

2 курс

ПЛАН – КОНСПЕКТ
проведения практического занятия № 19 по дисциплине
«Информатика»

часть 1

**Раздел 4. «Сетевые технологии обработки информации и
автоматизированные информационные системы (АИС)»**

**Тема 4.1:
«Классификация компьютерных сетей»**

Подготовил: преподаватель
В.Н. Борисов

Практическое занятие № 19 «Поиск информации в Интернете» (Поиск информации профессионального содержания в Интернете с использованием поисковых систем.). Сервисы Интернета» по Теме № 4.1. «Классификация компьютерных сетей.»

Цели занятия: изучить со студентами основы работы в Интернет, службы и сервисы Интернета, поисковые системы, электронную коммерцию, цифровые сервисы государственных услуг, практическое применение полученных знаний – поиск информации профессионального содержания в Интернет с использованием поисковых систем, по адресу.

Вид занятия: классно-групповое, комбинированное (по проверке знаний, умений по пройденному материалу, по изучению и первичному закреплению нового материала, применению на практике полученных знаний).

Методы проведения занятия: доведение теоретических сведений, выполнение практических заданий.

Время проведения: 2 ч

Основные вопросы:

1. Службы и сервисы Интернета (электронная почта, видеоконференции, форумы, мессенджеры, социальные сети).
2. Поиск в Интернете.
3. Электронная коммерция.
4. Цифровые сервисы государственных услуг.
5. Достоверность информации в Интернете.
6. Применение на практике изученного материала (выполнение практического задания – поиск информации профессионального содержания в Интернете с использованием поисковых систем, по адресу).

Литература:

1. 5 учебник: Босова, Л. Л. Информатика. 11 класс. Базовый уровень : учебник / Л. Л. Босова, А. Ю. Босова. — М. : БИНОМ. Лаборатория знаний, 2022. — 200 с., глава 4, параграфы 15,16.

Примерный расчет времени (по каждому занятию):

1. Вступительная часть – 20 мин.
2. Основная часть – 60 мин.
3. Заключительная часть – 10 мин.

Вступительная часть:

Занятия начать с объявления темы занятия, основных рассматриваемых вопросов, времени изучения темы (нового материала), закрепления на практике полученных знаний, перечисления литературы.

Основная часть (доведение теоретических сведений):

Первый вопрос: Службы и сервисы Интернета (электронная почта, видеоконференции, форумы, мессенджеры, социальные сети).

Интернет играет важную роль в жизни современного человека, являясь:

- 1) средством доступа к общим информационным ресурсам;
- 2) средством коммуникации (общения) между удалёнными пользователями.

Благодаря Интернету, ставшему в наше время мобильным, пользователь, где бы он ни находился, всегда имеет доступ к важным документам, почте, может получить информацию по любому интересующему его вопросу.

Средства обеспечения определённых услуг для пользователей сети Интернет принято называть **службами** (сервисами).

Для каждой службы Интернета существует своя программа-сервер, клиентская программа и свой протокол, обеспечивающий взаимодействие программы-клиента с сервером. Воспользоваться какой-либо службой Интернета можно только в том случае, если на компьютере установлено соответствующее программное обеспечение (клиентская программа).

15.1. Информационные службы

Информационные службы предоставляют пользователям возможность доступа к разнообразным информационным ресурсам (файлам, документам), хранящимся в Интернете.

Основными информационными службами Интернета являются Всемирная паутина, служба передачи файлов, служба файлообменников.

Всемирная паутина (англ. World Wide Web, WWW) представляет собой распределённую по всему миру информационную

систему, состоящую из миллиардов взаимосвязанных электронных документов — веб-страниц.

Каждый ресурс (страница, документ, файл) в Интернете имеет свой уникальный адрес или URL (англ. Universal Resource Locator — универсальный указатель ресурсов).

Адрес документа в Интернете (URL) состоит из следующих частей:

- 1) название протокола со знаками `://` в конце названия;
- 2) доменное имя сервера со знаком `/` в конце имени;
- 3) полное имя файла на сервере, где он находится.

Например:

<http://www.etudes.ru/data/models/sumofsquares/03.jpg>

протокол	адрес сервера	имя файла
----------	------------------	-----------

Ещё одной информационной службой является **служба передачи файлов**, предоставляющая пользователям услуги по хранению и обеспечению доступа к большому количеству файлов: системному и прикладному программному обеспечению, электронным книгам, музыке, видео и т. п. Передача файлов осуществляется по протоколу FTP (англ. File Transfer Protocol — протокол передачи файлов), именем которого названа и сама служба.

FTP-сервер — программа, позволяющая хранить файлы и передавать их по протоколу FTP.

FTP-клиент — программа, позволяющая подключаться к удалённому FTP-серверу и получать/передавать файлы по протоколу FTP. Существует множество бесплатных и платных FTP-клиентов. Встроенными возможностями для работы по протоколу FTP обладают браузеры, менеджеры файлов и некоторые другие программы.

На серверах службы FTP хранятся FTP-архивы — большие хорошо структурированные коллекции объединённых общей тематикой файлов. Среди них можно выделить: 1) коммерческие серверы и серверы ограниченного доступа, которые доступны только для зарегистрированных пользователей; 2) серверы с открытым доступом для всех желающих.

