

Kontur Labs' Mix

Pavel Braslavski

About myself

2000: PhD thesis on genres and search

2001 – present: Ural Federal University

2005-2012: Yandex

Feb 2012: Kontur labs



IR&NLP: query log mining, automatic summarization & snippet generation, CQA, thesauri

<http://kanas.ru/pb>



SKB Kontur

- Based in Yekaterinburg, Russia
- Founded in 1988
- Web services (SaaS) for businesses
- A core service – online tax filing for businesses
- 1M+ customers in 10 time zones
- 1000+ staff members (200+ developers)
- #5 software developer in Russia

<http://kontur.ru>

Current Projects

- IE from legal / accounting documents
- Questions and answers
- Analysis of the database of Russian companies

2 Pavels

6 interns

INFORMATION EXTRACTION

Task

Documents types: acts, накладные, invoices, agreements, ...

No standard templates

MS Excel (so far), scans to come

Fields of interest:

- Document type, No, Date
- Sum, Taxes, Reason
- Customer, Vendor
- ...

ООО "Росси-С"

ИНН/КПП 8604031207/860401001

628309 РОССИЯ ХМАО Тюменская обл. г. Нефтеюганск мкр. 2 дом 32 офис 215

☎ (3463) 22-74-71, 23-18-39

Счет № С-2543 от 30.11.2011**на оплату по Акту № 2543 от 30.11.2011****Покупатель:** ЗАО "ПФ"СКБ Контур"**ИНН покупателя:** 6663003127/660850001

ия платежного поручения

"Запсибкомбанк" ОАО г.Тюмень

БИК 047130639**Сч.№** 30101810100000000639

Банк получателя

ИНН 8604031207

КПП 860401001

Сч.№ 40702810900080002616

ООО "Росси-С"

Получатель

Основание:

№	Наименование товара, работы, услуги	Цена	Кол-во	Ед. изм.	Сумма
1	Агентское вознаграждение по договору № (86223/10 от 11.01.2010г.)	68,954.28	1	шт.	68,954.28

Акт № К-0002/11/11 от 30 ноября 2011 г.**Исполнитель:** Общество с ограниченной ответственностью "ПКП"**Заказчик:** ООО «Сертум-Про»

№	Наименование работ, услуг	Кол-во	Ед.	Цена	Сумма
1	Агентское вознаграждение по договору 250925/1102	1	шт	102,935.00	102,935.00
				Итого:	102,935.00
				Без налога (НДС)	-

Всего оказано услуг 1 на сумму 102,935.00 руб.

Сто две тысячи девятьсот тридцать пять пять копеек

Вышеперечисленные услуги выполнены полностью и в срок. Заказчик претензий по объему, качеству и срокам оказания услуг не имеет.

18.01.2013

Исполнитель**Заказчик**

Tagging

Mark :: 5 / 4054 :: AllDocuments \ 1005_OOO МЦ Право счет март 2012.xls

Файл Разметка

Пауза Продолжить Пред Применить След Удалить файл Открыть файл (Excel) К неразмеченному NextFormula NextError

Тип Документа	СЧЕТ	✗	N/A
Номер Документа	479	✗	N/A
Дата	31 Марта 2012 г	✗	N/A
Начало периода	Не используется	✗	N/A
Конец периода	Не используется	✗	N/A
Основание		✗	N/A
Имя Покупателя		✗	N/A
ИНН Покупателя		✗	N/A
КПП Покупателя		✗	N/A
Имя Продавца		✗	N/A
ИНН Продавца		✗	N/A
КПП Продавца		✗	N/A
Итоговая Сумма		✗	N/A
Сумма без НДС		✗	N/A
Сумма с НДС		✗	N/A
НДС		✗	N/A

Выделять без значка номера.

ООО "Межрегиональный центр "Право"

Адрес: 354340, г.Сочи, ул. Кирова, 30а, тел.: 41-11-81, 40-04-74

Образец заполнения платежного поручения

ИНН 2317037959

КПП 231701001

Получатель

ООО "Межрегиональный центр "Право"

Сч. № 40702810063090002259

Банк получателя

БИК 040396703

Филиал ОАО "УРАЛСИБ" в г. Сочи г. Сочи

Сч. № 30101810100000000703

ВНИМАНИЕ!!!!!!

С 1 ИЮНЯ 2010 ГОДА ИЗМЕНИЛИСЬ БАНКОВСКИЕ РЕКВИЗИТЫ

СЧЕТ № 479 от 31 Марта 2012 г.

