

УДК 004.457+ 330.43

UDC 004.457+330.43

СТАТИСТИКА ПОИСКОВЫХ ЗАПРОСОВ

SEARCH QUERY STATISTICS

Кириченко Евгений Владимирович
магистрант

Kirichenko Evgeniy Vladimirovich
undergraduate

Сытников Демид Анатольевич
магистрант

Sytnikov Demid Anatolevich
undergraduate

Петухов Александр Валерьевич
магистрант

Petuhov Aleksandr Valeryevich
undergraduate

Кацко Игорь Александрович
д.э.н., профессор
*Кубанский государственный аграрный
университет, Краснодар, Россия*

Katsko Igor Aleksandrovich
Dr.Sci.Econ., professor
Kuban State Agrarian University, Krasnodar, Russia

В статье дан обзор имеющихся статистических сервисов в современных поисковых системах. Обсуждаются статистические результаты на примере вступления России в ВТО

This article surveys the available statistical services in modern search engines. The statistical results are discussed on the example of Russia's accession to the WTO

Ключевые слова: ПОИСКОВЫЙ ЗАПРОС, ВТО, ПОИСКОВАЯ СИСТЕМА, ФУНКЦИОНАЛЬНЫЙ СЕРВИС

Keywords: WTO, SEARCH ENGINES, SERVICE FUNCTIONALITY

В последнее время особую актуальность приобретает изучение интересов населения к тем или иным событиям в жизни общества и государства через интернет. Целью данной статьи является иллюстрация возможностей использования статистического сервиса различных поисковых систем для изучения динамики изменения интересов населения на примере вступления России в ВТО.

Классификация знаний

Согласно теории искусственного интеллекта *знания* подразделяются на *отделимые (кодифицированные)*, которые можно представить в одной из естественных или искусственных семиотических (знаковых) систем, и *неотделимые*, носителем которых является человек. Кодифицированные знания обычно формализованы и достаточно легко могут храниться, копироваться и распространяться. Неотделимые знания накапливаются через личный опыт, обучение в процессе деятельности, социальные взаимоотношения и т.д. Как следствие, они трудно поддаются количественному

определению, хранению или передаче. Они выходят далеко за рамки технического прогресса и инноваций, находящих материальное воплощение в продуктах, услугах или процессах.

В свою очередь переносимые (кодифицированные) знания подразделяются на: структурированные, слабоструктурированные, неструктурированные. В первом случае знания имеют определенную последовательность удобных для восприятия форм: уравнения, формулы, графики, схемы, таблицы типа «объект-свойство», а также связи, которые позволяют облегчить их обработку и передачу – теоретической основой изучения таких знаний являются математические и статистические дисциплины, а также современные информационные технологии (например, *KDD* и *Data Mining*). Слабоструктурированные знания сегодня являются предметом изучения теории когнитивного моделирования. Неструктурированные же знания являются более сложными для восприятия и обработки, поэтому обычно предлагаются различные подходы к их структуризации.

Одним из примеров неструктурированной формы знаний, которые поддаются обработке, являются всевозможные тексты или же *WEB*-контент. Развитию последнего способствовало появление глобальных каналов связи, в частности сети Интернет.

Автоматические программные средства поиска и анализа информации

Рост объема доступных через Интернет данных, хранимых в слабо структурированном виде, способствовал появлению автоматических программных средств поиска информации и получения данных об использовании определенных ресурсов. Возник целый ряд интеллектуальных систем, основная задача которых состоит в эффективном извлечении знаний из сети Интернет.

Большинство систем мониторинга сети Интернет предоставляют возможность фильтрации и получения статистической информации о запросах пользователей. Подобные инструменты помогает определять количество обращений к разным файлам и серверам, адресам отдельных ресурсов.

Статистика запросов фактически представляет собой механизм, позволяющий проводить исследования, которые невозможно провести никаким другим способом. Подобная статистика является наиболее достоверным источником современного языка, в отличие от анализа поисковых результатов, которые являются приблизительными, в силу того, что информация в интернете быстро устаревает. Кроме того, запросы к поисковой системе считаются одним из наиболее репрезентативных источников живого языка [3].