Если вы хотите поделиться файлами (например, фотографиями, сделанными во время летнего отдыха) со своими друзьями, то можно воспользоваться **файлообменником** — службой, предоставляющей пользователю место под его файлы и круглосуточный доступ к ним, как правило, по протоколу http. Загрузив свои файлы на сервер файлообменника, вы получите ссылку, которую

сможете переслать кому-то конкретно или опубликовать для широкого круга лиц.

В последнее время всё большую популярность среди пользователей приобретают **облачные хранилища** данных (англ. cloud storage). С точки зрения пользователя данные хранятся и обрабатываются на одном большом виртуальном сервере, в так называемом «облаке», которое физически представляет собой многочисленные серверы, удалённые друг от друга географически, вплоть до расположения на разных континентах.

15.2. Коммуникационные службы

К основным коммуникационным службам относятся электронная почта, форум, чат, IP-телефония.

Электронная почта — e-mail (от англ. *electronic mail*) — одна из самых первых коммуникационных служб Интернета. Она предоставляет возможность передавать электронные письма (текстовые сообщения и прикрепленные к ним файлы) от пользователя-отправителя одному или группе адресатов.

Почтовый клиент — программа, помогающая составлять и посылать электронные сообщения, получать и отображать письма на компьютере пользователя.

Почтовый сервер — программа, пересылающая сообщения из почтовых ящиков на другие серверы или на компьютер пользователя по запросу его почтового клиента. На почтовом сервере создают почтовые ящики для пользователей с определённым именем и паролем для доступа. Адрес электронной почты имеет вид:

<имя_пользователя>@<имя_сервера>

Почтовая служба основана на следующих протоколах:

- SMTP (англ. Simple Mail Transfer Protocol — простой протокол передачи почты) — почтовый протокол, служащий для отправки сообщений с компьютера-клиента на почтовый сервер, а также для пересылки почты между серверами;
- POP3 (англ. Post Office Protocol — протокол почтового офиса версия 3) — почтовый протокол для получения доступа к почтовому ящику на сервере и пересылки сообщений на компьютер-клиент;
- IMAP (англ. Internet Message Access Protocol — протокол доступа к сообщениям Интернета) — протокол для доступа к почтовому ящику на сервере, позволяющий управлять корреспонденцией на сервере.

Коммуникационные службы предоставляют пользователям сети возможность обмениваться новостями, обсуждать проблемы, проводить дискуссии и т. п.

Обсуждение определённой темы группой собеседников, находящихся на значительном расстоянии, называют **телеконференцией**. Видеоконференция предусматривает использование средств передачи видеобразов.

Долгосрочные (постоянно действующие) телеконференции, в ходе которых собеседники посылают и читают текстовые сообщения в удобное для них время, называют **форумами**.

Форум предлагает набор разделов для обсуждения. Зарегистрированные на сайте пользователи, посылая свои сообщения, могут создавать внутри разделов темы и вести обсуждения в рамках этих тем. Сообщение и все ответы на него образуют «ветку» форума. Незарегистрированные пользователи получают статус гостей, которые могут просматривать ветки форума, но не имеют права принимать участие в обсуждениях. За соблюдением правил следят модераторы, имеющие право редактировать, перемещать и удалять чужие сообщения в определённом разделе или теме.

Службы интерактивного общения предоставляют возможность двум или группе пользователей обмениваться текстовыми сообщениями через Интернет в реальном времени. Первой такой службой была служба IRC (англ. Internet Relay Chat — ретранслируемый интернет-разговор) или просто **чат** (от англ. *chat* — болтать). В отличие от телеконференций, где обсуждение темы открыто всем, в системе IRC общение происходит только в пределах одного канала, в работе которого принимают участие обычно лишь несколько человек. Подключившись к каналу, пользователь видит на экране сообщения других участников. Сообщение, введённое любым участником, практически немедленно появляется на канале, и все остальные участники обсуждения могут «высказаться» в ответ. Для работы службы разработан одноимённый протокол IRC. IRC-сервер — программа, обеспечивающая работоспособность системы IRC и хранящая информацию о каналах и подключённых пользователях. IRC-клиент — программа для подключения к IRC-серверу и ведения беседы.

Некоторые службы интерактивного общения дают возможность вести обсуждение только двум собеседникам. Каждому пользователю при регистрации в таких службах предоставляется индивидуальный код, и каждый сам создаёт список контактов — лиц, с которыми он желает общаться.

IP-телефония — служба, обеспечивающая передачу телефонных разговоров абонентов по сети Интернет. Передача данных в IP-телефонии осуществляется на основе набора протоколов VoIP (англ. Voice over Internet Protocol — голос поверх протокола Интернета). Skype (скайп) — одна из самых популярных программ, позволяющих обмениваться мгновенными сообщениями, использовать голосовую связь и видеозвонки. С помощью Skype можно: общаться с одним или сразу с несколькими людьми, пересылать файлы, вместо изображения с веб-камеры передавать изображение с экрана монитора. Часто эта программа используется для проведения видеоконференций.

В последнее время среди пользователей Интернета широкое распространение получили **социальные сети** — интерактивные многопользовательские веб-сайты, содержание (контент) которых создаётся самими участниками сети. Такие сайты представляют собой автоматизированные социальные среды, позволяющие общаться группам пользователей, объединённых общими интересами.

15.3. Сетевой этикет

В сети Интернет существуют негласные правила поведения, так называемый сетевой этикет. Кратко, суть сетевого этикета может быть выражена одной фразой: «Уважайте своих невидимых партнёров по Сети!».

