

СТАТИСТИЧЕСКИЙ И ТЕМАТИЧЕСКИЙ ПРОФИЛЬ «ЖИВОГО ЖУРНАЛА»



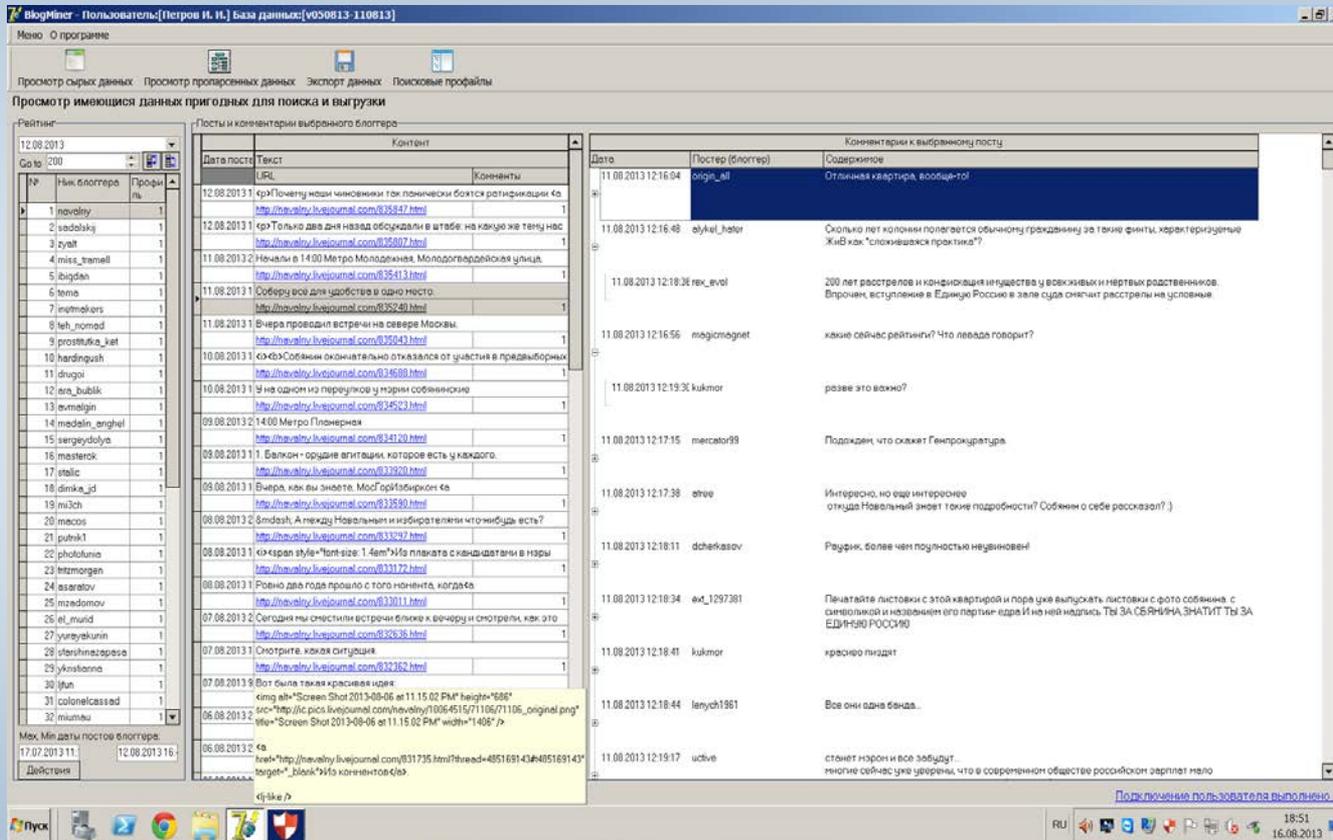
Олеся Кольцова, Сергей Кольцов

Конференция «Интернет и современное общество»
ИТМО, Санкт-Петербург, 9-11 октября 2013

ЗАДАЧИ

- Научиться получать выводы об онлайн-общественном мнении
- Собрать описательную статистику по российской блогосфере для этой и других задач
- Выработать систему основных показателей для описания блогосферы или ее значимой части

СБОР ДАННЫХ



BlogMiner - Пользователь: [Петров И. И.] База данных: [v030813_110813]

Меню О программе

Просмотр сырых данных | Просмотр пропаренных данных | Экспорт данных | Поискковые профили

