

Чайка Е.В.

студент

*Научный руководитель: Гурова Е.А., старший преподаватель
Армавирский Государственный Педагогический Университет*

**ДОКУМЕНТАЛЬНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ В
ИНТЕРНЕТЕ.**

Аннотация: Сегодня существует множество поисковых систем для поиска необходимой информации в интернете. Современные информационные системы в интернете учитывают особенности работы в интернет-пространстве, успешно используют практически неограниченные ресурсы хранения серверов, а также высокую производительность компьютера и современные достижения в области управления знаниями.

Ключевые слова: информационно-поисковая система / information retrieval system / поисковый робот / search robot / google / релевантность / relevance

Chaika E. V.

student

*Supervisor: E. A. Gurova, Senior Lecturer
Armavir State Pedagogical University*

DOCUMENTARY INFORMATION SYSTEMS ON THE INTERNET.

Annotation: Today, there are many search engines to find the necessary information on the Internet. Modern information systems on the Internet take into account the peculiarities of working in the Internet space, successfully use almost unlimited server storage resources, as well as high computer performance and modern achievements in the field of knowledge management.

Keywords: information retrieval system / information retrieval system / search robot / google / relevance / relevance

Согласно Net Applications в ноябре 2011 года, одним из наиболее распространенных инструментов поиска является Google, который дает пользователям самые полезные результаты.

Давайте посмотрим на пример Google о том, как искать и получать информацию. Материал, представленный ниже, основан на информации о ресурсах.

Сбор информации: сканирование

Google использует поисковые роботы, такие как Googlebot, для поиска общедоступных страниц. Роботы просматривают веб-страницы и нажимают на ссылки, которые они находят, так же, как пользователи интернета. В этом случае роботы отправляют данные на сервер Google через веб-страницы.

Сканирование начинается со списка веб-адресов из предыдущих сканирований и Sitemaps (Sitemaps) владельцев сайтов. При сканировании этих страниц для поисковых систем искать ссылки на другие страницы. Особое внимание уделяется новым страницам, изменениям существующих страниц и дефектным ссылкам. Google автоматически определяет, какие страницы сканировать, как часто и в какой степени. Однако большинству страниц не нужны специальные настройки для сканирования, индексирования и просмотра, поэтому ваши страницы будут отображаться в результатах поиска без предварительной подготовки.

Систематизация информации: индекс.

Чтобы пользователи могли найти необходимую информацию в огромной и постоянно растущей библиотеке интернета, Google сканирует все страницы и создает индекс, похожий на индекс в конце книги. Google index включает более чем 100.000.000 Gigabyte и потребовалось более миллиона часов компьютерных расчетов для компиляции.

В качестве поискового изображения документа компилируется словарь с частотой используемых слов и фраз. Подходы к управлению знаниями также используются при использовании мета-описаний и с учетом логических отношений между понятиями.

На первом этапе поиска для поиска слов запроса в индексе используются специальные алгоритмы для определения нужных страниц. Алгоритмы-это автоматические процессы и формулы, которые преобразуют пользовательские вопросы в ответы. Как правило, запрос соответствует тысячам, а иногда и миллионам веб-страниц. Современные алгоритмы Google используют более 200 различных сигналов или "ключей", чтобы понять, что ищет человек. Система учитывает такие параметры, как наличие слов на страницах сайта, релевантность информации, местоположение пользователя и индикатор ранжирования страницы (PageRank).

После этого процесс усложняется. Если вы ищете "собак", нет смысла показывать страницы, состоящие исключительно из этого слова.

Скорее всего, полезными результатами будут фотографии, видео или списки пород. Таким образом, системы индексации Google учитывают различные параметры страниц: дату публикации, наличие изображений или видео и многое другое. Сеть знаний-огромная база данных фактов, содержащая информацию о людях, местах, объектах и отношениях между ними-это еще один шаг на пути от простого сравнения ключевых слов к полному пониманию реальных интересов пользователя.

Вот некоторые типы информации, которые можно найти в интернете знать: описания и факты, которые являются общедоступными в сети Интернет; картинки из Интернета, для запросов на релевантные; аналогичные запросы, с помощью которых вы можете получить дополнительную информацию (например, при поиске Эйфелевой башни могут появиться другие достопримечательности Парижа); положение на карте или другие сведения для запроса (например, путем ввода имени известного исполнителя, вы можете получить план концерта и съемок его канал).

Поисковая система анализирует самые популярные результаты поиска и содержимое соответствующих веб-страниц. Если большинство страниц содержат такую информацию, то она дается в краткой, обобщенной форме. В

некоторых случаях вам может быть предложено выбрать один из нескольких вариантов, чтобы сузить свой поиск (например, если у вас есть книга под названием Золушка, а не фильм с тем же названием).

Основные Функции Поиска

Чтобы всегда получать необходимую информацию, вы должны следовать следующим рекомендациям по запросам в любой поисковой системе, включая Google.

1. Запросы должны быть простыми. Если вам нужна конкретная концепция, место или продукт, просто введите его название.

2. Добавьте уместные слова, когда простой запрос возвращает правильные

Результат'. Это может занять некоторое время, чтобы найти правильные слова для вашего поискового запроса.

3. Используйте слова, которые могут быть в поиске присутствует
Страница.

4. Используйте только значимые слова, вместо полной фразы или вопросы. Поскольку все слова в запросе используются для поиска, каждое последующее слово ограничивает диапазон результатов.

5. Используйте специальные функции. Для некоторых типов поиска необходимая информация отображается непосредственно под окном поиска.

Некоторые исключения из вышеперечисленных правил позволяют Google предоставлять пользователям наиболее релевантные результаты:

1. Часто используемые слова, такие как предлоги, статьи и союзы игнорируются, как правило. Но это исключение также имеет ограничение.

В запросе "а и Б сидели на трубе", буква "и", вероятно, является частью головоломки, и Google будет искать всю фразу с этой ассоциацией. А в запросе "с буквами Е и Е" буква "и" имеет гораздо меньшее значение, и в этом случае Google может игнорировать Союз.

2. Не все слова из поискового запроса могут быть на странице в результатах поиска присутствует. Например, если вы введете "вид сверху" в

поле поиска, вы можете увидеть красивые виды пулов сверху на веб-страницах, которые не содержат слова "сверху".

3. Синонимы могут заменять некоторые слова в исходном запросе.

Использованные источники:

1. Баширова М.М. Учебное пособие по дисциплине «Документальные информационные системы» Для направления подготовки 230700 «Прикладная информатика», Профиля подготовки «Прикладная информатика в образовании» – Махачкала: ДГИНХ, 2011.. – Махачкала: ДГИНХ, 2011 – 120 с.

2. Гагин А.В. Сервисы Интернет [Электронный ресурс]// Jet Info, №7 (Дата публикации: 1996). – URL: <https://www.computer-museum.ru/technlgy/ji96071.htm> (Дата обращения: 12.06.2021)

3. Информационные системы и технологии: учебное пособие / И.Л. Чудинов, В.В. Осипова; Томский политехнический университет. – Томск: Изд-во Томского политехнического университета, 2013. – 145 с.