Приведём основные правила сетевого этикета, которых следует придерживаться в почтовой переписке, а также при использовании других сервисов сети Интернет.

1. Ясно идентифицируйте себя.
2. Знайте и уважайте своего адресата.
3. Указывайте тему сообщения.
4. Пишите грамотно, кратко. Давайте чёткий ответ на поставленный вопрос.
5. В текстовых сообщениях можете выражать эмоции с помощью небольших рисунков, называемых смайликами.
6. Не запрашивайте подтверждения получения сообщения без необходимости.
7. Не допускайте спама — бессодержательных, навязчивых или грубых сообщений в адрес другого лица или группы лиц.
8. Не надейтесь на полную конфиденциальность переписки.

САМОЕ ГЛАВНОЕ

Средства обеспечения определённых услуг для пользователей сети Интернет принято называть службами (сервисами). Для каждой службы Интернета существует своя программа-сервер, клиентская программа и свой протокол, обеспечивающий взаимодействие программы-клиента с сервером. Воспользоваться какой-либо службой Интернета можно только в том случае, если на компьютере установлено соответствующее программное обеспечение (клиентская программа).

Информационные службы предоставляют пользователям возможность доступа к разнообразным информационным ресурсам (файлам, документам), хранящимся в Интернете. Основными информационными службами Интернета являются Всемирная паутина, служба передачи файлов, служба файлообменников.

К основным коммуникационным службам относятся электронная почта, форум, чат, IP-телефония.

В последнее время среди пользователей Интернета широкое распространение получили социальные сети — интерактивные многопользовательские веб-сайты, содержание (контент) которых создаётся самими участниками сети. Такие сайты представляют собой автоматизированные социальные среды, позволяющие общаться группам пользователей, объединённых общими интересами.

В сети Интернет существуют негласные правила поведения, так называемый сетевой этикет. Кратко, суть сетевого этикета может быть выражена одной фразой: «Уважайте своих невидимых партнёров по Сети!».

Всемирная паутина.

WWW (World Wide Web) — это попытка объединить в одном информационном инструменте возможности всех указанных средств, да еще добавить к ним передачу (помимо текстов и программ) графических изображений, звуков, видео. Все эти информационные объекты связываются структурой гипертекста.

Гипертекст — это прежде всего система документов (как правило, очень объемных) с перекрестными ссылками. Поскольку система WWW позволяет включить в эти документы не только тексты, но и графику, звук и видео, гипертекстовый документ превратился в гипермедиа-документ. В документах содержатся ссылки на другие документы, связанные по смыслу, например углубляющие понимание данного текста. Со ссылками могут быть связаны картинки, звуковые заставки, видеофрагменты. Картинки или их части также могут включать ссылки на текст, новые картинки или звук. Документы, на которые сделаны ссылки, могут находиться на удаленных компьютерах. По ссылкам можно значительно удалиться от первоначального источника информации, но к нему можно легко вернуться. Таким образом, читая статью о художественной галерее, вы сразу можете просматривать ее картины, а изучая музыкальные инструменты слышать их звучание.

Гипермедиа-документы хранятся на WWW-серверах сети Internet. Для работы с гипермедиа-документами разработано много различных программ-клиентов, называемых *программами просмотра WWW* или *браузерами*. Программы просмотра позволяют по известному точному адресу вызывать нужные вам документы, накапливать их, сортировать, объединять, редактировать, печатать.

Наиболее популярными программами просмотра являются Microsoft Internet Explorer и Netscape Navigator. Эти браузеры имеют много общего. Поэтому, освоив один из них, легко переключиться на работу с другим.

Если точный адрес интересующего вас документа вам не известен, необходимо обратиться к программам поиска. Для поиска информации в WWW имеются международные программные системы AltaVista, Lycos, Yahoo и др. Для русскоязычного поиска более удобными являются отечественные поисковые системы Rambler, Yandex и Aport. При работе с поисковыми пользователем задаёт *поисковый образ* – ключевые слова интересующей его темы, и система выдаёт списки и адреса тех документов, в которых эти слова встречаются. Заметим, что несмотря на наличие большого количества хороших программ поиска, лучше всего иметь точный адрес. Способ задания адреса определяется системой унифицированных *URL-адресов* (URL = Uniform Resource Locator — унифицированный указатель ресурсов).

Программа поиска для выбора нужных адресов обращается к серверам поиска, доступным через интерфейс Web. Основной функцией этих серверов является обработка информации из документов различных серверов (Web, FTP, Usenet и

др.), занесение ее в базу данных и предоставление адресов этой информации по запросам пользователей поисковых программ.

Существуют десятки поисковых серверов. Доступ к поисковым системам этих серверов обеспечивается через программы, указанные в браузере, только в том случае, если имеется соглашение между фирмой, содержащей поисковый сервер, и фирмой-производителем браузера. В браузере указывается, с каким сервером поиска устанавливается связь при выборе поисковой программы. Гипермедиа-документы создаются на языке *HTML* — *HyperText Markup Language*. Язык этот, по сути, является простым языком разметки текста и связывания страниц. Основная идея связывания страниц очень проста. На странице выделяется одно или несколько словосочетаний (ссылок), которые ссылаются на адреса новых страниц. Браузеры при щелчке мыши на таком словосочетании выбирают адрес и выполняют запрос на получение соответствующей страницы. Для повышения производительности при подготовке гипертекста используются специальные HTML-редакторы и средства конвертирования в HTML-формат документов, подготовленных в среде таких популярных текстовых редакторов, как Microsoft Word. Многие браузеры также включают редакторы, которые позволяют создавать и редактировать гипермедиа-документы.

