копеечной и ...
www.rian.ru/economy/20061121/55860420.html · 68 КБ
[Сохраненная копия](#) · [Еще с сайта](#) 4592 · Рубрика: [Информационные агентства](#)
-  2. [Riconomics " Почему трата Стабфонда **приведет к инфляции**?](#)
Почему же увеличение денежной массы в экономике - а в нашем случае к такому увеличению как раз **приводит** трата денег из Стабфонда, - **приводит к** ...
riconomics.com/2006/04/24/pochemu-trata-stabfonda-privedet-k-inflyatsii · 677 байт
[Сохраненная копия](#) · [Еще с сайта](#) 205 · Рубрика: [Экономика, менеджмент](#)
-  3. [Снижение НДС **приведет к** росту **инфляции** в России / Бухгалтерские новости / Все ...](#)
Главная / Все новости / Бухгалтерские новости / Снижение НДС **приведет к** росту **инфляции** в России
www.klerk.ru/news/?20805 · 41 КБ
[Сохраненная копия](#) · [Еще с сайта](#) 1363 · Рубрика: [Аудит, бухгалтерский учет](#)
-  4. [Lenta.ru: Экономика: Дополнительные траты на пенсионеров **приведут к инфляции**](#)
... правительство, чтобы успокоить общество после отмены льгот, могут **привести** в 2005 году **к инфляции** до 16 процентов (при запланированных 10 процентах) ...
www.lenta.ru/economy/2005/01/19/inflation/ · 21 КБ
[Сохраненная копия](#) · [Еще с сайта](#) 2773 · Рубрика: [Бизнес](#)

Что у других? (2)

[Инфляция - Бизнес-план](#)

Люди заранее предвидят **инфляцию** и уже повышают цены, что снова **приводит к инфляции**. Возможны следующие механизмы развития гиперинфляции: ...
www.businessproekt.ru/content/document_r_78BC6A12-4296-4A11-AA6E-6B162392D940.html - 42k - [Сохранено в кэше](#) - [Похожие страницы](#)

[Инфляция — Википедия](#)

Кроме того, на уровень **инфляции** существенное влияние оказывает объём денежной ... при стабильном уровне денежной массы **приводит** к росту темпов **инфляции**, ...
ru.wikipedia.org/wiki/Инфляция - 59k - [Сохранено в кэше](#) - [Похожие страницы](#)

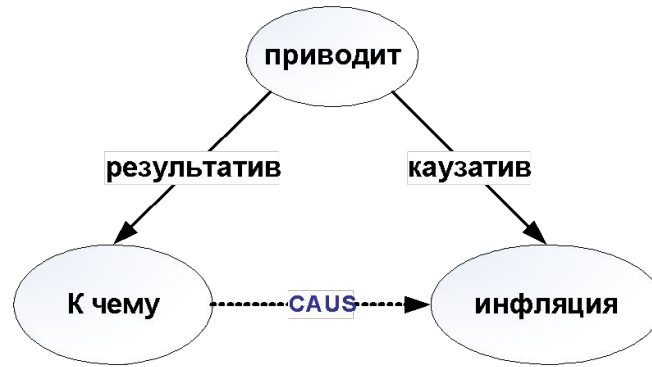
[ВОЙНА и МИР: К чему приводит "выдавливание" инфляции монетарными ...](#)

Ее предвыборной платформой было «выдавливание **инфляции** из британской экономики». Но Тэтчер и ее близкое окружение из современных Адамов Смитов – идеологов ...
www.warandpeace.ru/ru/exclusive/view/18619/ - 36k -
[Сохранено в кэше](#) - [Похожие страницы](#)