Просмотр имеющихся данных пригодных для поиска и выгрузки

Рейтинг:

| № | Имя блоггера | Прогресс |
|----|----------------|----------|
| 1 | nevalny | 1 |
| 2 | zadolskiy | 1 |
| 3 | zyvat | 1 |
| 4 | miss_tamell | 1 |
| 5 | bigdan | 1 |
| 6 | tema | 1 |
| 7 | intelmokers | 1 |
| 8 | leh_nompad | 1 |
| 9 | prostotka_net | 1 |
| 10 | hardingush | 1 |
| 11 | drugoi | 1 |
| 12 | ara_bublik | 1 |
| 13 | emselgin | 1 |
| 14 | madalin_angel | 1 |
| 15 | sergeydylova | 1 |
| 16 | maslenc | 1 |
| 17 | italic | 1 |
| 18 | dimka_gd | 1 |
| 19 | mi3ch | 1 |
| 20 | meos | 1 |
| 21 | putnik1 | 1 |
| 22 | phokulino | 1 |
| 23 | frizmorgan | 1 |
| 24 | asaratov | 1 |
| 25 | mzedamov | 1 |
| 26 | el_mind | 1 |
| 27 | yulyayukunin | 1 |
| 28 | stanislavzappa | 1 |
| 29 | yanklanna | 1 |
| 30 | lun | 1 |
| 31 | colonelcassid | 1 |
| 32 | miutaku | 1 |

Мак. Мин. даты постов блоггера: 17.07.2013 11:11:11 | 12.08.2013 15:15:15

Действия

Посты и комментарии выбранного блоггера

| Дата поста | Текст | Комменты |
|--------------|--|----------|
| 12.08.2013 1 | крэ! Почему наши чиновники так панически боятся ратификации €6... http://nevalny.livejournal.com/835847.html | Комменты |
| 12.08.2013 1 | крэ! Только два дня назад обсуждали в штабе: на какую же тему нас... http://nevalny.livejournal.com/835807.html | Комменты |
| 11.08.2013 2 | Начали в 14:00 Метро Молодежная, Молодогвардейская улица. http://nevalny.livejournal.com/835413.html | Комменты |
| 11.08.2013 1 | Соберу все для удобства в одно место. http://nevalny.livejournal.com/835240.html | Комменты |
| 11.08.2013 1 | Вчера проводил встречи на севере Москвы. http://nevalny.livejournal.com/835043.html | Комменты |
| 10.08.2013 1 | с-к-с Собянин окончательно отказался от участия в предвыборных... http://nevalny.livejournal.com/834688.html | Комменты |
| 10.08.2013 1 | Я на одном из переулков у мэрии собрание http://nevalny.livejournal.com/834523.html | Комменты |
| 09.08.2013 2 | 14:00 Метро Пловнерия http://nevalny.livejournal.com/834120.html | Комменты |
| 09.08.2013 1 | 1. Балкон - орудие агитации, которое есть у каждого. http://nevalny.livejournal.com/833920.html | Комменты |
| 09.08.2013 1 | Вчера, как вы знаете, Москва забирок ка http://nevalny.livejournal.com/833580.html | Комменты |
| 08.08.2013 2 | 8mdash. А между Новальными и избирателями что-нибудь есть? http://nevalny.livejournal.com/833297.html | Комменты |
| 08.08.2013 1 | с-к-с-к-с-к-с "Font-size: 1.4em">Ито плавает с каньдатеми в мэри http://nevalny.livejournal.com/833122.html | Комменты |
| 08.08.2013 1 | Ровно два года прошло с того момента, когда http://nevalny.livejournal.com/833011.html | Комменты |
| 07.08.2013 2 | Сегодня мы посетили встречи блогга в вечеру и смотрели, как это http://nevalny.livejournal.com/832838.html | Комменты |
| 07.08.2013 1 | Смотрите, какая ситуация http://nevalny.livejournal.com/832652.html | Комменты |
| 07.08.2013 9 | Вот была такая красивая идея img alt="Screen Shot 2013-08-06 at 11.15.02 PM" data-bbox="380 725 450 745"/> src="http://i.prst.livejournal.com/nevalny/10064515/71106/71106_original.png" data-bbox="380 745 450 765"/> file="Screen Shot 2013-08-06 at 11.15.02 PM" width="1406" /> | Комменты |
| 06.08.2013 2 | ка link="http://nevalny.livejournal.com/831735.html#read=4851691-4�-4" data-bbox="380 765 450 785"/> target="#block">ко комментов:€6. | Комменты |