Платательщик: ЗАО "ПФ "СКБ Контур"

Грузополучатель: ЗАО "ПФ "СКБ Контур"

№	Наименование товара	Единица изме- рения	Коли- чество	Цена	Сумма
1	Агентское вознаграждение по договору 231501/10 от 01.10.2010г за месяц март 2012 года	шт	1	324160.84	324160.84
				Итого:	324160.84
				Без налога (НДС).	-
				Всего к оплате:	324160.84

Всего наименований 1, на сумму 324160.84

Триста двадцать четыре тысячи сто шестьдесят рублей 84 копейки

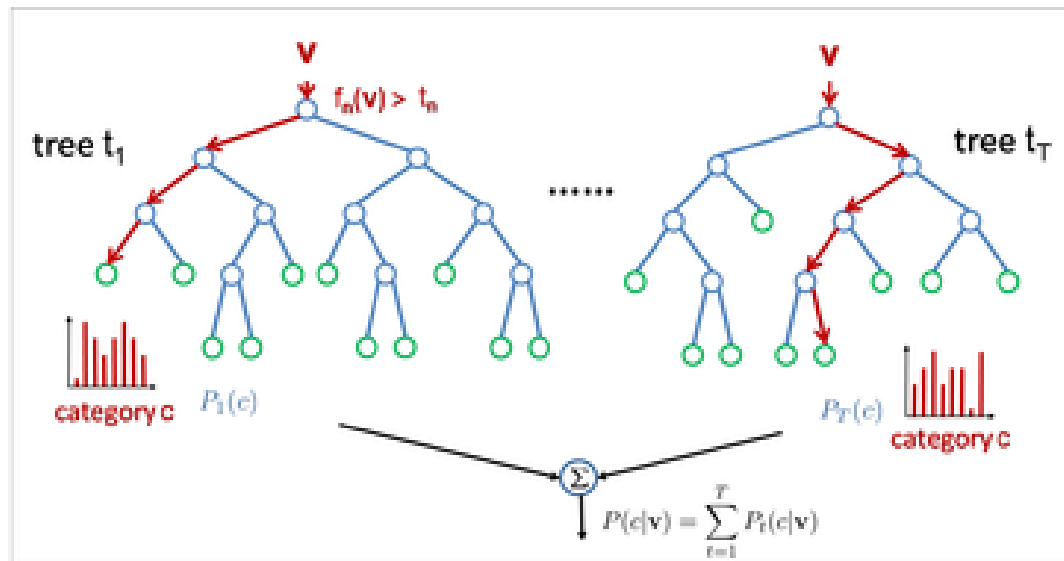
Руководитель предприятия _____ (Ичева Е.В.)

Главный бухгалтер _____ (Ичева Е.В.)

Features

- Location on the page
- Font size
- Bolded
- Distance & direction to keywords
- Location within text block (cell)
- Alignment
- Date – how far from now?

Method



random forest

COMMUNITY QUESTION ANSWERING

Otvety@Mail.Ru

categories

asker's profile

question

answers

answerer's profile

Категории

- [Вопросы друзей](#)
- [Авто, Мото](#)
- [Бизнес, Финансы](#)
- [Города и Страны](#)
- [Гороскопы, Магия, Гадания](#)
- [Досуг, Развлечения](#)
- [Еда, Кулинария](#)
- [Животные, Растения](#)
- [Знакомства, Любовь, Отношения](#)
- [Искусство и Культура](#)
- [Компьютеры, Интернет](#)
- [Красота и Здоровье](#)
- [Наука, Техника, Языки](#)
- [Образование](#)
 - [Домашние задания](#)
 - [ВУЗы, Колледжи](#)
 - [Детские сады](#)
 - [Школы](#)
 - [Дополнительное образование](#)
 - [Образование за рубежом](#)
 - [Прочее образование](#)
- [Общество, Политика, СМИ](#)
- [Путешествия, Туризм](#)
- [Работа, Карьера](#)
- [Семья, Дом, Дети](#)
- [Спорт](#)
- [Стиль, Мода, Звезды](#)
- [Темы для взрослых](#)
- [Товары и Услуги](#)
- [Философия, Непознанное](#)
- [Фотография, Видеосъемка](#)
- [Юридическая консультация](#)
- [Юмор](#)
- [О проектах Mail.Ru](#)

Главная » Образование » ВУЗы, Колледжи » Выбирается лучший ответ

 **Катя Фирсова**  **Гуру** (3243)
На голосовании: 7 месяцев
 Подарить стикер! **NEW**

Что изучает и в чем особенности фундаментальной и прикладной лингвистики?

