

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГБОУ ВПО «Кубанский государственный аграрный университет»

Факультет прикладной информатики
Кафедра компьютерных технологий и систем

Д. Ю. Жмурко, В. И. Лойко

МИРОВЫЕ ИНФОРМАЦИОННЫЕ РЕСУРСЫ

Учебно-методическое пособие

*Рекомендовано Учебно-методической комиссией
факультета Прикладной информатики
Кубанского государственного аграрного университета
в качестве учебно-методического пособия для студентов, обучающихся
по специальности «Прикладная информатика»*

Краснодар
2012

УДК 004.6 (076)
ББК 32.973.2
Ж 77

Р е ц е н з е н т ы:

Е. В. Луценко – доктор экономических наук, профессор
(Кубанский государственный аграрный университет)

Жмурко Д. Ю.

Ж 77 **Мировые** информационные ресурсы: учебно-методическое пособие (по направлению «Прикладная информатика») / Д. Ю. Жмурко, В. И. Лойко. – Краснодар: КубГАУ, 2012. – 158 с.

Учебно-методическое пособие по дисциплине «Мировые информационные ресурсы» подготовлено для специалистов, обучающихся по направлению «Прикладная информатика» (бакалавров, специалистов и магистров) разработано на кафедре компьютерных технологий и систем (КТС) факультета прикладной информатики (ФПИ) КубГАУ. Оно соответствует требованиями государственного образовательного стандарта (ГОС) высшего профессионального образования по направлению «Прикладная информатика», утвержденного 14.03.2000г. (регистрационный № 52 мжд/сп) Министерством образования РФ.

Рассмотрено и рекомендовано к изданию на заседании кафедры компьютерных технологий и систем КубГАУ 22 октября 2012 года, протокол № 2.

Рекомендовано к печати Советом факультета прикладной информатики Кубанского государственного аграрного университета 24 декабря 2012 года, протокол № 13.

УДК 004.6 (076)
ББК 32.973.2

© Д. Ю. Жмурко, В. И. Лойко, 2012
© ФГБОУ ВПО «Кубанский государственный аграрный университет», 2012

ОГЛАВЛЕНИЕ

I ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ	5
II ТРЕБОВАНИЯ К УРОВНЮ ОСВОЕНИЯ СОДЕРЖАНИЯ ДИСЦИПЛИНЫ	6
III СОДЕРЖАНИЕ ДИСЦИПЛИНЫ.....	9
IV КУРС ЛЕКЦИЙ	12
Лекция 1. ИНФОРМАЦИОННЫЕ РЕСУРСЫ И ИХ РОЛЬ В ЭКОНОМИКЕ	12
1.1. Основные понятия и определения	12
1.2. Информация и бизнес.....	13
Лекция 2. ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ	19
2.1. Истоки развитие информационных систем и технологий.....	19
2.2. Компоненты информационных систем	20
2.3. Технологии передачи и обмена информационными ресурсами	23
Лекция 3. ГЛОБАЛЬНЫЕ КОМПЬЮТЕРНЫЕ СЕТИ.....	28
3.1. Основные принципы и организационная структура Интернет.....	29
3.2. Инфраструктура сети Internet.....	29
3.3. Эталонная модель OSI	31
3.4. Протоколы сети Интернет	36
3.5. Протоколы, службы, сокеты и порты TCP/IP	40
3.6. Адресация в сети Интернет	42
3.7. Система доменных имен DNS	50
3.8. Сервисы сети Интернет.....	52
3.9. Язык разметки HTML	66
Лекция 4. ДОСТУП К ИНФОРМАЦИОННЫМ РЕСУРСАМ	67
4.1. Общие принципы поиска информационных ресурсов	67
4.2. Топология Web-пространства	75
4.3. Средства и методы навигации в Internet	78
4.4. XML – язык разметки и модель данных.....	88
Лекция 5. МИРОВОЙ РЫНОК ИНФОРМАЦИОННЫХ УСЛУГ	92
5.1. Характеристика мирового рынка информационных услуг	92
5.2. Основные секторы рынка информационных продуктов и услуг.....	92
5.3. Профессиональные базы данных и ресурсы Internet	95
Лекция 6. ГОСУДАРСТВЕННЫЕ ИНФОРМАЦИОННЫЕ РЕСУРСЫ.....	102
6.1. Общая характеристика государственных информационных ресурсов	102
6.2. Органы и организации, ответственные за формирование и использование государственных информационных ресурсов	102
6.3. Основные правовые акты, определяющие порядок формирования и использования информационных ресурсов	103
6.4. Информационные ресурсы министерств и ведомств.....	104
6.5. Информационные ресурсы региональных и муниципальных органов власти.....	104
6.6. Статистическая информация	105
6.7. Научная и техническая информация	107
6.8. Информационные ресурсы библиотечной сети России.....	108
6.9. Информационные ресурсы социальной сферы	109
6.10. Информация о природных ресурсах, явлениях, процессах	109
Лекция 7. ПРАВОВАЯ ИНФОРМАЦИЯ	111
7.1. Источники правовой информации.....	111
7.2. Основные агентства, обеспечивающие потребителя правовой информацией.....	112
7.3. Информационные справочные правовые системы	113
7.4. Технологии доступа к правовым базам данных	114
Лекция 8. БИРЖЕВАЯ И ФИНАНСОВАЯ ИНФОРМАЦИЯ.....	117

8.1. Российский рынок биржевой и финансовой информации	117
8.2. Информационные агентства биржевой и финансовой информации, предоставляемые ими услуги	118
8.3. Виды биржевой и финансовой информации, ее анализ.....	119
8.4. Программные средства анализа	120
V ПРАКТИЧЕСКИЕ ЗАДАНИЯ	122
ПЗ-1. Знакомство с поисковыми машинами	122
ПЗ-2. Расширенный язык запроса поисковых машин.....	124
ПЗ-3. Почтовые системы (электронная почта)	126
ПЗ-4. Поиск материалов для самостоятельной работы	127
ПЗ-5. Знакомство с предприятиями электронной торговли.....	128
VI ТЕСТОВЫЕ ЗАДАНИЯ.....	130
VII ВОПРОСЫ ДЛЯ ПОДГОТОВКИ К ЗАЧЕТУ	138
VIII КОНТРОЛЬНЫЕ РАБОТЫ	140
ГЛОССАРИЙ.....	143
СПИСОК ЛИТЕРАТУРЫ.....	157

ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

Цель дисциплины, ее место в учебном процессе

Данная дисциплина обеспечивает знакомство студентов с объектом дисциплины – мировыми информационными ресурсами (МИР), их местом и ролью в современных *информационных технологиях* (ИТ) и предметом дисциплины – *технологиями работы с МИР*, включающими в себя: принципы функционирования рынков информационных ресурсов; использование информационных ресурсов в информационных и коммуникационных системах областях специализации; приобретение студентами *знаний, умений и навыков* (так называемых ЗУН-ов) по дисциплине в соответствии с государственным образовательным стандартом высшего профессионального образования по специальности 230700.65 – "Прикладная информатика (по областям)".

Дисциплина "Мировые информационные ресурсы" входит в цикл специальных дисциплин федерального компонента предметов специальности 230700.65 – "Прикладная информатика (по областям)". По решению совета факультета прикладной информатики этот учебный курс преподается на третьем курсе в течение 5 семестра, так как предполагает определенный уровень подготовки к работе с ИТ и обеспечивает формирование теоретической базы, умений и навыков для усвоения последующих дисциплин профессиональной образовательной программы.

Цель преподавания дисциплины – знакомство с принципами и технологиями работы с МИР, их коммерческого использования, государственной политики в области ИТ; приобретение ЗУН-ов, необходимых для решения задач по профилю будущей специальности.

Задачи дисциплины:

- ознакомить с понятием МИР: формами, видами и классификацией *информационных ресурсов* (ИР); представлением ИР в глобальных информационных сетях;
- ознакомить с понятием государственной политики в области ИР: основными направлениями информационной политики РФ и документами, определяющими ее направления; мерами по обеспечению *информационной безопасности* (ИБ) государства;
- ознакомить с законодательными актами, регулирующими область обращения ИР: причинами их выделения в собственную правовую область, действующим законодательством РФ, отдельными положениями мировой правовой системы в этой области; влиянием международного законодательства на развитие сферы обращения ИР в РФ и мире;
- освоить работу с ИР в глобальной информационной сети Internet: рассмотреть особенности представления информации в глобальных сетях, методы поиска информации с помощью каталогов ИР и специализированных поисковых систем;
- овладеть методами оптимального поиска информации с помощью поисковых систем;
- ознакомить с методами получения информации по электронной почте, помощью тематических рассылок, конференций, по запросу пользователя;
- ознакомить с применением ИР в бизнесе, методами оценки их вклада в конечный продукт бизнеса, методами оптимизации работы с деловыми ИР;
- изучить особенности торговли ИР, видами электронного денежного обращения, спецификой информационных товаров;
- ознакомить с вопросами авторского права применительно к ИР и товарам, влиянием этого фактора на развитие информационной экономики;
- ознакомить с проблемами обеспечения ИБ в Internet и Intranet, принципами достаточной защиты, понятиями электронной подписи, электронных сертификатов;
- приобрести практический опыт работы с ИР, выполняя творческие задания по курсу;

- повысить уровень логической культуры и дисциплины мышления студентов, их эрудицию и понимание необходимости использования ИТ на старших курсах, в учебно-исследовательской работе, курсовом и дипломном проектировании, профессиональной деятельности по специальности.

II ТРЕБОВАНИЯ К УРОВНЮ ОСВОЕНИЯ СОДЕРЖАНИЯ ДИСЦИПЛИНЫ

Выписка из государственного образовательного стандарта

В соответствии с требованиями к обязательному минимуму содержания основной образовательной программы подготовки информатика по специальности 230700.65 (ранее 351400 - "Прикладная информатика /по областям/") в дисциплину "Мировые информационные ресурсы" включены следующие вопросы:

Индекс	Наименование дисциплины и ее основные разделы	Всего часов
СД.Ф.03	МИРОВЫЕ ИНФОРМАЦИОННЫЕ РЕСУРСЫ Информация и бизнес. Рынки информационных ресурсов: особенности спроса, предложения, рыночного равновесия. Мировые информационные ресурсы: определение, классификация и характеристика основных структур (баз данных, сетей) по различным признакам. Мировые информационные сети: структура информации, правила поиска, практикум. Технология и практика взаимодействия индивидуального и коллективного пользователя с мировыми ресурсами (по отраслям) через специализированные сетевые структуры; комплексная оценка эффективности использования мировых ресурсов	100

Квалификационная характеристика выпускника

Специальность утверждена приказом Министерства образования Российской Федерации № 686 от 02 марта 2000 года. Квалификация выпускника: информатик (квалификация в области). Нормативный срок освоения основной образовательной программы подготовки по специальности 230700.65 - "Прикладная информатика (по областям)" при очной форме обучения – 5 лет.

Перечень областей применения: *экономика, юриспруденция, политология, психология, социология, политология, психология, экология, гуманитарно-социальные и другие области*, в которых применяются профессионально-ориентированные информационные системы для реализации информационных технологий в соответствии со спецификой этой области.

Объектами профессиональной деятельности информатика (с квалификацией в области) являются:

- *информационные процессы*, которые определяются спецификой предметной области;
- *события, функциональные процессы и базы данных (БД) в предметной области, действия по выработке управленческого решения или по разработке экспертного заключения, информационные потоки, ресурсы* (материальные, информационные и иные нематериальные, денежные и др.) в организациях, характерных для предметной области: органы государственного и муниципального управления, финансовые и экономические учреждения, органы налогообложения, органы правопорядка и социальной защиты, воспитательные и образовательные учреждения, суды, органы юстиции, таможня, образовательные и воспитательные учреждения, информационные центры, архивы, фонды и библиотеки, органы государственной статистики, органы управления на предприятиях различных организационно-правовых форм (администрация, бухгалтерия, экономические отделы, служба юрисконсульта и др.);

- *новые направления деятельности в области применения, которые требуют внедрения компьютерного оборудования, локальных вычислительных сетей и (или) средств выхода*

в глобальные информационные сети для осуществления сбора, хранения, анализа обработки и передачи информации, необходимой для обеспечения функциональных процессов;

- профессионально-ориентированные информационные системы (ПОИС);

- информационные системы (ИС) в административном управлении, ИС в банковском деле, ИС в страховом деле, ИС в налогообложении, ИС в бухгалтерском учете и аудите, ИС фондового рынка, ИС в антикризисном управлении, ИС в таможенном деле, ИС в оценочной деятельности, ИС в маркетинге и рекламе;

в области юриспруденции:

- информационные и коммуникационные системы в оперативной деятельности, ИС в судебной экспертизе, ИС распознавания образов (в юридической деятельности), правовые информационно-справочные системы по видам юридической деятельности, ИС в арбитражном судопроизводстве.

Требования к уровню освоения содержания дисциплины

В результате изучения дисциплины "Мировые информационные ресурсы" студенты, выпускаемые по специальности 230700 – "Прикладная информатика" должны быть готовы к использованию международных информационных ресурсов и решению задач:

- обеспечения *информационной безопасности* (ИБ) функционирования информационной системы при взаимодействии с информационными рынками по сетям или с использованием иных методов обмена данными,

- оценки эффективности приобретаемого программного обеспечения и информационных баз данных для предметной области; а также,

иметь представление:

- о способах представления текстовой и нетекстовой информации в ИС, использовании средств мультимедиа и тенденциях их развития;

знать:

- об общей характеристике процессов сбора, передачи, обработки и накопления информации; о технических и программных средствах реализации информационных процессов;

информационные закономерности, специфику информационных объектов и ресурсов, информационных потребностей в предметной области;

- рынки ИР и особенности их использования;

уметь:

- находить и использовать информацию, необходимую для ориентирования в основных текущих проблемах экономики;

- использовать информационно-поисковые средства локальных и глобальных вычислительных и информационных сетей; ИТ и знания общей информационной ситуации, ИР в предметной области;

- применять полученные знания при изучении дисциплин по специальности, в курсовом и дипломном проектировании, в НИРС и в профессиональной деятельности по окончании вуза;

обладать навыками:

- поиска нужной информации в вычислительных сетях и проведения самостоятельных научных исследований.

К аудиторным занятиям по предмету "Мировые информационные ресурсы" относятся лекции и практические занятия, проводимые в соответствии с календарно-тематическим планом изучения дисциплины. *Самостоятельная работа студентов* (СРС) под контролем преподавателя осуществляется путем выполнения конкретных заданий в рамках программы самостоятельной работы студентов.

На лекциях (Лк) студентам сообщаются новые и систематизируются уже имеющиеся у них знания по предмету, детализируются узловые моменты дисциплины применительно к специальности. Практические занятия (ПЗ) проводятся в форме выполнения конкретных заданий по основным (ключевым) и наиболее сложным вопросам дисциплины.

ны. ПЗ нацелены на углубление и уточнение знаний студентов, полученных ими на лекциях. На ПЗ и в ходе самостоятельной работы с ПК, учебной и научной литературой, студенты вырабатывают умения и отрабатывают нужные им навыки работы с ИР.

Качество усвоения всех разделов и тем дисциплины проверяется путем осуществления текущего и итогового контроля.

Таблица 1 – Перечень дисциплин, освоение которых необходимо для изучения курса

Индекс	Наименование дисциплины и ее основные разделы
<i>ЕН. Общие математические и естественнонаучные дисциплины</i>	
ЕН.Ф.02	ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ
	Основные понятия информатики; технические и программные средства реализации информационных процессов; модели решения функциональных и вычислительных задач; алгоритмизация и программирование; языки программирования высокого уровня; основы и методы защиты информации; компьютерный практикум; информационная технология; структура программного обеспечения с точки зрения пользователя, средства и алгоритмы представления, хранения и обработки текстовой и числовой информации; программные среды; организация и средства, человеко-машинного интерфейса, мультисреды и гиперсреды; назначение и основы использования систем искусственного интеллекта; понятие об ИТ на сетях; понятие об экономических и правовых аспектах ИТ. Основные этапы компьютерного решения задач; модульные программы; объектно-ориентированное программирование; критерии качества программы; диалоговые программы; дружелюбность; основы программирования в телекоммуникациях и распределенной обработки информации; постановка задачи и спецификация программы; способы записи алгоритма; стандартные типы данных; представление основных структур: итерации, ветвления, повторения; процедуры: построение и использование; типы данных, определяемые пользователем; записи; файлы; динамические структуры данных; списки: основные виды и способы реализации; программирование рекурсивных алгоритмов; способы конструирования программ; основы доказательства правильности; архитектура и возможности семейства языков высокого уровня
<i>ОПД. Общепрофессиональные дисциплины</i>	
ОПД.Ф.01	ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ, СЕТИ И ТЕЛЕКОММУНИКАЦИИ Общие принципы построения и функционирование компьютера, его архитектура. Информационно-логические основы ЭВМ, ее функциональная и структурная организация, память, процессоры, каналы и интерфейсы ввода вывода, периферийные устройства, режим работы и ПО. Типовые вычислительные структуры, их ПО, режимы работы. Классификация и архитектура типовых вычислительных структур и сетей. Техническое, информационное и программное обеспечение вычислительных сетей (глобальных, региональных, локальных), организация их функционирования. Системы телекоммуникаций: коммутация и маршрутизация, цифровые сети связи. Эффективность функционирования ЭВМ, вычислительных систем и сетей телекоммуникаций; пути ее повышения. Перспективы развития вычислительных средств. Технические средства человеко-машинного интерфейса

III СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Раздел I. КОНЦЕПЦИЯ ДИСЦИПЛИНЫ

Информации, ее определение, виды и свойства. Информационные процессы. Информационные революции. Объект, предмет и задачи дисциплины.

Информатизация общества. Информатизация как совокупность процессов и явлений, связанных с целенаправленной обработкой информации, связанной с применением средств вычислительной техники и связи, информационных технологий и программного обеспечения. Окинавская Хартия Глобального информационного общества.

Мировые информационные ресурсы: определение, классификация и характеристика основных структур (баз данных, сетей) по различным признакам.

Определение информационных ресурсов (ИР), представление информации в виде ИР. Формы, виды, классификация, измерение ИР. Представление ИР в глобальных информационных сетях. ИР России. Мировые информационные ресурсы. Предмет и задачи курса.

Раздел II. ЗАКОНОДАТЕЛЬСТВО, РЕГУЛИРУЮЩЕЕ ОТНОШЕНИЯ В ОБЛАСТИ МИР

Государственная политика в области ИР. Понятие. Средства обеспечения государственной политики. Информационная политика РФ. Федеральная целевая программа «ЭЛЕКТРОННАЯ РОССИЯ на 2002–2010 годы». Национальный доклад «Информационные ресурсы России». Информационное неравенство. Информационная безопасность государства. Меры по обеспечению информационной безопасности. Доктрина информационной безопасности Российской Федерации.

Законодательство РФ в области ИР. Предпосылки государственного регулирования в области информационных технологий. Федеральные законы: «О праве на информацию», «Об информации, информатизации и защите информации», «Об электронной цифровой подписи», «О правовой охране программ для ЭВМ и баз данных», «Об обязательном экземпляре документов».

Законодательство ЕС и США в области ИР. Особенности правового регулирования мировых ИР. Влияние международного законодательства на развитие сферы обращения ИР в РФ и мире. Национальное и международное законодательство. Конфликты юрисдикции, связанные с характером мировых ИР.

Проблемы правового регулирования. Авторские и смежные права. Законы США об авторских правах в цифровую эпоху. Влияние международных и национальных коммерческих организаций на обращение ИР. Проблемы налогообложения сетевых коммерческих организаций. Информационное пиратство.

Раздел III. МИРОВЫЕ ИНФОРМАЦИОННЫЕ СЕТИ: ИХ СТРУКТУРА, ПРАВИЛА ПОИСКА

Глобальные информационные сети. Причина возникновения и история развития. Классификация информационных сетей. Корпоративные сети. Образовательные сети. Общедоступные и коммерческие сети. Обзор технологий размещения и доступа к информации в информационной сети. Сети BBS, FIDO, Bitnet, SprintNet, AOL, Internet 2, Next Generation Internet.

Сеть Internet. История образования. Принципы организации. Преимущества и недостатки. Способы размещения информации в Internet. Файловые архивы. Сайты. Конференции. Рассылки. Способы доступа к информации в Internet. Организация информационного трафика. Открытая и закрытая информация в Internet. Обращение к специализированным базам данных и информационным сетям. Получение информации по запросу, подписка на рассылки. Настройка динамических страниц. Обмен информацией в сети. Peer-to-peer.

Поиск информации в сети Internet. Каталогизация информации в сети Internet. Адресация в сети Internet. Проблема поиска информации и сортировки найденного по релевант-

ности. Особенности поиска текстовой информации. Кодировки, форматы записи. Поиск файлов, визуальной, аудио и мультимедийной информации. Каталоги ИР. Рейтинг ИР.

Поисковые системы (ПС). Принципы работы ПС. Понятие запроса к ПС. Вид запроса в ПС. Обработка, каталогизация, индексирование, сортировка, анализ информационных ресурсов, выдача информации по запросу, выделение главного и отсеечение второстепенного, повторяющегося (резюмирование) первичная аналитическая обработка запроса, уточнение требований к выборке информации. Возможности ПС. Классификации ПС. Сравнение ПС, действующих в русскоязычной части Internet. Международные ПС. Меры по оптимизации поиска.

Электронная почта. Принцип работы. Виды предоставляемых услуг. Получение информации при помощи почтовых служб. Уязвимости работы с почтой: получение вирусов, spam.

Раздел IV. ИНФОРМАЦИЯ И БИЗНЕС

Торговля в сети Internet. Торговля традиционными товарами. Преимущества и недостатки сетевого (Internet) магазина. Особенности товарного предложения. Способы организации торговли. Доставка товара. Система оплаты. Клиент сетевого магазина. История развития сетевой торговли. Сетевая торговля в РФ. Кризис DOT-COMов.

Реклама в сети Internet. Виды рекламы. Особенности Internet рекламы. Эффективность Internet рекламы.

Оплата в сети Internet. Способы оплаты в сетевых магазинах. Платежные системы в Internet, электронные деньги. WebMoney, PayCash, Yandex. Деньги. Особенности денежного обращения в сети Internet.

Рынки информационных ресурсов: особенности спроса, предложения, рыночного равновесия. Особенности торговли ИР: информация, музыка, документы, программные продукты.

Виды предоставления пользователю прав на программный продукт. Платный продукт, бесплатный продукт (freeware), пробный продукт (demo, shareware), с рекламной нагрузкой (bannerware), со сбором информации о покупателе (spyware).

Ограничение покупателя в правах пользования приобретенным виртуальным товаром: по времени, по функциональности, по технической поддержке, по числу использований, по передаче другому пользователю, по изменению и совершенствованию, по географическому, национальному признаку. Товары для личного и коммерческого использования.

Виртуальные товары и услуги. Особенности виртуальных товаров и услуг. Виртуальная собственность.

Авторское право и коммерция в сети Internet. Вопросы отчуждения собственности на информационный продукт. Виды лицензий. GNU.

Раздел V. ОБЕСПЕЧЕНИЕ БЕЗОПАСНОСТИ В РАБОТЕ С ИР

Несанкционированное использование программных продуктов и ИР (информационное пиратство) в РФ. Понятие информационного пиратства. Причины и следствия пиратства. Государственная политика по борьбе с пиратством. Борьба корпорации Microsoft с пиратством.

Несанкционированный доступ к ИР и информационному трафику. Определение несанкционированного доступа. Меры по защите информации. Аппаратные и программные средства защиты ИР.

Электронная цифровая подпись (ЭЦП). Анонимность в сети Internet. Потребность в авторизации потребителя сетевых услуг. Понятие ЭЦП. Сфера обращения ЭЦП. Международное и российское законодательство об ЭЦП. Примеры использования ЭЦП в РФ.

Защита от сетевых вирусов. Особенности компьютерных вирусов в информационных сетях. Классификация сетевых вирусов. Сетевые черви, трояны, шпионские модули.

Меры безопасности при работе с ИР. Меры безопасности при работе с электронной почтой. Программные средства защиты информационных ресурсов.

Раздел VI. ИСПОЛЬЗОВАНИЕ ИР В ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ

Технология и практика эффективного взаимодействия индивидуального и коллективного пользователя с ИР через специализированные сетевые структуры. Особенности работы индивидуального и коллективного пользователя с ИР. Использование специализированных, корпоративных ИР. Учет использования ИР в профессиональной деятельности. Оценка затрат на использование ИР на конечную стоимость выполнения работ. Комплексная оценка эффективности использования мировых ИР. Планирование использования ИР. Средства управления проектами. Обзор. Назначение. Сфера применения.

Оценка экономической эффективности. Технология microsoft.net. Определение. Предпосылки возникновения. Особенности работы на платформе microsoft.net. Альтернативные разработки фирмы Sun.

IV КУРС ЛЕКЦИЙ

Лекция 1. ИНФОРМАЦИОННЫЕ РЕСУРСЫ И ИХ РОЛЬ В ЭКОНОМИКЕ

Вопросы:

- 1.1. Основные понятия и определения
- 1.2. Информация и бизнес

1.1. Основные понятия и определения

В период информатизации общества формирование и производство информации становится важнейшим для ее полноценного применения. Понятие «*информация*» произошло от латинского слова "informatio", что означает изложение, разъяснение какого-либо факта, события, явления. В широком смысле информация определяется как сведения о той или иной стороне материального мира и происходящих в нем процессах. Информацию как продукт производства и применения отличает, прежде всего, предметное содержание. Она очень разнообразна и подразделяется по виду обслуживаемой ею человеческой деятельности: научная, техническая, производственная, управленческая, экономическая, социальная, правовая и т.п. Каждый из видов информации имеет свои технологии обработки, смысловую ценность, формы представления и отображения на физическом носителе, требование к точности, достоверности, оперативности отражения фактов, явлений, процессов.

Важнейшие свойства экономической и управленческой информации:

- достоверность и полнота;
- ценность и актуальность;
- ясность и понятность.

Информация *достоверна*, если она не искажает истинные факты. Недостоверная информация приводит к неправильному пониманию или принятию неверных решений. Информация является *полной*, если ее достаточно для понимания и принятия решений. Неполная информация сдерживает процесс принятия решений и может повлечь ошибки.

