

ХАРАКТЕРИСТИКА ПОИСКОВЫХ СИСТЕМ, ПРОБЛЕМЫ И ИХ ВОЗМОЖНОСТИ

Шарипова Биржан Дабаевна, к.э.н., доцент, Алматинский технологический университет, Казахстан, г. Алматы, мкр. Мамыр-1. Тел: 8 702 201 01 49, e-mail: Birjan2103@mail.ru

Муканова Аксулу, к.т.н., доцент, Алматинский технологический университет, Казахстан, г. Алматы.

Абдылдаев Эркинбек Кыянович, доктор технических наук, профессор, АО «Университет Нархоз», Казахстан, г. Алматы

Аннотация. Характеристика поисковых систем, заключается в том, что разработка алгоритма автоматического поиска научной и учебной литературы в сети Интернет и составление списка URL-ссылок на найденные документы. Для выполнения этой работы был выбран локальный сервер DENWER- Джентльменский набор web-разработчика,

включающий в себя язык программирования PHP, сервер APACHE, система управления базами данных MYSQL, каскадные таблицы стилей CSS, язык гипертекстовой разметки HTML. Было достигнуто сокращение время для поиска URL-ссылок в сети Интернет. Основные характеристики работы, возможность индексаций практически любого веб-сайта для скачивания с него ссылок на нужную литературу, возможность указания плохих ссылок, аутентификация пользователя.

Ключевые слова: поисковая система, программа-робот, программирование, denwer, php, html, spider, crawler, indexer, аутентификация, url.

CHARACTERISTICS OF SEARCH SYSTEMS, PROBLEMS AND THEIR OPPORTUNITIES

Sharipova Birzhan Dabayevna, PhD Econ., associate professor, Almaty technological university, Kazakhstan, Almaty, µR. Mamyр-1. Ph.

Mukanova Aksulu, PhD in Technological Sciences, associate professor, Almaty technological university, Kazakhstan, Almaty.

Erkinbek Abdylдаev, prof., doctor of technical sciences, JSC «University Narxoz», Kazakhstan, Almaty

Abstract. The characteristic of search engines is that development of an algorithm of automatic search of scientific and educational literature on a network the Internet and compilation of spiskaurl-links to the found documents. For execution of this operation the local DENWER-server the Gentlemen's set of the web developer including the PHP programming language, the APACHE server, MYSQL database management system, CSS cascading style sheets, language of a hypertext marking of HTML was selected. Abbreviation time for search of URL links on the Internet was reached. The main characteristics of operation, a possibility of indexations practically any web site for downloading from it of links to the necessary literature, a possibility of specifying of bad links, authentication of the user.

Keywords: search engine, programma-robot, programmirovanie, denwer, php, html, spider, crawler, indexer, autentifikatsiya, url.

Каждый из нас ежедневно пользуется интернетом. Но с обилием информации достаточно трудно найти искомое. Именно поэтому и были созданы поисковые системы. Каждый предпочитает тот или иной поисковик, но никто не задумывается, а какой самый популярный. Ниже приведён список популярности поисковых систем.

Основные мировые поисковые системы :

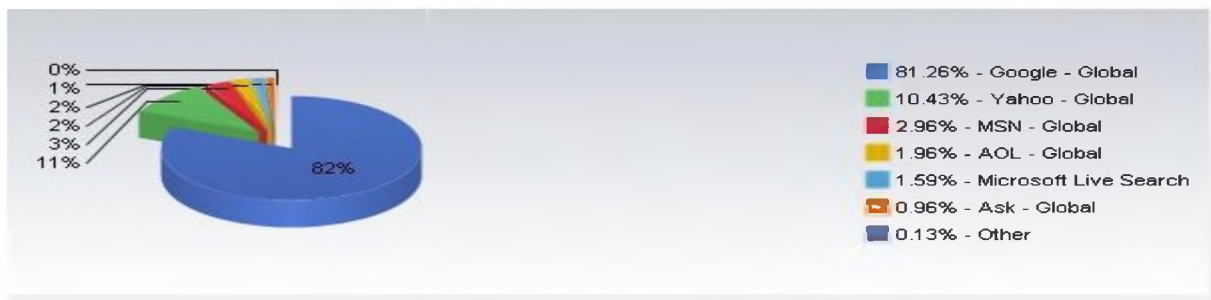


Рисунок 1-Мировой рейтинг поисковых систем

Начиная поиск чего-либо в Internet и имея минимум информации, а также пытаюсь ограничить потери времени, для получения наиболее общей информации возможно обращение к следующей базе данных [1,27 с].