Комментарии к выбранному посту

| Дата | Постер (блоггер) | Содержимое |
|---------------------|------------------|---|
| 11.08.2013 12:16:04 | olga_all | Отличная квартира, вообще-то! |
| 11.08.2013 12:16:48 | afkai_lesha | Сколько лет колонии полагается обычному гражданину за такие фанти, характеризующие Живёх как "сплошившая практика"? |
| 11.08.2013 12:18:36 | tek_evil | 200 лет расстрелов и кондиция имущества у всех никак и нервных родственников. Впрочем, vstuplenie в Единку России в зале суда снижает расстрел на услове. |
| 11.08.2013 12:16:56 | madamagnet | какие сейчас рейтинги? Что левада говорит? |
| 11.08.2013 12:19:31 | kuklor | разве это важно? |
| 11.08.2013 12:17:15 | mesaio99 | Подождем, что скажет Генпрокуратура |
| 11.08.2013 12:17:38 | eltee | Интересно, но еще интереснее откуда Новальный знает такие подробности? Собянин о себе рассказал?) |
| 11.08.2013 12:18:11 | dchekasov | Резанки, более чем полностью неуважены! |
| 11.08.2013 12:18:34 | ek_1297361 | Печатайте листовки с этой квартирой и поре уже выпускайте листовки с фото собянина, с символикой и названием его партии: едра и на ней надпись: Ты ЗА СБЯНИНА, ЗНАТИТ Ты ЗА ЕДИНКУ РОССИЮ |
| 11.08.2013 12:18:41 | kuklor | красиво падает |
| 11.08.2013 12:18:44 | lenuch1961 | Все они одна банда... |
| 11.08.2013 12:19:17 | isive | станет мэром и все забудут, многие сейчас уже уверены, что в современном обществе россиянин зарплат мало |

Подключение пользователя выполнено.