Прежде чем перейти к системам, позволяющим просмотреть статистику запросов пользователей, необходимо раскрыть понятие «статистика запросов».

Поисковый запрос – это информация, с помощью которой осуществляется поиск специальной системой, такой как: *Yandex, Google, Rambler* и др. Как правило, поисковый запрос задаётся в виде фраз или слов. Бывают и запросы в виде изображений. Формат ключевых запросов зависит от типа информации для поиска и устройства конкретной поисковой системы.

Ключевое слово – это слово, которое в совокупности с другими ключевыми словами, представляет текст сайта. Ключевое слово используется для поиска. Содержание текста, представленное ключевыми словами, анализируется лингвистическими и математическими методами. Например, анализ частоты появления слова в тексте [4].

Статистика запросов – это информация об обращениях пользователей к поисковой системе по «ключевым словам» [3].

Другими словами статистика запросов – это количество поисковых запросов пользователей по «ключевым словам» за определенный промежуток времени.

В большинстве случаев при работе с сервисом статистики имеется возможность фильтровать результаты по территориальному признаку, языку, а также в хронологической последовательности. При этом обычно, сервис показывает не только данные об искомом запросе, но также и о словосочетаниях, синонимах и близких темах.

В данной статье рассмотрены наиболее интересные и функциональные сервисы, предоставляемые такими поисковыми системами как *Yandex*, *Google*, *Rambler* и проанализированы полученные с их помощью данные по запросу о вступлении России в ВТО.

Прежде чем приступить к поиску интересующей нас информации, рассмотрим некоторые преимущества и недостатки предоставляемых сервисов каждой из поисковых систем.

Yandex

Яндекс предоставляет доступ к своей статистике всем желающим в рамках системы по продаже рекламы Яндекс.Директ. Кроме стандартной информации о количестве запросов в месяц, а также словосочетаниях и близких темах, поисковик предоставляет возможность отсеивать результаты по регионам, городам в хронологической последовательности.

Учитывая тот факт, что Яндекс является самой популярной в Рунете поисковой системой [5], подобная статистика является наиболее репрезентативной при оценке положения дел в Рунете.

В Яндекс *Wordstat* статистика запросов представляется в несколько упрощенном виде — объединяются все возможные словоформы (падежи, числа и т.п.), в большинстве случаев не учитываются предлоги, а так же вопросительные формы, например, «что такое» и т. п.

При помощи специальных операторов можно добиться конкретизации статистики Яндекса именно по интересующей словоформе поискового запроса. Обычно для этого достаточно бывает заключить нужный поисковый запрос в кавычки. При этом учитываться будут только эти слова запроса, но в любой допустимой словоформе или же вместе с кавычками можно будет дополнительно поставить восклицательные знаки перед каждым из слов, обязав тем самым статистику Яндекса учитывать только эти слова и только в выбранной вами словоформе.

Следует отметить, что в статистике поисковых запросов Яндекса приводятся не только производные от введенных вами слов (в левой колонке как раз будут показаны эти самые расширенные варианты запросов с добавлением других слов), но еще дополнительно в правой колонке будут показаны ассоциативные запросы, которые набирали те же самые пользователи в Яндексе вместе с введенными вами словами за одну и ту же сессию поиска.

Rambler

Система статистики имеется и у Рамблера. Она менее репрезентативна в силу меньшей популярности поисковой системы, чем статистика Яндекса, но её преимуществом является более подробная информация. К примеру, сервис выдает информацию о количестве запросов не только с заглавной страницы, но также и со всех остальных. Кроме того, статистика Рамблера позволяет использовать несложный язык запросов для уточнения или, наоборот, расширения результата.

Данный механизм отличается от статистики запросов в Яндексе тем, что в ней не объединяются результаты для разных словоформ. Т.е. можно без дополнительных операторов получить статистику частотности запроса именно по словам в нужном падеже и требуемом числе.