База данных: предмет ведет к Межсетевым ресурсам, построенным библиотекарями.

Содержание: Свыше 2,500 связей с Internet ресурсами, кратко аннотируемыми, с указанными предметными заголовками.

Поиск: Поиски могут быть ограничены названием ресурса, описанием его, или с указанными предметными заголовками.

Результаты: Результаты показаны в алфавитном порядке названиями ресурса.

Адрес: <http://sunsite.berkeley.edu/InternetIndex/>

Частота Модернизации: ежедневно

YAHOO!

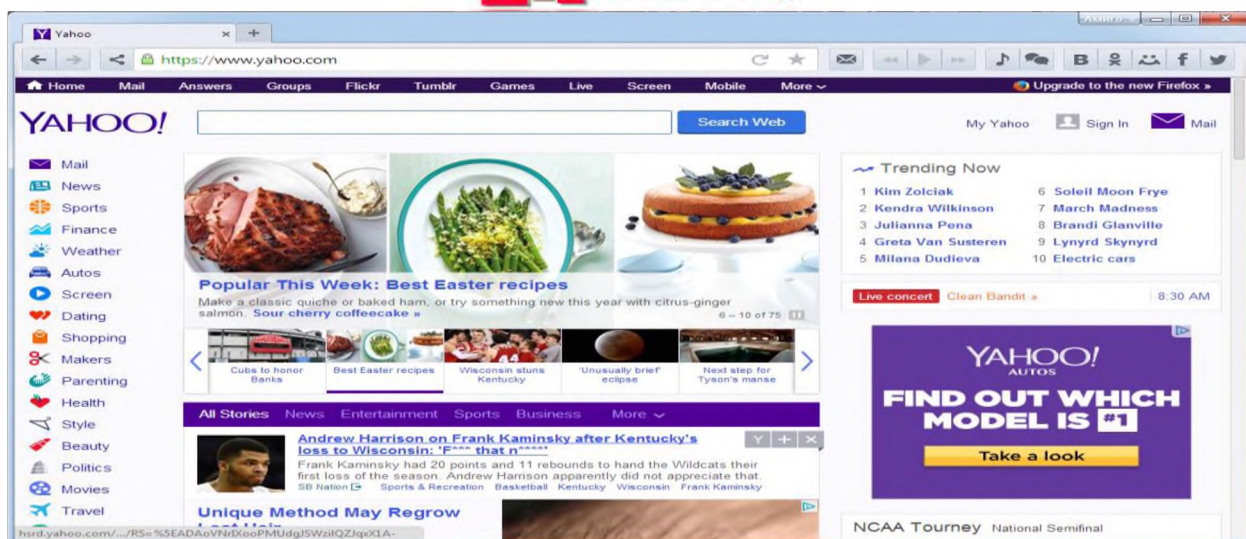


Рисунок 2 - Поисковик YAHOO

Yahoo! – самая известная поисковая машина. Её сайты разбиты по категориям и ключевым словам. Она содержит полезную информацию на своей домашней странице. Может подключаться к другим поисковым машинам

Поиск: Все Yahoo страницы предлагают не только простое поисковое окно, но и опции для этого поиска, а так же поиск Usenet или Email-адреса. Поиск может ограничиваться указанием определённого промежутка времени. Boolean операторы (и, или) и последовательный поиск также поддерживаются. Отметим: если поиск в Yahoo! не привёл к положительному результату, то процесс поиска автоматически переходит на Alta Vista, которая продолжает поиск, и в случае положительных результатов автоматически возвращает найденную информацию в Yahoo!.

Средством, облегчающим поиск, является наличие “tip search”(TS) - поиск с помощью “намёка”: Yahoo! Является подчиненным справочником, что означает, что система не имеет так много страниц, как поисковые машины, однако задание наиболее общих ключевых слов позволит найти необходимую тему на странице высокого уровня (первая страница, которая возникает перед пользователем при посещении сайта) для организации или компании [2,91 с].

Результаты: Связи отображаются в соответствии с очередностью задаваемых слов последовательностью поиска наряду с их описательным текстом и подчиненной иерархией.