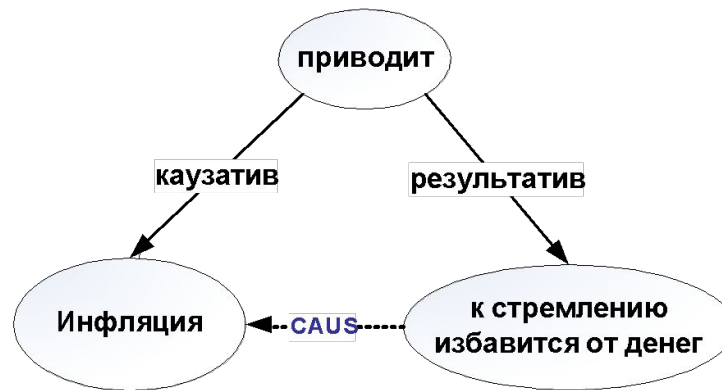
[ForUm :: Украинская инфляция приводит европейцев в ужас](#)

Оперативные новости Украины. Все о политической и экономической жизни страны.
www.for-ua.com/ukraine/2008/05/17/160142.html - 42k -
[Сохранено в кэше](#) - [Похожие страницы](#)

Как Exactus нашел ответ



Семантическая сеть запроса



Семантическая сеть предложения документа

Как Exactus нашел ответ(2)

1. На этапе индексации производится преобразование документов к внутреннему формату Exactus, обсчет TF*IDF весов.
2. Производится синтаксический и семантический анализ текстов (выявление синтаксем и их значений).
3. Полученные в результате анализа данные укладываются в линейные упорядоченные списки вхождений слов в документы с весами и значениями
4. Поиск представляет собой слияние линейных упорядоченных списков.

Особенности архитектуры Exactus

- Модули расположены на узлах кластерной установки.
- Управление задачами осуществляется посредством PVM-машины (Parallel Virtual Machine).
- Модули разделены на два типа: основные (лингвистические процессоры, индексаторы и т.д.) и вспомогательные (агрегаторы, синхронизаторы и т.д.).
- Система является кросс-платформенной, код написан на C и C++.
- Экспериментальная установка состоит из 8-и узлов кластера пиковой производительностью 100 Gigafllops.
- В качестве вычислительных узлов используются обычные персональные компьютеры, объединенные в стойку.
- Для взаимодействия узлов используется Gigabit Ethernet.