Оценка: 5 **Рейтинг: 5** (1 голос)

 Ускорить выбор лучшего ответа  Нарушение!

 Поделиться +1  Класс +1  Подписаться

 **Владимир Савчинский**  **Ученик** (205) 7 месяцев назад [\(ссылка\)](#)

 Нарушение!  Подарить стикер! **NEW**

Надо выпить чая))

Оценка: 0 **Рейтинг: 0** [3 комментария](#)

 **Некто в тапках**  **Оракул** (58541) 7 месяцев назад [\(ссылка\)](#)

 Нарушение!  Подарить стикер! **NEW**

Кстати такая дифференциация встречается во многих науках. Например есть прикладная математика распространенные разделы, которые не так часто носят прикладной характер. Фундаментальная лингвистика нацелена на описание и изучения законов языка. Прикладная же лингвистика как инструмент в иных предметных областях. Например есть компьютерная лингвистика построение систем, распознавание речи, извлечение данных.

Источник: Не проф. лингвист. Просто мнение.

Оценка: 5 **Рейтинг: 5** (1 голос)

Data

- \approx 313 000 questions with answers
- *Health and Beauty* Category
- 1 April 2011 – 31 March 2012

Health and Beauty

Красота и Здоровье

Баня, Массаж, Фитнес

Болезни, Лекарства

Врачи, Клиники, Страхование

Детское здоровье

Загар, Солярий

Здоровый образ жизни

Коррекция веса

Косметика, Парфюмерия

Маникюр, Педикюр

Салоны красоты и СПА

Уход за волосами

Прочее о здоровье и красоте

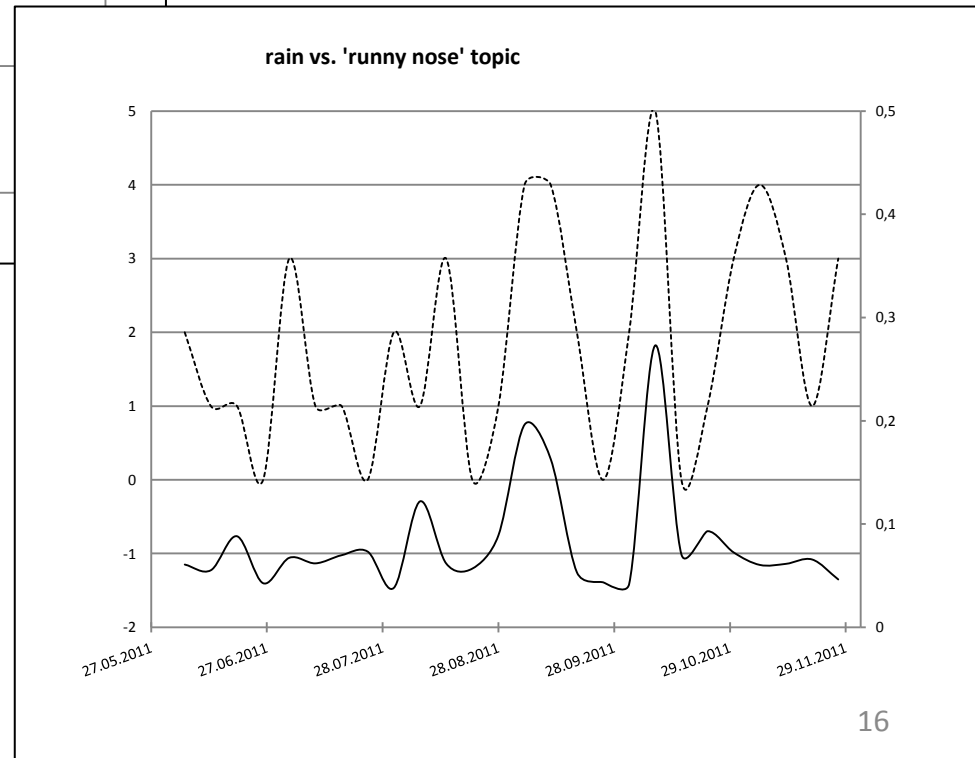
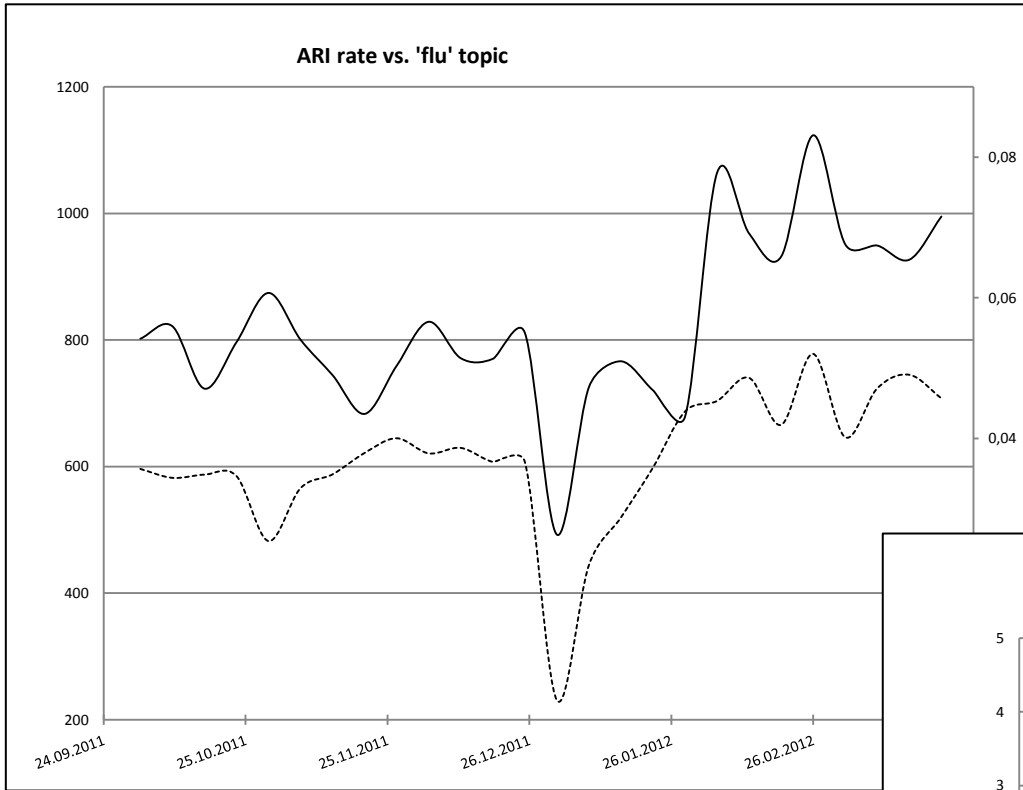
Отвечает врач

~120K questions

Latent Dirichlet Allocation (LDA)

128,000 questions → 70 topics