В настоящее время многими фирмами разработаны Web-серверы. Назовем некоторые из них:

- Internet Information Server фирмы Microsoft,
- Enterprise Server фирмы Netscape Communications,
- Server/Secure Server фирмы IBM,
- Web-сервер NetWare фирмы Novell.

WWW-серверы имеют все ведущие университеты и исследовательские центры мира, всемирно известные корпорации и небольшие фирмы, государственные учреждения и различные общественные организации, средства массовой информации. В России насчитывается несколько сотен общедоступных WWW-серверов. В них можно получить основные сведения о многих университетах, институтах РАН, коммерческих фирмах, банках, узнать новости по экономике и финансам, получить доступ к правовым справочным системам, газетам и журналам.

Однако, несмотря на доступность многочисленных средств поиска, решение задачи эффективного поиска остается не простым. Для того чтобы поиск приносил удовлетворительные результаты, нужно хорошо изучить возможности выбранной программы поиска, правила формулирования запросов. Слова запроса должны точно, полно и кратко характеризовать предмет вашего поиска. Очевидно, что чем больше слов использовано в запросе, тем больше сужается поиск. Целесообразно воспользоваться советами по организации поиска, содержащимися на страницах поисковых серверов. Следует обратить внимание на возможности локального поиска на серверах крупных организаций, часто

хранящих огромное количество ссылок, обеспечивающих доступ к тематически связанным серверам.

Одним из перспективных направлений развития Internet является доступ через Web-интерфейс к базам данных, в которых накоплена обширная ценная информация. Пользователи хотят извлекать информацию из баз данных и составлять отчеты в заданной форме. Реализуются такие задачи программами, встраиваемыми в Web-страницы, и выполняющимися в среде Web-браузера на вашем компьютере. Разработка программ в Web может быть произведена, в частности, на языке программирования Java, созданного фирмой Sun Microsystems. Выполнение Java-программ, встроенных в Web-страницы, обеспечивается практически всеми современными браузерами.

Программы просмотра Web-страниц

Программы просмотра (браузеры) предназначены для получения из сети запрошенных пользователем Web-документов и представления текстовой, графической, аудио, видео и другой информации в удобном виде на экране монитора. Среди многочисленных программ просмотра наиболее широкое распространение в настоящее время получили Netscape Navigator и Microsoft Internet Explorer. Поскольку различные браузеры обладают общими основными чертами, поняв принципы и овладев средствами работы одного из них, вы без труда сможете освоить другой. Большинство современных браузеров обеспечивают легкий доступ не только к страницам Web-серверов, но и многим другим видам услуг сети Internet. Они включают возможности обработки электронной почты, телеконференций UseNet, позволяют работать с сервисом FTP, Gopher и др. В браузеры встраиваются редакторы Web-страниц.

Открытие страницы Web

Для открытия страницы (документа Web) пользователь должен сообщить браузеру адрес этой страницы. Адрес задается в стандартном формате, разработанном для указания ссылок на любые доступные в Internet ресурсы. Он называется URL-адресом (см. выше).

Формат URL можно представить в следующем виде:

Вид_информационного_ресурса://доменное_имя_хост-
компьютера/имя_каталога/имя_подкаталога/имя_файла

URL состоит из двух частей. Первая его часть определяет вид ресурса, с которым вы хотите начать работу. Вид ресурса задается наименованием протокола, используемого системой для реализации доступа к этому ресурсу.

Используются следующие наименования протоколов:

- http — (HyperText Transfer Protocol — протокол передачи гипертекста) определяет переход к работе с Web-сервером;

- ftp — сервис FTP;
- gopher — сервис Gopher;
- wais — сервер индексированных баз данных WAIS;
- telnet - указывает на открытие сеанса связи по протоколу Telnet;
- file — если далее стоит (например) //c:, то указывает на обращение к файлу на локальном диске; если //, то это обращение к FTP серверу; (буква “с” может заменяться на любую другую букву, которой именуется локальный диск);
- news — определяет запуск программы просмотра новостей и открытие определенной группы новостей телеконференций Usenet. URL, использующий этот протокол, имеет другой формат:

news:имя_группы_новостей

- mailto — определяет запуск программы электронной почты для отправки сообщения по определенному адресу в Internet. URL, использующий этот протокол, имеет другой формат:

mailto:имя_пользователя@доменное_имя_хост_компьютера

Вторая часть URL-адреса указывает доменное имя хост-компьютера, на котором хранится требуемый документ, и через / может указывать точное местоположение и имя файла, в котором хранится документ. Хост-компьютером называется компьютер, который предоставляет в распоряжение удаленного пользователя некоторые свои ресурсы. В данном случае речь идет об информационных ресурсах, которые предоставляются пользователям, работающим на удаленном компьютере с программой-клиентом, сервисными службами (серверами) этих компьютеров – Web, FTP, Gopher.

Электронная почта.

Одним из средств взаимодействия пользователей в сетях является электронная почта (e-mail). С электронной почты начиналось создание Internet и она остается самым популярным видом деятельности в ней.