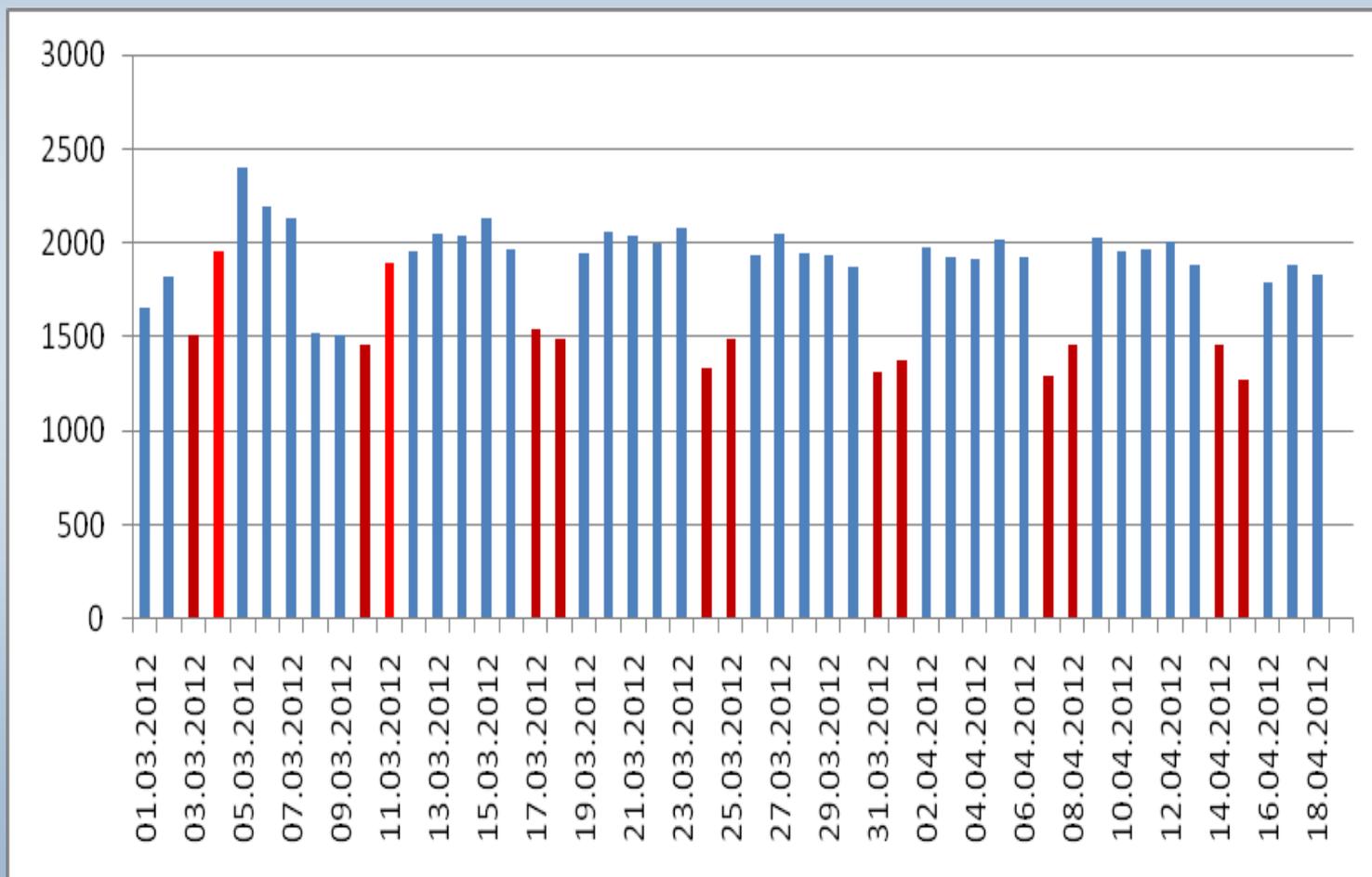
RU 18:51 16.08.2013

Koltran BLOGMINER

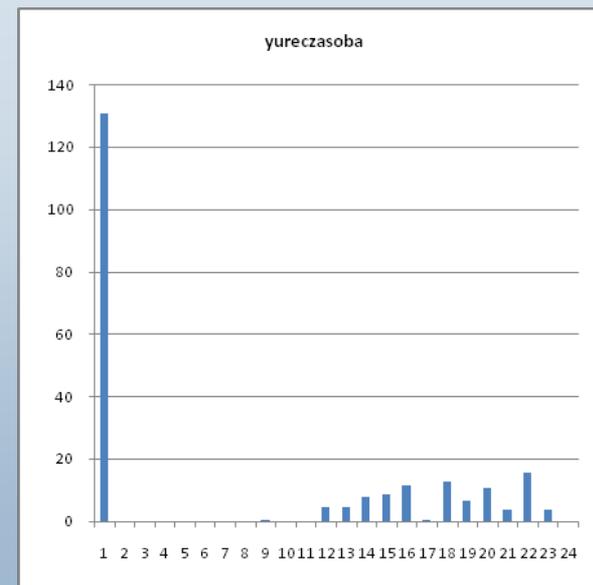
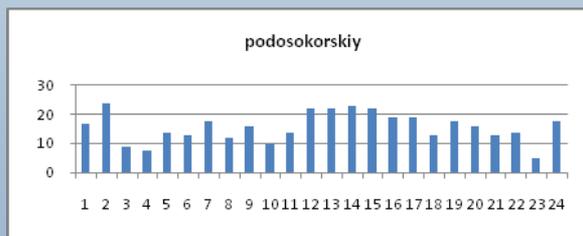
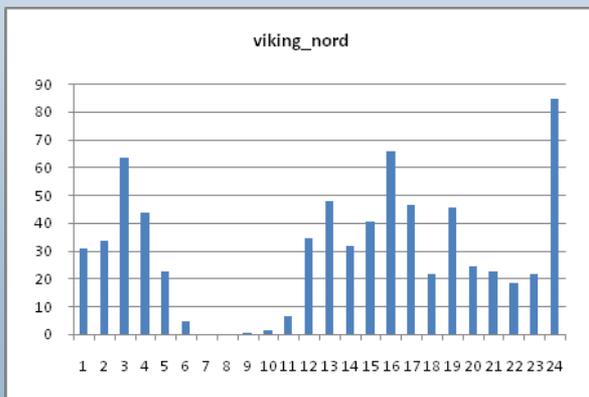
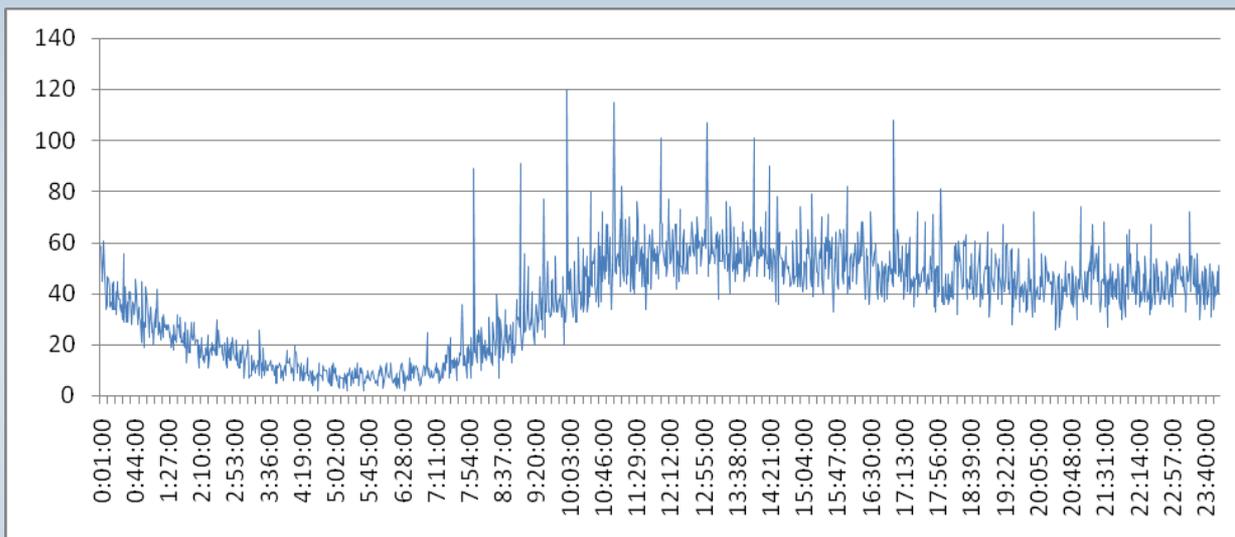
ДАННЫЕ В АНАЛИЗЕ