Google

Крупнейшая в мире поисковая система *Google* также предоставляет открытый доступ к своей статистике запросов. В отличие от двух предыдущих, количественная статистика доступна в формате *csv*. Визуально статистика представляется лишь относительно – в виде графика. Отчёты выделяются особой подробностью: например, кроме обычной статистики запросов пользователей, можно посмотреть степень конкуренции рекламодателей за конкретный поисковый запрос, просмотреть историю трафика для выбранных ключевых слов; предоставляется подсказка возможно полезных минус-слов.

В особом виде статистику отображают графики *Google Trends*. Сервис позволяет вводить до 5 разных запросов, изучать и сравнивать изменение интереса к ним в мире в виде графика за прошедшие 2-3 года.

Исследование предметной области

Нами анализировались данные в интернет ресурсах на тему вступления России во Всемирную Торговую Организацию. Анализировалось не наполнение *WEB*-ресурсов информацией, а частота и тип запросов, вводимых россиянами и связанными с вступлением России в ВТО.

Такой подход позволил проанализировать не мнение отдельно взятых лиц, а общую тенденцию интересов населения относительно вступления России в ВТО. Это позволило отразить наиболее реальную тематику вопросов, волнующих население страны.

О том, как применять подобные инструменты на практике и какую информацию мы можем получить, будет рассказано ниже.

Прежде всего, для проведения исследования по поисковым запросам нам необходимо составить список интересующих нас запросов – семантический словарь.

Для того что бы в различных сервисах мы получали наиболее актуальную информацию конкретизируем условия нашего поиска. Поиск будем производить по всем регионам России (Центр, Северо-Запад, Поволжье, Юг, Сибирь, Дальний Восток, Северный Кавказ, Урал) в период с 01 марта 2011 по 01 марта 2013 года.

В таблице 1 представлен используемый нами семантический словарь. Он составлен на основе «рейтингов запросов» за март 2013 г. со словом «ВТО».

Таблица 1 – Семантический словарь

Слово	Количество запросов в месяц
ВТО	42226
Вступление + ВТО	9953
ВТО плюсы + минусы	935
ВТО + влияние	1099
ВТО + пошлины	707



Рисунок 1 – Количество поисковых запросов в системе Яндекс

Проанализируем результаты, полученные от поисковых системам при вводе слов из семантического словаря. На рисунке 1 и в таблице 2 представлены результаты работы поисковой системы Яндекс.

Таблица 2 – Количество поисковых запросов в системе Яндекс

Месяц	ВТО	ВТО + вступление	ВТО плюсы + минусы	ВТО + влияние	ВТО + пошрины
2011.03	82508	18697	3485	299	853
2011.04	82523	18326	3181	368	769
2011.05	69606	13090	2212	453	426
2011.06	56350	10429	1804	238	643
2011.07	19309	3287	579	60	320
2011.08	20479	3645	495	44	373
2011.09	37534	6705	1262	191	249
2011.10	93905	24176	4909	299	1726
2011.11	249031	90306	19942	706	12694
2011.12	235793	82589	18740	936	11322
2012.01	105774	34032	6768	623	5564
2012.02	84969	25856	4064	558	3221
2012.03	128316	41040	5072	924	5334
2012.04	129286	40327	4499	1241	4362
2012.05	122649	35902	3573	1268	3253
2012.06	119297	37315	3363	776	3570
2012.07	257671	95112	11175	217	12009
2012.08	219650	89573	11935	575	23636
2012.09	129119	48705	7412	1535	8940
2012.10	125561	36757	5377	2316	3420
2012.11	124178	34512	4219	1659	2558
2012.12	122652	34225	4257	1909	2145
2013.01	89905	22958	2124	997	1972
2013.02	88834	22677	2469	1422	1572

По полученным результатам следует отметить, что поведение всех графиков практически одинаково и в определенные промежутки времени наблюдается как возрастание интересов, так и затухание. А именно, в период с марта 2011 до июля наблюдается спад интереса, а с июля до конца года по всем запросам активность пользователей возросла и достигла свое-

го максимума в ноябре 2011 года. Далее по всем запросам пошел спад и пользователи практически с постоянной частотой интересовались исследуемыми запросами. Следующий скачок активности россиян приходится на июль – август 2012 года.