В общем случае электронная почта – это многозначный термин, используемый для определения процесса передачи сообщений между компьютерами. Различают электронную почту, применяемую в локальных и глобальных сетях. Далее речь пойдет о глобальных системах электронной почты.

К преимуществам электронной почты относятся: скорость и надежность доставки корреспонденции; относительно низкая стоимость услуг; возможность быстро ознакомить широкий круг корреспондентов с сообщением; посылка не только текстовых сообщений, но и программ, графики, аудиофайлов; экономия бумаги и т.д.

Рассмотрим принципиальную схему, лежащую в основе работы различных систем электронной почты.

Для отправки почтового сообщения с помощью вашего компьютера вы вызываете почтовую программу, указываете получателя сообщения, создаете сам текст сообщения и даете указание программе, чтобы она выполнила его отправку. По сигналу на передачу сообщения устанавливается связь вашего компьютера с почтовым хост-компьютером, непосредственно включенным в ту или иную глобальную сеть. Сообщение, попадая на хост-компьютер отправителя, далее передается по каналам связи на машину получателя и там помещается в область дисковой памяти, принадлежащую адресату и называемую почтовым ящиком. Пользователь-получатель забирает поступившую почту из почтового ящика на свой компьютер и обрабатывает ее.

Любая система электронной почты состоит из двух главных подсистем:

1. клиентского программного обеспечения, с которым непосредственно взаимодействует пользователь;
2. серверного программного обеспечения, которое управляет приемом сообщения от пользователя-отправителя, передачей сообщения, направлением сообщения в почтовый ящик адресата и его хранением в этом ящике до тех пор, пока пользователь-получатель его не возьмет оттуда.
3. Различные почтовые программы могут быть классифицированы по разным критериям. Например, в какой операционной системе они могут работать. Сейчас получили наиболее массовое распространение продукты, работающие в ОС Windows . Широко используются программы обработки почты, входящие в состав браузеров Microsoft Internet Explorer , Netscape Navigator . *Браузер* (от англ. browser)– это программа, производящая поиск в сети Internet. (Подробнее о браузерах см. ниже в п. “Всемирная паутина WWW”).

. Существуют программы для пользователей систем UNIX и OS/2.

Для работы электронной почты необходимы специальные программы. Существуют два основных стандарта e-mail:

- SMTP (Simple Mail Transfer Protocol);
- X.400.

Стандарт SMTP привлекателен простотой, дешевизной, множеством сервисных функций и вследствие этого получил широкое распространение, в частности, в Internet. Существует также стандарт POP-3, отличающийся от SMTP в основном тем, что в этом стандарте клиент работает с программой, установленной на компьютере провайдера, а не на своём компьютере.

Стандарт X.400 отличается строгостью, жесткой стандартизацией, наличием коммерческих операторов с гарантированным уровнем сервиса, поддержкой большого количества национальных кодов. Этот стандарт ввиду названных особенностей пользуется большой популярностью среди государственных

организаций всего мира при работе, в частности, по правительственным телекоммуникационным линиям.

Из множества программ e-mail, работающих под управлением Windows в стандарте SMTP, можно назвать, например:

- Outlook Express, используемая в браузере MS Internet Explorer;
- Netscape Mail, входящая в состав браузера Netscape Navigator;
- Mail, HotMail, Hotbox и другие бесплатные программы в Internet;
- MSMail, входящая в состав офисного приложения Outlook;
- Eudora Pro компании Qualcomm и многие другие.

Несмотря на многообразие клиентских программ различных систем электронной почты, все они имеют общие функции:

- оповещение о прибытии новой почты;
- чтение входящей почты;
- создание исходящей почты;
- адресация сообщений
- использование адресной книги, содержащей список абонентов, которым часто посылают почту;
- отправка сообщений;
- обработка сообщений и их сохранение. К обработке сообщений относятся такие функции, как печать, удаление, переадресация письма, сортировка, архивирование сообщений, хранение связанных сообщений. Особо следует выделить программы, позволяющие работать с папками, создавать свои папки для хранения в них сообщений по различным темам. Это очень удобно и помогает быстрее и эффективнее обрабатывать почту.

Второй вопрос: Поиск информации в Интернете.

Для поиска информации в Интернете используются специальные поисковые системы, которые содержат в базах данных постоянно обновляемую информацию о веб-сайтах, файлах и других информационных ресурсах Интернета. Разные поисковые серверы могут использовать различные механизмы поиска, хранения и предоставления пользователю информации.

Поисковые системы общего назначения. Интерфейс поисковых систем общего назначения содержит поле поиска и список разделов каталога.

Заполнение баз данных поисковых систем. Заполнение баз данных поисковой системы осуществляется с помощью специальных программ-роботов, которые периодически «обходят» веб-серверы Интернета. Программы-роботы читают все встречающиеся документы, выделяют в них ключевые слова и заносят в базу данных, содержащую URL-адреса документов.

Так как **информация в Интернете постоянно** **меняется** (создаются новые веб-сайты и страницы, удаляются старые, меняются их URL-адреса и т. д.), поисковые роботы не всегда успевают отследить все эти изменения. Информация, хранящаяся в базе данных поисковой системы, может отличаться от реального состояния Интернета, и тогда пользователь в результате поиска может получить адрес уже не существующего или перемещённого документа.