Ценность информации зависит от того, какие задачи решаются с ее помощью. Чем выше уровень управления, тем выше цена принимаемого решения, а, следовательно, и ценность информации.

Актуальную информацию важно иметь при работе в постоянно изменяющихся условиях. Актуальность информации зависит от временных параметров ее получения. Информация становится *ясной* и *понятной*, если она выражена языком, на котором говорят те, кому предназначена эта информация. Если ценная и актуальная информация выражена в неясной и непонятной форме, то она *бесполезна*.

Управленческая информация обслуживает процессы производства, распределения, обмена и потребления, материальных благ и обеспечивает решение задач организационно-экономического управления народным хозяйством и его звеньями. Она представляет собой разнообразные сведения экономического, технологического, социального, юридического, демографического и другого содержания. В информационном процессе, каким является управленческая деятельность, информация выступает одним из важнейших ресурсов наряду с энергетическими, материальными, трудовыми и финансовыми ресурсами.

Экономическая информация является важнейшей составляющей управленческой информации, представляющая собой совокупность сведений экономического характера, которые можно фиксировать, передавать, обрабатывать, хранить и использовать в процессе планирования, учета, контроля, анализа, прогнозирования и принятия решений на всех уровнях отраслевого и регионального управления. Экономическая информация включает

сведения о состоянии объекта управления на определенный момент времени. Экономическая информация отражает деятельность предприятий и организаций посредством различных показателей: натуральных, стоимостных и безразмерных (индексы и коэффициенты).

Для экономической информации характерны:

- большие объемы;
- многократное повторение циклов ее получения и преобразования в установленные периоды (месяц, квартал, год, и т.д.);
- многообразие источников и потребителей;
- значительный вес логических операций при ее обработке.

Информация выступает одним из *главных ресурсов жизнедеятельности* в больших и сложных человеко-машинных системах, составляющих инфраструктуру современного общества. Являясь связующим звеном между разными видами интеллектуальной и материальной деятельности коллективов людей, между управлением и производством, информация, в отличие от других видов ресурсов (природных), не убывает со временем, а наоборот постоянно увеличивается, создавая условия для накопления опыта, способствуя выработке обоснованных управленческих решений.

Информационные ресурсы – отдельные документы и отдельные массивы документов, а также документы и массивы документов в информационных системах (библиотеках, архивах, фондах, банках данных, других информационных системах). Настоящее определение приведено в законе «Об информации информатизации и защите информации».

Расширяя определение информационного ресурса, данное в законе «Об информации информатизации и защите информации», можно сказать, что *мировые информационные ресурсы* – это документы и массивы информации, находящиеся в глобальных компьютерных сетях, доступ к которым не имеет территориальных ограничений и контролируется только источником информации.

Понятно, что документы и массивы информации, о которых говорится в этом законе, не существуют сами по себе. В них в разных формах представлены знания, которыми обладали люди, создававшие их. Таким образом, *информационные ресурсы – это знания, подготовленные людьми и зафиксированные на материальном носителе.*

Таким образом, накопленная и систематизированная информация с соответствующими средствами ее хранения, накопления, просмотра становится объектом купли продажи и, как любой другой *производственный ресурс*, получает оценку своей потребительской полезности в виде стоимости.

1.2. Информация и бизнес

Информация всегда была необходима для эффективного управления, но революция, произошедшая в коммуникационных системах, увеличила объем доступной информации и сделала процесс управления информацией более сложным и важным для фирмы. Задача бизнес-информации заключается в предоставлении менеджерам некоторой основы знаний как о внутренней среде фирмы, связанной с ее работой, так и о внешней, в которой функционирует фирма. Основная цель сбора информации состоит в том, чтобы на ее основе как можно точнее, с минимальным уровнем неопределенности, формировать знания и принимать решения. Соответственно при получении бизнес-информации акцент должен быть сделан на сборе реальных, экономических, детерминированных и безусловных фактов. Однако этот акцент может привести к сокращению числа альтернатив (а следовательно, и потенциала информации) за счет исключения из рассмотрения всей информации, не отвечающей требованию строгой «релевантности». Один из парадоксов современных взглядов на информацию состоит в том, что чем выше неопределенность, тем больше возможностей для дивергенции и направлений дальнейшего развития, вытекающих из этих возможностей; фактически тогда больший объем информации приводит в некоторых случаях к сокращению числа таких направлений. Эта проблема хорошо известна, однако вопрос о том, «какую информацию использовать?», остается очень непростым.

Другая проблема, возникающая при работе с бизнес-информацией, заключается в том, что, хотя сама информация находится под влиянием детерминистской экономической парадигмы, решения часто принимаются на основе впечатлений и невероятных примеров. Исследовательская работа с предпринимателями, менеджерами, специалистами по инвестированию и другими сотрудниками выявила большое доверие к «интуиции» и «внутреннему чувству». Таким образом, в некоторых случаях, возникает несоответствие между бизнес-информацией (которая основана на фактах) и деловым знанием (которое часто основано на интуиции).

Последняя группа вопросов связана с временными ограничениями. Бизнес-информация в основном описывает то, что произошло в недавнем и не очень далеком прошлом; даже новости, описывающие «настоящее», относятся, строго говоря, к самому недавнему прошлому. Полезность официальной статистики особенно уменьшается из-за временного лага между описываемыми событиями и публикацией информации. Тем не менее менеджерам все еще не хватает информации о том, почему произошли какие-то события или почему они должны произойти.

Виды бизнес-информации. Бизнес-информацию обычно классифицируют одним из двух способов.

Первый способ – по размещению источника информации, различие состоит в том, расположен источник информации внутри фирмы или же он внешний по отношению к ней.

Вторым способом информацию можно классифицировать по ее предназначению. Это самая распространенная таксономия, так как компании, которые намереваются собирать информацию, делают это для достижения какой-то конкретной цели.

Внешняя информация

1. *Рыночная информация.* Это огромная и сложная область, и рыночная информация сейчас по праву является основным товаром для фирм, начиная от рекламных агентств и заканчивая банками (а также для специализированных компаний, занимающихся рыночными исследованиями), конкурирующих за право предоставить информацию фирмам. Рыночная информация может быть очень разной – от конкретной (как, например, анализ структуры магазинов в узком сегменте розничного рынка) до самой общей (как анализ тенденций на мировом рынке какого-нибудь товара). От рыночной информации зависят разработки компаниями новых продуктов, маркетинговое планирование, планирование коммуникаций и рекламы, решения об экспорте, а также общие стратегические решения. Природа рыночной информации меняется в зависимости от цели и от того, какой компании предназначена эта информация; категории, по которым обычно ищется информация, включают размер и рост рынка, покупательную способность, привычки, спрос и поведение потребителей.

2. *Информация о конкурентах.* Ее иногда рассматривают как часть рыночной информации, но она заслуживает отдельного рассмотрения, поскольку может влиять на принятие стратегических решений, даже если рыночные условия не учитываются непосредственно. Компании могут, например, заинтересоваться, где конкуренты находят источники сырья и специалистов, для того чтобы участвовать в конкурентной борьбе за эти источники или сохранить уже имеющиеся. Известно, что точную информацию о конкурентах получить трудно, и эта область удостоилась пристального внимания благодаря таким неэтичным действиям некоторых компаний, как промышленный шпионаж.

3. *Макроэкономическая и геополитическая информация.* Хотя информация этого вида редко напрямую воздействует на компании, она может сыграть решающую роль при разработке долгосрочной стратегии. Например, информация о разработках в Китае помогла компаниям решить, выходить на китайский рынок или нет; примечателен тот факт, что первые иностранные компании, которые вышли на китайский рынок, обладали лучшими и наиболее достоверными источниками информации о политических и экономических переменах в Китае.

4. *Информация о поставщиках.* Эта информация играет важную роль и при разработке новых продуктов, и при расчете практических результатов. Информация о поставщиках обычно концентрируется на таких аспектах, как издержки, надежность, качество и время доставки.

5. *Внешняя финансовая информация.* Компаниям, особенно крупным, со сложной структурой инвестиций и (или) со сложной финансовой структурой, требуется разнообразная информация по таким вопросам, как, например, валютные курсы, динамика курсов акций, движение на рынке капитала и т. д. Причем, желательно получать такую информацию в режиме реального времени. Основная сложность в управлении такой информации – ее структуризация, ассимиляция с корпоративными данными и анализ очень больших и чрезвычайно быстрых информационных потоков.

6. *Информация о регулировании и налогообложении.* Компаниям нужна информация об условиях регулирования для соблюдения правовых норм. Оценка условий регулирования также нужная задача, предшествующая принятию любого решения о выходе на иностранные рынки. Информация о налоговых системах важна не только для соблюдения правовых норм, но и для определения эффективного с точки зрения налоговых выплат способа ведения операций. Например, могут быть какие-то налоговые стимулы к размещению производства в одной стране, а не в другой, где планируется введение более высоких налогов.

Внутренняя информация

1. *Информация о производстве.* Характеризуется тем, что ее источники находятся внутри фирмы или корпорации. Руководству компании нужна информация о таких вещах, как эффективность производства и производительность, издержки, отходы производства и качество. Информация о производстве важна для финансового планирования, но многие компании также опираются на нее и при маркетинговом планировании; им важно знать, способно ли производство поставлять продукцию определенного качества в объеме, достаточном для выполнения службами маркетинга и сбыта своих обязательств перед потребителями.

2. *Информация о трудовых ресурсах.* На сегодняшний день это «серая зона» в управлении информацией; хотя многие фирмы наладили сложную систему получения информации о рынке, конкурентах и производстве, лишь немногие имеют схожую по сложности систему для получения информации о трудовых ресурсах. Информация о трудовых ресурсах обычно сфокусирована на таких моментах, как обучение персонала и уровень квалификации, моральное состояние персонала и расходы на обеспечение кадрами.

3. *Внутренняя финансовая информация.* Эта категория информации описывает то, что обычно называют «показателями». Она включает основную информацию из бухгалтерского баланса о прибыли и издержках, об имуществе и обязательствах, а также большой спектр финансовых показателей. Такая информация обычно дает исходную картину финансового состояния и рентабельности компании.

Классификация источников бизнес-информации. Классифицировать всю бизнес-информацию по источникам можно двумя способами. Классификация первым способом предполагает, что информация может быть либо первичной, полученной в результате исследования или анализа, инициированного фирмой, либо вторичной, полученной из каких-то уже имеющихся источников. Вторичные источники разделяются по характеру доступа к ним.

Стратегии поиска информации могут быть либо официальными, использующими специализированных поставщиков информации, либо неофициальными, использующими различные каналы; они также могут быть активными (когда ищут источники информации) или пассивными (анализ информации из уже известных источников). Комбинация нескольких таких стратегий может быть сведена в таблицу и оказаться вполне достаточной, но это зависит от объекта управления. При поиске источников информации особое внимание нужно уделять таким вопросам, как стоимость и конкретность получаемой информа-

ции, надежность и гарантии достоверности информации. К сожалению, имеется прямая зависимость (причем, обычно очень тесная) между степенью конкретности информации и стоимостью ее получения.

	Формальные	Неформальные
Активные	Торговые ассоциации Правительственные учреждения Коммерческие учреждения Библиотеки Поставщики коммерческой информации	Поставщики Заказчики Торговые ярмарки Выставки Конференции
Пассивные	Действующая государственная система, банки, статистические органы, инструкции, информационные бюллетени, каталоги торговых организаций	Знакомства Добровольно представленная литература и рассылка писем

Оценка надежности информации (с заданием допустимой погрешности) должна включать источник, срок, в течение которого информация остается актуальной, и ее анализ. Репутация источника – это одно, а использование доказательств, подтверждающих ее, или «триангуляция» информации, – совсем другое. Наиболее распространенными источниками бизнес-информации являются:

- правительственные учреждения. Правительственные учреждения могут служить хорошим источником необработанных данных, особенно рыночных и макроэкономических, а также информации, касающейся регулирования. В большинстве случаев эти данные нужно интерпретировать перед использованием в анализе.
- библиотеки – разнообразная, редко меняющаяся информация статического характера;
- торговые ассоциации – периодические каталоги товаров;
- компании, занимающиеся частными исследованиями и информацией. Число компаний, занимающихся частными исследованиями и информацией, за последние годы выросло в несколько раз, они специализируются на сборе информации о рынках и конкурентах.
- газеты и журналы. Газеты и журналы представляют собой источник бесплатной (обычно), но не фильтрованной информации по широкому кругу вопросов.
- службы бизнес-информации. Службы бизнес-информации чаще всего предоставляют услуги по поиску специальной и детализированной информации на основе подписки или продажи. Обычно это коммерческие организации, но такие услуги предоставляют и академические учреждения (бизнес-школы, исследовательские институты).
- базы данных в режиме он-лайн. Развитие информационных технологий увеличило число и повысило уровень баз данных на CD-ROM и в режиме он-лайн, предоставляющих как количественную, в частности статистическую и подробную финансовую информацию, так и качественную, такую как информация о рынках и конкурентах, которую предоставляют службы новостей.
- Интернет со своей всемирной паутиной предлагает невероятное количество информационных сайтов. Для подписчиков количество услуг в Интернет, а также в частных сетях, таких как Microsoft Network, продолжает расти, обеспечивая доступ к конкретной, очень подробной информации. Группы определенных интересов, блоки новостей и доски объявлений в Интернете позволяют выделять пучки узкоспециализированной информации и таким образом увеличивать полезность доступной информации. Обмен информацией по всему миру в режиме реального времени способствует установлению контактов и поддержанию взаимоотношений.

Роль бизнес-информации. Потребность в информации неявно присутствует в каждом действии менеджера.

Существует представление, что вся работа управленческого персонала в современном офисе делится на шесть основных частей:

- ввод, сбор необработанных данных;
- коммуникации, перемещение информации от одного источника к другому;
- обработка, преобразование информации из одной формы в другую;
- сохранение обработанной информации;
- поиск – процесс доступа к хранящимся данным;
- вывод, формирование данных или информации в виде, удобном для пользователя.

Современный взгляд на управление как на способность контролировать информацию и использовать знание является общепринятой точкой зрения. Прагматический подход гласит, что информация – это функция, основная цель которой заключается в развитии рынков и увеличении прибыли. Организации используют информацию в четырех основных целях:

- для создания возможности функционирования или для роста прибыли и расширения рынков;
- для снижения риска и уменьшения неопределенности;
- для получения власти и средств воздействия на других;
- для контроля и оценки производительности и эффективности своей фирмы.

Из этих целей самой важной является снижение риска. Теоретически можно принимать решения, вообще не основываясь ни на какой информации: так поступают некоторые биржевые спекулянты, хотя опытные биржевые спекулянты при принятии решений используют такую информацию, как расчет вероятностей. Информация может быть разрозненной и очень неполной, но она, тем не менее, должна иметься в наличии. В то же время многие организации вынуждены решать, когда информации для эффективного управления уже достаточно; способность управлять информацией так, чтобы получать от нее пользу является важным навыком для любого менеджера.

Отсюда следует, что чем больше полезной информации есть у компании, тем меньше уровень риска в ее последующих действиях. Например, большинство маркетинговых исследовательских программ нацелено на получение информации о покупателях, моделях потребления, идентификации как можно больше индивидов на целевом рынке, т.е. получении как можно большего количества единиц информации.

Ситуация, когда вся доступная информация есть у всех участников рынка, известна как эффективный рынок; на таком рынке нет неопределенности, а так как ввиду отсутствия конкуренции все участники рынка действуют рационально, на нем нет риска. Рынками, наиболее приближенными к эффективному рынку, являются рынки капитала, а особенно рынки акций. Однако даже на рынках капитала нет информационной эффективности, а на рынках потребительских товаров и услуг эффективный рынок невозможен чисто логически. Даже в случае полной доступности информации для всех участников рынка остается проблема ее эффективного использования. Но пока существует асимметрия информации, фирмы, обладающие большим количеством информации, чем их конкуренты, будут господствовать над ними. Информация позволяет компаниям действовать первыми, а если у них есть более надежное знание, чем у конкурентов, то информация дает им более высокий шанс завоевать рынок.

Эффективность использования информационных ресурсов. Главной движущей силой сбора и анализа бизнес-информации является повышение конкурентоспособности. Для эффективного использования информации фирмы чаще всего используют компьютерные и телекоммуникационные технологии. Потенциальные выгоды от эффективных информационных и коммуникационных систем включают те, что приведены Фулером и Дженкинсом {Fuller & Jenkins, 1995} в таблице 2. Пользу от улучшения процесса обмена информацией можно описать как улучшение взаимоотношений с клиентами и поставщи-

ками, оптимизацию операций в бизнесе, а также более быстрый доступ к ресурсам и услугам экспертов.

Повышение качественного уровня управленческих решений	За счет высокого качества информации, передаваемой и получаемой от внутренних и внешних собственников (клиентов, поставщиков, филиалов и т. д.). Характеристики качества включают скорость, распределение во времени, физические свойства, безопасность, доступ, ясность и т. д.
Ускорение оборачиваемости ресурсов	Транзакции осуществляются либо скоростью, либо с меньшими издержками
Денежные потоки	Ускорение за счет, например, переводов средств, процедур управления средствами и процедур подписания контрактов
Рентабельность	Увеличивается из-за меньших издержек. Доступ к большему количеству клиентов, доступ к более выгодным клиентам, большие возможности, предоставляемые переговорами по телекоммуникационным средствам связи
Конкурентная стратегия	За счет разработки более качественных, своевременных и выгодных предложений своим клиентам благодаря использованию информационных технологий и телекоммуникаций
Повышение компетентности в профилирующей области	За счет получения возможности опереться через электронные средства массовой информации на компании, лучшие в своей области
Улучшение доступа	К информации о возможностях на рынках, поставщиках, ценах, предложениях и ресурсах

Однако сейчас все чаще стали признавать, что одной информации недостаточно. Для того чтобы использовать информацию эффективно, фирмы должны обладать механизмами трансформации информации в знание и использования этого знания как имущества. Несмотря на то, что десять лет назад учебники по менеджменту и статьи в журналах провозглашали информацию как самый ценный товар, сейчас акцент делается на знания. Успех компании в будущем, по-видимому, зависит от того, насколько правильно компания ими управляет

Контрольные вопросы к лекции 1

1. Что из себя представляет информация?
2. Каковы основные понятия и определения мировых информационных ресурсов (МИР)?
3. Каковы место и роль МИР в экономике и юриспруденции?
4. Как информация способствует развитию бизнеса?

Лекция 2. ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Вопросы:

- 2.1. Истоки развитие информационных систем и технологий
- 2.2. Компоненты информационных систем
- 2.3. Технологии передачи и обмена информационными ресурсами

2.1. Истоки развитие информационных систем и технологий

Индустрия информации и знаний предоставляет компоненты, необходимые для создания информационных и основанных на знаниях систем, а также поставляет данные, информацию и знания этим системам и их клиентам.

Начиная с 1950-х гг. эта индустрия быстро росла, двигаясь от специализированной арены, доступной только крупным компаниям и государственным департаментам, к массовым мировым рынкам коммерческих, общественных и персональных систем. Росту индустрии информации и знаний способствовали следующие факторы:

1. *Перемещение экономики от промышленного производства к производству услуг.* Часто встречающиеся фразы «индустриальная эпоха» и «информационная эпоха» характеризуют различие в сущности экономики первой и второй половины XX в. Индустрия обслуживания, такая как банковское дело и страхование, гораздо больше зависит от информационных систем, чем от традиционных производственных систем.

2. *Глобализация рыночной экономики* вызвала растущую потребность в информации и в более быстрых и надежных методах коммуникации. Некоторые отрасли промышленности преуспели в этом отношении. Например, функционирование глобального фондового рынка, товарного рынка, инвестиционного рынка стимулировало усовершенствование передачи информации по всему миру.

3. Создание экономических союзов и заключение глобальных торговых соглашений способствовало росту глобальной конкуренции и расширению региональных рынков, например, расширение Единого европейского рынка, Североамериканское соглашение о свободной торговле (North American Free Trade Agreement (NAFTA)), Создание Всемирной торговой ассоциации (ВТО). Глобальная конкуренция создала еще более высокую потребность в точной и своевременной информации, быстро передаваемой в различные географические районы.

4. Вышеупомянутые факторы и высокая потребность в усовершенствованных технологиях и системах стимулировали спрос на информационные системы. Одновременно со спросом интенсивно росло предложение, вызванное стремительным прогрессом в области телекоммуникаций, компьютерных сетей, аппаратных средств, программного обеспечения и методов программирования.

5. Развитие систем искусственного интеллекта, экспертных систем и нейронных сетей, содержащих инструментальные средства и методы для создания интеллектуальных систем для принятия решений и обработки информации. Это переместило индустрию информации из пассивной роли поставщика информации в область формализованных, сохраняющихся и активно применяющихся знаний для решения задач и процессов принятия решения.

6. Повышение уровня знаний и навыков работы с компьютером новых поколений коммерческих и индивидуальных потребителей информации и знаний по сравнению с предыдущими поколениями. Они ожидают большего от качества и стоимости всех видов услуг, включая информацию и знания. Эти ожидания стимулируют создание технологий, которые бы удовлетворяли и превосходили их.

7. Удаление организационных культур от централизованного, иерархического и бюрократического подхода к распределенному, единообразному и предпринимательскому

стилю управления. Это изменение стиля требует обширных информационных систем, которые объединяют различные организационные единицы и собирают информацию, обрабатывают ее и распространяют по территориально разбросанным единицам.

8. Логистическая координация стала одним из способов получения конкурентного преимущества на рынке. Логистической организации бизнеса во многом способствует новаторское использование информационных систем. Например, концепция логистической интеграции предприятий стала возможной только благодаря новаторскому использованию информационных систем и инновационным бизнес-процессам. В таких организациях поставщику доступна информационная система клиентов, а могут быть доступны и их производственные планы и данные об уровнях запасов. Это позволяет решить, когда производить и поставлять материалы заказчику. Таким образом, организации заказчика не надо хранить излишки запасов.

9. Возросшая потребность в высокопрофессиональных знаниях и навыках привели к тому, что специалистов стало очень дорого (а иногда и невозможно) заменить, а неопытные работники стали менее полезными. Это создало потребность в компьютерных системах, которые могут аккумулировать знания профессионалов и имеют достаточный интеллект для руководства неопытными работниками, чтобы приблизить их производительность к уровню производительности специалиста.

10. Идея об информационной супермагистрале предполагает объединение информационной технологии с технологией развлечений и образования, таким образом открывая новые горизонты для проникновения индустрии информации и знаний в повседневную жизнь потребителей.

2.2. Компоненты информационных систем

Структура информационных систем основана на следующих компонентах:

- Данные, информация и знания.
- Аппаратные средства.
- Прикладное программное обеспечение.
- Языки программирования.
- Средства анализа и проектирования.
- Коммуникации.
- Системные разработчики и обслуживающий персонал.

Данные, информация и знания. Данные составляют необработанный вход в информационные системы. Информационные системы, обрабатывая данные, преобразуют их в удобную для использования информацию. Раньше данные в информационных системах были количественными, а их обработка ограничивалась вычислением, сводкой и представлением в виде заранее установленных форм отчетов, позднее преобразованных в запросы с управлением через меню в режиме реального времени.

Характер входных данных изменился во многих отношениях. Сейчас привлекает внимание идея о знаниях как о важной составляющей информационных систем. На смену старым пришли системы с закодированными знаниями, называемые системами, основанными на знаниях, или экспертными системами. Данные для этих систем, в сущности, являются и количественными, и качественными и уже не ограничиваются числами и текстами. Графика, рукописные тексты и чертежи, звуки, изображения и видео представляют

собой совсем другие средства для передачи и преобразования данных в информацию. Хранение, обработка и использование информации и знаний, созданных с помощью этих новых форм, расширяют требования к ресурсам компьютера далеко за пределы традиционных масштабов. Это, в свою очередь, предъявляет большие требования к аппаратным средствам, коммуникационным возможностям и скорости, требует гораздо более сложного программного обеспечения и нуждается в новых способах анализа и проектирования информационных систем. Более того, новые носители размывают границы между бизнесом, развлечением и индивидуальными услугами.

Аппаратные средства. В настоящее время в архитектуре аппаратных средств преобладает система клиент-сервер. Системы клиент-серверы состоят из большого количества рабочих станций или персональных компьютеров (РС), которые подсоединяются (как клиенты) через компьютерную сеть к одной или нескольким мощным рабочим станциям или РС (называемым серверами) с гораздо более высокой емкостью запоминающего устройства и вычислительной мощностью. Клиентские рабочие станции или РС могут работать независимо на локальном уровне, но также могут иметь доступ к программному обеспечению и системам, расположенным на серверах. В настоящее время серверы представляют собой растущий рынок в индустрии аппаратных средств, потому что компании готовы платить за мощные серверы, которые могут быстро и рационально обслуживать больше клиентов.

С ростом доступности возможных периферийных устройств, объединяющих разнообразные типы доступа пользователей появилась возможность обработки разнородной информации в одной компьютерной системе. Присоединяемые периферийные устройства, такие как сканеры, графические принтеры с высоким разрешением, устройства для записи CD, распознаватели голоса или речи, zip-драйверы и устройства для архивации увеличивают возможности рабочих станций и РС. Таким образом, успехи в области технологии аппаратных средств РС и рабочих станций сделали их мощным и доступным компонентом обработки информации и на коммерческом, и на персональном уровне. Наряду с рабочими станциями сегодня широко используются портативные компьютеры. Портативные компьютеры позволяют работникам информационной системы находиться в постоянном контакте с системой и работать в любое время и в любом месте. К настоящему времени еще не использованы все возможности и не оценено все влияние портативных компьютеров на рабочее окружение. Поскольку информационные системы охватывают все аспекты коммерческих функций и индивидуальных задач, мобильные компьютеры начали менять вид офисов и рабочего окружения, а также человеческих взаимоотношений, беспрецедентно и непредсказуемо.

Прикладное программное обеспечение. Программное обеспечение является другой составляющей информационных систем и либо приобретается для системы, либо создается с нуля (т.е. изготавливается на заказ для конкретной системы). Современные программные продукты для информационных систем увеличили их мощность, практичность, интеграцию и разнообразие.