Стоит отметить, что представленный рисунок 1 был построен средствами MS Excel, так как сервис от Яндекс позволяет строить графики только по отдельно взятому запросу, что не всегда является удобным при анализе нескольких запросов.

Рассмотрим данные полученные с помощью сервисов от компании Google на рисунке 2.

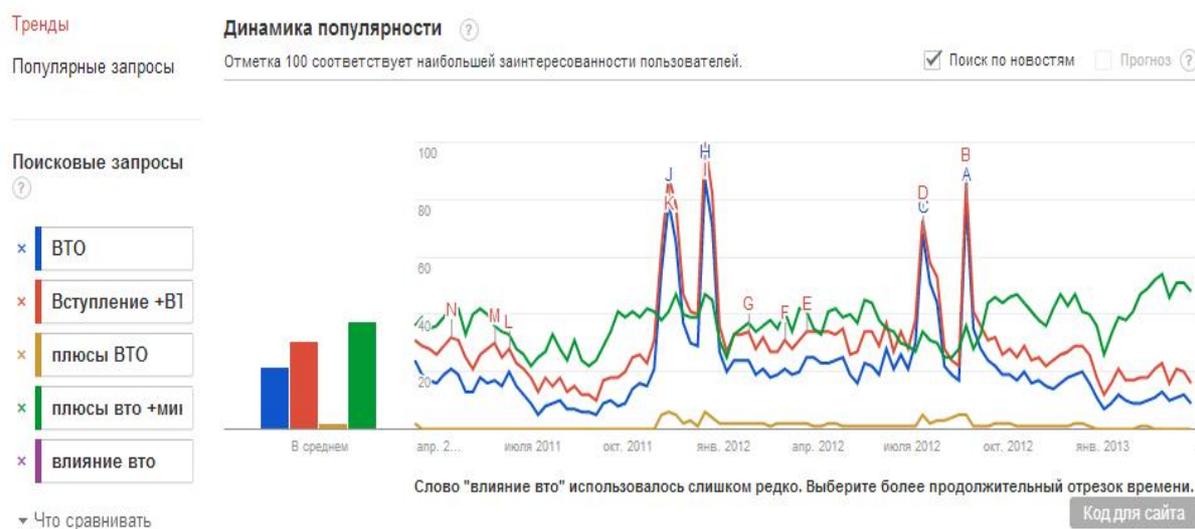


Рисунок 2 - Количество поисковых запросов в системе *Google*

Сервис от компании *Google* представляет данные в относительной частоте. Числа на графике показывают долю запросов по ключевым словам в общем числе запросов, выполненных в *Google* за определенное время. Они являются не абсолютным выражением объема поисковых запросов, а относительным, в масштабе от 0 до 100. Каждая точка на графике соотносится с максимальным значением. При отсутствии достаточного количества данных отображается значение равное нулю.

Период пиковых величин представлен в таблице 3.

Запросов «ВТО +пошлины» практически не было, поэтому мы не включали данные результаты в таблицу пиковых значений.

Таблица 3 – Периоды наибольшей активности пользователей, %

Даты	ВТО	ВТО + вступи- ление	ВТО плюсы	ВТО плюсы + минусы
6 – 12 ноября 2011	79	88	6	41
11 – 17 декабря 2011	87	100	6	47
8 – 14 июля 2012	68	73	5	34
19 – 25 августа 2012	79	86	5	36

В отличие от сервиса «Яндекс статистика», «Google Тренды» не предоставляет данные в табличном виде.

Одной из интересных возможностей Google Тренды является функция поиска по новостям. Данная возможность позволяет отобразить на графике в его пиковых точках наиболее популярную ссылку.