В целях обеспечения большего соответствия между содержанием базы данных поисковой системы и реальным состоянием Интернета большинство поисковых систем разрешают автору нового или перемещённого веб-сайта самому внести информацию в базу данных, заполнив регистрационную анкету. В процессе заполнения анкеты разработчик сайта вносит URL сайта, его название, краткое описание содержания сайта, а также ключевые слова, по которым легче всего будет найти сайт.

Поиск по ключевым словам. При поиске по ключевым словам в поле поиска вводится одно или несколько ключевых слов, которые, по мнению пользователя, являются главными для искомого документа. Можно также использовать сложные запросы, использующие логические операции, шаблоны и т. д.

Через некоторое время после отправки запроса поисковая система вернёт аннотированный (с короткими комментариями содержания документа) список URL-адресов документов, в которых были найдены указанные вами ключевые слова. Для просмотра этого документа в браузере достаточно активизировать указывающую на него ссылку.

Если ключевые слова были выбраны неудачно, то список URL-адресов документов может быть слишком большим (содержать десятки и даже сотни тысяч ссылок). Чтобы уменьшить список, можно в поле поиска ввести дополнительные ключевые слова или воспользоваться каталогом поисковой системы.

Язык построения запросов. Поиск по одному ключевому слову всегда производится однозначно (ключевое слово либо имеется в рассматриваемом тексте, либо отсутствует). При поиске же по нескольким ключевым словам возможны различные способы их комбинирования:

- в искомом тексте должны присутствовать все заданные ключевые слова («И то, И другое»);
- достаточно, чтобы в искомом тексте имелось хотя бы одно из заданных ключевых слов («то ИЛИ другое»);
- в тексте должно иметься одно ключевое слово, но обязательно должно отсутствовать другое («то, но НЕ другое»);
- ключевые слова составляют фразу, которая обязательно должна присутствовать в тексте «как есть»;
- ключевые слова обязательно должны быть взаимосвязаны (т. е. должны располагаться близко друг от друга, — например, в фразе могут быть разделены союзом или прилагательным).

Чтобы «объяснить» поисковой системе, как нужно понимать заданную последовательность ключевых слов, используется особый язык построения запросов.

Принципы построения языка запросов, как правило, универсальны для большинства поисковых систем, но для той или иной системы язык запросов может иметь свои особенности.

Пример языка построения запросов в поисковой системе Яндекс:

- заяц & кролик — поиск текстов, в которых есть оба заданных ключевых слова (и «заяц», и «кролик»);
- заяц | кролик — поиск текстов, в которых имеется хотя бы одно из заданных ключевых слов («заяц» или «кролик»);
- заяц -кролик — знак минуса перед словом «кролик» предписывает искать только такие тексты, в которых есть слово «заяц», но нет слова «кролик»;
- заяц +кролик — знак плюса перед словом «кролик» предписывает искать только такие тексты, в которых есть слово «заяц» и обязательно есть слово «кролик»;
- "братец Кролик" — ключевые слова, заключённые в кавычки, ищутся в тексте как цитата (обязательная фраза);
- кровать /2 диван — ключевые слова в искомом тексте могут следовать в любом порядке и должны быть разделены не более чем одним любым словом (например, это может быть текст «диван и кровать» или «кровать, тахта, диван»).

Кроме того, большинство современных поисковых систем являются «интеллектуальными» — при поиске учитываются все возможные формы (падежи, склонения, спряжения, единственное и множественное число) заданных ключевых слов. Например, при задании ключевого слова заяц поисковая система будет искать все варианты ключевых слов — «заяц», «зайца», «зайцами» и т. п.

Поиск в иерархической системе каталогов. Каталоги составляются редакторами, просматривающими каждый новый сайт до его включения в иерархическую систему каталогов. Каталоги обычно организованы в соответствии с предметной классификацией.

Поиск информации в каталоге сводится к выбору определённого каталога, после чего пользователю будет представлен список ссылок на URL-адреса наиболее посещаемых и важных веб-сайтов. Каждая ссылка обычно аннотирована.

Современные поисковые системы. Одной из наиболее полных и мощных поисковых систем является Google (www.google.ru).

В Рунете (русской части Интернета) обширные базы данных имеет поисковая система Яндекс (www.yandex.ru).

Сайты в базе данных ранжируются по количеству их посещений в день, неделю или месяц. Посещаемость сайтов определяется с помощью специальных

счётчиков, которые могут быть установлены на сайте. Счётчики фиксируют каждое посещение сайта и передают информацию о количестве посещений на сервер поисковой системы.

Современные поисковые системы часто являются информационными порталами, которые предоставляют пользователям не только возможности поиска документов в Интернете, но и доступ к другим информационным ресурсам (к новостям, к информации о погоде, о валютном курсе, к интерактивным географическим картам и т. д.). Пример: Российский образовательный портал: <http://edu.ru>.

16.2. Поиск информации в сети Интернет

Поиск нужного документа во Всемирной паутине может происходить разными способами:

- указанием адреса документа;
- путём использования поисковых систем.

Поисковая система — это программно-аппаратный комплекс, предназначенный для поиска информации во Всемирной паутине.
Поисковая машина — программная часть поисковой системы; комплекс программ, предназначенный для поиска информации.

Поисковые системы располагаются на специально выделенных компьютерах с мощными каналами связи. Ежеминутно они обслуживают огромное количество поисковых запросов клиентов.