Exactus – вид сбоку

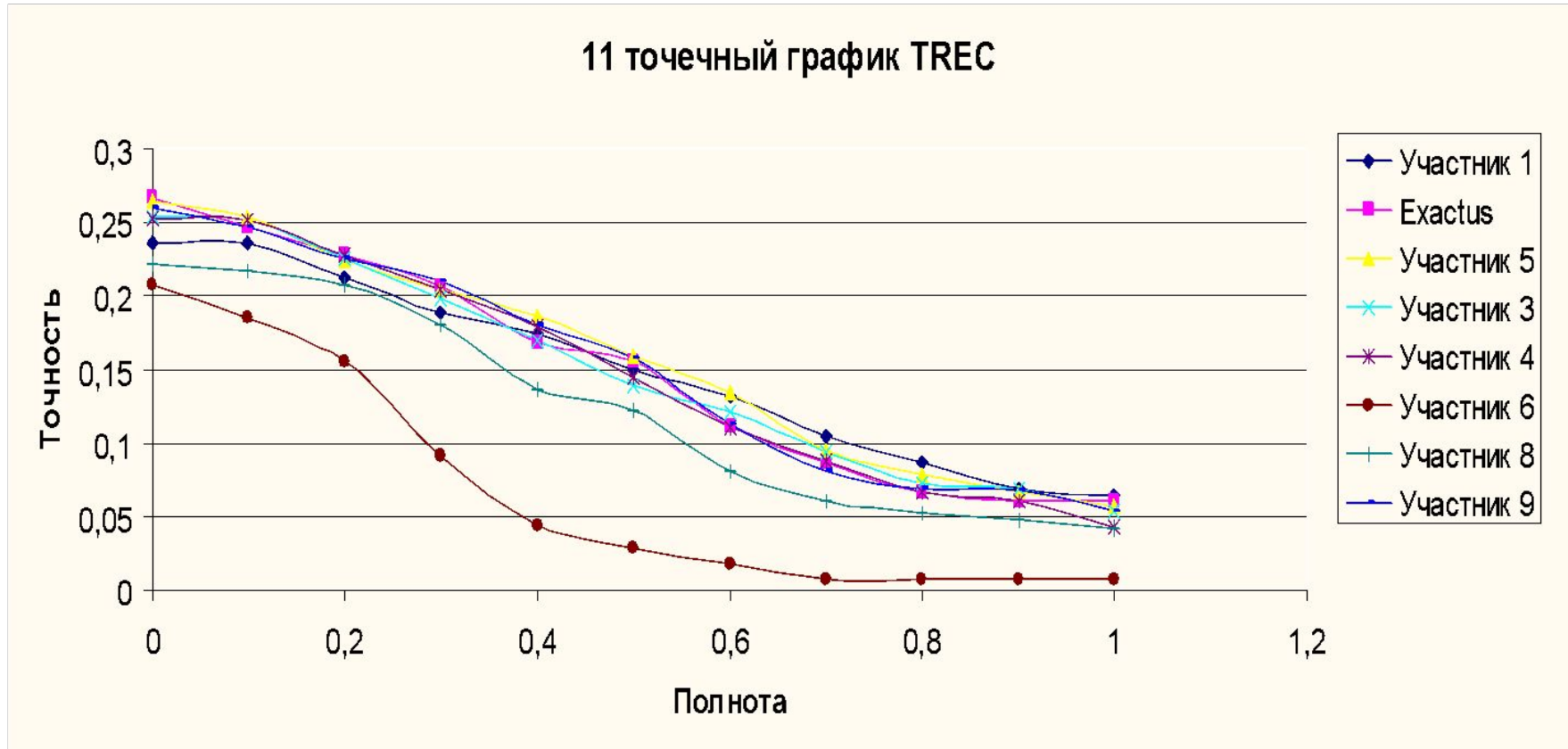


Экспериментальная проверка алгоритма Exactus (1)

- Принципы оценки результатов поиска
 - Эксперт оценивает соответствие документов исходному запросу на основе расширенного описания информационной потребности
 - Используются следующие шкалы оценки релевантности:
 - точно релевантно;
 - возможно релевантно;
 - вероятно релевантно;
 - не релевантно;
 - невозможно оценить.
 - Результат считается релевантным, если он получил оценку по одному из первых двух пунктов шкалы

Экспериментальная проверка алгоритма Exactus (2)

- Для выставления оценки результата используются два способа:
 - Строгая оценка AND – документ получает оценку релевантен или нерелевантен, если все оценщики выставили соответствующую оценку.
 - Нестрогая оценка OR - результат получает оценку релевантен, если хотя бы один оценщик выставил соответствующую оценку.



Краткий анализ результатов участия в РОМИП-2007 (1)

- Exactus принимал участие в поиске по коллекции белорусского Интернета.
- Наилучшие результаты достигнуты системой в AND-оценке по точности.
- Хорошие оценки достигнуты по другим показателям.
- Отсутствие Page Rank в алгоритме Exactus не привело к отставанию от других систем в точности и полноте поиска.

Краткий анализ результатов участия в РОМИП-2007 (2)

- Полученные на РОМИП результаты показывают перспективность симбиоза лингвистических и статистических алгоритмов поиска и возможность их применения в реальных условиях.
- Скорость поиска Exactus сравнима с современными поисковыми машинами на больших объемах данных (не более 2х секунд на любой запрос по коллекции РОМИП).
- Индексация и лингвистический анализ, по-прежнему, остаются узким местом Exactus. Единственный путь преодоления барьеров скорости анализа - использование современных вычислительных систем и параллельных вычислений.

СПАСИБО ЗА ВНИМАНИЕ!

Вопросы и замечания принимаются.
СПАСИБО ЗА ВНИМАНИЕ!

Вопросы и замечания принимаются.