
КЛЮЧЕВЫЕ СЛОВА НА РУССКОМ ЯЗЫКЕ В СИСТЕМАХ ПОИСКА ИНФОРМАЦИИ В ИНТЕРНЕТЕ (опыт семантического и культурологического анализа)

П.В. Морослин

Международный славянский институт
ул. Годовикова, 9, Москва, Россия, 129085

Статья посвящена анализу лексико-семантических и культурологических особенностей ключевых слов, функционирующих в информационно-поисковых системах Интернета. Рассматриваются частотность слов, система принципов номинации, семантические и ассоциативные связи ключевых слов в русскоязычном Интернете.

Ключевые слова: Интернет, система поиска, компьютерная коммуникация, ключевые слова.

Функционирование русского языка в Интернете вызывает значительный интерес исследователей. В научной литературе в последние годы появилось несколько монографий, десятки статей об особенностях русскоязычного Интернета (часто как синоним употребляется и слово Рунет), о новых речевых жанрах появившихся в Интернете на русском языке, об особенностях компьютерной коммуникации на русском языке [1; 2; 3; 4]. К сожалению, работы в данной области пока не позволяют более или менее объективно представить реальное функционирование разных текстов в Сети. Это, на наш взгляд, связано прежде всего с тем, что еще не был проведен более или менее подробный анализ особенностей функционирования электронных текстов в зависимости от задач общения, выполняемых функций, лингвистических особенностей систем хранения и поиска информации, способов создания и функционирования того или иного электронного текста. На наш взгляд, исследование русскоязычного Интернета должно опираться на ряд весьма сложных факторов: особенности самого Интернета как сложной многофункциональной иерархической системы; социолингвистическую характеристику основных групп пользователей, их коммуникативных потребностей, языкового вкуса; учет значительного влияния на развитие русскоязычного Интернета английского языка и культуры и традиций общения на английском языке; тенденции развития и изменения современного русского языка, которые активизировались в конце XX — начале XXI в.

Важную роль в оценке эффективности текстов, функционирующих в Интернете, играет такой показатель, как ключевые слова, т.е. слова, которые используются при поиске тех или иных веб-текстов, тех, которые пользователи набирают в поисковых системах для того, чтобы найти необходимую информацию. Понятие «ключевые слова» обозначает запросы, по которым тот или иной сайт может быть найден в поисковиках.

Поисковыми системами Интернета проводится статистика запросов — информация об обращениях пользователей к поисковой системе с помощью ключевых слов. В большинстве случаев при работе с сервисом статистики имеется

возможность классифицировать результаты поиска по датам, географии запросов. При этом, как правило, сервис показывает не только данные об искомом запросе, но также и о словосочетаниях, синонимах и близких темах («ищут также»).

Очевидно, что если сайт посвящен туризму и отдыху, то он должен находиться в системах поиска (например, Яндекс, Rambler, Google) на основе лексического запроса «отдых», «курорты», «горящие путевки», названия туристических фирм и т.п.

Эти ключевые слова несут важную информацию об интересах аудитории, в том числе дают представление о национально-ориентированной специфике актуальности электронных текстов на том или ином языке. Ключевые слова косвенным образом характеризуют и саму аудиторию, например, такие часто встречающиеся в 2006 г. ключевые слова российского Интернета, как *реферат*, *рефераты*, *банк рефератов*, *гороскоп*, *погода* подтверждают положение, что основными пользователями Интернета являются молодые люди в возрасте от 18 до 32 лет, в большинстве студенты, которых часто интересует информация образовательного характера. В 2008 г. интересы аудитории несколько изменились, на первый план вышли следующие слова: *марафон* (название известной фирмы в области мобильной связи), *переводчик* (система электронного перевода), *новости*, *работа*, *сбербанк* и др.

Сравнительный анализ поисковых запросов в англоязычной части Интернета показывает, что среди поисковых запросов первые места занимают слова, связанные с известными актерами, музыкантами, а также с покупками через интернет-магазины.

Намечается ярко выраженная тенденция в отражении системой ключевых слов тех или иных внешних событий, причем в различных жанрах Интернета эти интересы могут отражаться специфическим образом, например, в системе наиболее частотных поисковых запросов в новостных сайтах Интернета, а также в наиболее часто употребляемых словах и темах в блогах. Изменение интересов создателей блогов и тематики блогов в зависимости от тех или иных событий в Яндексе отражается специальными графиками, которые получили название «пульс блогосферы».

Система формирования ключевых слов является составной частью так называемой контекстной рекламы, которая представляет собой текст рекламного содержания или текст информационного характера. Отличительной чертой контекстной рекламы является то, что она появляется там, где с помощью ключевых слов ищут ту или иную информацию. Контекстная реклама возникает в связи с конкретным тематическим запросом в системе поиска с помощью ключевых слов. Например, если вводится словосочетание *русский язык*, то контекстная реклама указывает коммерческие курсы, которые предлагают обучение русскому языку. Если в систему поиска вводится ключевое слово *школа*, то в контекстной рекламе появляются сообщения об адресах школ, о сайтах школ, справочниках.

