

3.6. Информационно-поисковые системы

Понятие информационно-поисковой системы. Виды поисковых средств в Интернете. Характеристика поисковой системы Интернета. Информационно-поисковый язык.

В современном мире, который буквально пронизан постоянно нарастающими объемами информации, для человека, использующего эту информацию с целью ее превращения в знания, встает проблема ориентации. Чтобы не захлебнуться в информационном потоке, нам нужны техники отбора, фильтрации и оценки [10, 18].

Традиционными способами фильтрации и отбора информации человеком являются:

- поиск «сверху» (по оглавлению);
- поиск «снизу» (с помощью различных указателей);
- поиск с помощью гипертекстовых связей (перекрестных ссылок);
- полнотекстовый поиск путем просмотра всего текста [6, 70].

Последний вид поиска является наиболее точным, но и наиболее трудоемким, требующим больше всего времени и усилий.

Организация поиска предполагает следующие составляющие и этапы:

- 1) множество документов (текстов или их фрагментов), по которым следует производить поиск;
- 2) коммуникативная потребность в информации, выражающаяся в информационном запросе пользователя;
- 3) удовлетворение коммуникативной потребности, состоящее в выборе той части текстов исходного массива, которая соответствует информационному запросу [4, 197].

Упорядоченная совокупность документов и информационных технологий, предназначенных для хранения и поиска информации, представленной в виде текстов или их частей (фактов), получила название *информационно-поисковой системы* (ИПС) [19, 3].

Для экономии усилий человека с 1950-х годов осуществляются попытки создания автоматизированных ИПС. При этом в первых ИПС анализ и описание содержания документов (индексирование) выполнялись вручную, а поиски по этим документам проводились автоматически [19, 8].

Сегодня с развитием компьютерной техники и созданием высокоскоростных телекоммуникационных сетей в деле автоматизации поиска достигнуты значительные успехи, кратко и емко выразившиеся в знаменитой формуле Б. Гейтса «информация на кончиках пальцев» (*information at your fingertips*) [цит. по: 10, 16]. Данное выражение можно понимать таким образом: информация всегда находится в распоряжении человека, нужно лишь сделать несколько нажатий клавиш, чтобы получить доступ к ней.

Так, для поиска информации в Интернете служат различные классы поисковых средств [6, 71]:

- каталоги (*directories*);
- подборки ссылок (*bookmarks*);
- поисковые машины (*search engines*);
- базы данных адресов электронной почты и т.д.

Каждый вид поискового средства имеет свои особенности, так, если человек имеет недостаточно точное представление о цели поиска, ему целесообразнее использовать каталоги веб-ресурсов. Применение поисковых машин эффективно, если пользователь представляет, какие ключевые слова характеризуют нужные ему ресурсы.

Каталог веб-ресурсов — это постоянно обновляемая и пополняемая система ссылок на ресурсы, распределенные по иерархической структуре категорий. На верхнем уровне каталога представлены самые общие категории (рубрики), например «наука», «бизнес», «развлечения» и т.д. На нижележащих уровнях рубрики имеют более частный характер [6, 71]. Например, рубрика «наука» может делиться на категории «точные науки», «естественные науки» и «гуманитарные науки», последние — на философию, социологию, психологию, педагогику и т.д. Русскоязычный каталог сайтов можно найти, например, по адресу www.ru.

Коллекция ссылок представляет собой еще один способ организации информации во Всемирной паутине. Такая коллекция обычно составляется специалистом в определенной теме, постоянно обновляется и не содержит ненужной информации. Печатный аналог такой коллекции ссылок по использованию информационных технологий в лингвистике можно найти после библиографического списка в нашем пособии. Некоторые примеры коллекций ссылок по обучению английскому языку приводит С.В. Титова [45, 27—28].

Поисковые машины (или поисковые системы) — это специальные веб-страницы, позволяющие находить веб-ресурсы, текстовое содержание которых соответствует запросу пользователя. В Международном каталоге поисковых машин (www.searchenginecolossus.com) зарегистрировано свыше 2300 поисковых систем из 232 стран. По данным этого каталога, каждый день выполняется до 450 млн поисковых запросов [6, 72; 38, 364].

К наиболее известным поисковым машинам относятся [6, 72—73]:

- AltaVista (www.altavista.com);
- Excite (www.excite.com);
- Yahoo! (www.yahoo.com);
- AOL (<http://search.aol.com>);

- MSN (<http://search.msn.com>);
- Google (www.google.ru);
- Яндекс (www.yandex.ru);
- Rambler (www.rambler.ru);
- Апорт (www.aport.ru).

Рассмотрим, как осуществляется поиск в поисковой системе. Пользователь вводит свой поисковый запрос в специальную строку. Этот запрос, сформулированный на естественном языке, программой поиска преобразуется в *информационно-поисковый язык (ИПЯ)* — формальный язык, предназначенный для описания содержания документов, хранящихся в ИПС, и запроса [4, 201]. Информационно-поисковые языки представляют собой знаковые системы со своим алфавитом, лексикой, грамматикой и правилами пользования. О специфике ИПЯ каждой поисковой системы, особенно о его «синтаксисе» (т.е. о правилах сочетания ключевых слов, вводимых в строку поиска) можно узнать на отдельных вкладках соответствующей поисковой системы. Например, в Яндекс такая вкладка называется «Помощь — Как искать».

Процедура описания документа на ИПЯ называется *индексированием*. В результате индексирования каждому документу приписывается его формальное описание — *поисковый образ документа*. Аналогичным образом индексируется и запрос, которому приписывается поисковый образ запроса или *поисковое предписание*. Алгоритмы информационного поиска основаны на сравнении поискового предписания с поисковым образом запроса [4, 201].

Степень соответствия документа запросу задается категорией *релевантности*. При этом в процессе информационного поиска можно получить в выдаче значительный *информационный шум* — множество документов, формально релевантных, но не являющихся релевантными по смыслу [4, 197—198].

Чтобы получить меньше информационного шума, пользователю следует уточнять свой запрос, используя для этого дополнительные настройки поисковой системы. Так, в *Google*, нажав вкладку «Расширенный поиск», можно задать поиск целых словосочетаний (а не отдельных составляющих их слов), ограничить язык выдачи, дату