**2. Поисковые системы и машины в Internet.**

Обратимся к истории возникновения сети Internet, которая была создана в связи с возникшей необходимостью совместного использования информационных ресурсов, распределенных между различными компьютерными системами. Большинство первых приложений, включая FTP и электронную почту, были разработаны исключительно для обмена данными между хост-компьютерами Internet. Другие приложения, такие как Telnet, создавались для того, чтобы пользователь получил возможность доступа не только к информации, но и к рабочим ресурсам удаленной системы. По мере развития Internet (увеличения пользователей и хост-компьютеров) прежние методы обмена данными перестали отвечать возросшим потребностям пользователей. Возникла необходимость разработки новых способов поиска сетевых ресурсов и доступа к ним, которые позволяли бы использовать информацию независимо от ее формата и расположения. Для удовлетворения таких потребностей сначала были созданы поисковая система Archie, решающая задачу локализации ресурсов на FTP-сервере, и система Gopher, упрощающая доступ к различным сетевым ресурсам. Затем были разработаны сетевые информационные системы WWW и WAIS, предлагающие абсолютно новые методы получения информации. Принципы работы этих систем позволяют легко ориентироваться в огромном количестве информационных ресурсов без необходимости предоставления механизмов работы самой сети Internet. Такой подход позволяет говорить уже не просто о ресурсах взаимосвязанных компьютерных систем, а об особых информационных пространствах сети.

**FTP** (File Transfer Protocol, Протокол передачи файлов) регламентирует процедуру передачи файлов между компьютерами. Сервис, предоставляемый FTP, отличается необходимостью непосредственного соединения между компьютерами, участвующими в процессе обмена данными, и необходимостью идентификации пользователя для определения его прав доступа к каталогам и файлам удаленной системы. Одной из проблем доступа информации с помощью FTP является невозможность определить, какие файлы можно найти на FTP-сервере без предварительного установления с ним соединения с ним и просмотра структуры каталогов.

Система **UseNet** была разработана для поддержки телеконференций. Для их ведения организуются специальные рубрики, названия которых отражают общее содержание телеконференции. Основным достоинством UseNet является возможность получения практически любой информации в достаточно короткие сроки. Единственная проблема, возникающая у пользователя, заключается в выборе телеконференции, в которой он может получить ответ на свой вопрос. Эту задачу помогает решить иерархическая система имен телеконференций в UseNet. Кроме того, система UseNet была разработана для обмена исключительно текстовой информацией, поэтому для передачи бинарных файлов необходимо предварительно воспользоваться программой, конвертирующей данные.

Система **Archie** представляет собой комплекс программных средств, работающих со специальными базами данных. В этих базах данных содержится постоянно пополняющаяся информация о файлах, к которым можно получить доступ через сервис FTP. Пользуясь услугами системы Archie, можно осуществить поиск файла по шаблону его имени. При этом пользователь получит список файлов с точным указанием места их хранения в сети, а также с информацией о типе, времени создания и размере файлов. Доступ к информационно-поисковой системе Archie может осуществляться различными путями, начиная от запросов по электронной почте и с помощью сервиса Telnet и заканчивая использованием графических Archie-клиентов.

Система **Gopher** была разработана для упрощения процесса локализации FTP-ресурсов Internet и для более удобного представления сведений о содержании хранящихся на FTP-серверах файлов. Система Gopher дает возможность в удобной форме (в виде меню) представлять пользователям об имеющихся файлах и их содержании. Меню Gopher-серверов могут содержать ссылки на другие Gopher- и FTP-серверы. Ссылка на FTP-сервер позволяет как просмотреть его каталоги, так и автоматически получить файл. Таким образом, пользователь получает возможность “путешествовать” по Internet, не обращая внимания на местонахождение интересующих его ресурсов, и получать доступ к этим ресурсам.