| Период, за который сделана выгрузка из базы | 2011 г. | | 2012 г. | |
|---|-------------------|-------------------|-------------------|-------------------|
| | 10.08.- 10.09. | 27.11.- 27.12. | 03.02.- 03.03. | 12.03.- 12.04. |
| Количество постов в выгрузке | 54182 | 57282 | 61868 | 62067 |
| Количество блоггеров, имеющих посты в выгрузке | 1822 | 1482 | 1604 | 1563 |
| Кол-во постов, вошедших в тематический анализ | 49811 | 49811 | 57177 | 57835 |
| Кол-во блоггеров с постами, взятыми в анализ | 1798 | 1476 | 1593 | 1551 |
| Количество комментариев в выгрузке | 291209 | 445819 | 655437 | 730980 |

Распределение количества постов по дням недели, март-апрель 2012



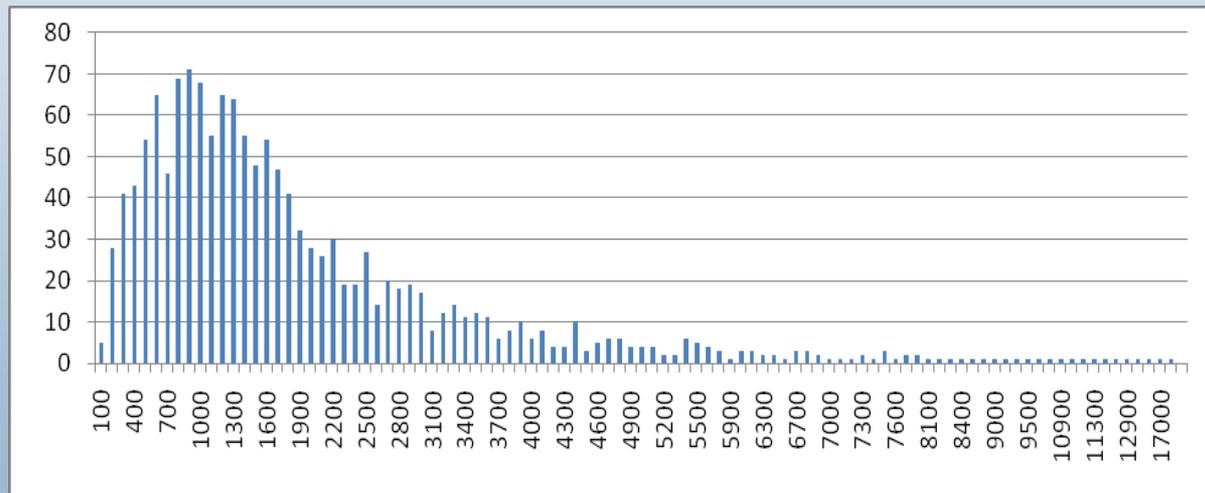
Распределение количества постов по времени суток (пример), декабрь 2011, среднесуточные данные