Адрес: <http://www.yahoo.com/>

Частота Модернизации: ежедневно

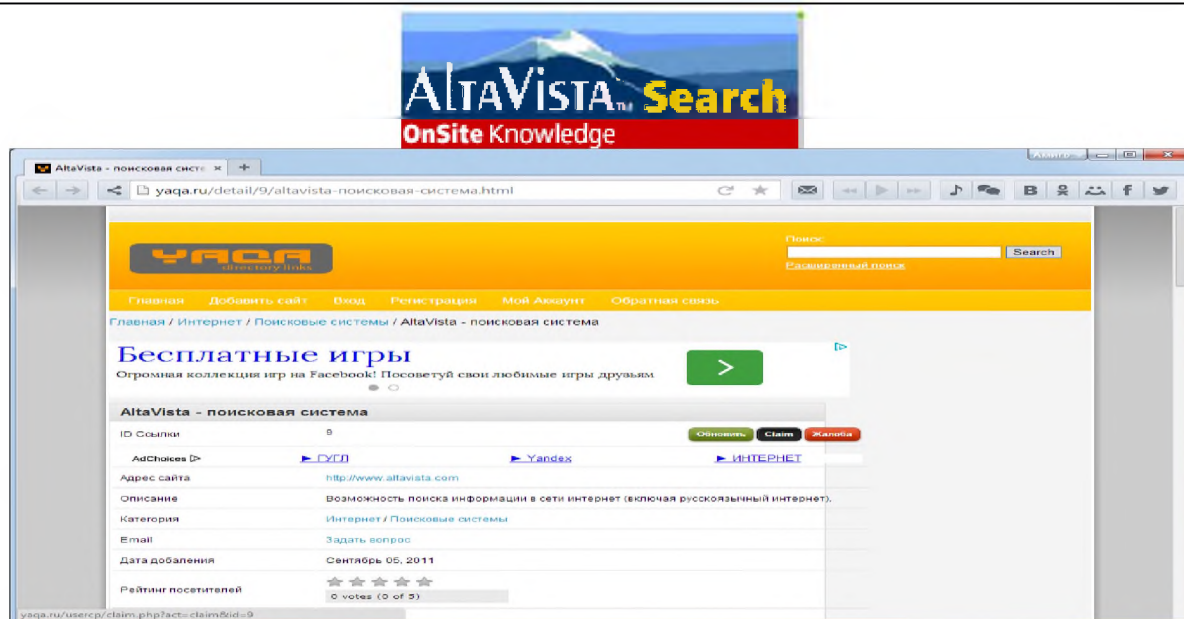


Рисунок 4 - Поисковая система AltaVista

Alta Vista поддерживает поиск по ключевому набору слов и для определения языка конкретной страницы использует методы искусственного интеллекта. Пользователи могут настроить опции поиска и выбирать тип поиска – сложный или упрощенный, а также воспользоваться различными способами предоставления информации. В отличие от машин, которые индексируют только ключевые слова, она индексирует весь текст, что позволяет осуществлять полный поиск. Однако из-за этого пользователь может просто утонуть в информации.

Базы данных: Расположенные по всему миру WWW-страницы и Usenet News (новости).

Содержание: 31 миллион WWW-страниц (на май 1997 г.) и полный текст более чем 14,000 newsgroups обновленный в реальном масштабе времени.

Поиск: Предлагает простой (simple (S)) поиск или (much more advanced (MMS)), т.е. более передовой, способ. S - поиск стоит в основном использовать для общих вопросов, MMS - поиск использует специфический поисковый синтаксис. Для облегчения выполнения процедуры имеется подсказка (Simple Search Help). MMS - поиск, используя булевы (boolean), т.е. с помощью ключевых союзов, используя (and, or, not - (и, или, не)) и простую смежность (near - (около)) позволяет употреблять несколько слов, чередование слов, словосочетание в качестве ключевых для проведения поиска [3,37 с].

TS - поиск: Введением ключа типа: " Ваша Фраза " как первое направление поиска, который будет ограничивать число найденных WWW - документов с заголовками типа " Ваша Фраза ".

Результаты: Предлагает три выбора результатов (но два дают тот же самый результат):

1) "Стандартные" ("Standard") - результаты, полученные машиной в виде списка параграфов, резюмируемые, с наличием URL - адреса, размером файла и последней датой модернизации. Результаты возвращаются как десять пунктов на экране,

2) "Компактный" ("Compact") помещают каждый пункт в одной строке с последней датой модернизации картотеки,

3) "Детальный" ("Detailed"), который является таким же самым, как и "Стандартный".