Реляционные базы данных упростили процесс создания и проектирования программного обеспечения. Базы данных и электронные таблицы увеличили скорость, емкость и эффективность, они также приобрели больше возможностей, таких как графический интерфейс, автоматическая генерация кода, статистический анализ данных и графический вывод данных. Сейчас на рынке РС, рабочих станций, клиентских серверов и универсальных вычислительных машин большинство баз данных – реляционные.

Достижения в технологиях и методологиях программного обеспечения изменили сущность программного обеспечения информационных систем. Объектно-ориентированные базы данных и метасистемы баз данных дополнили традиционное программное обеспечение для информационных систем новыми размерностями и уровнями сложности.

Языки программирования. Языки программирования развивались от полностью машинно-ориентированных языков, в которых код был близок к машинным методам обработки, к языкам, более ориентированным на человека, в которых машинный код приблизился к способу работы людей. Языки программирования подразделяются на четыре поколения, формируя спектр, на одном конце которого находится машинная ориентация, а на другом – ориентация на человека. Традиционные языки, такие как FORTRAN, COBOL, BASIC, Pascal и C, попадают в третье поколение, в то время как машинные языки принадлежат к первому поколению, а символьные языки, такие как ассемблер, – ко второму. К языкам четвертого поколения (называемым 4GL) относятся те из них, которые сокращают

требования к кодированию посредством добавления макрокоманд для задач общего характера, делая их, таким образом, более удобными для кодирования, например, языки, входящие в пакет Visual Studio 2003 – C++, C#, Visual Basic.Net, J# частично отвечают этим требованиям.

В развитии языков программирования существуют несколько основных тенденций. Первая тенденция представляет собой парадигму программирования не просто как кодирования, а как основы для моделирования реального мира, как можно более близкой к нему. Такую основу содержит объектно-ориентированный подход.

Вторая тенденция направлена на усовершенствование стандарта для языка SQL (языка структурированных запросов), разработанного исключительно для обработки информации в базах данных. SQL является языком поиска и запросов, впервые разработанным IBM для обработки реляционных баз данных. Постепенно он был расширен и фактически стал Индустриальным стандартом.

Третья тенденция – использование языка XML. Для эффективного представления и последующего поиска знаний в Internet необходимо связать воедино информационные ресурсы, позволить программным приложениям устанавливать связи между объектами и обмениваться информацией на одном языке. Именно для решения проблем внешнего и смыслового описания документов был разработан новый стандарт хранения и обеспечения доступа к информации, который получил название XML – "расширенный язык разметки". Благодаря своей простоте и расширяемости, XML практически сразу получил широкое распространение и послужил основой для создания других специализированных языков разметки. XML – это удобный формат обмена данными, язык описания документов, при этом спецификацию XML можно рассматривать как реализацию иерархической модели данных. Такие компании как Oracle, IBM, Microsoft разрабатывают технологии, улучшающие возможности работы их реляционных СУБД. Но XML позволяет создавать СУБД нового поколения, где все данные хранятся в XML-формате.

Средства анализа и проектирования. Информационные системы представляют собой дорогостоящие проекты, которые требуют значительных финансовых и людских ресурсов. Для создания информационной системы необходимо проанализировать существующие бизнес-процессы и процедуры, и сделать вывод о необходимости их автоматизации. Этот процесс называется системным анализом. Если системный анализ показал необходимость создания новой системы, ее надо спроектировать. Проектирование имеет два этапа: логический и физический. На этапе логического проектирования проект системы создается без рассмотрения конкретных аппаратных средств, программного обеспечения или людских ресурсов. Физическое проектирование подробно определяет, как нужно создать и реализовать систему.

На этапах анализа и проектирования разработки системы используют CASE-инструменты (автоматизированное проектирование программного обеспечения). Это программные продукты, содержащие ряд методов, которые помогают системным компоновщикам при анализе и проектировании системы. Некоторые наиболее передовые продукты CASE автоматически генерируют все или почти все необходимые компьютерные коды непосредственно из диаграмм системы (такие коды обычно имеются в одном из языков программирования).

Большее значение при разработке информационных систем стало придаваться анализу требований пользователя и созданию удобных для него интерфейсов. Методы анализа требований разрабатывались для выявления предпочтений пользователей и установления их отношений с техническими описаниями информационных систем. Новые концепции анализа и проектирования заставили разработчиков и поставщиков информационных услуг поставить потребителей во главу угла. Средства анализа и проектирования информационных системы сыграли выдающуюся роль как средство распространения инновационных разработок бизнес-процессов.

Коммуникации. Коммуникационный компонент информационных систем влечет за собой создание ряда технологий, включающих топографию связи (соединительные линии, такие как телефонные линии, микроволны, спутники и беспроводные технологии), аппаратные переключатели и продукты и стандарты программного обеспечения коммуникаций. Коммуникационный охват может ограничиваться относительно маленькой областью в несколько километров (локальная сеть или LAN), обширной региональной областью, охватывающей сотни и тысячи километров (широкомасштабная сеть или WAN) или быть глобальным (глобальная сеть или GAN).

Пропускная способность сетевой связи зависит от каналов связи и от скорости, с которой компьютер, подключенный к сети, может передавать данные. Каналы связи используют оптоволоконные, спутниковые, микроволновые и беспроводные технологии, и в настоящее время их пропускная способность не ставит перед связью никаких ограничений. Основное ограничение накладывают преобразователи сигналов, потому что обычно они работают на заданной скорости передачи (бод или байты в секунду). Сравнительно новой технологией в этой области является асинхронный режим передачи (АТМ), который имеет гибкость переключения в широком диапазоне (от 9600 бит/с до 4 Мбит/с). Технология АТМ делает возможной отправку различных типов данных и информации (текст, звук, изображение и видеоданные) по одному и тому же коммуникационному маршруту. Последним ограничением в связи является скорость, с которой компьютер может отправлять и принимать данные, что обычно зависит от компьютерной шины (внутренний канал связи в компьютере). Вследствие возрастающей мощности компьютера это ограничение также становится все меньше.

Системные разработчики и обслуживающий персонал. Работники информационных систем представляют собой человеческий капитал, инвестированный в информационные системы. При этом самые крупные инвестиции делаются в приобретение навыков системными разработчиками и обслуживающим персоналом, которые создают и обслуживают информационные системы. В связи с быстрыми темпами изменений в ИТ постоянно растет потребность в оснащении новым оборудованием и переобучении работников информационных систем. Поэтому образование в этой области развилось в крупную индустрию, которая предлагает мастерские, семинары, короткие курсы и другие программы для совершенствования навыков работников информационных систем. Компании стали лучше осознавать необходимость обучения своих работников информационных систем основным навыкам бизнеса, характерного для этой компании, включая коллективную работу и ориентацию на покупателя.

2.3. Технологии передачи и обмена информационными ресурсами

Мультимедиа. Ввод в употребление графических выходных устройств изменил и характер ввода, и характер вывода информации. С началом применения манипулятора мышь и операций с окнами графический ввод и вывод информации переместился за пределы нескольких специализированных систем программного обеспечения в область универсального метода коммуникации с использованием компьютера. Графический интерфейс пользователя (GUI) позволяет направлять и выполнять задачи посредством щелчка мышью на значках, изображенных на экране.

Графические носители информации не ограничились GUI. Сейчас пользователи могут создавать и передавать собственные графические объекты. Более того, изображения могут вводиться в компьютер и обрабатываться как часть базы данных. Среди графических изображений наибольший интерес для бизнеса представляет обработка печатных и рукописных текстов и изображений. Обработка, хранение и распознавание изображений повышает уровень сложности систем и нуждается в более значительных компьютерных ресурсах. Для процесса распознавания используются интеллектуальные системы. Другие средства коммуникации с помощью компьютера – голос и звук. Сначала возможности компьютера в этом отношении ограничивались способностью преобразовывать символы в звуковые выходные данные. Позднее технология расширилась до ввода, хранения и

управления звуковыми данными. Более сложный процесс – сопоставление звуков с теми, что хранятся. Как и распознавание символов, распознавание звука и голоса требует сложной методологии, и все еще находится на раннем этапе развития. Видеоданные как носитель информации требуют еще более значительных компьютерных ресурсов и добавляют еще более высокий уровень сложности. Сверхбыстродействующий характер отображения видеоданных требует намного большей пропускной способности аппаратных средств.

Концепция мультимедийных систем позволяет использовать множественные носители для сбора, хранения, обработки и вывода информации. Это означает, что системы должны иметь устройства ввода и вывода для обработки мультимедийных носителей информации и программного обеспечения.

Компакт-диски. Другой средой информационного обмена являются компакт-диски и интерактивные компакт-диски. Большинство этих дисков принадлежит к типу постоянного запоминающего устройства (ROM). CD-ROM содержит постоянно записанную на нем информацию, которую нельзя изменить или удалить. Ввод в употребление CD-ROM (и компьютерного дисковод, который читает его) сделал возможным хранение большого количества информации и предоставление ее клиентам. Многие библиотечные источники, публикации большого объема, видеоинформация продаются на CD-ROM. Благодаря большой емкости запоминающего устройства на одном и том же носителе информации можно хранить графики, изображения и видео.

Интернет и информационная супермагистраль. Интернет начал свое существование в 1969 г., когда ARPA (Advanced Research Project Agency – Управление перспективного развития научно-исследовательских работ, которое позднее стало Оборонным управлением перспективного развития научно-исследовательских работ – Defense Advanced Research Project Agency (DARPA), частью Министерства обороны) соединило четыре компьютера для проведения эксперимента с сетью. Позже пятьдесят университетов с исследовательскими целями присоединились к ARPANET. Затем начали свое существование ряд других сетей, таких как NSFNET (сеть, основанная Национальным научным фондом), UUCP (сеть компьютеров Unix), USENET (сети пользователей), BITNET (Because It's Time Network) и расширения ARPANET, названные DARPANET и MILNET. Позднее все это было соединено в то, что сейчас называется Интернетом. По оценкам Интернет объединяет более 6000 общественных, частных и международных сетей и миллионы компьютеров, число которых продолжает расти. Интернет управляется группой волонтеров, называемой Интернет-обществом (Internet Society, ISCO), его стандарты устанавливаются другой группой, Специальной комиссией по проектированию Интернета (Internet Engineering Task Force, IETF). Идея Интернета в том, что компьютеры, подключенные к сети, должны быть способны осуществлять связь по любому доступному пути, а не по фиксированному маршруту. Это ввело в употребление концепцию протокола связи, метода коммуникации компьютеров и разрешения коммуникационных конфликтов, как, например, если два или более компьютеров хотят отправить или получить сообщение одновременно по одному и тому же маршруту. ARPANET и ее последующее расширение привели к разработке протокола управления передачей/межсетевым протокола (Transmission Control Protocol/Internet Protocol, TCP/IP), расширению протокола связи. Существует ряд других протоколов связи, но TCP/IP применяется наиболее широко.

Пользователи используют Интернет для электронной почты, передачи файлов, выполнения программ на различных компьютерах, поиска баз данных и библиотек, обмена идеями в виде групп новостей, а также для того, чтобы общаться и играть. Начиная с 1991 г. коммерческие организации также предлагают Интернет-пользователям свои услуги. Доказано, что изобилие информации в Интернете является ценным источником, который только начали использовать в бизнесе. С появлением Интернета стала стираться граница между доступом к частной и профессиональной информации. Интернет продемонстрировал важность информационной супермагистрали для национального и глобального обмена информацией и коммуникации. Это побудило правительства многих стран поддерживать

идею супермагистрали для информационного обмена. Благодаря успехам телекоммуникационной технологии и доступности супермагистрали характер многих видов социальной и индивидуальной деятельности изменится до невероятной степени.

Распределенные системы. Системы, которые функционируют и доступны более чем из одного местоположения, называются распределенными системами. В распределенных системах возможно использование данных, размещенных в различных местах. В этом случае каждый раз, когда пользователь запрашивает у системы определенную информацию, распределенная система может собрать данные из различных мест, чтобы ответить. Также существует возможность территориального распределения прикладных программ. В этом случае запрос пользователя может активизировать операции обработки данных более чем в одном местоположении, прежде чем будет получен ответ. Распределенные системы можно обновить более чем из одного места.

Открытые системы. В информационной индустрии существуют многочисленные платформы. Платформа представляет собой комбинацию аппаратных средств и программного обеспечения (главным образом операционных систем и сервисного программного обеспечения, способствующего функционированию аппаратных средств). Чаще всего информационные системы разрабатываются для конкретной платформы. Перемещение системы с одной платформы на другую обычно требует больших затрат времени и денег.

Многие компании имеют филиалы, которые работают со своими информационными системами на различных платформах. Заставить эти системы осуществлять связь друг с другом – трудно, а иногда и невозможно. Под концепцией возможности взаимодействия систем понимается способность осуществлять связь через различные платформы. Открытая система позволяет программному обеспечению и информационным системам работать и осуществлять связь через платформы. В компаниях по производству программного обеспечения, особенно тех, что разрабатывают операционные системы (программное обеспечение, управляющее операциями в компьютере), существует тенденция к переходу от патентованных операционных систем к открытым системам, в которых могут работать программные продукты, разработанные на других операционных системах. Открытая система является весьма подходящей концепцией, и значительная доля будущих разработок программного обеспечения для информационной индустрии будет базироваться на концепции возможности взаимодействия и открытых систем.

Интеграционные процессы на основе современных информационных технологий.

Глобализация экономических связей. Возросшие возможности компьютеров с мультимедийными носителями информации добавили изображение, звук и видеоданные в качестве дополнительных форм доставки информации. В настоящее время программное обеспечение баз данных дает возможность сохранять и отыскивать изображения и звуковые данные вместе с текстовой и цифровой информацией. Графические пакеты позволяют добавлять изображения, видео- и звуковые данные. Компьютеризованные библиотеки имеют приставки для видеоданных как формы представления информации. Однако вычислительные возможности были бы бесполезными, если бы данные (в форме текста, изображения, звука и видео) не могли передаваться через компьютерные сети. Добавление видеоданных для передачи в сети значительно увеличивает сложность обработки информации и требования к ресурсам. С другой стороны, это добавление имеет возможность коренным образом изменить информационную индустрию и объединить много видов услуг в компьютеризованные системы. Мы уже являемся свидетелями коммуникации офисов посредством электронной почты с вложениями электронных документов. То, что раньше было различными каналами коммуникаций (телефонные звонки или бумажные документы), сейчас объединяется в корпоративные информационные системы. Добавление телеконференций, электронной почты с видеоизображением и видеотелефонов продвигает это объединение еще на шаг вперед, внедряя в информационные системы встречи «лицом к лицу». Любое деловое общение в фирме можно связать с общей информационной систе-

мой. Это предусматривает гладкую и единообразную глобальную, континентальную, региональную и локальную связь. Другими словами, географическое расстояние больше не будет существенно влиять на бизнес-коммуникации, что, в свою очередь, будет способствовать глобализации рыночной экономики.

Интеграция домашних компьютеров. В США (и многих других промышленно развитых странах) существуют два канала связи, подключенных к большинству домов: телефонный и телевизионный. Изучение возможности дополнить эти каналы способностью передавать данные и информацию вызывает большой интерес. Такая возможность зависит от способности передавать видеоданные в дома и из домов и от компьютеров внутри домов для получения и обработки информации.

С помощью домашних компьютеров, отправляющих и получающих информацию в форме видеоданных, текста и изображений, люди (и фирмы) могут иметь доступ к товарам и услугам, которые обычно предоставляются за пределами дома, таким как видеопрокат, розничная торговля, образовательные материалы (включая видео, библиотечные материалы и информационные сводки), услуги в области здравоохранения, такие как первая помощь, помощь в чрезвычайных обстоятельствах или медицинская консультация для тех, кто не может покинуть свой дом. Таким образом, с представлением видео как передаваемых данных и изображений индивидуальные каналы связи также будут объединены через системы компьютеров.

Это еще не воплотилось в жизнь, но части необходимой технологии и систем уже разрабатываются. Асинхронная технология делает возможной коммуникацию компьютерных сетей с различными пропускными способностями. Телефонные компании развили возможность дополнения телефонов видеоизображениями. Кабельные компании разрабатывают двустороннюю телевизионную передачу, предоставляя клиентам возможность взаимодействовать со своими телевизорами. Экспоненциально увеличивается число людей, которые с различными целями общаются по Интернету. По-видимому, общество готовится к новой форме взаимодействия, в которой мы будем общаться в киберпространстве вместо того, чтобы нуждаться в физически близком расположении для информационного обмена.

Интеграция информационных услуг. Поскольку количество информации и ее доступность через компьютерную связь возрастает, пользователи все чаще сталкиваются с проблемой определения местоположения источника информации и доступа к нему через паутину запутанных и не очень удобных для пользователя компьютерных сетей. Ряд провайдеров продают доступ к данным, новостям и информации в формате, удобном для пользователей. Информация и данные, поставляемые в рамках таких услуг, либо собраны и всем доступны, либо закодированы провайдерами. Пользователи этих услуг имеют доступ к удобным экранам, которые ориентируют их в направлении нужной информации или данных. Некоторые провайдеры информационных услуг дают пользователю возможность доступа через систему к другим видам услуг, таким, например, как заказ и покупка товара. Эта комбинация запроса информации и действия (такого, как поиск наиболее дешевого товара, и в случае нахождения такового его приобретение) положила начало концепции «агентского» программного обеспечения, которое проектируется для поиска различных источников информации для данной задачи, и затем проведения соответствующих действий. Концепция агентского программного обеспечения все еще находится на этапе разработки, но она имеет большой потенциал и может стать основным коммерческим и индивидуальным компьютерным продуктом.

Контрольные вопросы к лекции 2

1. Что представляет собой Российский рынок биржевой и финансовой информации?
2. Каков механизм работы биржевых и финансовых информационных ресурсов?
3. Каковы истоки развитие информационных систем и технологий?
4. Каковы компоненты информационных систем?

5. В чем состоит суть технологии передачи и обмена информационными ресурсами?
6. Что представляют собой информационные агентства биржевой и финансовой информации?
7. Перечислите виды услуг, предоставляемые информационные агентства биржевой и финансовой информации
5. Как осуществляется анализ биржевой и финансовой информации?

Лекция 3. ГЛОБАЛЬНЫЕ КОМПЬЮТЕРНЫЕ СЕТИ

Вопросы:

- 3.1. Основные принципы и организационная структура Интернет
- 3.2. Инфраструктура сети Internet
- 3.3. Эталонная модель OSI
- 3.4. Протоколы сети Интернет
- 3.5. Протоколы, службы, сокет и порты TCP/IP
- 3.6. Адресация в сети Интернет
- 3.7. Система доменных имен DNS
- 3.8. Сервисы сети Интернет
- 3.9. Язык разметки HTML

Процессы взаимодействия вычислительных средств достаточно быстро переросли рамки отдельных фирм и организаций. Тенденции к интеграции и глобализации в современном мире получили адекватное отражение в виде совокупности вычислительных машин, объединенных коммуникационной средой, охватывающей значительные по расстоянию территории, получили название глобальных компьютерных сетей. Среди сетей, получивших общемировую известность, могут быть названы:

ARPANET – в течение многих лет (1969-1984) являлась наиболее развитой глобальной компьютерной сетью. Сейчас – одна из крупных подсетей Интернета. Ориентирована на исследовательские задачи.

NASA Science Internet (NSI) – объединяет несколько компьютерных сетей по космическим исследованиям, физике космоса и другим научным направлениям в общую сеть глобального распространения.

BitNet – также является одной из старейших глобальных сетей. BitNet располагает сетевым доступом к распределенным базам научно-исследовательских данных.

Earn (Западная и Центральная Европа) – в состав сети входят компьютеры национальных исследовательских центров Франции, Англии, Германии и других стран.

EVnet – одна из наиболее крупных европейских компьютерных сетей, действует с 1982 года, имеет региональные части практически во всех европейских странах, включая страны Балтии и Россию. Российскую часть EVnet представляет АО Релком, имеющее шлюзы, соединяющие EVnet с Интернетом; высокоскоростные выделенные линии связи; прикладные службы сети – электронную почту, списки рассылки, архивную службу.

Глобальные сети обеспечивают связь и предоставление сервисных услуг большому количеству абонентов, расположенных практически в любой точке земного шара. Типичными абонентами глобальных компьютерных сетей являются локальные сети предприятий и учреждений, расположенные в разных городах и странах.

Глобальные сети создаются крупными телекоммуникационными компаниями для оказания услуг абонентам. Компании, поддерживающие работоспособность сетей называют *операторами сети*, а поставщиков сетевых ресурсов абонентам – *провайдерами*.

Владелец, оператор и провайдер могут быть представлены как одной компанией, так и различными. Глобальная сеть может быть создана отдельной крупной корпорацией. Примером служат сети компаний Dow Jones или «Транснефть», созданные для внутренних потребностей.

В России на базе многих отраслевых сетей сформированы коммерческие организации, взявшие на себя обеспечение информационных услуг. Это привело к появлению общедоступных коммерческих специализированных систем, обеспечивающих выход на международные сети: Sprint, Bizlink, Infonet, PIENet. Несколько позже появились общедоступные специализированные телекоммуникационные сети Relcom, «Мир», «Ремарт» и др.

3.1. Основные принципы и организационная структура Интернет

Сеть Интернет – это глобальная информационная структура. Она является главным компонентом и механизмом распространения информационных ресурсов, средой взаимодействия между пользователями и компьютерами вне зависимости от их географического месторасположения.

Возникновение Интернета базировалось на идее, предполагавшей возможность объединения нескольких независимых компьютерных сетей с различной архитектурой на основе сети ARPANET. Сегменты новой сети, должны были функционировать независимо от временного нарушения связи в одном из них. Были сформулированы основные принципы построения новой сети:

- принцип открытой топологии – для включения в Интернет отдельной сети не должно производиться никаких дополнительных изменений;
- принцип интегрированной доставки информации – информация в Интернете передается пакетами, если пакет не смог достигнуть пункта назначения, то через короткое время он должен быть передан снова;
- принцип маршрутизации – для объединения сетей используются специальные устройства – маршрутизаторы, которые должны максимально упростить прохождение потока пакетов;
- принцип децентрализованного управления – не должно существовать единого центра управления объединенной сетью.

Важно отметить, что одной из ключевых концепций Интернета было не создание одного или нескольких приложений для работы с сетевыми устройствами, а создание общей инфраструктуры, в которой могли бы существовать и эволюционировать новые приложения.

Организационные структуры сети Интернет. В конце 70-х годов стало понятно, что рост размеров Интернета сопровождается ростом числа заинтересованных групп и, таким образом, необходима координация. Было сформировано несколько координирующих органов. IAB (Internet Architecture Board) занимается координацией деятельности в области развития структуры сети Интернет. IETF (Internet Engineering Task Force) состоит из нескольких рабочих групп, разрабатывающих и утверждающих стандарты для сети Интернет. В настоящее время в IETF существует 75 рабочих групп, изучающих различные аспекты развития Интернета. Впоследствии основным органом, осуществляющим информационную поддержку и регулирование в Интернете, стал Internet Society (ISOC) ISOC является общественной организацией, базирующейся на взносах и пожертвованиях спонсоров ISOC проводит ежегодные конференции, выпускает информационные материалы, поддерживает информационные серверы.

3.2. Инфраструктура сети Internet

В настоящее время основу сети Интернет составляют высокоскоростные магистральные сети. Независимые сети подключаются к магистральной сети через точки сетевого доступа NAP (Network Access Point) Независимые сети рассматриваются как автономные системы, то есть каждая из них имеет собственное административное управление и собственные протоколы маршрутизации. Деление сети Интернет на автономные системы позволяет распределить информацию о топологии всей сети и существенно упростить маршрутизацию. Автономная система должна состоять не менее чем из 32 меньших по размеру сетей. Обычно в качестве автономных систем выступают крупные, независимые, национальные сети. Примерами подобных сетей являются сеть EUNet, охватывающая страны центральной Европы, сеть RUNet, объединяющая университеты России. Автономные сети могут образовывать компании, специализирующиеся на предоставлении услуг доступа в сеть Интернет, – провайдеры. Такими провайдерами, например, являются компания UUNET в США и компания Relcom в России.

Внутри автономной системы данные предаются от одной сети к другой, пока не достигнут точки сопряжения с другой автономной системой. Обмен данными возможен только в том случае, если между автономными системами существуют соглашения о предоставлении транзита. Сети, включенные в автономные системы, представляют собой региональные сети, сети университетов, исследовательских центров и коммерческих фирм, а также сети более мелких региональных провайдеров. Скорость доступа определяется пропускной способностью канала связи внутри автономной системы и между автономными системами. Для модемного соединения, которое используют большинство домашних пользователей персональных компьютеров, пропускная способность канала невелика – от 19.2 до 57.6 Кбит/с, для выделенных телефонных линий, часто используемых для подключения к сети Интернет небольших локальных компьютерных сетей, – от 64 Кбит/с до 2 Мбит/с; для спутниковых и оптоволоконных каналов связи, которые в основном используются для создания автономных сетей, – от 2 Мбит/с и выше.