Распре-
деление
длин
постов

Распреде-
ление средних
длин постов
у блогера по
блогерам



ТЕМАТИКА ПОСТОВ: МЕТОД ОПРЕДЕЛЕНИЯ

TopicMiner ver. 41

Stemming: **Gibbs LDA sampling** Kullback Leibler Distance

Parameters of simulation: Alpha: 0,5 Beta: 0,1 Save step: 50

Number of documents: Label1

Perplexity



Status: Calculation completed

Words with high probability

| | 1 | 2 | 3 | 4 | 5 | 6 |
|----|-------------------|------------------|----------------------|--------------------|---------------------|------------------|
| 1 | федерация: 0,0141 | верба: 0,007033 | мрамор: 0,015868 | mdash: 0,016396 | танк: 0,018219 | японский: 0,015 |
| 2 | российский: 0,012 | ад: 0,005830 | соотечественник: 0,0 | мост: 0,008951 | т: 0,006409 | фонд: 0,010658 |
| 3 | трудовай: 0,00778 | хусаиновна: 0,00 | ваниль: 0,007334 | остров: 0,008124 | мм: 0,006021 | япония: 0,00916 |
| 4 | к: 0,007614 | женщина: 0,0044 | стручок: 0,004287 | тоннель: 0,006304 | вооружение: 0,00544 | только: 0,008497 |
| 5 | работодатель: 0,0 | ангелин: 0,00422 | месторождение: 0,00 | остер: 0,005642 | машин: 0,005247 | вот: 0,008331 |
| 6 | район: 0,006578 | террорист: 0,003 | гражданство: 0,0040 | поезд: 0,005477 | пушка: 0,004666 | посольство: 0,00 |
| 7 | статья: 0,006233 | ветка: 0,003226 | камень: 0,004084 | строительство: 0,0 | название: 0,004472 | деньга: 0,006834 |
| 8 | работник: 0,00606 | земля: 0,003025 | разный: 0,003271 | железнодорожный | башня: 0,004279 | пономарев: 0,00 |
| 9 | расход: 0,005715 | дерево: 0,002825 | нация: 0,003068 | шоколад: 0,002995 | польский: 0,003504 | нужный: 0,00650 |
| 10 | стоимость: 0,0057 | почка: 0,002625 | мраморный: 0,00286 | км: 0,002995 | армия: 0,003504 | американский: (|
| 11 | север: 0,005369 | освящать: 0,0022 | гражданин: 0,002865 | бисквит: 0,002829 | украинский: 0,00331 | британский: 0,00 |
| 12 | отпуск: 0,005369 | душ: 0,002024 | РВП: 0,002662 | стоимость: 0,0028 | снаряд: 0,003117 | правда: 0,00550 |
| 13 | компенсация: 0,00 | веточка: 0,00182 | россия: 0,002662 | шоколадный: 0,002 | комплекс: 0,003117 | японец: 0,00517 |
| 14 | конституционный: | век: 0,001823 | документ: 0,002458 | тут: 0,002664 | орудие: 0,002924 | лекин: 0,005005 |
| 15 | крайний: 0,005024 | вход: 0,001823 | квота: 0,002255 | линия: 0,002498 | платформа: 0,00234 | восток: 0,005005 |
| 16 | договор: 0,004852 | праздник: 0,0018 | получение: 0,002255 | ччасток: 0,002333 | таварс: 0,002149 | весчсд: 0,004835 |

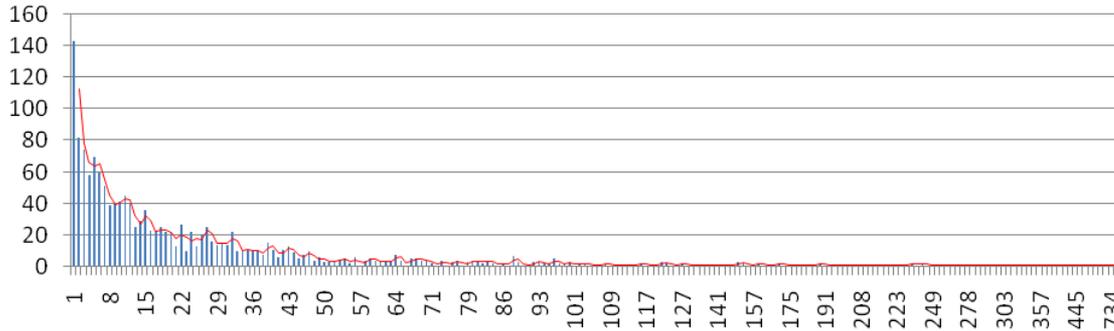
Export to Excel... Number of words for export: 100