нервный	глаз	гормон	горло
стресс	зрение	железа	ангина
нерв	очки	эндокринолог	полоскать
попить	линза	гормональный	сода
успокоительный	видеть	организм	раствор
пустырник	капля	щитовидка	вода
успокаивать	окулист	нарушение	соль
система	носить	проблема	полоскание
таблетка	глазной	щитовидный	болеть
глицин	комп	влиять	теплый



Disease – Drug

thrush		angina		herpes	
flucostat [*]	155	iodine	130	aciclovir ^{***}	307
candid ^{**}	92	chamomile	127	zovirax ^{***}	138
clotrimazole ^{**}	89	nitrofuril	111	wax	95
fluconazole [*]	89	lugol	93	fenistil	41
diflucan [*]	77	salvia	70	valtrex	34

CQA: current tasks

- Semiautomatic Content Quality Estimation
- User behavior analysis
- New domain: accounting and legal questions

FOCUS DATABASE



ЗАО "Аргус СФК"

Закрытое акционерное общество "Аргус СФК"


Действующее предприятие

ИНН 6632030272
КПП 663201001
ОГРН 1096632000662
ОКПО 62021449

Дата образования: 4 мая 2009

624975
Свердловская обл
р-н Серовский
пгт Восточный
[ул Заводская, 1](#)  [на карте](#) 
04.05.2009

Тел: [4-47-79](#) 

Генеральный директор
[Борисенко Валерий Васильевич](#) 
20.04.2012

- 20.1 Распиловка и строгание древесины; пропитка древесины
 - 02.01 Лесоводство и лесозаготовки
 - 20.2 Производство шпона, фанеры, плит, панелей
- [Все виды деятельности \(11\)](#)

Учредители

Уставный капитал: 105 010 000 руб.
04.05.2009

100% [ЗАО "Аргус-Холдинг"](#)
105 010 000,00 руб. 07.04.2011

Арбитражные дела


[В качестве ответчика](#) – 10 (2,4 млн. руб.)