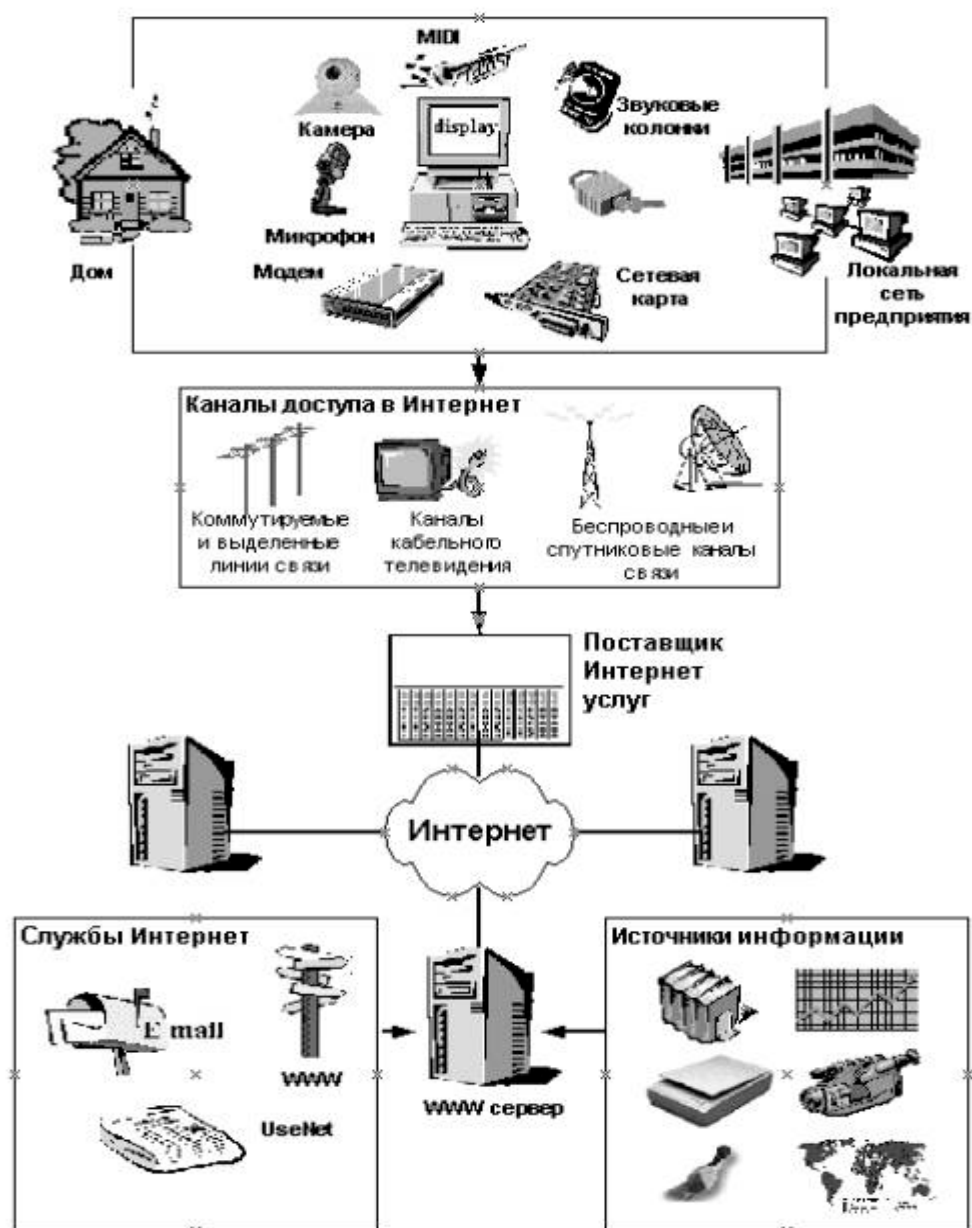


Рисунок 3.1 – Инфраструктура сети Интернет

ISP-провайдеры имеют в ряде регионов точки присутствия (POP – Point Of Presence), обеспечивающие подключение локальных пользователей. Все клиенты провайдера ISP_1

(см. рис.3.1) могут взаимодействовать с клиентами провайдера ISP_2 при подключении каждого из провайдеров к точкам сетевого доступа (NAP), обеспечивающим трафик между сетями.

Подключение домашнего компьютера к Сети осуществляется с помощью модема. Для соединения с провайдером (ISP – Internet Service Provider) используются различные каналы: телефонные линии, выделенные линии связи, сети кабельного телевидения, спутниковая и беспроводная связь. До настоящего времени наиболее распространенным среди индивидуальных пользователей является сеансовое подключение – соединение с набором модемов (модемным пулом) провайдера по телефонной линии. После подключения к модемному пулу персональный компьютер становится частью Сети и имеет доступ к сервисам, предоставляемым провайдером.

Локальная сеть организации подключается к сети провайдера через службу защиты информации в большинстве случаев по выделенной линии связи. Для соединения компьютеров в сети создается коммутируемый канал физической связи. При использовании такой технологии все передаваемые в сети сообщения разбиваются на части, которые называются пакетами. Каждый пакет снабжается заголовком, в котором указывается адрес назначения пакета. Коммутаторы, используя адрес, передают пакеты друг другу до тех пор, пока она не достигнут места назначения. Если какой-либо коммутатор слишком загружен и не может передать пакет в течение некоторого времени, он помещает его в очередь пакетов и передает позже.

3.3. Эталонная модель OSI

Взаимодействие устройств в вычислительной сети является сложным процессом, реализация которого требует решения многих взаимосвязанных проблем и задач. В начале 80-х годов международная организация по стандартизации (ISO) при поддержке других организаций по стандартизации разработала модель взаимодействия открытых систем (Open System Interconnection) модель OSI. Схема модели приведена на рис.3.2.

Модель OSI очень быстро стала одной из основных моделей, описывающих процесс передачи данных между компьютерами. Рассматриваемая модель разработана, исходя из принципа разделения задач. Она описывает системные средства взаимодействия, реализуемые операционной системой, системными утилитами, системными аппаратными средствами. Модель не описывает взаимодействия приложений с конечным пользователем.

В контексте эталонной модели уровни существуют для того, чтобы инкапсулировать и изолировать отдельные типы функциональных возможностей. В общем смысле, они призваны обеспечивать обслуживание более высокого уровня и доставлять данные на нижележащие уровни или принимать данные от более низкого уровня (для входящего трафика).

Процесс взаимодействия компьютеров может быть представлен как набор взаимодействий одинаковых уровней. Правила, определяющие формат сообщений между одинаковыми уровнями называются *протоколом*. Каждый уровень модели обрабатывает *модули данных* соответствующего протокола (*PDU* - Protocol Data Unit).

Модули данных протокола помещаются в «информационные конверты», начинающиеся специальной служебной информацией – *заголовком* (header) и завершающиеся так называемым *концевиком* (trailer). Заголовок представляет характерную для данного уровня метку, какой бы модуль данных он не передавал. Аналогично концевик (присутствие которого необязательно для некоторых уровней) содержит информацию, касающуюся обнаружения и исправления ошибок, заключительные знаки окончания данных и любые другие данные, предназначенные для ясного указания на завершение модуля данных протокола. Каждый уровень при передаче блока информации нижележащему уровню снабжает его своими заголовками. Заголовки вышележащих уровней воспринимаются нижележащими уровнями как данные блока информации. На каждом уровне информация обрабатывается в соответствии с назначенными функциями. При получении блока информации от

нижележащего уровня заголовки и другая служебная информация текущего уровня отбрасываются и данные обрабатываются.



Рисунок 3.2 – Модель OSI

Средства каждого уровня обрабатывают протокол своего уровня и интерфейсы с соседними уровнями. Структура эталонной модели ISO/OSI точно изображает множество реализуемых наборов протоколов. Набор протоколов, аппаратных и программные компоненты, достаточный для организации взаимодействия в сети и действующих на отдельно взятом компьютере принято называть *стеком коммуникационных протоколов*. Таким образом, в системе Windows сетевую интерфейсную плату (*NIC* – Network Interface Card), драйвер, позволяющий операционной системе взаимодействовать с ней и различные программные компоненты, составляющие другие уровни можно именовать стеком протоколов для данного компьютера.

Каждый уровень эталонной модели относительно независим и описывает строго определенные функции взаимодействия сетевых устройств. Все уровни образуют иерархическую систему, в которой запрос вырабатывается на каком-либо уровне, передается на исполнение нижележащему уровню. Результаты обработки запроса передаются на выше лежащий уровень. Для описания взаимосвязей между двумя уровнями устанавливаются правила или соглашения, которые называются *интерфейсом*. Интерфейс определяет набор сервисов, предоставляемый уровнем соседнему уровню.

Физический уровень (уровень 1). Физический уровень содержит материальные среды передачи (кабели или беспроводные средства), которые необходимы в любой сети для отправки и получения сигналов, представляющих физическое выражение сетевых коммуникаций. На физическом уровне определяются характеристики электрических сиг-

налов, передающих данные; физического канала связи; типы разъемов с назначением каждого контакта. Его функция состоит в приведении в действие, поддержке и свертывании сетевых соединений. Физический уровень занимается аппаратурой и соединениями, позволяющими ей получать доступ к какой-либо сетевой среде передачи данных. Он координирует отсылку и получение сигналов по среде передачи и устанавливает, какие кабели, соединители и сетевые интерфейсные карты (NIC) необходимы для обеспечения доступа к определенному участку сети.

Физический уровень описывает передачу компьютерных битов исходящей информации и контролирует их преобразование в аналоговые сигналы, используемые в сетях. Имея дело с входящими сообщениями, физический уровень осуществляет обратный процесс, превращая получаемые по сетевой среде сигналы в битовые комбинации, которые должны быть отправлены компьютеру через сетевой интерфейс.

Модули данных протокола на физическом уровне состоят из определенных серийных наборов сигналов, соответствующих битовым комбинациям для *кадров (фреймов)* на канальном уровне.

Канальный уровень (уровень 2). *Канальный уровень* расположен между Физическим и Сетевым уровнями. Его работа – сделать возможной надежную передачу данных через Физический уровень на отправляющем конце и проверить надежность при получении на принимающем конце. Канальный уровень управляет двухточечной передачей (point-to-point transmission) через сетевую среду от одного компьютера к другому по одному логическому или физическому *сегменту кабеля*. Он распознает отдельные устройства в локальной среде посредством специального адреса, который идентифицирует каждый интерфейс. Этот уровень обеспечивает соединения в локальной сети (LAN – Local Area Network) между компьютерами, к которым эти интерфейсы подключены. В ходе управления соединениями Канальный уровень осуществляет упорядочивание данных. С этой целью он может контролировать скорость, на которой данные передаются от отправителя к получателю – этот процесс называется управление потоками в среде передачи. Канальный уровень реагирует на локальные условия перегрузки и помогает предотвратить переполнение среды передачи локальным трафиком. Он запрашивает передачу данных при готовности модулей данных протокола (PDU) к отправке, а также отвечает за прием и составление входящих модулей данных для поступающей информации.

Модули данных на канальном уровне должны уместиться в определенных битовых комбинациях, соответствующих пропускной способности сети. Поэтому обмен данными осуществляется определенными порциями, которые называются *кадрами* или *фреймами*. В каждый кадр (фрейм) добавляются определенные последовательности бит в начало и конец для выделения кадра, адрес компьютера-отправителя, адрес компьютера получателя. Кроме всего прочего в каждый кадр добавляется вычисляемая контрольная сумма, которая необходима для проверки корректности передачи кадра. Исправление обнаруженной ошибки возможно за счет повторной передачи кадра. Протоколы канального уровня на компьютере реализуются сетевыми адаптерами и их драйверами. Протоколы канального уровня учитывают топологию сети. *Топология* – это конфигурация графа, вершинами которого являются компьютеры сети или другие коммуникационные устройства, а ребрами – физические связи между ними. Компьютеры называются *узлами сети*. Протоколы канального уровня поддерживают топологии «общая шина», «звезда», «кольцо».

Использование протокола канального уровня в глобальных сетях ограничено применением между компьютерами, соединенными индивидуальной линией связи (что практически невозможно). Для обмена данными между конечными узлами разных сетей используются средства следующего сетевого уровня.

Сетевой уровень (уровень 3). На сетевом уровне решаются вопросы объединения сетей с разными топологиями, с разными принципами передачи данных между конечными узлами для образования единой транспортной системы. Сетевой уровень решает вопросы обмена данным между сетями. С ним связано понятие сетевого адреса, здесь происходят

сложные процессы направления модулей данных протокола от отправителя к получателю. Сетевой уровень заведует *логическими адресами*, ассоциируемыми с отдельными компьютерами в сети и позволяющими службе доменных имен (*DNS – Domain Name System*) соотносить удобные для человеческого восприятия имена этих машин с уникальными машиночитаемыми числовыми адресами. Сетевой уровень использует информацию об адресации, чтобы определить, как послать модуль данных протокола (PDU) от отправителя к получателю, когда источник и пункт назначения трафика располагаются в разных физических сегментах сети.

Сетевой уровень поддерживает множественные одновременные соединения между различными IP-адресами, чтобы несколько приложений могли поддерживать сетевое соединение в одно и то же время. Он имеет возможность определять, какое соединение относится к отдельному процессу или приложению на компьютере и, поэтому, может направлять трафик к нужному получателю, или доставлять входящие данные нужному процессу или приложению на принимающей машине. Это объясняет, почему одновременно на одном компьютере могут быть запущены сеансы Web-браузера и почтовой программы, и данные доставляются к ним без смешения в разных информационных потоках. Модуль данных протокола, связанный с сетевым уровнем называется *пакетом* (packet). Методика, применяемая для пересылки или ретрансляции отдельных модулей данных протокола (PDU) от отправителя к получателю называется *коммутацией пакетов*. Каждый пакет, кроме адреса компьютера, снабжается адресом сети как получателя, так и отправителя. Сетевой уровень способен опознавать и использовать многочисленные *маршруты* между отправителем и получателем во время передачи сообщений. Для соединения сетей используется специальное устройство – *маршрутизатор*, который собирает информацию о топологии межсетевых соединений. На пути передачи пакета между конечными узлами, находящимися в разных сетях, могут находиться другие промежуточные сети, через которые необходимо сделать транзитные передачи. Таким образом, пакет проходит несколько маршрутизаторов, которые образуют маршрут. Пакеты одного сообщения могут достигать пункта назначения разными маршрутами. Проблема выбора наилучшего маршрута является главной задачей сетевого уровня, решение которой возлагается на маршрутизатор.

Критерии выбора могут быть следующими: время передачи пакета, надежность передачи. Есть два принципиально различных способа работы сетевого уровня.

Первый – это *метод виртуальных каналов*. Он состоит в том, что канал связи, устанавливается при вызове (начале сеанса (session) связи), по нему передается информация, и по окончании передачи канал закрывается (уничтожается). Передача пакетов происходит с сохранением исходной последовательности, даже если пакеты пересылаются по различным физическим маршрутам, т.е. виртуальный канал динамически перенаправляется. При этом пакеты данных не включают адрес пункта назначения, т.к. он определяется во время установления связи.

Второй – метод дейтаграмм. *Дейтаграммы* – независимые модули, они включают всю необходимую для их пересылки информацию. В то время, как первый метод предоставляет следующему уровню надежный канал передачи данных, свободный от искажений (ошибок) и правильно доставляющий пакеты в пункт назначения, второй метод требует от следующего уровня работы над ошибками и проверки доставки нужному адресату.

Транспортный уровень (уровень 4). Само название транспортного уровня отражает его назначение – обеспечить надежную сквозную передачу модулей данных протокола от отправителей к получателю. Транспортный уровень регламентирует пересылку пакетов сообщений между процессами, выполняемыми на компьютерах сети. Завершает организацию передачи данных: контролирует на сквозной основе поток данных, проходящий по маршруту, определенному сетевым уровнем: правильность передачи блоков данных, правильность доставки в нужный пункт назначения, их комплектность, сохранность, порядок следования. С этой целью транспортный уровень присоединяет к пакету данные для выяв-

ления и исправления ошибок. Такие данные обычно занимают концевые части пакетов, где специальные значения, называемые *контрольными суммами* (checksums), высчитываются перед и после доставки данных, а затем сравниваются для подтверждения безошибочности передачи. Если контрольная сумма, подсчитанная при отсылке, совпадает с полученной, значит передача прошла без ошибок, в противном случае протоколы транспортного уровня запрашивают повторную передачу модулей данных. Объем данных, который может быть послан от отправителя к получателю, произволен, но контейнеры, пересылающие эти данные, характеризуются фиксированным максимальным размером – максимальной единицей передачи (*MTU* – Maximum Transmission Unit). Транспортный уровень отвечает также за деятельность по сегментации и повторной сборке пакетов.

Сегментация заключается в преобразовании большого сообщения в пронумерованную последовательность порций данных, называемых сегментами. Каждый сегмент имеет размер максимальной полезной нагрузки, которую поддерживает сетевая среда.

Повторная сборка пакетов означает процесс размещения сегментов в изначальном порядке и использования их для воссоздания данных в том виде, в котором они существовали до сегментации. Транспортный уровень имеет все необходимые средства для осуществления запросов на повторную передачу ошибочных или отсутствующих модулей протокола уже в процессе повторной сборки. Таким образом, гарантируется надежная доставка данных от отправителя к получателю. Протоколы транспортного уровня реализуются программными средствами узлов сети, компонентами сетевых систем.

Сеансовый уровень (уровень 5). На сеансовом уровне устанавливаются, поддерживаются и завершаются, а при необходимости обрываются текущие передачи между отправителем и получателем. Сеансовый уровень определяет механизмы, позволяющие отправителям и получателям требовать начала или окончания диалога, а также его сохранения даже тогда, когда информационный поток между сторонами может не осуществляться. Этот уровень обладает механизмами, предназначенными для надежной поддержки текущих диалогов, известными под названием *контрольных точек* (checkpoint). Они определяют крайнюю точку, при которой возможно признание передачи успешной, и подобную же точку, при которой необходим откат передачи для воспроизведения или восстановления отсутствующих или поврежденных элементов. Кроме того, сеансовый уровень определяет совокупность механизмов, посредством диалоги, выпадающие из синхронизации, могут быть синхронизированы повторно. Этот же уровень ответственен за картографию сети – он преобразовывает региональные (доменные) компьютерные имена в числовые адреса, и наоборот. Он координирует не компьютеры и устройства, а процессы в сети, поддерживает их взаимодействие – управляет сеансами связи между процессами прикладного уровня. Модули данных протокола на этом уровне так и называются – *сеансовые модули данных протокола* (*SPDU* – Session PDU). Отдельно протоколы сеансового уровня не используются. Его функции реализуются в протоколах прикладного уровня.

Уровень представления (уровень 6). Уровень представления отвечает за подачу (представление) данных в сеть (на пути вниз по стеку протоколов). Он управляет преобразованием данных из характерных форм, применяемых в сети, в более специфические, предназначенные для использования на определенных платформах и наоборот. Именно такая схема делает возможным сетевое взаимодействие в корне различных типов компьютеров, которые могут по-разному представлять числа и символы. На уровне представления должно присутствовать специальное компьютерное устройство – *редиректор* (redirector) в терминологии Microsoft или *сетевая оболочка* (network shell) в терминологии Novell NetWare и Unix. Задача этого устройства состоит в том, чтобы различать запросы на сетевые и локальные ресурсы и перенаправлять их на подходящие локальные или удаленные подсистемы. Таким образом, компьютеры используют для доступа к ресурсам одну подсистему независимо от того, где размещены эти ресурсы – на локальной или удаленной машине. В результате становится возможным проектирование и разработка кроссплатформенных распределенных приложений, скрывающих для пользователя различие между ло-

кальными и удаленными ресурсами. Пользователи могут просто запрашивать ресурсы, а способы удовлетворения этих запросов возлагаются на редиректор.

Уровень представления, кроме всего прочего, может поставлять приложениям специальные функции обработки данных, включая шифрование или сжатие данных (для исходящих сообщений), их расшифровку или распаковку (для входящих сообщений), перекодировка текстовой информации и изображений и т.д.

В итоге, функции уровня представления сводятся к преобразованию формы представления данных полученных от прикладного уровня одной системы в форму, необходимую для восприятия прикладным уровнем другой системы. Что бы ни делал Уровень представления на стороне отправителя, обратное действие необходимо совершить на стороне получателя, с тем, чтобы оба участника передачи на определенном этапе располагали данными в одинаковом виде.

Прикладной уровень (уровень 7). Приложения всегда принимают участие в запросе доступа к сети, но Прикладной уровень определяет не сами приложения, а интерфейс, который они могут применить для запроса сетевых служб. Таким образом, Прикладной уровень характеризует типы служб, которые приложения могут запрашивать из сети, и обуславливает формы, принимаемые данными при получении или отправке сообщений. Он определяет, что именно приложения могут требовать от сети и какие виды деятельности данная сеть поддерживает. Здесь регламентируются полномочия доступа к определенным файлам и службам, а также права пользователей на выполнение соответствующих действий с конкретными элементами данных.

На этом уровне реализуется, по крайней мере, пять прикладных служб: передача файлов, удаленный терминальный доступ, электронная передача сообщений, служба каталогов и управление сетью. В качестве единицы информации протоколы этого уровня используют *сообщение* (message).

3.4. Протоколы сети Интернет

Для включения в сеть большого числа компьютеров необходимо было выработать единый набор правил, определяющих способ взаимодействия узлов сети (последовательность передачи, формат сообщений) – *протокол*. Ключом к объединению сетей стал протокол, поддерживающий межсетевое взаимодействие. Первая версия нового протокола появилась в 1973 году. Протокол был назван TCP (Transmission Control Protocol) и обеспечивал достаточно широкий диапазон транспортных услуг.

Первоначальная версия TCP хорошо работала при решении таких задач как передача файлов или удаленная регистрация, но в некоторых случаях (например, при передаче речи) она не могла исправлять передачу пакетов. Этот факт привел к разделению TCP на два протокола: IP для адресации и передачи отдельных пакетов и TCP для разделения сообщений на пакеты, обеспечения целостности и восстановления потерянных пакетов. Объединенный протокол стали называть TCP/IP.

В основном в сети Интернет используется семейство протоколов TCP/IP. Если рассматривать более подробно структуру протоколов TCP/IP в сравнении с моделью OSI, то в отличие от полной модели OSI, которая имеет семь уровней, TCP/IP использует лишь четыре – уровень доступа к сети, межсетевой, транспортный и прикладной. Эти комплекты уровней похожи, но не идентичны. Причиной этого является то, что некоторые функции Сеансового и представительского уровней эталонной модели OSI/OSI заключены в Прикладном уровне TCP/IP, а функции Канального и Физического уровней модели OSI включены в Уровень доступа к сети модели TCP/IP (рис.3.3.).

Уровень доступа к сети. *Уровень доступа к сети* (в данном случае самый нижний) соединяет узел с каналом и определяет способ соединения. Результатом работы этого уровня является передача сигнала, состоящего из пакетов в канал. Из функций канального уровня модели OSI TCP/IP поддерживает многие из существующих стандартов, определяющих среду передачи данных. В локальных сетях Ethernet используется стандарт FDDI (Fibre Distributed Data Interface), в сетях Token Ring – стандарт PPP (Point to Point Protocol).

В крупных территориальных сетях это могут быть X.25 и ISDN. Программное обеспечение для канального уровня называется *драйверами* и взаимодействует с сетевыми картами.

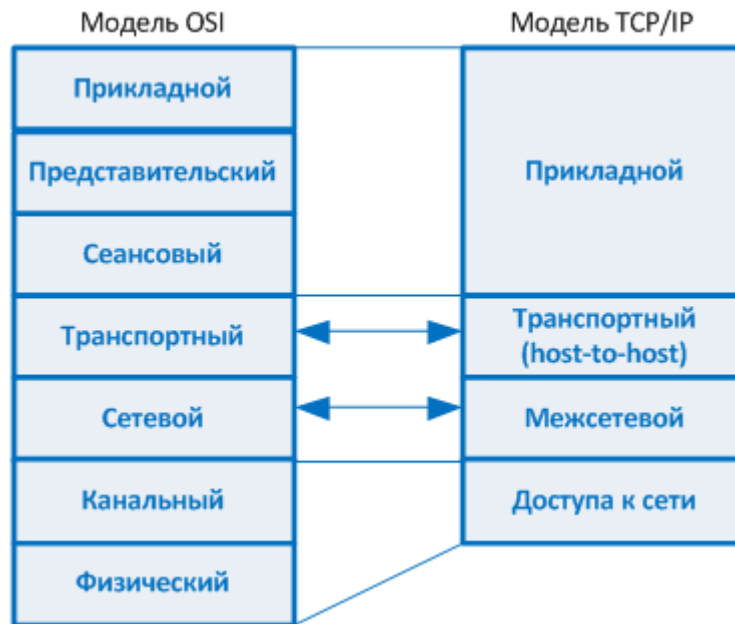


Рисунок 3.3 Сравнение уровней эталонной модели OSI и сетевой модели TCP/IP

Протоколы уровня доступа к сети модели TCP/IP. Самые важные протоколы этого уровня – *SLIP* и *PPP*.

SLIP – наиболее старый и элементарный протокол для последовательного канала. Он начал свое существование в системах Unix. Протокол SLIP не содержит встроенных возможностей по безопасности и не может обеспечить усовершенствованную доставку данных. На сегодняшний день он уже редко применяется для установления сетевых соединений с Internet и с частными сетями.

PPP – это более современный протокол для последовательного канала (телефонные линии), широко применяемый для соединений в сетях Internet и частных сетях TCP/IP. Он нейтрален по отношению к другим протоколам и может задействовать протоколы нескольких типов одновременно в течение одного соединения по последовательному каналу. Реализация протокола PPP в системах Windows обеспечивает в пределах одного соединения поддержку всех основных Windows-протоколов: TCP/IP, IPX/SPX, NetBEUI (расширенный пользовательский интерфейс базовой системы ввода вывода), а также туннельный протокол PPP (point-to point protocol) и другие протоколы виртуальных частных сетей (VPN).

В настоящее время PPP как протокол для последовательного канала наиболее предпочтителен. причина этого в том, что оно поддерживает множество возможностей, касающихся обеспечения безопасности, например, шифрование регистрационной информации или всего трафика, проходящего по последовательному каналу.

Межсетевой уровень. *Межсетевой уровень* является основой семейства протоколов TCP/IP. Этот уровень предоставляет адресное пространство, обеспечивает перемещение пакетов в сети, а также управляет их маршрутизацией. На межсетевом уровне выполняются три важнейшие для TCP/IP задачи:

- *Фрагментация MTU* (максимальных единиц передачи). Когда по определенному маршруту данные передаются из сети одного типа в сеть другого типа, то их максимальные единицы передачи могут различаться. при переходе из среды, поддерживающей большую максимальную единицу передачи в среду с меньшей единице передачи, данные должны быть разбиты на меньшие блоки, соответствующие меньшей единице. Требуется,

хотя бы однонаправленное преобразование, поскольку меньшие пакеты необязательно должны объединяться для прохождения по сети с большей максимальной единицей передачи.