LINIS TopicMiner

ГРУППИРОВКА АВТОМАТИЧЕСКИХ ТЕМ

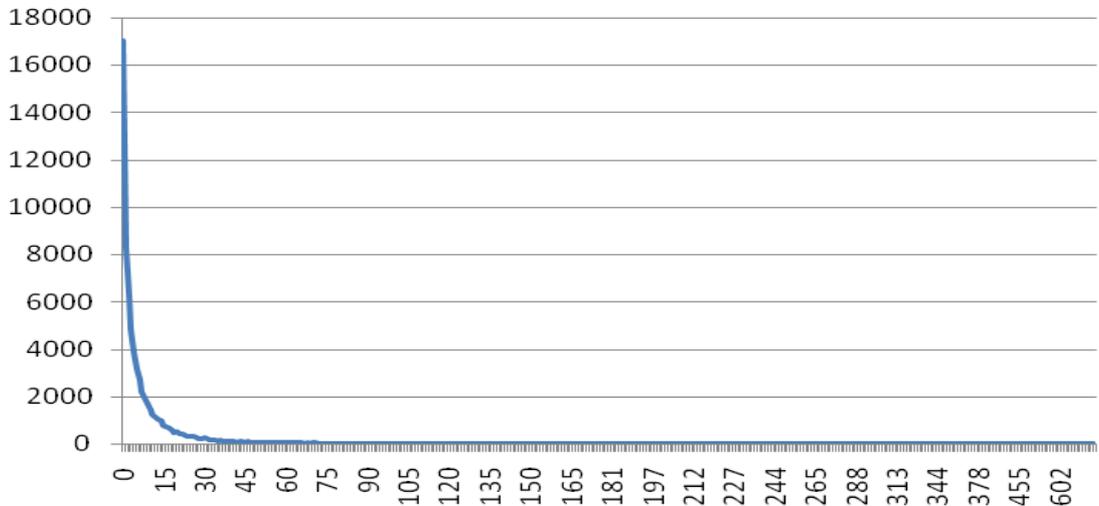
| № группы | Содержание группы тем |
|----------|---|
| 1 | Внутр. политика России, особ. выборы, протесты и права человека |
| 2 | Др. политика (вкл. внешнюю, регион. управление и историю СССР) |
| 3 | Социальные темы |
| 4 | Экономика |
| 5 | Культура |
| 6 | Рекреативная деятельность (отдых, туризм, творчество) |
| 7 | Приватная сфера (семья, дети, межличн/ отношения, секс) |
| 8 | Потребление |
| 9 | Посты на других языках (украинский, английский) |
| 10 | Посты, объединенные по специфической лексике |
| 11 | Не поддающиеся интерпретации темы. |

03 февраля - 03 марта 2012



Распределение количества постов на блогера

Распределение кол-ва комментариев на пост



Распреде-
ление кол-ва
комментов
на блогера
по блогерам



Распреде-
ление среднего
количества
комментов на
пост блогера
по блогерам

Самые активные

| Ник блоггера | Кол-во постов |
|--------------|---------------|
| Carabaas | 967 |
| Ivoropaeva | 790 |
| prosto-telo | 762 |
| Ibigdan | 734 |
| Uzoranet | 700 |
| viking-nord | 690 |
| Dgz | 624 |
| prikolizmus | 624 |
| bloggmaster | 611 |
| Pryf | 584 |

Самые комментируемые

| Ник блоггера | Кол-во получ. комментариев |
|--------------|----------------------------|
| Navalny | 18532 |
| teh-nomad | 17304 |
| Ibigdan | 16408 |
| Tema | 13072 |
| putnik1 | 8759 |
| Avmalgin | 8315 |
| Drugoi | 8215 |
| Zyalt | 7563 |
| md-prokhorov | 6376 |
| Dolboeb | 6309 |

Самые эффективные

| Ник блоггера | Кол-во постов | Кол-во комментариев на пост |
|-----------------|---------------|-----------------------------|
| sobchak-xenia | 6 | 398 |
| Dolgachov | 1 | 372 |
| pesen-net | 1 | 308 |
| navalny | 65 | 285 |
| zhgun | 1 | 258 |
| prostitutka-ket | 13 | 256 |
| md-prokhorov | 26 | 245 |
| fritz morgen | 30 | 189 |
| drugoi | 44 | 187 |
| naganoff | 11 | 175 |