По принципу действия различают несколько типов поисковых систем, а именно:

- поисковые каталоги, управляемые человеком;
- системы, использующие поисковых роботов;
- гибридные поисковые системы.

Поисковые каталоги (веб-каталоги или тематические каталоги) содержат базу данных ссылок на веб-сайты, распределённых по отдельным тематическим рубрикам. Такие каталоги заполняются специалистами вручную. Поиск в них осуществляется спуском по дереву каталога:

- определив тему, по которой будет выполняться поиск, пользователь выбирает соответствующую рубрику тематического каталога;
- прочитав описания ссылок на открывшейся странице, пользователь может перейти по ссылке, соответствующей его ожиданиям; если же нужных ссылок не обнаруживается, то можно уточнить тему и повторить поиск в этой же или другой поисковой службе.

В 1994 году Дэвид Фило и Джерри Янг из Стэнфордского университета (США) предприняли попытку упорядочить большое количество накопившихся у них ссылок на разнообразные информационные источники. Так появилась идея использования специализированной базы данных для эффективного поиска информации в сети. Очень скоро созданная ими система Yahoo! стала самым популярным и полным иерархическим предметно-ориентированным путеводителем по Интернету. В наши дни — это одна из наиболее известных поисковых систем.

Действие поисковых систем, использующих поисковых роботов, основано на постоянном, последовательном изучении всех страниц всех сайтов Всемирной паутины. Для каждого документа составляется его поисковый образ — набор ключевых слов, отражающих содержание этого документа. В связи с постоянным обновлением информации поисковые системы периодически возвращаются к ранее изученным страницам, чтобы обнаружить и зарегистрировать изменения. Информация о ключевых словах исследованных таким образом страниц сохраняется в поисковой системе.

При поступлении запроса от пользователя поисковая система на основании имеющейся в ней информации формирует список страниц, соответствующих критериям поиска. Найденные документы, как правило, упорядочиваются в зависимости от местоположения ключевых слов (в заголовке, в начале текста), частоты их появления в тексте и других характеристик.

Различные поисковые системы, использующие поисковых роботов, имеют схожую структуру, включающую:

- 1) **модуль индексирования**, состоящий из трёх программ-роботов (Spider или «паук» — скачивает веб-страницы; Crawler или «путешествующий паук» — переходит по всем ссылкам, имеющимся на странице, и ищет новые документы, ещё не известные поисковой системе; Indexer или «робот-индексатор» — разбивает на фрагменты страницы, которые скачали «пауки», анализирует их и составляет некоторое описание этих страниц);
- 2) **базу данных** — хранилище представленных в определённом формате всех скачанных и обработанных модулем индексирования документов;
- 3) **поисковый сервер** — система выдачи результатов поиска, определяющая, какие страницы и в какой степени удовлетворяют запросу пользователя.

Поисковая система, получив запрос на поиск, анализирует ту информацию, которая была ею проиндексирована. С одной стороны, это позволяет существенно повысить скорость обработки поискового запроса. С другой стороны, результаты поиска нельзя считать полными, т. к. ни одна поисковая система не может загрузить в свою базу данных информацию обо всех без исключения ресурсах. Кроме того, результаты поиска могут быть отчасти устаревшими — ситуация в сети Интернет меняется быстрее, чем происходит обновление сведений в базах данных поисковых систем.

Гибридные поисковые системы сочетают в себе функции систем, управляемых человеком, и систем, использующих поисковых роботов.

Несмотря на общие принципы работы, поисковые системы различаются по таким характеристикам, как: язык запроса, зона поиска, глубина поиска внутри документа, метод упорядочивания информации и др. На данный момент самой популярной в мире поисковой системой является Google, а крупнейшей отечественной поисковой системой — Яндекс.

В большинстве поисковых систем реализовано три основных типа поиска:

- 1) поиск по любому из слов — результатом является огромный список всех страниц, содержащих хотя бы одно из ключевых слов; применяется, когда пользователь не уверен в ключевых словах;
- 2) поиск по всем словам — в этом режиме формируется список всех страниц, содержащих все ключевые слова в любом порядке;
- 3) поиск точно по фразе — в результате составляется список всех страниц, содержащих фразу, точно совпадающую с ключевой (знаки препинания игнорируются).

Чтобы поиск стал более продуктивным, во всех поисковых системах предусмотрены специальные языки формирования запросов со своим синтаксисом. Эти языки во многом похожи. Выяснить особенности определённого языка можно с помощью справочной системы, входящей в состав поисковой машины.

Найдите информацию о правилах формирования поисковых запросов в поисковых системах Яндекс и Google. Сравните их между собой. Укажите общее и различия.

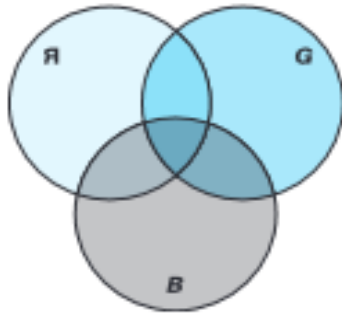
Пример. В языке запросов некой поисковой машины для обозначения логической операции ИЛИ используется символ «|», а для логической операции И — символ «&». Известны запросы и количество найденных по ним страниц некоторого сегмента сети Интернет:

Запрос	Найдено страниц (тыс.)
Яндекс & Google	145
Bing & Google	580
Яндекс & Bing & Google	85
(Яндекс Bing) & Google	x



Чему равно x , если все запросы выполнялись практически одновременно, т. е. во время выполнения запросов набор страниц, содержащих все искомые слова, не изменился?