- *Адресация.* Здесь определяется механизм, посредством которого все сетевые интерфейсы в сети TCP/IP должны быть поставлены в соответствие с конкретными уникальными битовыми комбинациями, идентифицирующими каждый интерфейс в отдельности, а также сети, которым эти интерфейсы принадлежат.

- *Маршрутизация.* Здесь определяется механизм пересылки пакетов от отправителя к получателю, в который могут быть вовлечены многочисленные промежуточные ретрансляторы. Эта функция объединяет не только процессы, делающие возможной успешную доставку, но также методы отслеживания ее эффективности и соглашения об ошибках при неудачной доставке.

Таким образом, межсетевой уровень осуществляет доставку данных от отправителя к получателю. При необходимости он переупаковывает данные в контейнеры меньшего размера, разрешает вопросы идентификации местонахождения отправителя и получателя и определяет способы доставки информации по сети из первого пункта во второй.

Протоколы межсетевого уровня. IP (Internet Protocol) представляет собой основу протоколов TCP/IP. Протокол IP относится к типу протоколов без установления соединения, то есть – никакой управляющей информации кроме той, что содержится в самом IP-пакете, по сети не передается. Кроме того, протокол IP не гарантирует надежной доставки сообщений. Поток данных протокол IP разбивает на определенные части – дейтаграммы и рассматривает каждую дейтаграмму как независимую единицу, не имеющую связи с другими дейтаграммами.

Дейтаграмма – общее название единицы данных, которыми оперируют протоколы без установления соединения. Основной задачей протокола IP является передача дейтаграмм между сетями. Дейтаграммы, передаваемые с помощью протокола IP, называют IP-пакетами.

- *IGMP* (Internet Group Multicast Protocol) отвечает за многоканальную передачу, при которой пакеты одного и того же сообщения передаются по разным маршрутам в зависимости от пропускной способности сети. Этот протокол также используется для передачи аудио- и видеоинформации.

- *ICMP* (Internet Control Message Protocol) – протокол контроля сообщений контролирует маршрутизацию на основе протокола IP и поведение сети.

- *PING* (Packet Internetwork Grouper) отправитель пакетов проверяет доступность и период кругообращения между IP-адресами отправителя и получателя.

- *ARP* (Address Resolution Protocol) протокол разрешения адресов осуществляет преобразование IP-адресов в MAC-адреса (адреса управления доступом к среде) в определенном сегменте кабеля (всегда применяемого для финального этапа доставки пакетов)

- *RARP* (Reverse Address Resolution Protocol) осуществляет обратное преобразование из MAC-адреса в числовые IP-адреса.

- *RIP* (Routing Information Protocol) протокол маршрутной информации – это базовый протокол маршрутизации для областей локальных и объединенных сетей.

- *BGP* (Border Gateway Protocol) пограничный межсетевой протокол – это протокол маршрутизации, осуществляющий соединение с обычными магистралями сети Internet или с другими маршрутными доменами в той же сети, где многочисленные стороны несут совместную ответственность за формирование трафика.

Транспортный уровень. *Транспортный уровень* отвечает за передачу данных по указанному адресу, обеспечивая своевременную и качественную передачу сообщения получателю информации, необходимую фрагментацию исходящих сообщений перед их передачей и их повторной сборкой перед доставкой на Прикладной уровень для последующей обработки.

Устройства, действующие в сети Internet, обычно называют *хостами* (hosts). Транспортный уровень иногда именуется двух-хостным, поскольку он подразумевает перемещение данных от одного хоста к другому.

На транспортном уровне модели TCP/IP присутствуют два протокола: *TCP* (протокол управления передачей) и *UDP* (User Datagram Protocol) протокол передачи дейтаграмм пользователя. Эти транспортные протоколы по-разному осуществляют доставку данных: *на основе соединений* – TCP и *без установления соединения* – UDP. Различие выражается в действиях протокола TCP по организации и поддержке соединения между отправителем и получателем перед пересылкой данных, получению положительного подтверждения об успешной их передаче или запроса на повторную пересылку отсутствующих или ошибочных данных. В отличие от TCP протокол UDP просто пересылает данные и не выполняет последующей проверки после их получения.

Так как протокол IP межсетевого уровня не гарантирует надежную доставку сообщений, эту задачу решает протокол TCP. Надежность протокола TCP заключается именно в том, что источник данных повторяет их посылку в том случае, если не получит в определенный промежуток времени от адресата подтверждения их успешного получения. Части, на которые протокол TCP разбивает поток данных, принято называть сегментами. Каждый сегмент предваряется заголовком. В заголовке сегмента существует поле контрольной суммы. Если при пересылке данные повреждены, то по контрольной сумме протокол TCP может это определить. Поврежденный сегмент уничтожается, а источнику ничего не посылается. Если данные не были повреждены, то они пропускаются на сборку сообщения приложения, а источнику отправляется подтверждение.

Операция, связанная с разбиением информации на пакеты и маршрутизацией пакетов, получила название *коммутации пакетов*. Это очень важное решение, определяющее производительность сетей. Поскольку региональные и глобальные сети охватывают значительные территориальные пространства и включают в свою архитектуру разнообразные устройства, то для повышения эффективности работы этих устройств необходимо, чтобы они были загружены равномерно, а информация передавалась бы с одной скоростью. Разбиение сообщений на пакеты позволяет исключить неравномерность передачи информации и предоставляет возможность, в случае сбоев, осуществить повторную посылку одного или нескольких пакетов. Кроме того, снижение неравномерности нагрузки в сети может быть выполнено более гибким выбором маршрута.

Прикладной уровень. *Прикладной уровень* модели TCP/IP также известен под именем *уровень процессов*, поскольку именно на нем стек протоколов взаимодействует с приложениями и процессами на хостах. Пользовательские интерфейсы процесса или приложения также определяются здесь. На этом уровне наблюдается перекрытие функций протоколов и служб. Например, протокол FTP (File Transfer Protocol) протокол передачи файлов и Telnet (сетевой теледоступ) – все это специальные протоколы, основанные на TCP/IP, они определяют службы передачи файлов, терминальной эмуляции и т.д. Большая часть высокоуровневых служб основана на TCP/IP и действует на прикладном уровне этой модели. Прикладной уровень объединяет все службы (Internet Information Services – IIS), которые система предоставляет пользователю. К наиболее важным прикладным протоколам относятся протокол передачи гипертекста HTTP, протоколы для работы с электронной почтой POP, SMTP, IMAP, MIME, новостной протокол NNTP, протокол удаленного управления Telnet, протокол передачи файлов FTP. На этом уровне работает система доменных имен DNS и протокол SNMP, предназначенный для управления сетевыми устройствами.

Работа служб TCP/IP находится в зависимости от двух основных элементов:

1. Специальный «процесс приемника» (в терминологии Unix), называемый *присоединенной процедурой* или иначе *демоном* (daemon), действует на сервере, обрабатывая входящие пользовательские запросы на определенные службы. В системах на платформе

Windows 2003 процесс INETINFO.EXE возникает при работе Web-сервера, IIS (Internet Information Services) или FTP-сервера.

2. Для каждой службы TCP/IP существует свой *адрес порта* 16-битное число, которого применяется для идентификации отдельного процесса или службы. Адреса в диапазоне от 0 до 1024 называются хорошо известными адресами портов (Well-know port address) и связывают определенный адрес порта с соответствующей службой. Например, хорошо известный адрес порта FTP – порт 21.

Любой демон или процесс приемника постоянно находится в ожидании попыток соединения с хорошо известным адресом порта, соответствующим какой-либо службе. После прибытия запроса на соединение выполняется процесс проверки приемника, чтобы убедиться в разрешении на его выполнение. Если выполнение, возможно, то либо создается другой временный процесс (Unix), либо порождается отдельный поток выполняемых задач (Windows 2003). Такой процесс или поток длится столько времени, сколько требуется для обслуживания данного запроса пользователя, при этом для его выполнения применяется временный адрес порта в диапазоне от 1025 до 65535 (иногда службы используют четыре адреса порта, чтобы обе стороны могли одновременно отправлять и принимать данные). Как только для выполнения запроса создается процесс или поток, процесс приемника или демон возвращается к выполнению первоочередной задачи по ожиданию сервисных запросов.

3.5. Протоколы, службы, сокеты и порты TCP/IP

На любом компьютере с TCP/IP в одно и то же время может работать множество приложений. Нормальной ситуацией является одновременная работа почтовой программы, браузера и FTP-клиента. В среде TCP/IP необходим механизм, позволяющий проводить различие между многочисленными приложениями, использующими транспортные протоколы TCP и UDP для обработки множества исходящих потоков данных перед их передачей протоколу IP для адресации и инструкций по доставке. Также необходимо обеспечить обратный процесс для входящих данных. Входящий поток модулей данных протокола транспортного уровня должен быть обследован и разделен, а результирующие сообщения доставлены соответствующему запрашивающему приложению.

Объединение различных источников исходящих данных в единый исходящий поток называется *мультиплексированием*. Дробление входящего потока данных на отдельные блоки, доставляемые соответствующим приложениям, называется *демультиплексированием*. Эта деятельность обычно осуществляется на Транспортном уровне, где исходящие сообщения также разделяются на блоки, соответствующие возможностям сетей, по которым они передаются и где осуществляется повторная сборка входящих сообщений в нужном порядке из входящих блоков данных. Чтобы упростить эту работу для идентификации отдельных протоколов, TCP/IP присваивает им соответствующие *номера протоколов*, а эти протоколы, в свою очередь, применяют номера портов для обозначения отдельных протоколов и служб Прикладного уровня. Эта методика зародилась в среде UNIX, где осуществлялась посредством ряда конфигурационных файлов. Этот факт объясняет механизм возникновения методик во всех реализациях TCP/IP.

Для идентификации *хорошо известных протоколов* (well-known protocols) резервируются многочисленные номера портов. Эти протоколы (иногда называемые *службами*) назначают ряд чисел, представляющих крупные коллекции сетевых служб TCP/IP, таких как передача файлов (FTP), эмуляция терминала (Telnet) и работа электронной почты (SMTP – Simple Mail Transfer Protocol, POP3 – Post Office Protocol и IMAP – Internet Message Access Protocol).

Номера протоколов TCP/IP. В заголовке IP-дейтаграммы номер протокола представлен своим десятым байтом. Это 8-битовое значение указывает на то, какой протокол Транспортного уровня должен принимать доставку входящих данных. Полный список номеров протоколов можно найти по адресу <http://www.iana.org>. В таблице 3.1. приведены несколько номеров хорошо известных протоколов.

Таблица 3.1 – Некоторые протоколы TCP/IP

№	Акроним	Имя протокола	
0	IP	Internet Protocol	Протокол Internet
1	ICMP	Internet Control Message Protocol	Протокол контроля сообщений в Internet
6	TCP	Transmission Control Protocol	Протокол управления передачей
7	UCL	User Control List	Пользовательская управляющая таблица
8	EGP	Extension Gateway Protocol	Внешний шлюзовой протокол
11	NVP-11	Network Voice Protocol	Сетевой протокол передачи речевых сообщений
17	UDP	User Datagram Protocol	Протокол передачи дейтаграмм пользователя
18	MUX	Multiplexing	Мультиплексирование

Номера портов TCP/IP. После того, как протокол IP передает входящие данные протоколам TCP или UDP, работающим на транспортном уровне, они должны выполнить свои обязанности, а затем передать данные *прикладному процессу* (application process), которому они предназначались, т.е. любой запущенной программе, принимающей данные от имени пользователя. Прикладные процессы TCP/IP иногда называют *сетевыми службами* (network services); они идентифицируются посредством номеров портов. *Номер порта источника* (source port number) является идентификатором процесса, отославшего данные, а *целевой номер порта* (destination port number) обозначает процесс, который должен принять их. Оба этих значения представляются в виде двух байт (16 бит) в первом слове заголовка любого сегмента TCP или пакета UDP. Поскольку номера портов являются 16-битовыми значениями, то в десятичном выражении они могут принять значение в диапазоне от 0 до 65535.

Изначально номера портов до 256 были зарезервированы для хорошо известных служб, таких как Telnet или FTP, а значения от 256 до 1024 – для служб UNIX. На сегодняшний день все адреса портов до 1024 представляют хорошо известные службы, но помимо них в диапазоне от 1024 до 65535 существует множество так называемых *зарегистрированных портов* (registered ports), соответствующих отдельным прикладным службам. Просмотреть, что находится в этом обширном диапазоне адресов можно на сайте <http://www.iana.org>

Сокеты TCP/IP. Хорошо известные службы имеют заранее установленные номера портов. Это облегчает процесс соединения «клиент-сервер», поскольку как отправитель, так и получатель соглашаются, что определенные службы связаны с конкретными адресами портов. В дополнение к таким номерам существуют *динамически назначаемые номера портов*. Они не устанавливаются заранее и применяются, если в этом есть необходимость, для обеспечения временного соединения, во время которого осуществляется ограниченный обмен данными. Это позволяет каждой системе поддерживать множество открытых соединений и присваивать каждому из них свой уникальный динамический назначаемый адрес порта. Они попадают в диапазон от 1024, где каждый неиспользованный номер порта может быть задействован для временного соединения. После применения клиентом или сервером для установления соединения хорошо известного адреса порта это соединение, оно называется сеансом (session), всегда передается временной паре адресов сокетов, которая обеспечивает отсылку и получение адресов портов для дальнейшей связи между отправителем и получателем. Сочетание особого IP-адреса (для сервера, на котором выполняется процесс) и *динамически назначаемого адреса порта* (где поддерживается соединение) называется *адресом сокета* или просто *сокетом* (socket).

Инкапсуляция данных в ТСР/IP. На каждом уровне стека протоколов ТСР/IP – Прикладном, Транспортном, Межсетевом и Уровнем доступа к сети исходящие данные упаковываются и передаются для доставки на более низкий уровень. С другой стороны, во входящих данных инкапсулированная информация, поступившая с более низкого уровня, не затрагивается вплоть до ее доставки на более высокий уровень. *Инкапсуляция* – процесс размещения блоков данных одного уровня в блоки данных другого. Таким образом, каждый модуль данных протокола обладает своим собственным открывающим компонентом – заголовком пакета (packet header), который идентифицирует используемый протокол, отправителя и предполагаемого получателя. Многие модули данных могут содержать завершающий компонент – *концевик* (trailer). Он осуществляет проверку целостности данных модуля, содержащихся в блоке *полезной нагрузки* (payload). Заключение полезной нагрузки между заголовком и концевиком определяет механизм инкапсуляции, в котором данные с более высокого уровня используются, а затем, перед передачей на нижележащий уровень или отправкой по сетевой среде, снабжаются соответствующим заголовком.

Анализ протоколов. *Анализ протоколов* или *сетевое планирование* – это процесс, предполагающий внедрение в систему сетевых сообщений, перехват проходящих по сети пакетов, отслеживание сетевой статистики и расшифровку пакетов в читаемую форму. Программы-анализаторы протоколов «подслушивают» сетевые сообщения. Многие анализаторы протоколов могут также пересылать пакеты – это полезная функция при тестировании сети или устройств.

Среди анализаторов протоколов, работающих в среде Windows, очень популярными являются два: EtherPeek for Windows (производитель WildPackets, Inc.) и Sniffer Network Analyzer (производитель – Network Associates).

Анализаторы протоколов часто применяются для поиска и локализации неисправностей в сетевых процессах передачи информации. Обычно анализаторы помещаются в узлах сети и настраиваются на захват проблемных последовательностей сообщений. Считывая пакеты, проходящие через кабельную систему, можно выявить дефекты и ошибки этого процесса. Например, если Web-клиент не может подсоединиться к определенному Web-серверу, анализатор протоколов можно применить для перехвата передачи. Экспертиза этой передачи выявляет процессы, используемые клиентом для распознавания IP-адреса Web-сервера, определения аппаратного адреса локального маршрутизатора и подачи Web-серверу запроса соединения.

3.6. Адресация в сети Интернет

Каждый компьютер, включенный в сеть Интернет, имеет уникальный IP-адрес, на основании которого протокол IP передает пакеты в сети. Людям понятнее *символические адреса*, например человек легче запомнит строку www.course.com, чем *числовой адрес* 199.95.72.8. Однако компьютеры имеют дело с *сетевыми адресами* в форме битовых комбинаций, соответствующим числовым адресам.

Таким образом, то что мы в десятичной системе счисления выражаем в виде 199.95.72.8, компьютер читает как 11000111 01011111 01001000 00001000.

Это обстоятельство объясняет, почему протокол IP использует несколько разновидностей схемы адресации:

- *Символическая.* Содержит имена, принимающие специфическую форму, например, support.dell.com. Такие имена называются *доменными* (domain names). Чтобы быть действительным любое доменное имя должно соответствовать хотя бы одному уникальному *числовому IP-адресу* (numeric IP-address). Однако доменные имена, лишь указывают на числовые адреса и не эквивалентны им. Важность и необходимость доменных имен определяется тем, что большинство пользователей запоминает и отождествляет их с определенными хостами в сети.

- *Логическая числовая.* Представляет набор из четырех чисел, разделенных точками, например – 172.16.1.10. Каждое из этих чисел должно быть в диапазоне от 0 до 255, чтобы его можно было представить в виде восьми двоичных разрядов (битов). В термино-

логии TCP/IP они называются октетами. Каждый числовой IP-адрес действует сетевом уровне эталонной модели OSI или на межсетевом уровне модели TCP/IP, и задача его состоит в присвоении уникального набора чисел каждому без исключения сетевому интерфейсу в данной сети, а в масштабе Internet для всех машин, видимых в этой сети. Еще одно название такого способа адресации – *точечное десятичное представление* (dotted decimal notation).

- *Физическая числовая (аппаратная)*. Представляет собой 6-байтный числовой адрес, которым производитель сетевых интерфейсов маркирует программно-аппаратные средства. Три первых байта, называемые организационно уникальным идентификатором (OUI), идентифицируют производителя любого применяемого интерфейса, а три заключительных байта обеспечивают другой уникальный числовой идентификатор, называемый *физическим числовым адресом*. Физический числовой адрес действует на подуровне Канального уровня сетевой модели OSI, называемым *уровнем управления доступом к среде* (MAC – Media Access Control). По этой причине он также известен под названием *MAC-адрес*. Задача подуровня управления логическим соединением канального уровня состоит в том, чтобы предоставить сетевому интерфейсу возможность установления двухточечного соединения с другим сетевым интерфейсом на одном сегменте кабеля. Протокол разрешения адресов (ARP) нужен для преобразования числовых IP-адресов в MAC-адреса, а протокол определения адреса по местонахождению (RARP), напротив, используется для преобразования MAC-адресов в IP-адреса.

Важно запомнить, что IP-адреса ссылаются на доменные имена, позволяя пользователям идентифицировать ресурсы сети и получать к ним доступ. При непосредственном совершении каждой сетевой передачи IP-адреса преобразуются в MAC-адреса, чтобы один сетевой интерфейс можно было идентифицировать как отправителя (sender), а другой – как получателя (receiver).

В соответствии с многоуровневой природой сетевых моделей MAC-адреса можно связать с Канальным уровнем эталонной модели OSI или Уровнем доступа к сети модели TCP/IP. Логические числовые IP-адреса связаны с Сетевым уровнем OSI или с Межсетевым уровнем TCP/IP.

На Канальном уровне один сетевой интерфейс организует передачу фреймов (кадров) от себя к другому сетевому интерфейсу, так, что вся передача проходит в одной физической или локальной сети. По существу, это означает ряд передач от интерфейса к интерфейсу, в ходе которых данные переходят от одного MAC-адреса к другому.

На Межсетевом уровне адрес изначального отправителя представлен в поле исходного IP-адреса в заголовке IP-пакета, а адрес конечного получателя – в поле IP-адреса назначения в том же заголовке. Несмотря на то, что MAC-адреса постоянно изменяются по мере того, как фрейм перемещается от интерфейса к интерфейсу, информация об исходном и конечном IP-адресах сохраняется.

Дублирование числовых IP-адресов запрещается, потому что, когда происходит дублирование, определить какому сетевому интерфейсу действительно принадлежит IP-адрес, очень трудно. В соответствии с соглашением из сети выкидываются все интерфейсы, совместно использующие один и тот же адрес. Таким образом, если после присвоения машине IP-адреса, она не может получить доступ в сеть, то разумно предположить, что произошло дублирование IP-адресов. А если выясняется, что примерно в то же время недоступной стала другая машина, то можно сказать – дело именно в дублировании.

При интерпретации числовых IP-адресов близость двух адресов (особенно, если различаются только один или два правых октета) иногда говорит о том, что машины, которым соответствуют эти адреса, находятся в одной общей сети или даже в одном физическом сегменте кабеля.

Классы IP-адресов. Изначально IP-адреса были разделены на пять классов: А, В, С, D, Е. Рабочими адресами, используемыми в обычных случаях являются адреса классов А, В и С. Адреса классов D и Е предназначены для специальных случаев.

Адреса класса D используются для многопунктовых передач, при которых отдельный адрес может связываться более, чем с одним главным компьютером. В диапазон адресов класса D входят значения от 224.0.0.0 до 239.255.255.255. Они применяются только в случаях, когда информация передается одновременно нескольким получателям, например, приложения для видео- и телеконференций применяют групповые адреса.

Адреса класса E предусмотрены исключительно для экспериментального применения. Адреса класса E могут появляться только в том случае, если в сети проводятся эксперименты или разработка, связанная с протоколом IP. Их диапазон от 240.0.0.0 до 247.255.255.255.

Рассмотри более подробно адреса первых трех классов. Известно, что конструкции IP-адресов классов A, B, C имеют следующие формы:

Класс A *n.h.h.h*

Класс B *n.n.h.h*

Класс C *n.n.n.h*

В данной системе *n* означает часть сетевого адреса, предназначенную для идентификации сети по числу, а *h* – часть адреса для идентификации хоста по числу.

Адреса класса A

Выраженные в двоичной системе адреса класса A всегда принимают следующую форму:

0bbbbbbb.bbbbbbbb.bbbbbbbb.bbbbbbbb

В качестве первой цифры всегда выступает 0, а все остальные цифры (символ b) могут быть либо нулями, либо единицами. Эта схема сокращает количество сетей, фиксируя самый старший разряд. Таким образом, количество сетей к которым можно обратиться находится в диапазоне от 00000000 до 01111111, в десятичном выражении от 0 до 128. Но в любой сети IP-адреса, состоящие только из нулей и только из единиц резервируются для специальных целей, следовательно из 128 возможных сетевых адресов используются только адреса в диапазоне от 00000001 до 01111110 или от 1 до 126. Кроме того, адрес 10 (00001010) зарезервирован для применения в частных сетях. Адрес 127.n.n.n резервируется для тестирования целостности и используемости стека протоколов TCP/IP, установленного на каждом компьютере. Таким образом, в общедоступной сети Internet максимальное количество адресуемых сетей класса A достигает 124. Оставшиеся три октета класса A предназначены для хостов, получается, что в пределах каждой сети класса A доступный диапазон адресов соответствует 24 битам. Число адресов можно подсчитать, возведя 2 в степень, равную количеству бит в адресе: $2^{24} = 16\,777\,216$, после чего вычесть 2 (адреса с сочетанием всех нулей и всех единиц). В результате получаем 16 777 214. Основные данные об адресах класса A представлены в таблице 3.2.

Таблица 3.2 – Адреса класса A

Максимальное количество сетей	$2^7 - 2$	126
Максимальное количество используемых сетей	$2^7 - 2 - 2$	124
Количество хостов в сети	$2^{24} - 2$	16777214
Частный IP-адрес	10.0.0.0	1
Диапазон адресов	1.0.0.0	126.0.0.0

Адреса класса B

Выраженные в двоичной системе, адреса класса B принимают следующую форму:

10bbbbbb.bbbbbbbb.bbbbbbbb.bbbbbbbb

Первые две цифры 1 и 0, а остальные (b) могут быть либо нулями, либо единицами. Эта схема сокращает общее количество возможных сетей, фиксируя два самых старших разряда. Первые два октета определяют номер сети, а дальнейшие – номер хоста. Поэтому в сетевой части на адресное пространство выделяется 14 бит. Таким образом, количество

используемых сетевых адресов равняется $2^{14} - 2 = 16\,382$. Кроме того, 16 адресов этого класса от 127.16.0.0. до 172.31.255.255, резервируются для частного применения. Получается, что общее количество общедоступных IP-адресов класса В равно $16\,382 - 16 = 16\,366$.

Два оставшихся октета адресов класса В выделяются под хосты, это означает, что адресное пространство для хостов в пределах каждой сети класса В составляют 16 бит. Число адресов в хостовой части равно $2^{16} - 2 = 65\,536 - 2 = 65\,534$. Основные данные об адресах класса В приведены в таблице 3.3.

Таблица 3.3 – Адреса класса В

Максимальное количество сетей	$2^{14} - 2$	16382
Максимальное количество используемых сетей	$2^{14} - 18$	16366
Количество хостов в сети	$2^{16} - 2$	65534
Частный IP-адрес	от 172.16.0.0 до 172.31.255.255	16
Диапазон адресов	128.0.0.0	191.255.0.0

Адреса класса С

Выраженные в двоичной системе, адреса класса С принимают следующую форму:

110bbbb.bbbbbbbb.bbbbbbbb.bbbbbbbb

Первые три цифры 1, 1 и 0, а все остальные, обозначенные буквой (b) могут быть нулями или единицами. Эта схема сокращает общее количество возможных сетей, фиксируя три самых старших разряда. Первые три октета адресов класса С определяют номер сети, а последний – номер хоста. Таким образом, в сетевой части на адресное пространство выделяется 21 бит. Поэтому, максимальное количество используемых сетевых адресов равняется $2^{21} - 2 = 2\,097\,150$. Определено, что 256 адресов класса С от 192.168.0.0 до 192.168.255.255, резервируются для частного применения. Получается, что общее количество сетевых адресов класса С равняется $2\,097\,150 - 256 = 2\,096\,894$.

Заключительный октет адресов класса С выделяется под хосты. По этой причине адресное пространство для хостов в пределах каждого сетевого адреса составляет 8 бит. Число адресов можно подсчитать возведением числа 2 в степень равную количеству бит в адресе, а затем вычесть 2. В результате получится 254. Данные об адресах класса С приведены в таблице 3.4.