ВЫВОДЫ

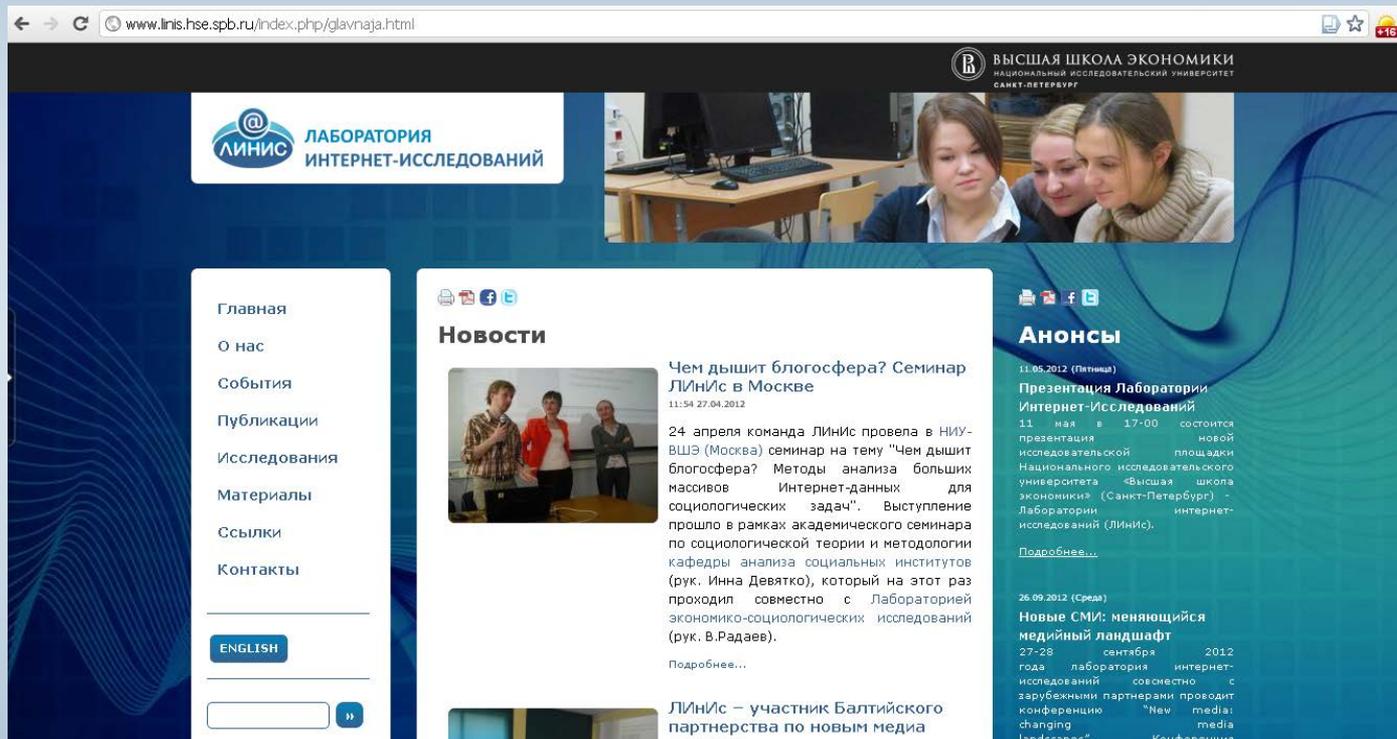
Надо больше исследовать 😊

Исследование не проверяет гипотез

Не отвечает на конкретный вопрос

Подготавливает дальнейшие исследования

СПАСИБО!



The screenshot shows the homepage of the LINIS website. At the top, there is a navigation bar with the LINIS logo and the text 'ЛАБОРАТОРИЯ ИНТЕРНЕТ-ИССЛЕДОВАНИЙ'. Below this is a large banner image showing three young women looking at a computer screen. The main content area is divided into three columns. The left column contains a vertical menu with links: Главная, О нас, События, Публикации, Исследования, Материалы, Ссылки, and Контакты. Below the menu is an 'ENGLISH' button and a search bar. The middle column is titled 'Новости' and features a news article with a photo of three people. The article title is 'Чем дышит блогосфера? Семинар ЛИНИС в Москве' and the date is '11:54 27.04.2012'. The text of the article describes a seminar held at the HSE Moscow campus on April 24th, focusing on the analysis of large-scale Internet data for sociological tasks. The right column is titled 'Анонсы' and features two short news items. The first is dated '11.05.2012 (Пятница)' and is titled 'Презентация Лаборатории Интернет-Исследований'. The second is dated '26.09.2012 (Среда)' and is titled 'Новые СМИ: меняющийся медийный ландшафт'.