[В качестве истца](#) – 38 (13,8 млн. руб.)

Финансовое состояние

 [Бухгалтерские формы](#)

Общедоступные данные из Росстата

 Финансовый анализ за [2010](#), [2011](#)

По методике сервиса «[Эксперт](#)» (PDF)

Государственные контракты


Выигранные: 1,2 млн. руб.

[Подробнее \(4\)](#)

Регистрация во внебюджетных фондах

ПФР 075 025 074719
ФСС 660116827666011
ФОМС 652380900016335

Выписка из ЕГРЮЛ

 [Скачать выписку](#)
от 21 сентября 2012

Запросить свежую

Предполагаемое время
ожидания 4 мин.

- 7M+ organizations
- 20M+ persons (many duplicates)

Relationships:

- Founded-by (co \rightarrow person, co \rightarrow co)
- Affiliated-with (person \rightarrow co)

Very sparse

GCC \sim 1.2M

Main directions

- De-duplicate, denser links
- Graph-based interface
- Data from external sources
 - Wikipedia
 - Stock exchange filings
 - News
 - Court decisions

People de-duplication

- Russian official names:
 - Alexander Sergeevich Pushkin
 - Lev Nikolaevich Tolstoy
 - Fedor Mikhailovich Dostoevskiy
- ~7M person mentions have unique id (INN)
- Task: predict the total number of namesakes for a given full name based on previous observations
- Language Modeling approach + Geo

Найдено 162 (0.23 сек.)

↓ Могут встретиться однофамильцы

1 [ОО "ГГФФ"](#)

Деятельность прочих общественных организаций, не включенных в другие...
 Общественная организация "Геленджикская городская Федерация футбола"
 Учредитель и Председатель: **Смирнов Сергей Иванович** · 230402391535
 353460, Краснодарский край, г Геленджик, ул Свердлова, 8, 81
 ИНН: 2304030603 ОГРН: 1032335035314

2 [ПАРРис НО НЦ ОВД](#)

Деятельность профессиональных союзов
 Первичная профсоюзная организация авиаработников Радиолокации,
 Радионавигации и Связи Новосибирского Центра ОВД филиала
 "ЗапСибазеронавигация"
 Учредитель и Председатель профсоюзного комитета: **Смирнов Сергей Иванович**
 630123, Новосибирская обл, г Новосибирск, ул Аэропорт
 ИНН: 5402456868 ОГРН: 1055400014570

3 [ООО "Э.Д.И.Т."](#)

Производство общестроительных работ по строительству мостов, надземных...
 Учредитель и Директор: **Смирнов Сергей Иванович** · 352501843661
 160000, Вологодская, г Вологда, ул Козленская, 2

ИНН: 3525116515 ОГРН: 1022500074404

Все регионы

[Москва](#)
[Санкт-Петербург](#)
[Свердловская обл](#)
[Еще ▶](#)

Все отрасли

[Торговля](#)
[Строительство](#)
[Услуги](#)
[Производство](#)
[Транспорт](#)
[ИТ и связь](#)
[Финансы](#)
[Гостиницы и рестораны](#)
[Еще ▶](#)

Найдено 9 (0.09 сек.)

↓ Могут встретиться однофамильцы

1 [ООО "СпецРесурс"](#)

Оптовая торговля прочими непродовольственными потребительскими...

Учредитель и Генеральный директор: **Довганич Роман Йозефович**

125252, г Москва, ул Зорге, 9

ИНН: 7714670902 ОГРН: 1067759792759

2 [ООО "Хорекс"](#)

Оптовая торговля прочими непродовольственными потребительскими...

Учредитель и Генеральный директор: **Довганич Роман Йозефович**

111397, г Москва, пр-кт Зеленый, 39, 1

ИНН: 7720569749 ОГРН: 1067760254231

3 [ООО "ТехноСтиль"](#)

Оптовая торговля прочими непродовольственными потребительскими...

Учредитель и Генеральный директор: **Довганич Роман Йозефович**

105066, г Москва, ул Ольховская, 27, стр 6

ИНН: 7708618502 ОГРН: 1067759603889

4 [ООО "ПромЭнерго"](#)

Оптовая торговля прочими непродовольственными потребительскими...