Таблица 3.4 – Адреса класса С

Максимальное количество сетей	$2^{21} - 2$	2097150
Максимальное количество используемых сетей	$2^{21} - 258$	2096894
Количество хостов в сети	$2^8 - 2$	254
Частный IP-адрес	от 192.168.0.0 до 192.168.255.255	256
Диапазон адресов	192.0.1.0	223.255.255.255

Если сеть, в которую включены компьютеры, является частью Интернета, то сетевой адрес выдает специальное подразделение Интернета – InterNIC (Internet Network Information Center) или его представители. Номер узла определяет администратор сети. Изначально присвоение IP-адресов сетям для общего применения осуществлялось присвоением им конкретного диапазона адресов. К настоящему времени практически не осталось не присвоенных адресов класса А и В. Количество доступных адресов класса С также невелико – на сегодня свободно менее 15% таких адресов. Во избежание полного исчерпания адресного пространства предпринимаются определенные мероприятия, такие как:

использование технологии бесклассовой междоменной маршрутизации; дефицит сетевых адресов породил торговлю существующими сетевыми IP-адресами; использование технологии преобразования сетевых адресов совместно с частными адресами, когда один общедоступный сетевой адрес выступает за произвольное количество частных адресов.

IP-сети, подсети и маски. Если два сетевых интерфейса находятся в одной физической сети, они могут взаимодействовать друг с другом прямо на уровне управления доступом к среде. Но как программное обеспечение «узнает», что сообщения иницируются между двумя машинами в одной физической сети? Ключом к решению проблемы такого распознавания является специальная битовая комбинация, называемая *маской подсети* (subnet mask), которая должна быть определена любого сетевого интерфейса, применяющего TCP/IP.

Каждый из трех основных классов IP-адресов – А, В и С – обладает своей собственной маской подсети по умолчанию. Исследование этих значений дает значительную информацию о маскировании подсетей.

Маска подсети – это специальная битовая комбинация, использующаяся в паре с IP-адресом и содержащая единицы в тех разрядах, которые должны в IP-адресе интерпретироваться как номер сети. Таким образом, маски по умолчанию для классов А, В и С вполне очевидны.

Из таблицы 3.5. видно, что каждый раз, когда в схеме адреса появляется *n*, обозначая сетевую часть адреса, это значение в маске подсети заменяется числом 255, соответствующее битовой комбинации 1111 1111. Каждое значение 255 маскирует один октет, составляющий сетевую часть адреса.

Таблица 3.5 – Маски по умолчанию для основных классов

Класс	Схема	Маска по умолчанию
Класс А	n.h.h.h	255.0.0.0
Класс В	n.n.h.h	255.255.0.0
Класс С	n.n.n.h	255.255.255.0

Здесь *n* – сетевая часть, *h* – хостовая часть адреса

Понятие *подсеть* возникает тогда, когда сетевые адреса подразделяются сверх того, что принято по умолчанию для класса, к которому эти адреса принадлежат. Происходит захват разрядов хостовой части адреса и использование эти захваченных разрядов для создания многочисленных областей в контексте одного сетевого адреса.

Таким образом, маска подсети, которая больше маски по умолчанию для данного адреса, разделяет сетевой IP-адрес на несколько подсетей. Например, для адреса класса В, имеющего по умолчанию маску подсети 255.255.0.0, маска подсети 255.255.192.0 захватывает два разряда хостовой части, чтобы применить их для идентификации подсети. Поскольку десятичное число 192 эквивалентно двоичному 1100 0000, получается, что два первых разряда хостовой части адреса будут использоваться для организации подсетей. Такая схема сетевого адреса называется расширенным сетевым префиксом (extended network prefix).

С помощью двухразрядной маски подсети, приведенной в примере, можно идентифицировать четыре подсети, т.к. доступны только четыре комбинации битов – 00, 01, 10 и 11. Но существующее правило присвоения адресов, резервирующее адреса, состоящие только из нулей и только из единиц уменьшает количество используемых подсетей на два значения.

Процесс захвата разрядов из хостовой част адреса для дальнейшего подразделения сетевой части адреса называется *фрагментацией сетевого адреса*, или *организацией подсетей*.

С точки зрения маршрутизации, организация подсетей позволяет сетевым администраторам сочетать подсети с фактическими областями маршрутизации сети, чтобы компьютеры, находящиеся в одной физической сети, могли взаимодействовать с помощью MAC-адресов. Другие компьютеры, находящиеся в другой физической сети, будут принадлежать другой подсети и, соответственно, их числовые адреса будут различаться в части, относящейся к подсетям. Когда компьютер в одной подсети собирается наладить сообщение с компьютером в другой подсети, трафик должен пересылаться от отправителя к ближайшему IP-шлюзу для того, чтобы переправить сообщение из одной подсети в другую.

IP-шлюз (IP gateway) – это устройство, обеспечивающее связь различных IP-сетей или подсетей. IP-шлюз часто называют *маршрутизатором* (router), поскольку он обычно хранит информацию о «достижимости» различных сетей, выбирает лучший (кратчайший, скорейший или характеризующийся меньшими издержками) путь или «маршрут» для каждого получаемого пакета, а затем отправляет его по адресу назначения.

Расчет масок подсетей. Есть несколько разновидностей масок подсети. Их реализация зависит от желаемой схемы сегментации адреса. Простейшая форма маски подсети – это *маска подсети постоянной длины* (CLSM – Constant-Length Subnet Masking). При реализации этой схемы каждая подсеть включает одно и то же количество станций и представляет собой простое разделение адресного пространства.

Другая разновидность маски подсети – это *маска подсети переменной длины* (VLSM – Variable-Length Subnet Masking). Она позволяет разделить адрес на несколько подсетей, каждая из которых необязательно равна по размеру другим. При разработке схемы маскирования подсетей необходимо учитывать количество устройств, которые будут находиться в каждой подсети. Если оно колеблется в пределах не более 20% в ту или иную сторону, то разумнее всего применить схему с маской постоянной длины. Но если на один или два сегмента приходится большое количество устройств, а на остальные сегменты – намного меньшее, то более эффективное распределение адресного пространства может обеспечить схема с маской переменной длины.

Разработка маски подсети постоянной длины. Чтобы создать маску подсети постоянной длины (CLSM), в которой каждая часть сети обладает одинаковым количеством адресов, необходимо выполнить следующие действия:

1. Определить требуемое количество подсетей.
2. Поскольку количество подсетей должно быть выражено битовой комбинацией, к нему необходимо прибавить 2 (1 для сетевого адреса, 1 для широковещательного адреса) и увеличить получившееся число до ближайшей степени двойки. Если число равно степени двойки, то его можно использовать.
3. Расписать разряды хостовой части адреса в обратном направлении.
4. Убедиться, что в каждой подсети осталось достаточное количество доступных ведущих адресов.
5. Вычислить доступное количество подсетей по формуле: 2^{b-2} , где b представляет собой количество разрядов в маске подсети, а 2 вычитается, чтобы исключить сетевой (все нули) и широковещательный (все единицы) адреса, которые необходимы в каждой сети и подсети.

Пример 1. Разработки маски подсети постоянной длины для адреса класса C.

Пусть некая организация ABC Inc. хочет создать 12 подсетей для своего адреса класса C – 200.10.10.0. В каждой подсети не должно быть более 10 ведущих адресов.

1. Требуемое количество подсетей равно 12 – определяется условиями задачи.
2. $12 + 2 = 14$. Ближайшая степень двойки равна 16 или 24. Таким образом, нужна четырехразрядная маска подсети.

3. Резервируем четыре старших разряда из последнего октета маски подсети класса C по умолчанию и получаем битовую комбинацию – 1111 0000 для маски подсети. Десятичное значение для этого двоичного числа $2^7 + 2^6 + 2^5 + 2^4 + 0 + 0 + 0 + 0 = 128 + 64 + 32$

$+ 16 = 240$. Тогда маска подсети по умолчанию класса С расширяется с 255.255.255.0 до 255.255.255.240.

4. При подсчете количества ведущих адресов в каждой подсети действует правило, что ни один разряд, применяемый для маски подсети, не должен быть использован для ведущих адресов. Подсчитав количество оставшихся нулей в маске подсети (1111 0000), получаем 4.

5. Формула $2b - 2$ справедлива и для подсчета количества хостов, только при этом b становится количеством разрядов в ведущем адресе, подсчитанных на предыдущем шаге. Таким образом, количество хостов в каждой подсети равно $24 - 2 = 14$.

Сравнив полученные результаты с условиями задачи, видим, что количество хостов в каждой подсети $14 > 10$, т.е. разработанная схема маскирования подсетей с четырехразрядной маской 240 (1111 0000) позволяет предоставить по 14 ведущих адресов для каждой подсети, а требуется не более 10.

Пример 2. Разработка маски подсети постоянной длины для адреса класса В. Пусть корпорация XYZ, крупная международная компания с сотнями филиалов по всему миру, желает организовать 300 подсетей для своего сетевого адреса класса В – 178.16.10.0. В каждой подсети должно быть не больше 100 ведущих адресов.

1. Вычисляем количество разрядов в маске подсети: $300 + 2 = 302$. Ближайшая степень двойки 512 или 29. Таким образом, нужна 9-разрядная маска подсети.

2. Зарезервировав 9 разрядов, начиная со старшего, получаем маску подсети с битовой комбинацией 11111111 10000000. Поскольку эта маска распространяется на два октета, то для каждого из них подсеть необходимо подсчитывать отдельно. Для первого октета результат получается равным 255 (во всех разрядах стоят единицы). Во втором октете единица находится только в старшем разряде. Переводим эту двоичную комбинацию 1000 0000 в десятичное число: $2^7 = 128$. Соответственно, маска подсети класса В по умолчанию 255.255.0.0 расширяется до 255.255.255.128.

3. Количество нулей в оставшейся части маски подсети определяет количество разрядов, возможных для использования в ведущем адресе – 7.

4. Вычисляем количество хостов в каждой подсети $2^7 - 2 = 128 - 2 = 126$. Таким образом, исходя из требований корпорации XYZ разработана маска подсети 255.255.255.128, позволяющая использовать в каждой подсети по 126 хостов, при необходимых 100. Это вполне удовлетворяет условиям поставленной задачи.

Расчет суперсетей. Суперсети захватывают разряды из сетевой части IP-адреса. Суперсети позволяют различным сетевым IP-адресам объединяться и действовать совместно, как если бы они представляли собой единую логическую сеть. В результате заметно повышается эффективность передачи сообщений внутри локальной сети, т.к. исчезает необходимость внутренней маршрутизации. В некоторых случаях это также делает возможным адресацию к суперсети большего количества хостов, чем при сочетании множества адресов.

Суперсети используются, в основном для адресов класса С, захватывают 3 разряда из сетевой части адреса и добавляют их в хостовую часть. Поэтому битовая комбинация третьего октета маски подсети по умолчанию класса С, имеющая вид 1111 1111 становится меньше на три разряда и принимает вид 1111 1000. Если выразить число разрядов, переданных в хостовую часть адреса в десятичном выражении, то получится $111 = 2^2 + 2^1 + 2^0 = 4 + 2 + 1 = 7$. Это означает, что в результате маска подсети по умолчанию 255.255.255.0 сокращается и принимает форму 255.255.248.0.

Теперь хостовая часть адреса имеет 11 разрядов – 111 0000 0000. Количество доступных хостов для данной суперсети подсчитывается по известной формуле $2^{11} - 2 = 2046$. Каждый отдельный адрес класса С по умолчанию может адресовать только 254 ($2^8 - 2$) хоста. Если сравнивать адресное пространство для хостов по умолчанию, например, для восьми адресов $254 \times 8 = 2032$ с адресным пространством, предоставляемым суперсетью

2046, можно обнаружить $2046 - 2032 = 12$ хостов, которые не обнаруживались при использовании каждого адреса по отдельности.

Еще одно важное обстоятельство состоит в том, что организация суперсети позволяет получить доступ к группе хостов через один адрес маршрутизатора, а сами эти хосты могут налаживать друг с другом прямое взаимодействие, обходясь без маршрутизации в данном адресном пуле.

Организация суперсетей используется поставщиками доступа в Internet, которые могут объединять множество адресов класса С, и таким образом обслуживать большее количество пользователей, чем это казалось возможным. Сочетание организации суперсетей с техникой маскирования подсетей маской переменной длины дает поставщикам доступа в Internet (ISP) возможность разделять и дробить адресные пространства с большой эффективностью.

Частные и общедоступные адреса. В пределах классов А, В и С выделяются отдельные адреса и адресные пространства для использования в качестве частных IP-адресов (private IP address). Это означает, что любая организация может применить эти частные адреса в пределах своих собственных сетевых доменов, не получая предварительного разрешения и не производя никаких расходов. Диапазоны частных IP-адресов представлены в таблице 3.6.

Таблица 3.6 – Диапазоны частот IP - адресов

Класс	Адрес (диапазон)	Кол-во сетей	Общее количество частных хостов
А	10.0.0.0	1	16777214
В	172.16.0.0	16	1048544
С	192.168.0.0-192.168.255.0	256	65024

При выборе типов частных IP-адресов можно руководствоваться необходимым количеством хостов на сеть, готовностью разделять на подсети большие классы IP-адресов (обычно класса В) или организовывать суперсети для меньших классов адресов (обычно класса С). Типовым для организации, имеющей несколько физических сетей, является решение применить адреса класса В и подразделить их на подсети в третьем октете, оставляя четвертый для адресов хостов (поскольку локальные сети в большинстве своем невелики, ограничение 254 хоста на подсеть не представляет никакой проблемы).

Один из недостатков частных IP-адресов – их маршрутизация в общедоступной сети Internet невозможна. Соответственно, если нужно подсоединить компьютеры к Internet, и при этом использовать частные IP-адреса в пределах локальных сетей, то на устройство, расположенное на границе между частной стороной сети и ее общедоступной стороной, необходимо установить дополнительное программное средство, которое может обеспечить работоспособность такой схемы. Оно должно обеспечивать преобразование частных IP-адресов в общедоступные для исходящего трафика и обратное преобразование для входящего трафика. *Имитация адреса* или *подстановка адреса* выполняется прокси сервером (proxy server).

Частные IP-адреса представляют ценный ресурс для всех организаций, потому что подавляющее большинство компьютеров в сетях TCP/IP – это клиентские рабочие станции. Поскольку они никогда не предоставляют сервисные услуги широкой аудитории Internet для них частные IP-адреса являются наиболее подходящими. Клиенты могут читать электронную почту, получать доступ к сети Internet и другим службам, а также пользоваться локальным сетевыми ресурсами. Ни одно из этих требований не предполагает применения общедоступных IP-адресов и не запрещает применение частных адресов.

Общедоступные IP-адреса являются средствами идентификации серверов и служб, которые должны быть доступны в сети Internet. Большинству организаций общедоступные IP-адреса нужны для двух видов оборудования:

- внешние интерфейсы пограничных устройств всех типов, включая маршрутизаторы, прокси-серверы и брандмауэры, выполняющие роль защитного периметра между внешней и внутренней частью сетей;

- серверы, предназначенные для обеспечения доступа к ним из сети Internet. Среди них общедоступные Web-серверы, почтовые серверы, FTP-серверы и новостные серверы.

Сюда могут входить любые службы прикладного уровня модели TCP/IP, которые организация может счесть нужными выставить для доступа к ним из сети Internet.

Технология *преобразования сетевых адресов* (NAT – Network Address Translation) и средства подстановки необходимы при использовании частных IP-адресов. Это объясняется тем, что взаимодействие клиентов с сетью в границах локальной сети без какой-либо формы «скрытия» адресов может привести к тому, что адресная структура такой внутренней сети станет известна посторонним людям, обладающим достаточной квалификацией. Они могут использовать эту информацию для проникновения внутрь организации – из общедоступной сети Internet во внутреннюю корпоративную сетевую среду, и это проникновение не сулит ничего хорошего. Вот почему для обеспечения безопасности IP считается уместным применять прокси-сервер, позиционирующийся на границе между внутренним и внешним трафиком. Когда исходящий трафик проходит через прокси-сервер из внутренней сети, он заменяет внутренние сетевые адреса одним или несколькими адресами, чтобы трафик, перемещающийся в общедоступную сеть Internet не мог стать источником информации об адресной структуре внутренней сети.

Прокси-сервер может обеспечить обратное посредничество (reverse proxying). Он получает возможность обращаться к серверам во внутренней сети, в то же время, демонстрируя окружающему миру только свой собственный адрес и пересылая внутренним серверам для дальнейшей обработки только допустимые запросы. Посторонние наблюдатели могут узнать при этом только адрес прокси-сервера, действующего от имени внутреннего устройства. Это помогает избежать утечки информации о фактических адресах внутренней сети. Поэтому, блокируя эту информацию, администраторы IP-сетей ограничивают потенциальную возможность проникновения в систему и других атак.

Организации получают общедоступные IP-адреса у своего провайдера, обеспечивающего доступ к сети Internet. Это одна из причин, которая удерживает организации от перехода от одного провайдера к другому. Поскольку все устройства, доступные в сети Internet должны иметь собственные публичные IP-адреса, перемена провайдера связана с трудоемкой процедурой, называемой IP-перенумерацией. При этом нужно заменить адреса на всех машинах, использующих адрес старого провайдера, на новые. Раньше всеми IP-адресами, номерами протоколов и хорошо известными адресами портов заведовало агентство IANA (Internet Assigned Numbers Authority) – Агентство по выделению имен и уникальных параметров Internet. Оно же отвечало за присвоение MAC-адресов сетевым интерфейсам. Теперь всем этим занимается организация ICANN (Internet Corporation for Assigned Names and Numbers) – организация по присвоению имен и номеров в сети Internet. Но, учитывая нехватку общедоступных IP-адресов, только поставщики доступа в Internet (ISP – Internet Service Provider) могут получить их в ICANN. В большинстве случаев организации, наряду с другими возможными Internet-услугами получают общедоступные адреса у своего провайдера.

3.7. Система доменных имен DNS

Человеку крайне неудобно использовать числовые IP-адреса, поэтому логичным представляется создание механизма, позволяющего ставить в соответствие IP-адресам символьные имена.

В сети Интернет для этой цели используется система доменных имен (DNS – Domain Name System), которая обеспечивает жизненно важную возможность преобразования символьных доменных имен в соответствующие числовые IP-адреса. Кроме того, DNS поддерживает другие службы, например она преобразует адреса электронной почты в IP-адреса серверов, обрабатывающих соответствующий трафик.

Совокупность всех баз данных DNS содержит полную схему соответствия всех действительных доменных имен действительным IP-адресам, и такие базы данных разбросаны по всей сети Internet (а значит и по всему земному шару). Каждая отдельная база данных функционирует независимо от других и, относительно самостоятельно, управляет содержащимися в ней данными. Несмотря на свою рассредоточенную и децентрализованную структуру, DNS прекрасно функционирует, обеспечивая мощную, надежную и стабильную базу для Internet-адресации.

Структура баз данных DNS. Структура базы данных DNS отражает структуру самого пространства присваиваемых имен доменов. Это иерархическая древовидная структура, у которой корень (root) изображается вверху схемы. Аналогичную структуру имеет жесткий диск, в корне которого изображается буква, представляющая сам диск, а от нее «ветвятся» многочисленные уровни каталогов и подкаталогов. В пространстве присваиваемых имен домены соприкасаются в корне, который обозначается точкой «.». Прямо под точкой находятся первичные домены (домены верхнего уровня). Система имен доменов верхнего уровня имеет четко выраженный региональный характер и определяется в стандарте 3166 Международной организации по стандартизации (ISO). Домены первого уровня назначаются для каждой страны, при этом принято использовать трехбуквенные и двухбуквенные аббревиатуры.

Например, **ru** – России, **ua** – Украина, **su** – бывший Советский Союз, **ca** – Канада, **us** – США, **uk** – Великобритания, **de** – Германия, **jp** – Япония.

Несмотря на отсутствие центра управления Интернетом, есть организации, занимающиеся проверкой и выдачей адресов, например InterNIC, контролирующая систему корневых доменов.

К настоящему времени зарегистрировано около 80 имен доменов первого уровня. Кроме того, несколько имен доменов первого уровня выделено для различных типов организаций и ресурсов:

- com – коммерческие организации (например, ibm.com);
- edu – образовательные организации (например, spb.edu);
- gov – правительственные организации (например, loc.gov);
- org – некоммерческие организации (например, w3.org);
- net – организации, поддерживающие сети (например, rtpn.net);
- mil – военные организации;
- biz – ресурсы для бизнеса;
- info - информационные ресурсы.

Название домена второго уровня обычно соответствует названию фирмы или организации, либо подчеркивает назначение сервера. Например, www.microsoft.com. Здесь в качестве имени домена второго уровня указано название корпорации Microsoft. Такая схема построения имен упрощает их запоминание, а иногда дает возможность просто угадать имя нужного сервера. Так, по аналогии с предыдущим именем можно выйти на Web-сервер компании IBM: www.ibm.com.

В основе системы доменных имен лежит иерархический принцип. Имя строится из нескольких элементов, между которыми ставятся точки. Имя читается справа налево. Так самым старшим элементом в имени www.gambler.ru будет домен *ru*, ему подчинен домен *rambler*. Младшая часть доменного имени соответствует конечному узлу сети. У одного узла может быть несколько имен. Совокупность имен, у которых несколько старших частей доменного имени совпадают, называется доменом. Например, имена mail.econ.ru.ru и www.econ.ru.ru принадлежат домену *econ.ru.ru*.

Чтобы система DNS могла работать, необходимо выполнение еще одного существенного условия – каждому уникальному доменному имени должен соответствовать хотя бы один действительный IP-адрес. Поддержание таких соответствий является первичной функцией службы DNS.

Для каждого имени домена создается свой DNS-сервер, который хранит базу данных соответствий IP-адресов и доменных имен, расположенных в данном домене, а также содержит ссылки на DNS-серверы доменов нижнего уровня. Таким образом, для того чтобы получить адрес компьютера по его доменному имени, приложению достаточно обратиться к DNS-серверу корневого домена, а тот, в свою очередь, перешлет запрос DNS-серверу домена нижнего уровня. Важно понимать, что каждое действительное доменное имя в конечном итоге располагается в определенной базе DNS, причем данной записью заведует один конкретный сервер. Благодаря такой организации системы доменных имен нагрузка по разрешению имен равномерно распределяется среди DNS-серверов.

Работа серверов доменных имен. Когда клиент TCP/IP отправляет запрос имени на DNS-сервер, адрес этого сервера он запрашивает из своих конфигурационных данных TCP/IP. Серверы запрашиваются в том порядке, в котором они перечислены в соответствующих записях реестра. Порядок следования серверов можно изменить и, тем самым, распределить нагрузку между несколькими DNS-серверами. Процесс ответа сервера на запрос доменного имени состоит из следующих действий:

- DNS-серверы извлекают данные об имени из общего пространства присваиваемых имен доменов, причем любой DNS-сервер может осуществить поиск по своему КЭШу и ответить на запрос.
- Если в базе данных или в КЭШе локального сервера не находится нужной информации, он может начать исполнять роль специального кэширующего сервера по отношению к другим известным серверам имен в его «соседстве».
- Если ни один из этих поисков не приводит к желаемому результату, сервер имен посылает запрос на разрешение имени к корневому серверу, который перенаправляет запрос на *ответственный сервер* (authoritative server) за данный сегмент базы данных, связываясь с корневым сервером соответствующего домена и следуя указателям серверов имен (NS) вплоть до получения доступа к ответственному серверу. Этот процесс всегда завершается каким-либо ответом – IP-адресом, соотносимым с именем, которое необходимо разрешить или сообщением об ошибке.

Такой процесс называется *разрешением доменных имен* (domain name resolution) или просто разрешением имен. В файлах баз данных, формирующих данные DNS, содержится информация, необходимая для самой системы DNS, записи маркировки источника полномочий, обработки почтовых записей, информация о хорошо известных службах. Файлы, связывающие имена хостов с адресами, обычно обозначаются *имя_домена.dns*, где *имя_домена* – локальное имя домена или субдомена для зоны, относящейся к DNS-серверу. Например, DNS-сервер для домена lanw.com, называется lanw.com.dns. Файлы, которые связывают адреса с доменными именами при *операции обратного поиска*, как правило, обозначаются как *addr.in-addr.arpa.dns*, где *addr* – это номер сети домена в обратном порядке без завершающих нулей. Например, для домена lanw.com, чей номер сети – 208.224.65.0, файл называется 65.224.206.in-addr.arpa.dns.

В таких файлах содержится статическая копия базы данных DNS в том виде, в котором она существовала в момент ее копирования на диск или перед включением DNS-сервера

3.8. Сервисы сети Интернет

Электронная почта: E-mail. Система электронной почты позволяет доставить сообщение на любой компьютер, включенный в сеть Интернет. Сообщение может содержать текст или файл практически любого формата – графику, музыку и т. д. Все пользователи электронной почты имеют уникальные адреса. Адрес пользователя зарегистрирован в определенном домене Интернета. С каждым доменом связан почтовый сервер, управляющий адресами пользователей. Пользователь набирает текст письма в специальной программе, которая называется почтовым клиентом. Эта программа, в зависимости от возможностей, позволяет создавать и редактировать новые письма, обрабатывать пришедшие, хранить и систематизировать переписку и т.д. Почтовый клиент помещает письмо в

«почтовый ящик» пользователя, расположенный на почтовом сервере. Сервер, в свою очередь, передает письмо на почтовый сервер адресата.

Адрес электронной почты. В Интернете принята система адресов, которая базируется на доменном адресе машины, подключенной к сети. Адрес пользователя состоит из двух частей, разделенных символом «@» («альфиус» или «собака»), например: Ivan@rol.ru.

В данном случае

Ivan – это имя пользователя,

rol.ru – доменное имя почтового сервера.