С подобными задачами вы встречались в курсе информатики основной школы и знаете, что их условие может быть представлено с помощью кругов Эйлера и записано на языке теории множеств.



Пусть $Я$, G и B — множества страниц, содержащих слова Яндекс, Google и Bing соответственно.

Тогда $Я \cap G$ — множество страниц, соответствующих запросу «Яндекс & Google», и его мощность равна 145 (тыс.): $|Я \cap G| = 145$.

Рассуждая аналогично, можем записать: $|B \cap G| = 580$, $|Я \cap B \cap G| = 85$.

Требуется найти мощность множества $(Я \cup B) \cap G$.

Преобразуем это выражение:

$$(Я \cup B) \cap G = (Я \cap G) \cup (B \cap G).$$

В справедливости такого преобразования вы можете убедиться, изобразив левую и правую части равенства с помощью кругов Эйлера.

Согласно принципу включений-исключений, для двух множеств имеем: $|X \cup Y| = |X| + |Y| - |X \cap Y|$.

В нашем случае получаем:

$$\begin{aligned} |(Я \cap G) \cup (B \cap G)| &= |Я \cap G| + |B \cap G| - |(Я \cap G) \cap (B \cap G)| = \\ &= |Я \cap G| + |B \cap G| - |Я \cap G \cap B| = 145 + 580 - 85 = 640. \end{aligned}$$

Итак, $x = 640$.

Как бы точно ни был составлен пользователем запрос к информационной системе, только малая часть из полученных по этому запросу документов будет релевантной, т. е. соответствующей зафиксированной в запросе информационной потребности.

Полнота поиска — это отношение числа выданных релевантных документов к общему числу релевантных документов, имеющих в базе данных поисковой системы. В идеале это число должно равняться 1; на практике может достигать значений 0.7–0.9.

Точность поиска — это отношение числа выданных релевантных документов к общему числу документов, выданных системой по данному запросу. Значение этого параметра колеблется от 0,1 до 1.

Полнота и точность определяют качество или эффективность поиска.

Библиотеки, энциклопедии и словари в Интернете.

Библиотеки. Электронные библиотеки в Интернете содержат электронные (цифровые) копии печатных книг, диссертаций и других документов. Для этого наиболее часто используются форматы файлов PDF, DjVu, FB2; реже — текстовые форматы TXT, RTF, DOC или формат веб-страниц HTML.

Публичная Российская государственная библиотека хранит электронные версии наиболее значительных произведений мировой и русской литературы. Фонд включает электронные копии книг, журналов, карт, нот, изобразительных материалов, диссертаций и авторефератов диссертаций (с разрешения авторов) по различным отраслям знаний. Документы находятся в свободном доступе через Интернет.

Российская государственная библиотека: <http://www.rsl.ru>

Библиотека Максима Мошкова — крупнейшая и старейшая в России электронная библиотека художественной литературы (с ноября 1994 г.). Она хранит прозу, поэзию, фантастику, детективы и другие жанры. Все произведения доступны для скачивания из Интернета.

Библиотека Мошкова: <http://lib.ru>

Научная электронная библиотека elibrary.ru — крупнейшая в России коллекция электронных журналов и баз данных по всем отраслям наук. Ко многим ресурсам доступ открыт всем пользователям Интернета.

Научная электронная библиотека: <http://elibrary.ru>

Энциклопедии и словари. Универсальные энциклопедии содержат сведения о природе и обществе, а также по всем отраслям науки и техники. Специализированные энциклопедии и словари посвящены какой-либо одной отрасли науки и техники. Электронные энциклопедии в Интернете могут быть копиями известных универсальных печатных энциклопедий (Энциклопедия Брокгауза и Эфрона, Большой Энциклопедический словарь и др.) или специализированных словарей (Толковый словарь В. И. Даля и др.).

Электронные энциклопедии: <http://dic.academic.ru>

Существуют также оригинальные электронные энциклопедии. Википедия — это проект свободной многоязычной энциклопедии, в которой каждый может изменить или дополнить любую статью или создать новую.

Википедия: <http://ru.wikipedia.org>

Современная энциклопедия «Кругосвет» содержит новейшие знания по всем отраслям науки и техники. Много интересного в ней можно найти и по проблемам информатики.

Энциклопедия «Кругосвет»: <http://www.krugosvet.ru>

Переводчики и словари. Онлайн-овые мультязычные переводчики позволяют переводить тексты, набранные в окне перевода или скопированные из буфера

обмена, либо даже целые веб-страницы, включая гиперссылки, с сохранением исходного форматирования, и электронные письма.

Онлайновый переводчик ПРОМТ: <http://www.translate.ru>

Интернет-версии электронных словарей позволяют получить точный и достоверный перевод слов с английского, немецкого, французского, итальянского и испанского языков на русский, и обратно. Словарные статьи содержат транскрипцию (для английского языка), все варианты перевода, примеры использования и устойчивые словосочетания. Онлайновые словари включают как общие, так и тематические словари, которые помогут подобрать адекватный перевод для специализированных терминов.

Онлайновый словарь Lingvo: <http://www.lingvo.ru>.