Поиск по ключевым словам оказывает огромное влияние на популярность того или иного сайта. Согласно данным различных рейтингов популярными являются следующие российские сайты: 1) V Kontakte, 2) <http://mail.ru/>, 3) Яндекс,

4) Одноклассники.ru, 5) Rambler, 6) Google (русскоязычная версия поискового сервера google.ru), 7) Narod.ru, 8) YouTube и др.

Весьма популярная рейтинговая система Top 100 Rambler в ноябре 2008 г. показывает несколько иную картину. В самые популярные ресурсы Интернета на русском языке входят 1) СМИ и периодика, 2) сервисы (почта, онлайн переводчик, программа просмотра фильмов ru.tube, ресурс для блогов «Живой журнал»), 3) сайты, связанные с товарами и услугами. Следует подчеркнуть, что наблюдается корреляция между популярностью тех или иных групп тематических сайтов и частотностью ключевых слов. Высокочастотными являются слова, связанные с поиском новостей, покупками, развлечениями.

Необходимо подчеркнуть, что отмечаются изменения в актуальности тех или иных ключевых слов в зависимости от внешних событий или обстоятельств. Яркий пример — изменения в количестве ключевых слов во время террористической атаки 11 сентября в США или чемпионата Европы по футболу. Следует отметить, что, несмотря на повышение актуальности тех или иных слов в связи с определенными событиями, общая тенденция выбора ключевых слов для данной лингвокультурной общности остается стабильной.

В анализе запросов Яндекса были опубликованы результаты анализа лингвистических особенностей ключевых слов. Около 2,5% запросов сформулированы как вопрос. Это обычные вопросы, и пользователи, которые их задают, общаются с поисковой системой так, как будто это живой человек. У Яндекса даже спрашивают (Как пройти в библиотеку?) — в среднем 41 раз в месяц, (Зачем Герасим утопил Муму?) — 53 раза и (Кто убил Лору Палмер?) — 107 раз. Вопросы, начинающихся со слова «как», задают больше, чем вопросов, начинающихся со всех остальных вопросительных слов вместе взятых. Интересно, что вопрос «что?» значительно популярнее вопроса «кто?», и в частности вопрос «что делать?» интересует пользователей больше вопроса «кто виноват?»

Самые короткие составляющие запросов к Яндексу — отдельные буквы и цифры. Пользователей интересуют все буквы русского алфавита, больше всего — предлоги и однобуквенные союзы (например, «и» в августе 2008 г. искали 4 385 966 раз), а меньше всего — букву «ъ» (9 тысяч запросов в месяц). Наиболее длинные осмысленные слова, как правило, сложные существительные, состоящие из нескольких корней. Самое длинное слово, заданное в качестве запроса к Яндексу в августе 2008 г., состоит из 37 символов — «гиппопотомомонстросесквипедалиофобия». По этому запросу находится 4583 страницы (на сентябрь 2008 г.). Среди самых длинных запросов, на которые существуют ответы в Интернете, преобладают различные химические соединения (этилоксиэтилпарафенилендиаминсульфат — 35 символов), названия компаний («Средневожжксельэлектросетьстрой» — 32 символа) и разного рода фобии (например, гексакосийгексеконтагексофобия — 31 символ).

В запросах на поиск картинок самые длинные запросы — это «электростеклоподъемник» и «электроводонагреватель» — по 22 символа. Для общения с поисковой машиной чаще всего используют существительные — эту часть речи со-

держат 75% запросов к поиску. Вторая по распространенности часть речи — прилагательные, они присутствуют в 16% запросов к веб-поиску. Глаголы используют только в 5% случаев, а наречия отмечены менее чем в 1%

Существующие системы анализа ключевых слов позволяют эффективно использовать различные инструменты, показатели для поиска и выбора ключевых слов. В основе сбора данных по ключевым словам лежат два принципа — принцип счетчика (специального скрипта, загружающегося пользователем сервером вместе с загрузкой веб-страницы) и принцип анализа логов (специальных файлов на сервере, фиксирующих все посещения). Оба способа сбора информации работают независимо друг от друга и каждый с определенной степенью погрешности. Из собранной с помощью счетчика или лог-анализатора информации можно формировать разные массивы данных, изучать отдельные срезы и тематические выборки. Подобные обобщенные данные часто можно встретить в исследованиях по глобальной статистике Рунета, например, HotLog и SpyLog.