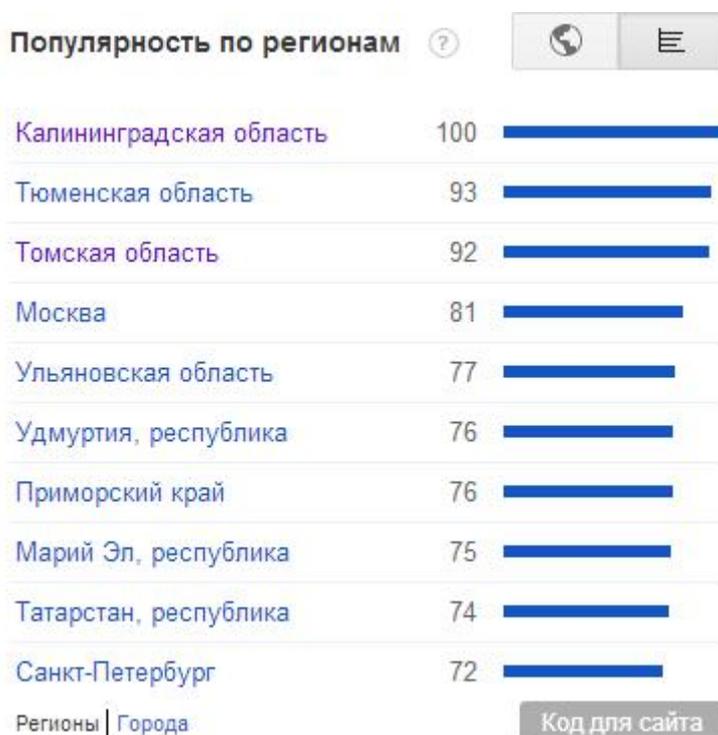


Рисунок 4 – Популярность запроса «ВТО» по регионам России

Еще одной интересной особенностью, отличающей «Google Тренды» от других сервисов, является возможность рассмотрения популярности за-

проса по регионам. Данные представлены в относительном виде. Данный интерфейс показывает, какой регион из общей массы был наиболее заинтересован исследуемым поисковым запросом (рисунок 4).

Для изучения активности пользователей конкретного региона по месяцам можно перейти по ссылке, представленной в списке (рисунок 5).

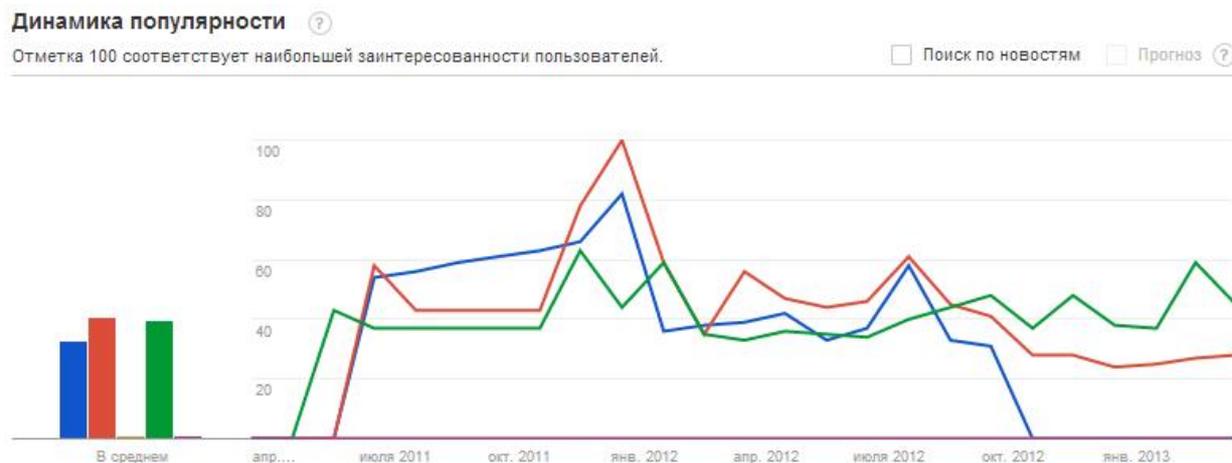


Рисунок 5 – Популярность запроса «ВТО» в Калининградской области

Данные, полученные с помощью сервисов поисковой системы Рамблер изображены на рисунке 6 и представлены в таблице 4.



Рисунок 6 – количества поисковых запросов в системе Рамблер

Сервис «Рамблер статистка» смог предоставить данные только в период с февраля 2012 по февраль 2013.