Учредитель и Генеральный директор: **Довганич Роман Йозефович**

105066, г Москва, ул Ольховская, 33

ИНН: 7708620460 ОГРН: 1067760254990

Все регионы

[Москва](#)

[Санкт-Петербург](#)

[Свердловская обл](#)

[Еще ▶](#)

Все отрасли

[Торговля](#)

[Строительство](#)

[Услуги](#)

[Производство](#)

[Транспорт](#)

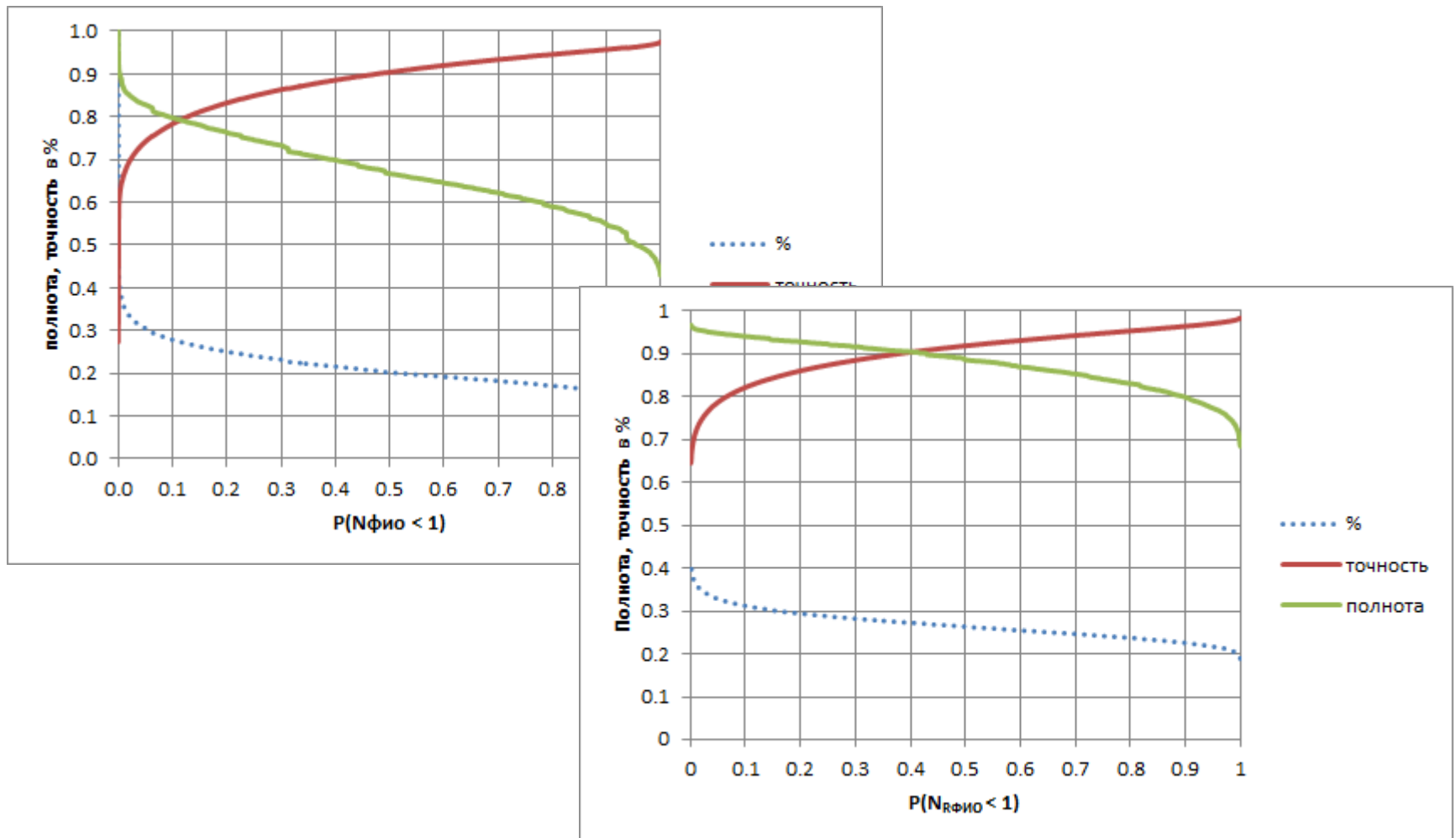
[ИТ и связь](#)

