

Задача о языке запросов

Т.Ю. Селихова

Автор: Селихова Т.Ю., учитель информатики и ИКТ средней школы № 4 села Монастырище Черниговского района Приморского края.

Предмет: Информатика и ИКТ.

Класс: 10.

Тема: Поиск информации в Интернете.

Профиль: Общеобразовательный.

Уровень: Профильный.

Текст задачи. В настоящее время персональные пользователи осуществляют поиск информации в глобальной сети с помощью поисковых систем. При этом одни отдают предпочтение Google, другие Яндекс, третьи выбирают Yahoo или Рамблер. Между тем, основ-

ная масса пользователей не видит принципиального различия в оформлении запроса в той или иной поисковой системе.

Можно ли утверждать, что язык запроса не зависит от используемой поисковой системы? Существует ли необходимость в знании языка запросов обычным пользователем?

а) *Выделите ключевые слова для информационного поиска.*

б) *Найдите и соберите необходимую информацию.*

в) *Обсудите и проанализируйте собранную информацию.*

г) *Сделайте выводы.*

д) *Сравните ваши выводы с культурным образцом.*

Возможные информационные источники

Web-сайты:

http://ru.wikipedia.org/wiki/%D0%AF%D0%B7%D1%8B%D0%BA_%D0%B7%D0%B0%D0%BF%D1%80%D0%BE%D1%81%D0%BE%D0%B2

<http://www.google.ru/intl/ru/help/refinesearch.html>

<http://www.runetica.com/profiles/google/cheatsheet.html>

<http://help.yandex.ru/search/?id=481939>

<http://help.rambler.ru/1/rsearch/#1>

<http://www.afort.ru/help/?p=3>

http://www.shtogrin.com/library/search_engines/advansed_search/yahoo/

Культурный образец

Российская национальная библиотека <http://www.nlr.ru>

РЕСУРСЫ

Поисковая система Google

http://www.nlr.ru/res/inv/ic_old/google.htm

Google автоматически вставляет между ключевыми словами оператор «AND», поэтому в списке найденных документов присутствуют только те, которые содержат все заданные слова. В начале списка находятся документы, в которых ключевые слова расположены рядом.

Поддерживается поиск по фразе. Фраза заключается в кавычки.

Google ищет документы с точными соответствиями заданных слов. Поэтому поиск документов на слово «hotel» и «hotels» даст разные результаты. То же правило относится и к словам, заданным во фразе.

Поддерживает поиск по части ключевого слова.

Google не делает различия между строчными и прописными буквами.

Чтобы исключить документы, содержащие какое-то слово, нужно в запросе поставить перед этим словом знак «-» (без пробела). Возможна постановка знака также и перед фразой.

Google не поддерживает логический оператор «OR». Нельзя задать поиск документов, в которых содержалось бы или первое, или второе слово. Вместо этого предлагается сделать несколько запросов, варьируя ключевые слова.

Помимо кавычек Google учитывает следующие знаки препинания, служащие для связи слов: дефисы (mother-in-law), косые черты, знаки равенства, апострофы (Bill's birthday). При поиске слова, связанные

этимися знаками, воспринимаются как фраза.

При поиске Google игнорирует некоторые слова и знаки, т.н. стоп-слова. Сюда входят одиночные цифры и буквы, а также слова типа http, the, to и проч. В тех случаях, когда учёт стоп-слов является необходимым для поиска, перед ними ставится знак «+».

Имеется возможность поиска документов на определённом языке.

Поисковая система Yahoo!

http://www.nlr.ru/res/inv/ic_old/yahoo.htm

База данных Yahoo! пополняется вручную отобранными ссылками на web-сайты. Каждая ссылка состоит из заголовка html-документа и краткой аннотации. Поиск в каталоге Yahoo! можно осуществлять как по разделам, рубриками и подрубрикам, так и с помощью ключевых слов.

Кроме поиска в разделах и рубриках каталога, предлагается поиск по ключевым словам. Поиск осуществляется сразу в пяти областях, на которые разбита база данных Yahoo! (Yahoo! Categories, Yahoo! Web Sites, Web Pages, News Stories, Yahoo! Net Events).

По умолчанию используется оператор OR. Документы, содержащие все ключевые слова, помещаются в начало списка. При прочих равных условиях выше ранжируются документы, которые содержат ключевые слова в заголовке. Поддерживается поиск по фразе. Фраза заключается в кавычки. «-» — исклю-

чить документы, содержащие это слово. Возможна постановка этих знаков также и перед фразой. Yahoo поддерживает дополнительные поисковые операторы:

t: — поиск ключевых слов только в заголовке html-документа;

u: — поиск ключевых слов только в URL-адресе html-документа.

Yahoo поддерживает поиск по начальной части ключевого слова с помощью знака «*», поставленного после слова без пробела.

Поисковая система Рамблер

http://www.nlr.ru/res/inv/ic_old/rambler.htm

Поисковая система Rambler содержит миллионы документов с более чем 42 тысяч сайтов (имен DNS) России и стран СНГ. Имеет развитый язык запросов и гибкую форму вывода результатов.

Запрос может состоять из одного или нескольких слов, разделённых пробелами. Могут быть использованы как русские, так и английские слова и словосочетания.