Адрес: <http://altavista.digital.com>

Частота модернизации: Постоянно WWW-роботом.



Рисунок 5 –Поисковая система Excite

Для анализа информации Excite использует поисковую технологию Intelligent Concept Extraction, что позволяет делать запросы по образцу. Это самая популярная поисковая система в Америке. Для каждой найденной страницы она оценивает степень соответствия запросу.

Базы данных: WWW-страницы по всему миру, новости, карты, "yellow pages" ("желтые страницы"), свободно распространяемое программное обеспечение, основные цитаты, программы телевидения, погоду, E-mail адреса, рейсы авиа линий.

electronic) AND library. (цифровой или виртуальный или электронный) И библиотека.

Результаты: Результаты показаны с названием документа, разряд уместности в процентах, URL - адрес, резюме программного обеспечением документа, и опция, чтобы восстановить "More Like This" ("Скорее Этот"), которая позволяет использовать документ как ваш вопрос [4].

Адрес: <http://www.excite.com/>

Частота Модернизации: Постоянно - WWW-роботом.

Для поиска в Интернете использует многопроцессорную параллельную обработку 10. млн. страниц ежедневно. Полезная сторона Hot Bot ограничение на тип страниц по средствам выбора кнопок.

База данных: Расположенные по всему миру WWW-страницы.

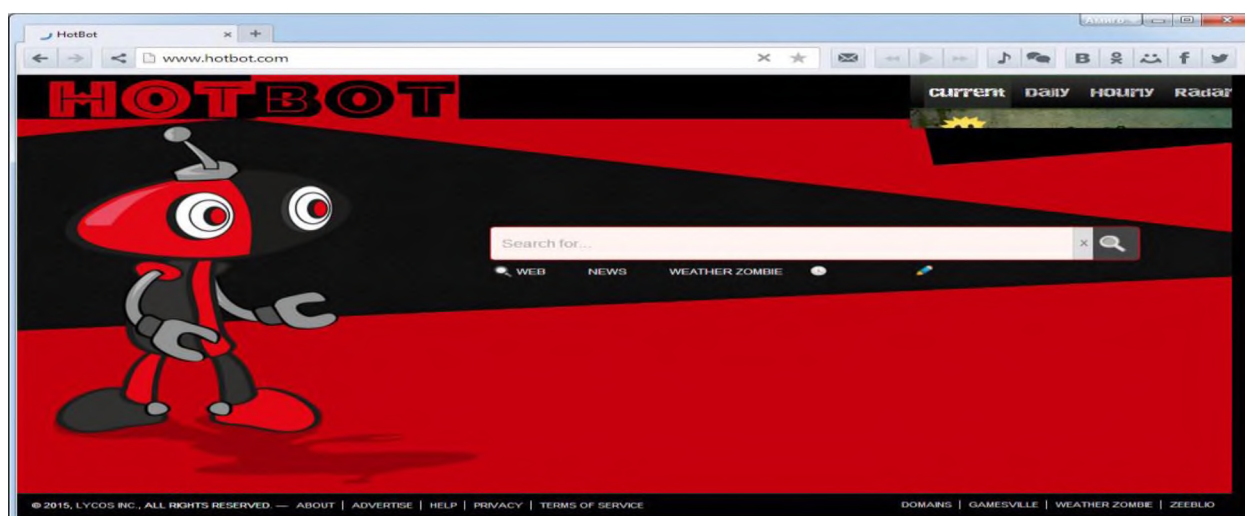


Рисунок 6 - Поисковая система HotBot

Содержание: 54 миллиона WWW-страниц (на сентябрь 1996 г.).

Поиск: Предлагает S - поиск и Эксперт(Expert (E))- поиск, поддерживает boolean-операторы (И и ИЛИ), поиск фразы, и выбор " человек "или" URL ". E-поиск также поддерживает задание даты, местоположение (страна и т.д.)

TS - поиск: использует заключение фразы в двойные кавычки (например, " слова фразы ").

Результаты: Результаты показаны с названием документа, разряд уместности в проценте, URL, размер документа.

Адрес: <http://www.hotbot.com/>

Частота модернизации: Постоянно WWW-роботом ("Slurp").



Базы данных: расположенные по всему миру WWW-страницы, новости, запасают цитаты, карты, желтые страницы("yellow pages"), e mail адреса, и т.д.