[Финансы](#)

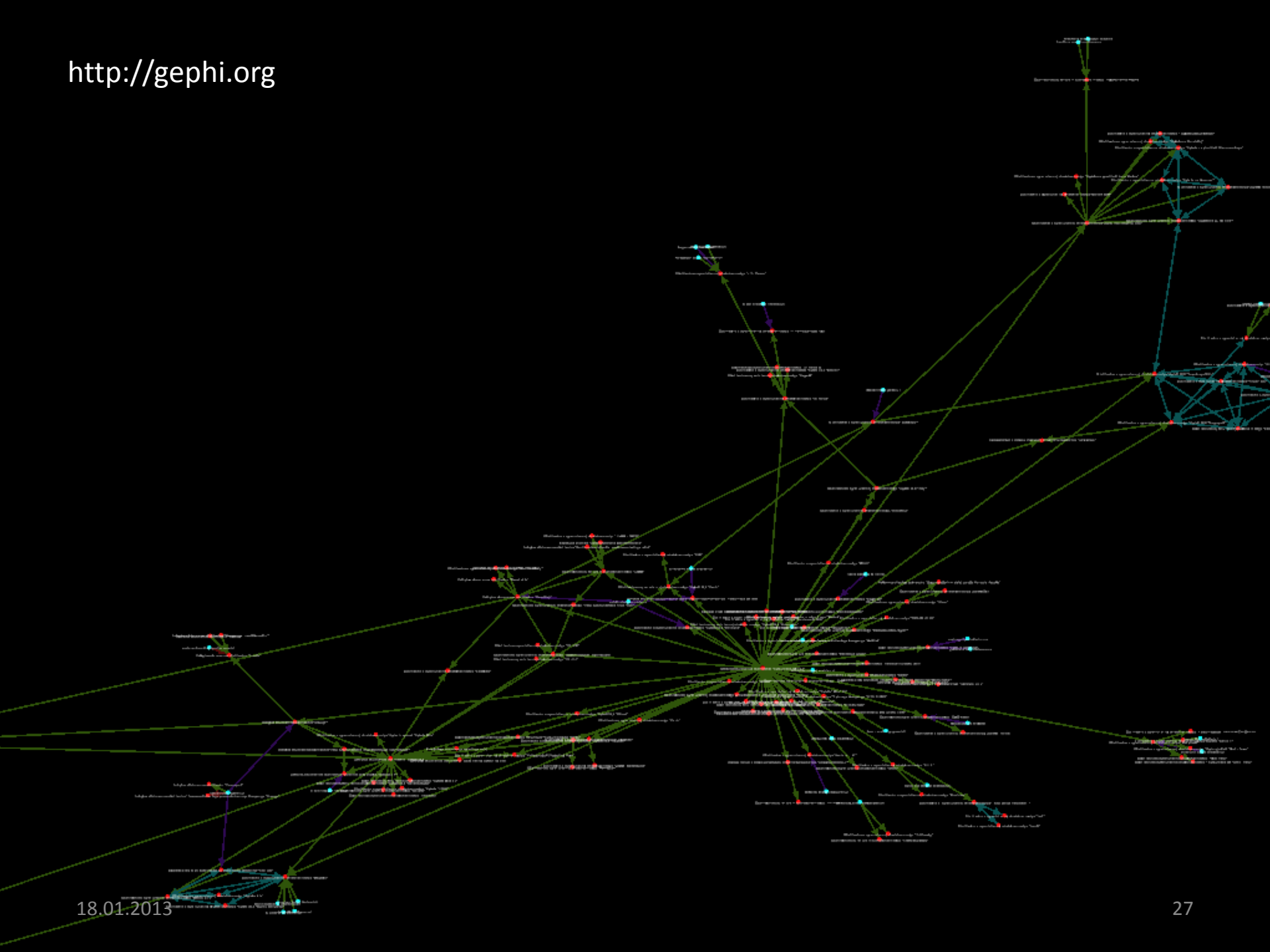
[Гостиницы и рестораны](#)

[Еще ▶](#)

Precision/Recall



<http://gephi.org>



18.01.2013

27

Questions?

pb@kontur.ru



Moscow, 24-27 March 2013

Yandex



<http://ecir2013.org>

7th RuSSIR 2013
Kazan, 16-20 September 2013
<http://russir.org>