По полученным данным, так же как и в других системах, наблюдается одновременный рост и спад активности пользователей по всем запросам. Следует отметить, что пользователи поисковой системы Рамблер проявляли больше интереса к вопросу о пошлинах после вступления России в ВТО, чем к положительным или отрицательным сторонам вступления России в ВТО.

Таблица 4 – Количество поисковых запросов в системе Рамблер

Месяц	ВТО	ВТО + вступление	ВТО плюсы + минусы	ВТО + пошлины
2012.02	3892	343	68	71
2012.03	8387	863	102	145
2012.04	5560	449	58	108
2012.05	5541	340	41	62
2012.06	4607	456	45	66
2012.07	6837	909	118	185
2012.08	6639	862	119	324
2012.09	3204	341	68	103
2012.10	4000	274	68	81
2012.11	4345	268	84	41
2012.12	6664	317	73	60
2013.01	4483	230	55	36
2013.02	5280	250	43	59

Сервис «Рамблер статистика» не отображает полученные результаты в графическом виде, поэтому для построения графика мы обратились к возможностям *MS Excel*.

Несмотря на то, что поведение кривой графика во всех рассматриваемых сервисах, практически одинаково, существуют и различия. А именно:

1. Сервис от Яндекс и Рамблер показывает нам довольно большую заинтересованность россиян вопросами пошлин после вступления России в ВТО, в то время как сервис от *Google* показал практически нулевую активность пользователей данного сервиса.

2. График, полученный с использованием сервиса от *Google*, показал нам резкие скачки, в периоды, представленные в таблице 3, в то время как Яндекс и Рамблер показал нам одноразовое возрастание и затухание кривой.

3. Что касается запроса «ВТО плюсы +минусы», то сервис от *Google* показывает стабильную заинтересованность россиян этим вопросом, без резких скачков, в то время как в Яндекс и Рамблер практически идентично с остальными кривыми рассматриваемых запросов.

Проанализируем результаты, полученные с помощью всех рассматриваемых сервисов, и сопоставим их с событиями, связанными со вступлением России в ВТО.

Наиболее важные события, связанные со вступлением в ВТО

1. В течение осени 2011 г. были согласованы остававшиеся вопросы на переговорах с США.

2. Одновременно в течение осени 2011 г. в результате многомесячных неформальных российско-грузинских консультаций при посредничестве Швейцарии удалось выработать приемлемое для обеих сторон решение по контролю за передвижением гражданских грузов по территории Абхазии и Южной Осетии. В результате со стороны Грузии были сняты возражения по созыву формального заседания Рабочей группы.

3. 10 ноября 2011 г. переговоры о присоединении России к ВТО были завершены. Рабочая группа одобрила пакет документов о присоединении РФ к ВТО для внесения на рассмотрение Восьмой министерской кон-

ференции ВТО. Таким образом, мандат Рабочей группы по присоединению России к ВТО был исчерпан, после чего она была распущена.

4. 16 декабря 2011 г. в ходе на 8-й Министерской конференции стран-членов ВТО в Женеве был одобрен пакет документов по присоединению России к ВТО. Пакет включал в себя: протокол о присоединении России к ВТО, содержащий Перечень тарифных уступок и перечень специфических обязательств по услугам; доклад Рабочей группы по присоединению РФ к ВТО.

5. В соответствии с правилами ВТО, России был предоставлен срок в 220 дней для ратификации пакета документов о присоединении к ВТО национальным парламентом.

6. 10 июля 2012 г. Государственная дума РФ 238 голосами против 208 и 1 воздержавшемся одобрила Протокол о присоединении России к Всемирной торговой организации.

7. 18 июля 2012г. Совет Федерации РФ ратифицировал Протокол о присоединении России к Всемирной торговой организации.

8. 21 июля 2012 г. Президент России В. Путин подписал федеральный закон "О ратификации Протокола о присоединении РФ к Марракешскому соглашению об учреждении Всемирной торговой организации от 15 апреля 1994 г."