Содержание: Главный база данных: 50 миллионов URL-адресов. (на сентябрь 1996 г.)

Результаты: Включает название документа, размер картотеки, URL, краткое резюме, извлеченное из документа, и разряд уместности в процентах.

Адрес: <http://www.infoseek.com/>

Частота Модернизации: Постоянно WWW-роботом.

Дополнительная информация: в случае большого количества информации см. <http://info.infoseek.com/>.

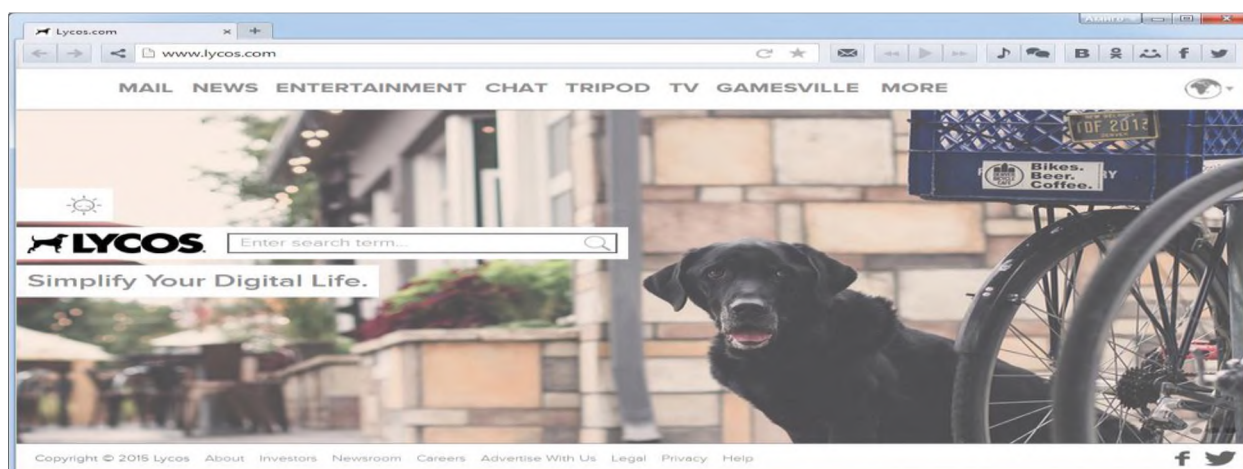


Рисунок 7 - Поисковая система

Lycos – одна из первых поисковых машин. Машина удобна для работы с поиском и для одновременного просмотра сайтов. При выводе информации показывает краткий обзор, и найденные адреса.

Базы данных: расположенные по всему миру WWW-страницы, звуки, картины, "top 5% sites"

Поиск: предлагает S - поиск и клиентурный(Custom (C)) поиск. C-поиск поддерживает boolean-операторы AND и OR (И и ИЛИ), также как некоторые другие назначения.

Результаты: результаты внесены в упорядоченный список; информация включает адрес документа (URL), название, размер файла, и выдержки из файла.

Адрес: <http://www.lycos.com/>

Частота модернизации: постоянно WWW-роботом.

ПРОБЛЕМЫ И ВОЗМОЖНОСТИ ПОИСКОВЫХ СИСТЕМ

Интеллектуальные агенты выполняют инструкции от имени пользователя, имеют некоторую самостоятельность. После поиска они оповещают пользователя о результатах. Агенты учатся в результате своей деятельности.

Интеллектуальность – обучение на основе обратной связи по примерам ошибками посредством взаимодействия с другими агентами.

Простота использования – можно тренировать агента используя естественный язык.

Индивидуальный подход – адаптация к предпочтениям пользователей.

Интегрированность – непрерывное обучение применение уже имеющихся знаний к новым ситуациям.

Автономность – ощущение окружающей среды, и анализ выводов.

Список литературы

1. Компьютерные технологии в жизненном цикле изделия: Учебное пособие/ Яблочников Е.И., Фомина Ю.Н., Саломатина А.А. СПб: СПбГУ ИТМО, 2010.
2. Возможности поисковых систем Google и Yandex / Н.Г. Соколов, Москва-2014.
3. Поисковые системы и продвижение сайтов в Интернете/Колисниченко Д.Н, 2007, 272-с.
4. www.citforum.ru - CIT forum, Поисковые системы в сети Интернет