Система **Veronica** используется для поиска информации в Gopher-пространстве по заголовкам пунктов меню. После ввода ключевого слова, система Veronica выясняет, встречается ли оно в меню на каком-либо Gopher-сервере, и в качестве результатов поиска выдает список заголовков пунктов меню, содержащих ключевое слово. Поскольку система Veronica не является автономной поисковой программой, а тесно связана с системой Gopher, она обладает тем же, что и система Gopher, недостатком: далеко не всегда по заголовку можно сказать, что собой представляет тот или иной информационный ресурс. Достоинства системы заключается в том, что нет необходимости узнавать, где расположена найденная информация, достаточно выбрать требуемую запись из списка.

Основой системы **WAIS** является комплекс глобальных баз данных, в которых содержится индексированная информация о ресурсах сети. Система WAIS обеспечивает распределенный поиск текста в этих базах данных по ключевому слову или строке. Таким образом, сервис WAIS прежде всего ориентирован на содержание текстовых документов, а не на названия и формат файлов, в которых они хранятся. В системе также реализованы связи с различными базами данных, хранящихся на других хост-компьютерах, и механизм получения необходимого документа. В дополнение к этому в системе WAIS достаточно просто реализован механизм добавления новых баз данных к уже существующим. Документы, в которых система WAIS осуществляет поиск, обычно представляют собой текстовые файлы, которые могут иметь любой размер и содержать любую информацию.

Самым новым и наиболее совершенным средством поиска, получения и отображения информации в Internet является система **World Wide Web** (WWW). Эта система предоставляет пользователю возможность, работая с графическим интерфейсом, просматривать документы, хранящиеся на хост-компьютерах в Internet, причем для перемещения между документами используются гипертекстовые связи (ссылки). Эти ссылки включаются в документы системы WWW и могут указывать на любые информационные ресурсы Internet. WWW-документы (WWW-страницы) подготавливаются на специально разработанном языке – HTML (Hypertext Markup Language, Язык разметки гипертекста), доступном обычным пользователям, и позволяет включать в WWW-страницы кроме текстовой информации изображения, ссылки на другие документы, другие дополнительные элементы интерфейса, такие как интерактивные формы для ввода данных при работе с документом, и многое другое. Важным достоинством системы WWW является то, что графический интерфейс многих программ просмотра WWW (WWW-браузеров) делает работу с этой системой весьма легкой для начинающих пользователей, не имеющих опыта работы в Internet. Пользователи системы WWW также не обязаны знать, где и в какой форме хранятся данные, поскольку WWW-браузер сам выполняет большую часть работы.

В системе WWW имеется также доступ к специально разработанным поисковым системам (поисковым машинам), которые выполняют поиск WWW-страниц, содержащих информацию на заданную с помощью ключевых слов тему. Поиск производится среди заранее проиндексированных WWW-страниц множества серверов Internet, и результатом поиска является автоматически сгенерированная новая WWW-страница, содержащая ссылки на найденные документы. Это чрезвычайно мощное средство делает весьма легкой и удобной для пользователей процедуру поиска информации в Internet, а результат поиска позволяет представить в наглядной форме. Наиболее мощными поисковыми машинами на сегодняшний день в мире являются: **Alta Vista** (**www.altavista.digital.com**), **Yahoo!** (**www.yahoo.com**), **Infoseek** (**www.infoseek.com**), **GoTo** (**www.goto.com**), **FTP search** (**ftpsearch.ntnu.no** – поиск в системе FTP), а также отечественные: **Rambler** (**www.rambler.ru**), **Yandex** (**yandex.ru**), **АПОРТ!** (**www.aport.ru**), **Следопыт** (**www.medialingua.ru/www/wwwsearc.htm**), **Compass** (**compass.tsu.ru**), **Russian FTP** **search** (**ftpsearch.city.ru** – поиск в системе FTP). Следует отметить, что результатом поиска могут быть не только HTML-документы, хранящиеся на WWW-серверах, но и другие типы файлов на других типах серверов, если информация об этих документах содержится в соответствующих базах данных поисковых машин.