9. 22 августа 2012 г. Российская Федерация официально стала 156-м членом Всемирной торговой организации.

Пики в интересах Россиян практически совпадают с перечисленными событиями, что говорит о политической активности и заинтересованности граждан в событиях, происходящих в стране.

Выводы

1. В результате проведенных исследований была выявлена прямая взаимосвязь роста активности пользователей в наиболее важные периоды государственной деятельности, связанной со вступлением России в ВТО.

Таким образом, анализ результатов, полученных из рассмотренных в работе сервисов, позволяет отразить наиболее точные интересы россиян, касающиеся вступления России во Всемирную Торговую Организацию. Следует отметить, что только совместное использование нескольких сервисов позволяет отразить наиболее верную картину.

Полученные результаты можно использовать для дальнейшего анализа *WEB*-контента. Например, можно более точно указывать период появления статей касающихся рассматриваемых вопросов или анализировать только те результаты запросов, которые наиболее сильно волновали пользователей в определенный период времени. Потенциально существует возможность привлечение средств *Web-Mining*, *Text-Mining* для более глубокого изучения не только запросов, но и текстовой информации по интересующему нас направлению.

2. Сегодня ни для кого не секрет, что тенденции изменений в обществе всегда начинаются «снизу» (да, конечно их можно инициировать с помощью СМИ и т.д.). Именно этот факт и является одной из причин по которой проводятся всевозможные социологические и эконометрические обследования. Из настоящей статьи, очевидно, что уже сегодня существует альтернатива всевозможным «выборочным» обследованиям, проводимым службами государственной статистики, социологами и политологами. Нужно понимать, что изучение социально-экономических процессов и тенденций в обществе требует комплексного подхода, опирающегося на системную методологию и адекватный инструментарий. Использование средств и инструментов, имеющихся на сегодня статистических сервисов, в современных поисковых системах позволяют поднять возможности изучения социально-экономических систем на новый качественный уровень.

Литература

1. Байков В. Интернет. Поиск информации. Продвижение сайтов / В. Байков - Санкт - Петербург, 2000. - 288 с.

2. Щербаков А. Ю. Интернет-аналитика. Поиск и оценка информации в web- ресурсах. Практическое пособие / А. Ю. Щербаков - Книжный мир, 2012. – 80 с.

3. Информационные ресурсы и услуги. Многоязычный проект по созданию полноценной и точной энциклопедии со свободно распространяемым содержимым: [сайт]. – Режим доступа: <http://ru.wikipedia.org/>

4. Информационные ресурсы и услуги. Блог Яремчук Романа по созданию сайтов и все что с этим связано: [сайт]. – Режим доступа: [http : http://www.delaydengi.com/](http://www.delaydengi.com/)

5. Информационные ресурсы и услуги. крупный сервис дневников и сообществ, авторитетный сервис статистики для сайтов: [сайт]. – Режим доступа: <http://www.liveinternet.ru/>

References

1. Bajkov V. Internet. Poisk informacii. Prodvizhenie sajtov / V. Bajkov - Sankt - Peterburg, 2000. - 288 s.

2. Shherbakov A. Ju. Internet-analitika. Poisk i ocenka informacii v web- resursah. Prakticheskoe posobie / A. Ju. Shherbakov - Knizhnyj mir, 2012. – 80 s.

3. Informacionnye resursy i uslugi. Mnogojazychnyj proekt po sozdaniju polnocennoj i tochnoj jenciklopedii so svobodno rasprostranjaemym sodержимым: [sajt]. – Rezhim dostupa: <http://ru.wikipedia.org/>

4. Informacionnye resursy i uslugi. Blog Jaremchuk Romana po sozdaniju saj-tov i vse chto s jetim svjazano: [sajt]. – Rezhim dostupa: [http : http://www.delaydengi.com/](http://www.delaydengi.com/)

5. Informacionnye resursy i uslugi. krupnyj servis dnevnikov i soob-shhestv, avtoritetnyj servis statistiki dlja sajtov: [sajt]. – Rezhim dostupa: [http:// www.liveinternet.ru/](http://www.liveinternet.ru/)