Формат сообщения электронной почты. Почтовое сообщение состоит из трех частей: конверта, заголовка и тела сообщения. Пользователь видит только заголовок и тело сообщения. Конверт используется только программами доставки. Заголовок всегда находится перед телом сообщения и отделен от него пустой строкой. Заголовок состоит из полей. Поля состоят из имени поля и содержания поля. Имя поля отделено от содержания символом ":". Минимально необходимыми являются поля «Date», «From», «cc» или «To», например:

From: Sergey V. Stahurlov

Date: 7 августа 2005 г. 15:41

To: ebook-free@mail.ru

Поле «From» определяет отправителя, поле «Date» определяет дату отправки сообщения, поле «To» – получателя. Если используется поле «Cc», то в нем указываются адреса нескольких получателей. Заголовок может содержать дополнительные поля:

Date: 27 Aug 76 0932

From: Ken Davis

Subject: Re: The Syntax in the RFC

Sender: KSecy@Other - host

Reply to: Sam_Irving@reg.organization

To: George Jones

cc: Important folks- Tom Softwood, "Sam_Irving"@Other – Host;

Message-ID: 4331.629.XYzi – What@Other - Host

В данном случае поле «Sender» указывает, что Ken Davis не является автором сообщения. Он только переслал сообщение, которое получил из KSecy@Other-host. Поле «Subject» определяет тему сообщения, «Reply to» – пользователя, которому отвечают. В поле «cc» указываются другие получателя этого сообщения. Поле «Message-ID» содержит уникальный идентификатор сообщения и используется программами доставки почты. Для описания тела почтового сообщения Интернета предназначен стандарт MIME (или, в нотации Интернета, RFC-1341) Все почтовые программы поддерживают это многоцелевое расширение Internet Mail (MIME), которое позволяет прикреплять к сообщению файлы различных форматов. В стандарте MIME зарезервировано несколько способов представления разнородной информации. Стандарт определяет семь типов данных, которые можно передавать в телеписьма:

- текст (text);
- смешанный тип (multipart);
- почтовое сообщение (message);
- графический образ (image);
- аудиоинформация (audio);
- видео (video);
- приложение (application).

Возможности электронной почты не ограничиваются только пересылкой корреспонденции. По почте можно получить доступ ко многим ресурсам Интернета, которые используют почтовых роботов, отвечающих на запросы пользователей.

Протокол SMTP. Протокол SMTP (Simple Mail Transfer Protocol), *простой протокол электронной почты* является самым распространенным протоколом отсылки электронной почты в сети Internet. Он отправляет и принимает сообщения электронной почты с помощью процессов SMTP-отправителя (Sender-SMTP) и SMTP-получателя (Receiver-SMTP), которые обеспечивают службы передачи и приема. Базовая структура SMTP-сообщения изображена на рис.3.3.

Процессы Sender-SMTP и Receiver-SMTP являются клиентскими или серверными приложениями, поддерживающими функциональные возможности протокола SMTP. Обычно Sender-SMTP передает запрос на инициализацию SMTP-соединения через порт 25 сервера Receiver-SMTP. После установления соответствующего TCP-соединения Receiver-SMTP отвечает посредством числового кода 220, обозначающего готовность сервера принимать команды и информацию протокола SMTP.

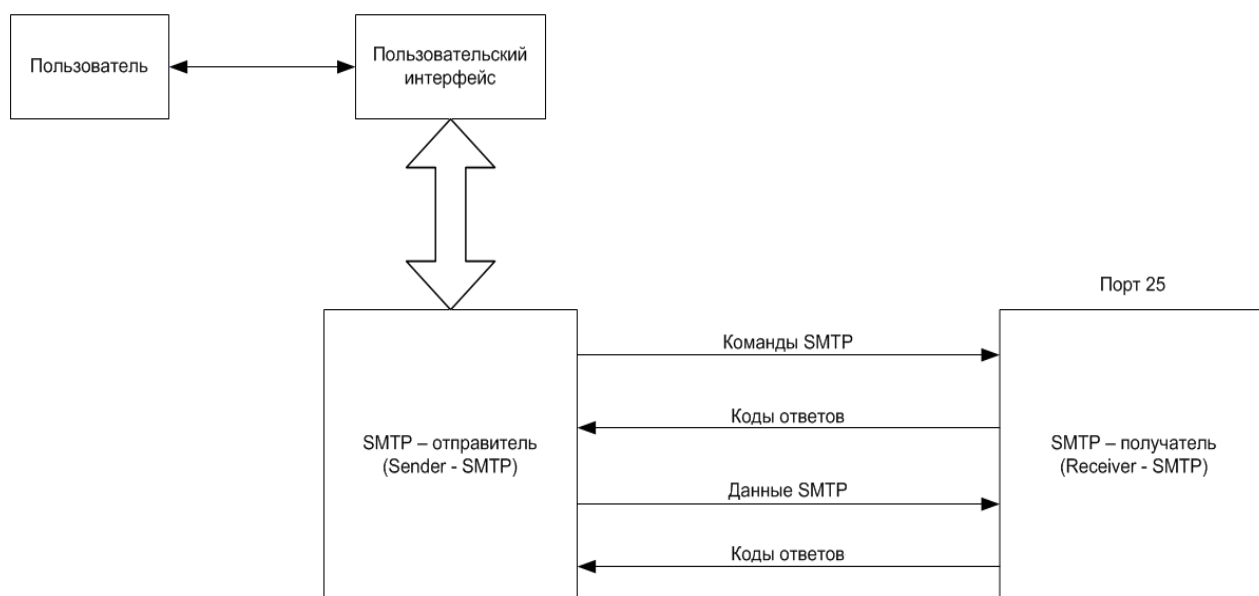


Рис.3.4. Передача электронной почты по протоколу TCP

После чего становится возможной отсылка почтовых команд и сообщения. Receiver-SMTP отсылает коды ответов, реагируя на отправленные команды и почтовые сообщения. Исходный набор команд протокола SMTP был определен в документе RFC 821 и постоянно расширялся по инициативе агентства IANA (агентство по выделению имен и уникальных параметров протоколов Internet). В таблице 3.7. приведен перечень некоторых команд протокола SMTP.

Таблица 3.7 – Команды протокола SMTP

Команда	Описание
1	2
HELO	Иницирует SMTP-сеанс
HELLO	Иницирует SMTP-сеанс с поддержкой расширения электронной почты
MAIL	Иницирует почтовую передачу, идентифицирует адрес электронной почты Sender- SMTP
RCPT	Идентифицирует получателя почты
DATA	Иницирует передачу почтовых данных
VRFY	Проверяет существование получателя
QUIT	Завершает сеанс связи по протоколу SMTP
SEND	Отправка на терминал пользователя

1	2
TURN	Изменение функций приложений протокола SMTP (отправитель – на получатель, получатель – на отправитель)
BBITIME	Использование 8-битных данных
SIZE	Объявление размера сообщения
CHECKPOINT	Контрольная точка/перезапуск
PIPELINING	Конвейерная обработка команд
DSN	Оповещение о состоянии доставки

Процесс Receiver-SMTP отправляет ответы состояния, содержащие числовой код ответа (таблица 3.8).

Таблица 3.8 – Коды ответов протокола SMTP

Код ответа	Описание
211	Системный ответ состояния или справочный ответ
214	Справочное сообщение
220	<домен> Служба готова
221	<домен> Служба закрывает канал передачи
250	Запрошенное действие принято и выполнено
251	Пользователь не локален: перенаправление по адресу <путь переадресации>
354	Запуск передачи почты – оканчивается <CRLF> <CRLF>
421	<домен> Служба недоступна; закрытие соединения
450	Почтовый ящик недоступен: запрошенное действие не выполнено
451	Локальная ошибка в ходе обработки; действие прекращено
452	Недостаточно системной памяти; действие не выполнено
500	Синтаксическая ошибка, команда не опознана
501	Синтаксическая ошибка в параметрах и аргументах
502	Команда не реализована
503	Некорректная последовательность команд
504	Параметр команды не реализован
550	Почтовый ящик недоступен; действие не выполнено
551	Пользователь не локален; попробовать путь <путь переадресации>
552	Распределение памяти превышено; действие прекращено
553	Запрашиваемое имя почтового ящика; действие не выполнено

Рассмотрим некоторые действия, выполняемые во время отправки электронной почты посредством протокола SMTP. В таблице 3.9 представлена последовательность таких действий, где *S* – обозначает SMTP-сервер, а *C* – SMTP-клиента.

Таблица 3.9 – Последовательность действий SMTP-сообщений

Процесс	Отправить /получатель	Последовательность пакетов
Запрос на SMTP-TCP-соединение	<i>C</i> > <i>S</i>	TCP SYN
Ответ сервера	<i>S</i> > <i>C</i>	TCP SYN ACK
Установка портов: Порт сервера 25; Порт клиента 1047	<i>C</i> > <i>S</i>	TCP ACK
Обмен исходными сообщениями	<i>S</i> > <i>C</i>	220: Сервер готов (дата/время)

Отправка почты на SMTP-сервер	C > S	HELO
	S > C	250: приветствие (IP-адрес клиента)
		250: BBITMIME – использование 8-битных данных
		250: SIZE – объявление размера сообщения
	C > S	MAIL: включает адрес электронной почты источника
	S > C	250: отправитель подтвержден
	C > S	RCPT: включает адрес электронной почты получателя
	S > C	250: получатель подтвержден
	C > S	DATA
	S > C	354: запуск подачи почты
Закрытие SMTP и TCP-соединения	C > S	Клиент отправляет сообщение электронной почты
	S > C	250: сообщение принято
	C > S	QUIT
	C > S	TCP FIN
	S > C	221: завершение соединения и подтверждение (ACK)
	C > S	Сброс (клиент закрыл ожидающий порт 1047)

Спецификация протокола SMTP содержится в документе RFC 821.

В некоторых клиентах электронной почты и межсерверных почтовых передачах в сети Internet эта служба работает скрытно. Большинство почтовых клиентов применяют протокол SMTP для отсылки исходящей почты через близлежащий SMTP-сервер, при этом предполагается хранение сообщений, если адресат принадлежит к тому же домену или их переадресация, если адресат принадлежит другому домену.

Система World Wide Web. В 1989 году гипертекст представлял новую многообещающую технологию, которая имела относительно большое число реализаций, с одной стороны, а с другой стороны, делались попытки построить формальные модели гипертекстовых систем, которые носили скорее описательный характер и были навеяны успехом реляционного подхода описания данных. Идея создания системы WWW заключалась в том, чтобы применить гипертекстовую модель к информационным ресурсам, распределенным в сети, и сделать это максимально простым способом. WWW заложил три краеугольных камня системы из четырех существующих ныне, разработав:

- язык гипертекстовой разметки документов HTML (HyperText Markup Language);
- универсальный способ адресации ресурсов в сети URL (Universal Resource Locator);
- протокол обмена гипертекстовой информацией HTTP (HyperText Transfer Protocol).

Позже команда разработчиков добавила к этим трем компонентам четвертый: универсальный интерфейс шлюзов CGI (Common Gateway Interface).

Концепция WWW. Архитектура системы WWW построена по хорошо известной схеме «клиент-сервер». Программа-клиент (браузер) выполняет функции интерфейса пользователя и обеспечивает доступ практически ко всем информационным ресурсам Интернета. Он объединяет различные сервисы, и в этом смысле, браузер выходит за обычные рамки работы клиента только с сервером определенного протокола. Браузер – это интерпретатор HTML.

И, как типичный интерпретатор, в зависимости от команд (разметки) выполняет различные функции. В круг этих функций входит не только размещение текста на экране, но и обмен информацией с сервером по мере анализа полученного HTML-текста, что наиболее наглядно происходит при отображении встроенных в текст графических изображений. При анализе URL-спецификации или по командам сервера браузер запускает дополнительные внешние программы для работы с документами в форматах, отличных от HTML, например GIF, JPEG, MPEG, Postscript и т.п.

Другую часть программного комплекса WWW составляет сервер протокола HTTP, базы данных документов в формате HTML, управляемые сервером, и программное обеспечение, разработанное в стандарте спецификации CGI. База данных HTML-документов – это часть файловой системы, которая содержит файлы в формате HTML и связанные с ними графику и другие ресурсы. Среди прикладного программного обеспечения, работающего с сервером, можно выделить программы-шлюзы. *Шлюзы* – это программы, обеспечивающие взаимодействие серверов различных протоколов, например, через FTP (рис.3.5).

Во второй половине 90-х годов произошли некоторые изменения в архитектуре сервиса World Wide Web. В дополнение к HTML стали активно применяться другие языки разметки. Рабочая группа W3 Consortium разработала спецификацию языка XML (Extention Markup Language) – расширяемого языка разметки. Он позволяет упростить соединения с базами данных, отделяя данные от интерфейса. Изменения коснулись и клиентской части технологии. В настоящее время происходит постепенный переход от простой классической архитектуры клиент-сервер к трехуровневой архитектуре с сервером приложений, в роли которого выступает программа-клиент.

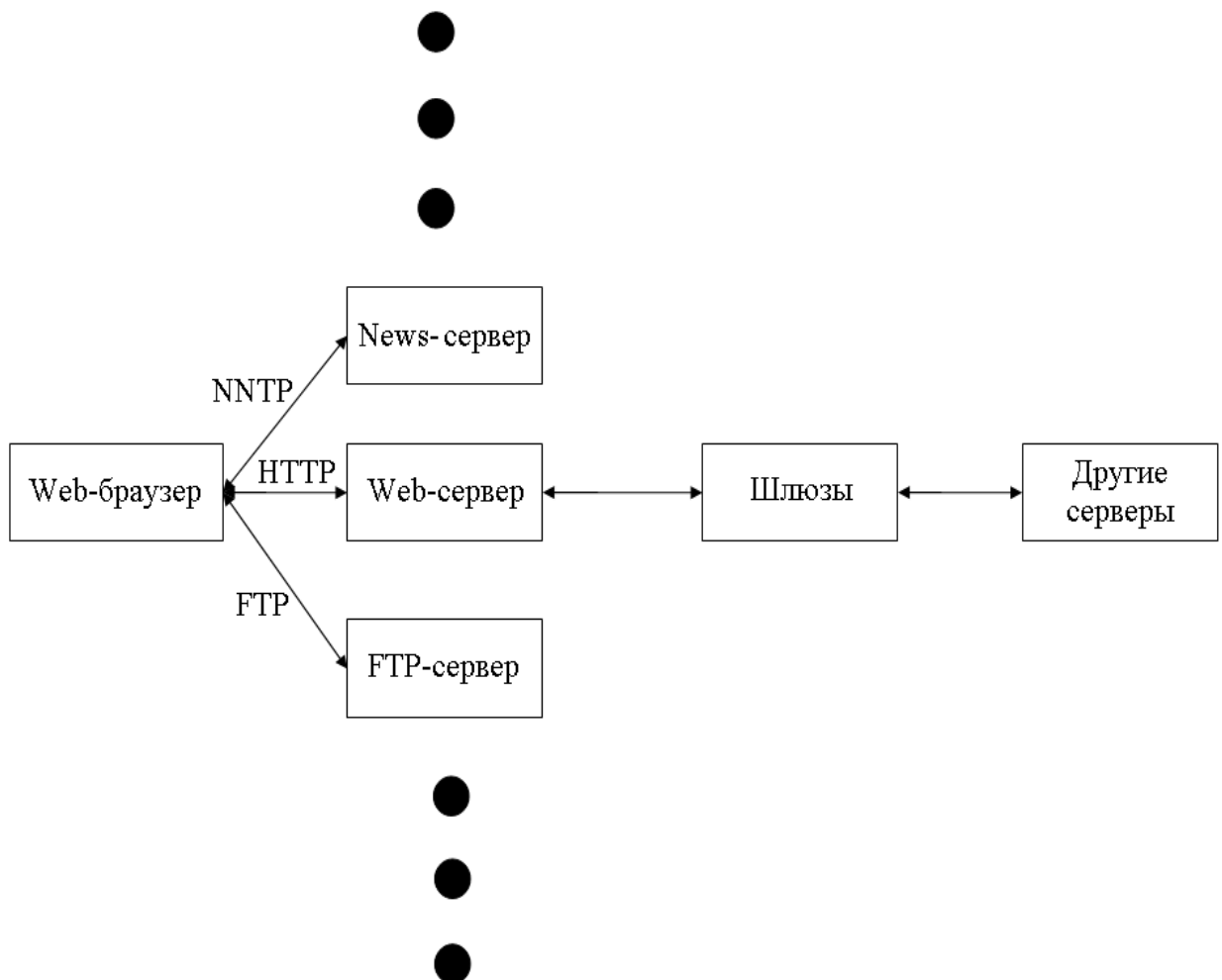


Рисунок 3.5 – Браузер как интерфейс с различными сервисами Интернета

Язык разметки гипертекста (HTML). Идея HTML – пример чрезвычайно удачного решения проблемы построения гипертекстовой системы при помощи специального средства управления отображением. В 1989 году активно обсуждалась проблема интерфейса гипертекстовых систем, то есть способов отображения гипертекстовой информации и навигации в гипертекстовой сети. Значение гипертекстовой технологии сравнивали со значением книгопечатания. Утверждалось, что лист бумаги и компьютерные средства отображения / воспроизведения серьезно отличаются друг от друга, и поэтому форма представления информации тоже должна отличаться.

Наиболее эффективной формой организации гипертекста были признаны контекстные гипертекстовые ссылки, а, кроме того, было признано деление на ссылки, ассоциированные со всем документом в целом и с отдельными его частями. В HTML гипертекстовые ссылки встроены в тело документа и хранятся как его часть. В WWW документы – это обычные ASCII-файлы, которые можно подготовить в любом текстовом редакторе. Таким образом, проблема создания гипертекстовой базы данных была решена чрезвычайно просто.

Большая часть документов в системе World Wide Web хранится в формате HTML – это язык гипертекстовой разметки, используемый для кодирования документов.

Язык HTML представляет собой набор команд, в соответствии с которыми браузер отображает содержимое документа, команды HTML не отображаются в языке HTML реализован механизм гипертекстовых ссылок, который обеспечивает связь одного документа с другими. Эти документы могут находиться на том же сервере, что и страница, с которой на них делается ссылка, а могут быть размещены на другом сервере (рис.3.6)

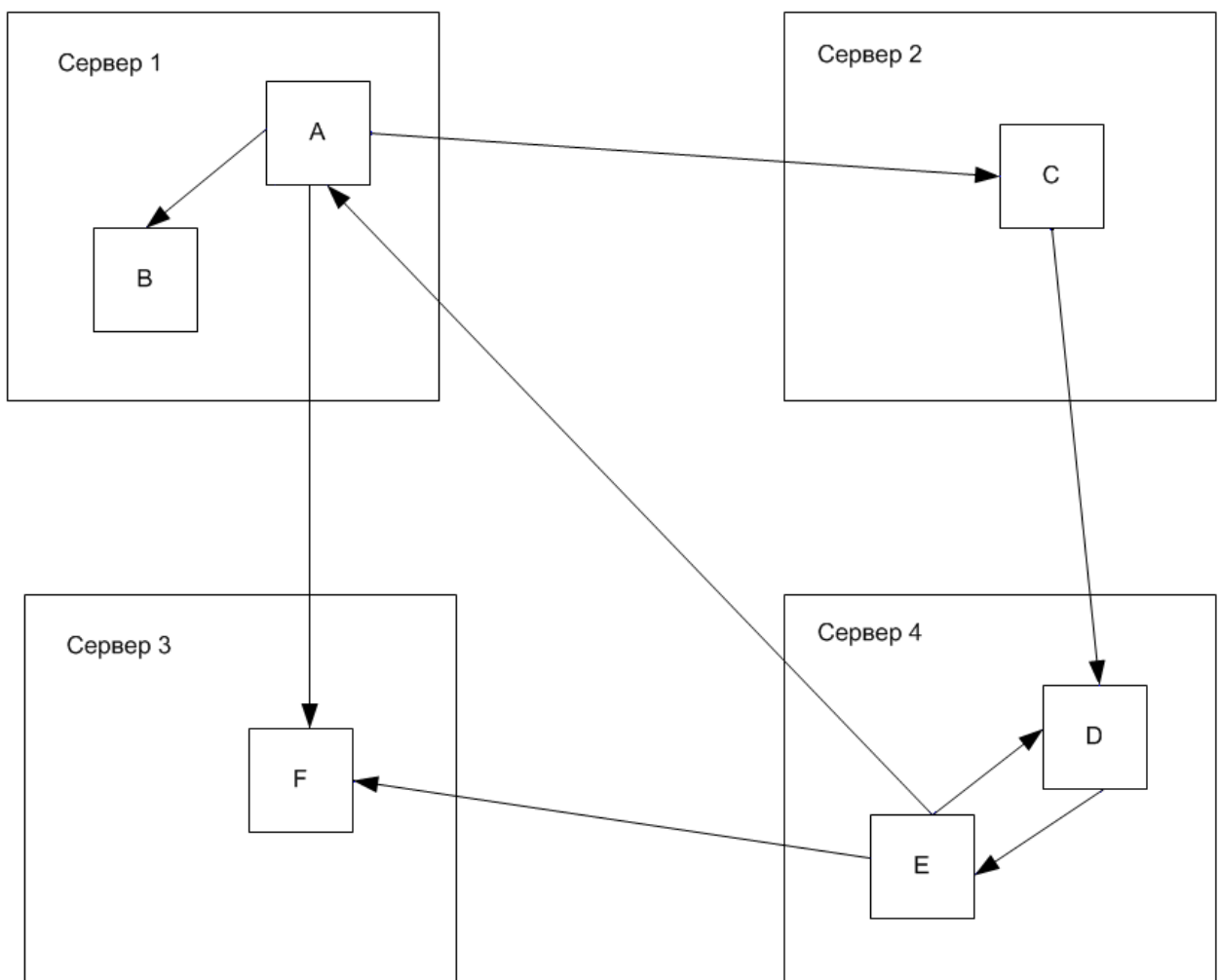


Рисунок 3.6 – Ссылки между гипертекстовыми документами

Вторым краеугольным камнем WWW стала универсальная форма адресации информационных ресурсов, представляющая собой систему, учитывающую способы адресации и идентификации различных сервисов Интернета.

Универсальный идентификатор ресурса (URI – Uniform Resource Identifier), так же известный под названием Web-адреса, представляет собой сочетание унифицированного адресатора ресурса (*URL – Uniform Resource Locator*) и унифицированного имени ресурса (*URN – Uniform Resource Name*). Согласно документу RFC 1945 «универсальные идентификаторы ресурсов – это просто форматированные строки, идентифицирующие ресурс по имени, местоположению или любой другой характеристике».

Но реально из всего, что описано в URI, для организации баз данных в WWW требуется только URL. Без наличия этой спецификации вся мощь HTML оказалась бы бесполезной. URL используется в гипертекстовых ссылках и обеспечивает доступ к распределенным ресурсам сети. В URL можно адресовать как другие гипертекстовые документы формата HTML, так и ресурсы E-mail, Telnet, FTP, Gopher, WAIS. Различные программы различным образом осуществляют доступ к этим ресурсам. Следует отметить, что программы обработки электронной почты в формате MIME также имеют возможность отображать документы, представленные в формате HTML. Для этой цели в MIME зарезервирован тип text/html.

Третьим в списке стоит протокол обмена данными в World Wide Web – HTTP (HyperText Transfer Protocol). Данный протокол предназначен для обмена гипертекстовыми документами и учитывает специфику такого обмена. Так, в процессе взаимодействия клиент может получить новый адрес ресурса сети, запросить встроенную графику, принять и передать параметры и т. п. Управление в HTTP реализовано в виде ASCII-команд.

Реально разработчик гипертекстовой базы данных сталкивается с элементами протокола только при использовании внешних программ или при доступе к внешним относительно WWW информационным ресурсам, например базам данных.

Последняя составляющая технологии WWW – это спецификация CGI (Common Gateway Interface). CGI была специально разработана для расширения возможностей WWW за счет подключения внешнего программного обеспечения. Эта технология соответствовала принципам простоты разработки, доступности и наращивания возможностей WWW. Предложенный и описанный в CGI способ подключения не требует дополнительных библиотек. Сервер взаимодействует с программами через стандартные потоки ввода/вывода, что упрощает программирование до предела. При реализации CGI чрезвычайно важное место заняли методы доступа, описанные в HTTP (GET и POST).

Универсальный адрес ресурса. Для того чтобы получить информацию из Интернета, необходимо знать адрес, по которому она расположена. Универсальный адрес ресурса (URL) – это адрес в системе World Wide Web, с помощью которого однозначно определяется любой документ. В общем случае универсальный адрес ресурса имеет следующий формат: протокол:/компьютер/ путь. Основным протоколом в системе World Wide Web является протокол HTTP – протокол передачи гипертекста, поэтому большая часть адресов начинается следующим образом: «**http://**». Тем не менее, могут быть использованы и другие протоколы передачи данных, например протокол передачи файлов – FTP или протокол передачи данных в формате Gopher (устаревший стандарт – прообраз www). Тогда на первое место в универсальном адресе ресурса ставится название используемого протокола.

Например:

ftp:// --> указатель на доступ через FTP

gopher:// --> указатель на ресурсы Gopher

file:// --> указатель на файл в файловой системе.

Компьютер – это адрес сервера, с которым необходимо установить соединение. Может использоваться как IP-адрес, так и имя сервера в доменной системе имен.

Например: http://www.econ.ru.ги или ftp://194.85.120.66.

Адреса большей части серверов в системе World Wide Web начинаются с префикса `www`. Этот префикс используется просто как удобное обозначение того, что на данном компьютере запущен Web-сервер. В современных версиях браузеров нет необходимости указывать имя протокола в начале каждого адреса ресурса. Если имя протокола не указано, то браузер попытается самостоятельно определить, какой протокол необходимо использовать. *Путь* представляет собой точное указание месторасположения документа на Web-сервере. Это может быть название директории и файла, как в следующем примере: `http://www.econ.pu.ru/info/history/jubilee.htm`. Если ввести в строке «адрес» браузера данный адрес, браузер установит связь с компьютером `www.econ.pu.ru` по протоколу HTTP и запросит у него документ с названием `jubilee.htm` из каталога `/info/history`. Если не указано имя файла, а только каталог, в котором он должен находиться, то пользователю будет передан файл, который администратор Web-сервера определил как файл, передаваемый по умолчанию. Обычно таковым является файл с названием `index.htm` (`index.html`) или `default.htm` (`default.html`). Если в каталоге нет файла по умолчанию, то будет выдано сообщение об ошибке. Последняя часть универсального адреса ресурса, разделенная знаком «?» может включать дополнительную информацию, которую обычно используют для того, чтобы передать Web-серверу параметры запроса пользователя в интерактивных страницах, а также путь и имя той программы на сервере, которая этот запрос будет обрабатывать. Например, `http://www.econ.pu.ru/sf/cgi-bin/main.bat?object=teachers&id=1`. Получив такой запрос, Web-сервер попытается найти исполняемый файл `main.bat` в каталоге `/sf/cgi-bin/`, запустить ее и передать ей параметры `object`, имеющий значение «`teachers`» и `id`, равный «1».