Предназначение любого интернет-ресурса заключается в эффективном достижении целей, определенных при его создании. Эффективность сайта зависит от его содержания, определяется числом посетителей сайта (в частности, приходящих из поисковых серверов), скоростью и удобством получения интересующей информации, количеством повторных возвращений на данный сайт. Эффективность сайта также зависит от того, насколько интересны статьи, удобна навигация, привлекателен дизайн и т.п. Для оценки эффективности сайта собираются статистические данные посещения данного ресурса, применяются различные методы математической статистики для получения и обработки результатов.

Интернет-статистика дает возможности анализировать различные особенности сайта. Во-первых, можно существенно улучшить дизайн, навигацию и размещение ссылок на сайте. Во-вторых, на основании полученных статистических отчетов возможно повысить рейтинг сайта на крупнейших поисковых системах Рунета (Яндекс, Google, Rambler).

Система сбора интернет-статистики используется для получения данных о посещениях сайта, последующей их обработки и получения различных отчетов, отображающих интересующую информацию. Все это позволяет оценивать эффективность интернет-сайта по разным параметрам, определять удобство пользования как сайтом в целом, так и его различными элементами навигации; характеризовать качество отдельных элементов сайта (баннеров, статей, элементов каталога и т.д.).

Среди основных функциональных возможностей систем статистики считается посещаемость сайта; учитываются все переходы посетителей с поисковых машин по каждой ключевой фразе; рассчитывается время просмотра каждым посетителем каждой страницы сайта; проводится анализ данных статистики посещений для выявления групп предпочтений среди посетителей сайта.

Существуют специальные сервисы поисковых систем, задачей которых является помощь в выборе наиболее подходящих ключевых слов. Чтобы выявить близкие по значению выражения, функционирующие в информационно-поисковых

системах Интернета, можно использовать специальные сервисы поисковых систем Rambler, Яндекс, Google. Эти системы представляет статистику поисковых запросов и работает по принципу «Те, кто ищут (запрос пользователя), ищут также...» Сервис предназначен в помощь пользователю, совершившему поиск по какому-либо запросу и не получившему нужной информации. Статистика «ассоциаций» помогает выяснить интересы пользователей, путем анализа списка запросов, схожих с основными. Статистика анализа запросов Яндекса на ключевое слов русский язык, которое на Яндексе в 2008 г. имело 95641 показ, показывает, что русский язык встречалось также в других словосочетаниях:

	Показов в месяц
Русский язык	95 641
ЕГЭ русский язык	9 467
Словарь русского языка	7 988
Скачать русский язык	5 750
Русский язык, тесты	5 261
Правила русского языка	3 810
Русский язык: результаты	2 999
Русский язык: результаты ЕГЭ	2 677
Уроки русского языка	1 510
Русский язык 2008	1 436
Толковый словарь русского языка	1 282
Курс русского языка	1 263
ЕГЭ русский язык 2008	1 147
Учебник русского языка	1 129
Фотошоп русский язык	1 090

Таким образом, предварительный анализ показывает, что ключевые слова вступают в Интернете друг с другом в особые отношения, на основе гипертекстовых связей, тематической связи слов, разного рода нового типа ассоциаций. Возникают, иногда непривычные с точки зрения классической лексикологии семантические связи слов. Например, к ключевому слову *виза* интернет-системой будут предложены названия фирм, адреса посольств, путеводители, карты дорог, адреса гостиниц и др. Ключевые слова выполняют различные функции: помогают найти текст с соответствующим содержанием; повышают посещаемость (тем самым и рейтинг того или иного сайта), если в текст включаются наиболее частотные или востребованные в данный момент ключевые слова.

ЛИТЕРАТУРА

- [1] *Атабекова А.А.* Лингвистический дизайн WEB-страниц (сопоставительный анализ языкового оформления англо- и русскоязычных WEB-страниц). — М.: Изд-во РУДН, 2003.
- [2] *Дедова О.В.* Теория гипертекста и гипертекстовые практики в Рунете. — М.: Изд-во МГУ, 2008.
- [3] *Какорина Е.В.* Язык Интернет-коммуникации // Язык массовой и межличностной коммуникации. — М.: Медиатека, 2007.
- [4] *Трофимова Г.Н.* Языковой вкус интернет-эпохи в России: Функционирование русского языка в Интернете: концептуально-сущностные доминанты. — М.: Изд-во РУДН, 2004.

**RUSSIAN KEY WORDS
IN THE INTERNET SEARCH ENGINES
(semantically and cultural approaches)**

P.V. Moroslin

International Slavonic Institute
Godovikova str., 9, Moscow, Russia, 129085

The research dealt with both linguistic and cultural approaches in considering Russian key words in the Internet search engines. In the article systems of collecting and saving key words were considered as well as a process and results for a new type of semantically and associative relations among key words in the Internet analyzed.

Key words: the Internet, the system of search, computer communication, the keywords.