По умолчанию находятся только те документы, в которых встретились все введенные слова, т.е. используется логический оператор AND.

Строчные и прописные буквы не различаются.

Поддерживается поиск по фразе. Фраза заключается в кавычки.

Поддерживается поиск с использованием булевых операторов And («и»), Or («или»). Вместо них (или в сочетании с ними) могут использоваться также символы '&', '|'.

Части запроса могут быть сгруппированы с помощью круглых скобок (). Например: (философия and культурология) or семиотика.

Рамблер умеет искать слова во всех формах (например, аминокислота, аминокислоты, аминокислотой и т.д.). Чтобы слово находилось во всех формах, перед ним надо поставить служебный символ '#'.

По умолчанию Рамблер ищет слова запроса так, как они введены, чтобы уменьшить «шум» в найденных документах.

'@' перед словом позволяет находить не только само это слово, но и однокоренные слова. Пример: #аминокислота and @цинк.

Поддерживаются символы «*» (для обозначения произвольной части слова) и «?» (для обозначения любого символа слова). Например: к?мпания= компания или компания.

Поисковая система Апорт

http://www.nlr.ru/res/inv/ic_old/aport.htm

Апорт — уникальная двуязычная поисковая машина. Имея те же основные черты, что и AltaVista, Апорт содержит намного больше современных ресурсов российского Интернета. Также Апорт входит в число ведущих поисковых систем, сертифицированных Microsoft как локальные поисковые системы для русской версии Microsoft Internet Explorer.

Одним из существенных преимуществ Апорта является англо-русский и русско-английский on-line перевод запросов и поисков резуль-

РЕСУРСЫ

тата, автоматическая проверка орфографических ошибок запроса, более информативный вывод результатов поиска для найденных сайтов (не только первое предложение!), возможность поиска в любой грамматической форме (что особенно важно для русского языка), поддержка пяти основных кодовых страниц (разных операционных систем) для русского языка.

По умолчанию используется оператор AND, т.е. релевантными считаются документы, содержащие все слова запроса. Поддерживается поиск с помощью логических операторов AND (варианты: И, &, +), OR (ИЛИ, |), NOT (НЕ, -). Круглые скобки задают порядок действия логических операторов. Поддерживается поиск по фразе. Фраза заключается в кавычки. Поддерживается поиск по начальной части ключевого слова с помощью символа «*», например: семиоти* = семиотика, семиотический.

Поисковая система Яндекс

http://www.nlr.ru/res/inv/ic_old/yandex.htm

Официально поисковая машина Yandex.ru была анонсирована 23 сентября 1997 года. На сегодняшний день Yandex имеет самую большую в русской сети поисковую базу (проиндексировано 20 млн документов объемом более 200 Гб).

Шестого июня 2000 г. компания «Яндекс» на сервере www.yandex.ru представляет новую версию своей поисковой машины. Теперь поиск осуществляется не только по веб-

страницам, но и по специализированным массивам данных.

Отличительные особенности Yandex — поиск на естественном языке, большая и актуальная база данных страниц русского Интернета, интеграция машины поиска с интернет-каталогом, изучение запросов пользователей.

Поиск происходит с учётом морфологии русского языка, например: «идёт»= идёт, идут, идти, шёл, шла и т.д. Если необходимо осуществить поиск по точной словоформе, то перед ним ставится знак «!» (без пробела).

Различаются слова, набранные строчными и прописными буквами.

По умолчанию ищутся документы, содержащие хотя бы одно слово запроса.

«+» делает наличие ключевого слова в документе обязательным.

«-» перед словом исключает документы, которые его содержат.

Поддерживается поиск по фразе. Фраза заключается в кавычки.

Яндекс поддерживает собственную систему обозначений булевых операторов.

Методический комментарий

Задача относится к профильному уровню, поскольку предполагает более глубокое изучение основ языка запроса нескольких поисковых систем и их сравнение. Для представления решения задачи необходимо не только найти и проанализировать теоретический материал, но и проверить на практике эффективность поиска с помощью языка запросов.

Синтаксис	Что означает оператор	Результат поиска
пробел или &	логическое И (в пределах предложения)	лечебная физкультура
&&	логическое И (в пределах документа)	рецепты && (плавленный сыр)
	логическое ИЛИ	фото фотография снимок
()	группирование слов	(технология изготовление) (сыра творога)
~	бинарный оператор И НЕ (в пределах предложения)	банки ~ закон
~~	бинарный оператор И НЕ (в пределах документа)	путеводитель по Парижу ~~ (агентство тур)
/(n m)	расстояние в словах (-назад +вперед)	поставщики /2 кофе музыкальное / (-2 4) образование вакансии ~ /+1 студентов
&&/(n m)	расстояние в предложениях (-назад +вперед)	банк && /1 налоги

Культурный образец не содержит прямого ответа на поставленный вопрос, а лишь демонстрирует основы языка запросов самых популярных поисковых систем.

Ключевыми являются фразы: «поисковая система», «язык запросов», «запросы на Google», «запросы на Yahoo», «запросы на Рамблер», «запросы на Яндекс», «запросы на Апорт».