Для указания документов на одном и том же сервере в HTML-документах часто используется сокращенное обозначение, называемое относительным адресом. Перед отправкой запроса на Web-сервер браузер преобразует относительный адрес в полный, например, если документ по адресу `http://www.econ.pu.ru/info/index.htm` содержит ссылку на документ `history/jubilee.htm`, то браузер преобразует эту ссылку в `http://www.econ.pu.ru/info/history/jubilee.htm`.

Протокол передачи гипертекста (HTTP). Протокол передачи гипертекста (HTTP) – это стандартный протокол Прикладного уровня для передачи документов между серверами и браузерами в системе World Wide Web. Этот протокол применяется в WWW с 1990 года.

Протокол HTTP позволяет установить соединение между клиентом и сервером. Соединение сохраняется только на время обработки сервером запросов клиента. Он использует модель запрос/ответ, в которой HTTP-клиенты отсылают запрос и ожидают пакет с ответом, содержащий код состояния. Запросы содержат универсальный идентификатор ресурса (URI), который идентифицирует желаемый Web-ресурс. По умолчанию HTTP-сообщения применяют порт 80.

Запрос клиента и ответ сервера образуют так называемую транзакцию. Транзакция состоит из четырех действий.

1. Клиент связывается с сервером, чтобы установить TCP/IP соединение.
2. Клиент посылает запрос серверу. В большинстве случаев это запросы на получение HTML-документов.
3. Сервер посылает документ, если это возможно, в противном случае возникает сообщение об ошибке.
4. После передачи запроса и получения ответа соединение закрывается.

HTTP-запросы, отсылаемые клиентами, содержат метод (команду), обозначающий потребность клиента. HTTP-серверы отвечают числовым кодом (кодом состояния) и данными (если они затребованы). На рис.3.7 показана схема взаимодействия этих элементов при передаче HTTP-данных.

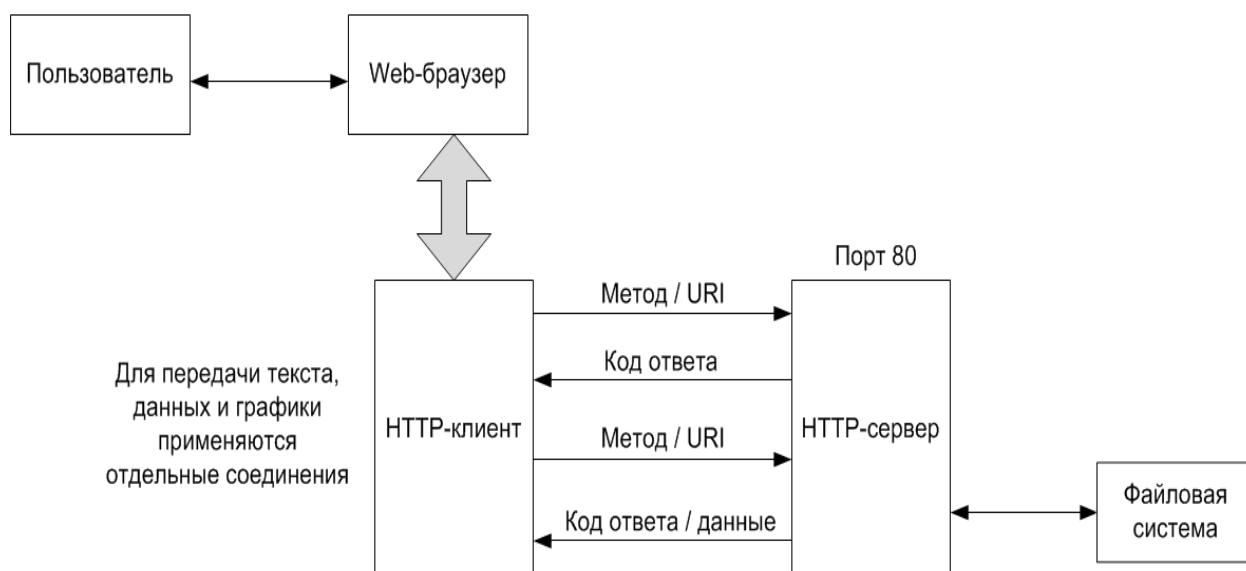


Рисунок 3.7 – Схема взаимодействия элементов при передаче HTTP-данных

Методы протокола HTTP – это команды, выдаваемые HTTP-клиентом HTTP-серверу. В протоколе HTTP v1.0 было определено всего три метода, но протокол HTTP v1.1 расширяет этот список (табл. 3.10).

Таблица 3.10 – Методы протокола HTTP v 1.0 и 1.1

Метод	Описание
GET	Извлекает информацию, которую идентифицирует URI, содержащийся в запросе.
HEAD	Извлекает метаинформацию, относящуюся к требуемому URI.
POST	Отправляет данные на HTTP-сервер (они должны быть размещены в виде новой единицы, подчиненной ресурсу, идентифицируемому в запросе посредством URI).
OPTIONS	Определяет опции и/или требование, связываемые с ресурсом или возможностями сервера.
PUT	Отправляет данные на HTTP-сервер (они должны храниться в URI, указанном в запросе POST).
DELETE	Удаляет ресурс, определяемый URI
TRACE	Запускает удаленную, относящуюся к Прикладному уровню возвратную петлю сообщения запроса: при этом клиент получает возможность увидеть, что от него получил сервер.
CONNECT	Применяется для присоединения устройства-агента и туннеля к конечной точке.

PUT Отправляет данные на HTTP-сервер (они должны храниться в URI, указанном в запросе POST)

DELETE Удаляет ресурс, определяемый URI

TRACE Запускает удаленную, относящуюся к Прикладному уровню возвратную петлю сообщения запроса: при этом клиент получает возможность увидеть, что от него получил сервер.

CONNECT Применяется для подсоединения устройства-агента и туннеля к конечной точке.

HTTP-сервер отправляет коды состояния, сообщая, успешно ли выполнен запрос или же он был неудачен. Основные группы кодов и значения некоторых из них приведены в таблице 3.11.

Таблица 3.11 – Коды состояний протокола HTTP v 1.1

Код состояния	Описание
1xx	Информационные
100	Продолжение
101	Переключение протоколов
2xx	Успешное выполнение
200	ОК
201	Создано
202	Принято
3xx	Переадресация
300	Перемещение на постоянной основе
301	Найдено
305	Использование агента
4xx	Клиентская ошибка
400	Не корректный запрос
403	Запрещено
404	Не найдено
414	Запрос слишком длинный URI
5xx	Ошибка сервера
500	Внутренняя ошибка
501	Не реализовано
502	Не корректный шлюз
503	Служба недоступна

Во время просмотра Web-сайта, например <http://www.iana.org>, с помощью протокола HTTP выполняется ряд взаимодействий между клиентом и сервером. Последовательность таких сообщений, где *S* означает HTTP-сервер, а *C* – HTTP-клиента приведена в таблице.

Таблица 3.12 – Последовательности сообщений HTTP

Процесс	Отправить /получатель	Последовательность пакетов
Установка TCP-соединение	C > S	TCP SYN
Подтверждение сервера	S > C	TCP SYN ACK
Порт сервера 80. Порт клиента 1881	C > S	TCP ACK
Отправка свойств клиента	C > S	HTTP GET Request (URI: www.yandex.ru) представление свойств клиента: Connection (соединение), User-Agent (агент пользователя), Host (хост), Accept (принятие), Accept-Encoding (кодировка), Accept-Language (язык), Accept-Charset (набор символов).
Предоставление свойств сервера	S > C	HTTP GET Reply: code 200. Представление свойств: Time (время), Server Type (тип сервера), List Modified Date/Time (последнее обновление дата/время).

Сообщения, приведенные в таблице 3.12, следуют одной общей схеме, которая заключается в открытии TCP-соединения, отправке GET-запроса на, получении файла и, наконец, по окончании его загрузки – закрытие соединения.

У протокола HTTP есть еще одна особенность – он может использовать свой безопасный вариант HTTPS, в которой протокол SSL (протокол защищенных сокетов) применяется для шифрования трафика между защищенным клиентом и защищенным сервером. Этот режим работы характерен для коммерческих сообщений в сети Internet всех типов – размещения заказов с помощью кредитных карт, просмотра конфиденциальных финансовых записей, осуществления биржевых торгов, денежных переводов и других видов деятельности, подразумевающих передачу секретной информации. Порт 443 протокола TCP закрепляется за протоколом HTTP, применяющим TCP вместе со средствами безопасности Транспортного уровня через протокол SSL. Порт 443 протокола UDP также закрепляется за протоколом HTTP для аналогичных целей.

Служба FTP. *FTP* – это протокол передачи файлов (File Transfer Protocol). Он представляет хранилище для информации, которая всегда доступна. FTP нужен для передачи файлов больших размеров (слишком больших для передачи по электронной почте). С помощью FTP пользователи могут подключаться к хранилищу и загружать информацию или приложения на свои компьютеры. Протокол FTP обеспечивает метод передачи файлов на основе соединений, т.е. TCP. Ключевыми компонентами FTP-сообщения являются пользовательский интерфейс (UI – User Interface), интерпретатор протоколов (PI – Protocol Interpreter), команды FTP, процесс передачи данных (DTP – Data Transfer Process), передаваемые файлы, командное соединение (command connection), а также соединение для передачи данных (data transfer connection) – оба на основе TCP. На рисунке 3.8. все эти базовые элементы изображены во взаимосвязи на рис.3.8.

Пользовательский интерфейс (UI – User Interface) – это графическая оболочка, которую видит пользователь. Файл `ftp.exe` в операционной системе Windows обеспечивает интерфейс командной строки, облегчающий взаимодействие пользователя с процессом FTP. Клиентские приложения сторонних производителей, такие как WS_FTP Pro корпорации Ipswitch, снабжены графическим интерфейсом.

Интерпретатор протоколов (PI – Protocol Interpreter) интерпретирует команды пользователя, инициирует управляющее соединение от динамического порта пользователя к FTP-порту сервера (номер порта 21), запускает FTP-команды и отслеживает процесс передачи данных.

Протокол FTP использует команды, которые следуют непосредственно за TCP-заголовком. Эти команды принимаются и обрабатываются принимающим интерпретатором протокола.

Файловая система на любом конце FTP-сообщения может состоять из различных форматов файлов, таких как ASCII или бинарных.

Процесс передачи данных (DTP – Data Transfer Process) – это процесс фактического перемещения данных по соединению, предназначенному для этих целей. Данные не должны перемещаться по командному соединению. Процесс передачи данных проходит по TCP-соединению между FTP-клиентом и FTP-сервером, которое устанавливается после создания командного соединения по протоколу FTP.

Командное соединение происходит на базе протокола TCP и создается автоматически при установлении FTP-соединения в ответ на команду открытия, исходящую от FTP-клиента, посредством номера порта 21 на FTP-сервере. Для управления другим концом командного соединения клиент применяет динамический номер порта. Командное соединение остается открытым в течение любой продолжающейся передачи данных, поскольку оно представляет отдельный путь обмена командной информацией параллельно с соединением для передачи данных.

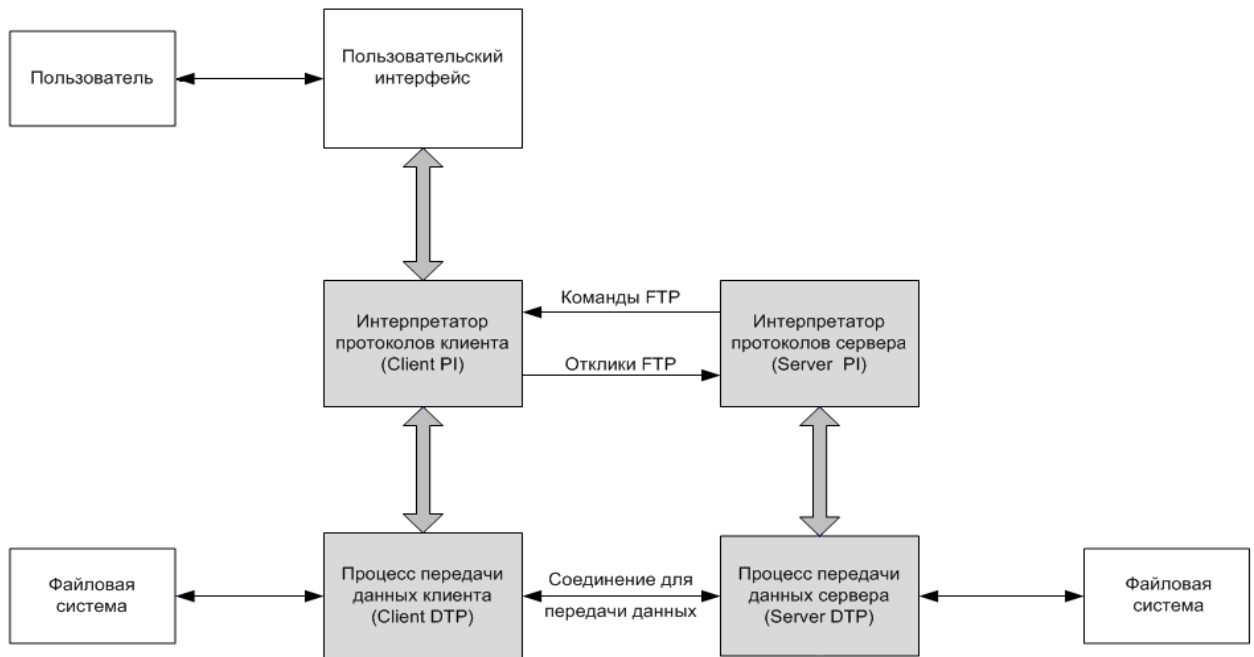


Рисунок 3.8 – Процесс взаимодействия клиента и сервера по протоколу FTP

Соединение для передачи данных также происходит на базе протокола TCP. По умолчанию такое соединение создается сервером. Если пользователь посылает команду PASV и сервер ее принимает, то это соединение может быть создано клиентом. Для установления соединения, отслеживания последовательности пакетов, определения и регулировки окна передачи данных в соответствии с правилами контроля и предотвращения перегрузок, а также восстановления от потери данных протокола FTP используется транспортный протокол TCP.

Служба Telnet. *Telnet* дает возможность пользователям подключаться к ресурсам удаленных компьютеров в режиме терминала. Пользователь может использовать эти ресурсы (корпоративные данные, приводы компакт-дисков, сканеры, принтеры и др.) так как будто они принадлежат его компьютеру.

Протокол Telnet обеспечивает *двунаправленное* (bidirectional) байтовое сообщение. Изначально разработанный как метод сообщения для *терминального доступа* (terminal access). Telnet в большинстве случаев использует хорошо известный порт 23 на стороне сервера (по соображениям безопасности он может быть изменен администратором) и динамический порт на стороне клиента.

Хосты, использующие протокол Telnet, обмениваются информацией об *опциях*, поддерживаемых ими во время установления соединения. Удаленный хост называется виртуальным терминалом сети (NVT – Network Virtual Terminal). Опции включаются в определенные структуры – DO, DON'T, WILL, и WON'T, предлагающие согласиться с использованием данной опции или отказаться от нее (рис.3.9).

Виртуальный терминал сети. Каждая сторона сообщения Telnet обозначается как виртуальный терминал сети (NVT). NVT-клиент обычно инициирует Telnet-соединение, в то время как NVT-сервер предлагает клиенту некоторые службы.

Структуры DO, DON'T, WILL, WON'T. В ходе процесса установления Telnet-соединения хосты предлагают и соглашаются (или отказываются) с использованием в данном сообщении определенных параметров. В этом процессе согласования применяются следующие управляющие структуры:

- DO – 253 (0xFD);
- DON'T – 254 (0xFE);
- WILL – 251 (0xFB);
- WON'T – 252 (0xFC).



Рисунок 3.9 – Передача данных по протоколу Telnet

Хост отправляет один из этих операторов, сопровождая его кодом опции. Эта опция считается принятой, когда второй хост отвечает оператором DO или WILL. Возврат оператора DON'T или WON'T означает, что хост не принимает предложенную опцию.

Опции. Опции представляют собой параметры или соглашения, применяемые в ходе Telnet-соединения. Опции протокола Telnet подробно описаны в документах RFC. В таблице 3.13 приведены первые двадцать позиций списка зарегистрированных для Telnet опций. Полный список опубликован на сайте <http://www.iana.org>

Несмотря на свое давнее прошлое, Telnet по сей день остается важной и широко распространенной службой TCP/IP. Все большее применение находят безопасные реализации протокола – STelnet (Secure Telnet). В этих реализациях интерфейс SSL (Secure Socket Layer) – протокол защищенных сокетов шифрует трафик службы Telnet, позволяя избежать «засвечивания» в сети имен учетных записей и паролей, которые в других условиях передаются в виде открытого текста.

Таблица 3.13 – Опция протокола Telnet

Код	Название	Перевод
0	Binary Transmission	Бинарная передача
1	Echo	Эхо – отображение полученных сигналов
2	Reconnection	Перестройка соединения
3	Supress Go Ahead	Запрещение продолжения
4	Approx Message Size Negotiation	Согласования приблизительного размера сообщения
5	Status	Состояние
6	Timing Mark	Отметка времени
7	Remote Controlled Trans & Echo	Удаленное управление транспортировкой и эхом
8	Output Line Width	Выходная ширина строки
9	Output Page Size	Выходной размер страницы
10	Output Carriage-Return Disposition	Выходное расположение возврата каретки
11	Output Horizontal Tab Stops	Остановки горизонтальной табуляции
12	Output Horizontal Tab Disposition	Позиции горизонтальной табуляции
13	Output Formfeed Disposition	Расположение перевода страницы
14	Output Vertical Tabstops	Вертикальные остановки табуляции
15	Output Vertical Tab Disposition	Вертикальные расположение табуляции
16	Output Linefeed Disposition	Расположение перевода страницы

17	Extended ASCII	Расширенный стандарт ASCII
18	Logout	Конец сеанса
19	Byte Macro	Байт макро (команды)
20	Data Entry Terminal	Терминал ввода данных

3.9. Язык разметки HTML

HTML – (*Hyper Text Markup Language*) язык разметки гипертекстовых документов. Документы, созданные с помощью HTML обладают свойством *интерактивности*, т. е. пользователь может перемещаться от документа к документу (или от раздела к разделу) при помощи *гиперссылок*. Важное правило HTML связано с наличием некоторых обязательных тегов. Обязательные теги должны присутствовать в любом HTML-документе. HTML-документ всегда начинается тегом `<html>` и заканчивается тегом `</html>`, он указывает браузеру, как интерпретировать содержание документа;

`<head> </head>` — «голова» документа. В нее помещается служебная информация, название документа, метатеги, ссылка на CSS-файл и некоторая другая информация;

`<title> </title>` — парный тег заголовка HTML-документа. Помещенный внутри этого тега текст отображается в поле заголовка окна браузера;

`<body> </body>` — парный тег «тела» документа. В него помещается информация, предназначенная для отображения на экране.

Важное обстоятельство связано с определенной жесткой последовательностью расположения некоторых тегов. Приведенные выше теги располагаются в четком порядке, предписываемом стандартом HTML. Простейший HTML-документ должен содержать следующие теги.

```
<html>
<head>
<title>Название документа</title>
</head>
<body>
<p>Текст документа</p>
</body>
</html>
```

Как видно из этого кода, документ начинается с открывающего тега `<html>`. Он сигнализирует браузеру, что этот файл является HTML-документом. Затем следует код «головы» документа, начинающийся открывающим тегом `<head>`. В данном случае в «голову» помещен код только названия документа, состоящий из открывающего тега `<title>`, текста «Название документа» (который будет отображен в поле заголовка браузера) и закрывающего тега названия `</title>`. Закрывающий тег `</head>` замыкает конструкцию «головы».

Далее следует код «тела» документа начинающийся тегом `<body>`. В парный тег параграфа помещен текст, который будет выведен на экран компьютера. Замыкает конструкцию «тела» закрывающий тег `</body>`, после которого находится закрывающий HTML-документ тег `</html>`.

Контрольные вопросы к лекции 3.

- 3.1. Каковы основные принципы и организационная структура Internet?
- 3.2. Из каких компонент состоит инфраструктура сети Internet?
- 3.3. Что отражает эталонная модель OSI
- 3.4. Для чего нужны протоколы сети Internet?
- 3.5. Каковы основные протоколы, службы, сокеты и порты TCP/IP?
- 3.6. В чем суть адресация в сети Internet?
- 3.7. Для чего используется система доменных имен DNS?

Лекция 4. ДОСТУП К ИНФОРМАЦИОННЫМ РЕСУРСАМ

Вопросы:

- 4.1. Общие принципы поиска информационных ресурсов
- 4.2. Топология Web-пространства
- 4.3. Средства и методы навигации в Internet
- 4.4. XML – язык разметки и модель данных

4.1. Общие принципы поиска информационных ресурсов

Средства взаимодействия компьютерных систем обеспечивают лишь «технические» возможности доступа к информации. Однако нетрудно заметить, что осуществление поиска необходимых данных требует решения обширного круга задач логического плана.

Основы информационного поиска. Основные принципы информационного поиска были сформулированы еще в первой половине этого века. Между 1939 и 1945 годами У.Е. Баттенон была разработана система для отыскания патентов. Каждый патент классифицировался в соответствии с понятиями, к которым он имел отношение. Для каждого понятия, использовавшегося в системе, была создана 800-позиционная перфокарта. При регистрации в системе нового патента находились карты, соответствующие тем понятиям, которые в нем рассматриваются, и в позиции пробивались номера патента. Чтобы найти патент, в котором рассматривается одновременно несколько понятий, необходимо было совместить карты, соответствующие этим понятиям. Номер нужного патента определялся из позиции просвета. Основные принципы информационного поиска с тех пор не изменились. На примере уже этой ИПС видно, как происходит процесс поиска. Во-первых, должен быть создан массив указателей на информационные ресурсы. Указатель (index) содержит в себе некое свойство документа и ссылки на документы, этим свойством обладающие. Указатели могут быть различных видов.

Широко распространен, например, авторский указатель. Такой указатель позволяет получить ссылки на работы интересующего нас автора. Также указатели могут быть составлены и по другим атрибутам документа. В системе Баттена использовался предметный указатель, то есть документы классифицировались по понятиям (предметам), которые в них затрагиваются.

Процесс создания указателей на документы называется индексированием, а термины, использующиеся для индексирования, называются *терминами индексирования*. В случае с авторским указателем роль терминов индексирования будут выполнять фамилии авторов хранящихся в фонде работ. Совокупность используемых терминов индексирования называется *словарем*.

Массив указателей, полученный после индексации информационных ресурсов, называется *индексом* (Index database). После создания индекса к нему обращаются посредством запросов. Так как процесс поиска заключается в сопоставлении запроса пользователя с имеющимися данными, полученный запрос также должен быть переведен на язык индексирования. В индексе выполняется поиск соответствующих запросу документов, пользователю выдается список ссылок на подходящие ресурсы. Для повышения скорости индексирования и поиска словарь и индекс должны быть упорядочены по системе, наиболее отвечающей задачам поиска в данной предметной области.

Предметное индексирование и механизм поиска. Когда говорят об *информационно-поисковой системе* (ИПС), подразумевают, что она использует предметный указатель. Предметный указатель позволяет отыскивать документы, касающиеся некоего «предмета». Для составления предметного указателя анализируется содержание документа и определяется «предмет» или «предметы», о которых в документе идет речь. Затем названия этих предметов переводятся на *информационно-поисковый язык* (ИПЯ). Таким образом, получается *поисковый образ документа* (ПОД). Проиндексировав (создав поисковые

образы) все информационные ресурсы, можно получить то, что принято называть индексом (index database) – основной массив данных ИПС.

Как видно на рис.4.1., поиск происходит не по тексту документов, а по их поисковым образам, составленным на ИПЯ.



Рисунок 4.1 – Типовая схема ИПС

Поэтому ИПЯ – основная часть информационно-поисковой системы, от которой в первую очередь зависит качество системы.

В состав информационно-поискового языка входят:

1. Словарь индексационных терминов – множество терминов индексирования.
2. Кодовый словарь – множество кодовых терминов.
3. Словарь входов – множество входных терминов.
4. Вспомогательные средства языка индексирования – средства, используемые совместно с индексационными терминами для расширения или сужения определенных понятий.
5. Правила использования языка индексирования.

Для повышения эффективности поиска словарь, используемый системой, должен быть контролируемым, то есть он должен быть организован таким образом, чтобы полнота и точность поиска была оптимальной. Очевидно, что организация словаря зависит от многих факторов – предметной области, в которой будет использоваться ИПС, характера интересов пользователей, степени их подготовки и т.д.

Для улучшения результатов поиска необходимо определить степень специфичности терминов, используемых при индексации. Принято использовать два принципа – выявление наиболее специфического термина, соответствующего объему и содержанию отражаемого понятия, и избыточное индексирование.

Под *избыточным индексированием* понимается дополнение поискового образа терминами, связанными с основным. При этом могут использоваться термины, связанные с основным как отношением обобщения или спецификации, так и ассоциативной связью.

Дополнение поискового образа терминами с ассоциативной связью может увеличить полноту поиска, но неизбежно понижает его точность. Недостатком избыточного индексирования является также увеличение объема поисковых образов. Для решения этой проблемы во многих ИПС используется избыточное индексирование не документов, а запросов. Использование предметного индексирования не исключает использования при создании поискового образа атрибутов документа. Это могут быть такие атрибуты, как данные об авторе, дата публикации, язык публикации и т. д.

Стратегии поиска. Точность и полнота поиска зависят не только от характеристик самой ИПС, но и от того, как создается запрос. Идеальный запрос может быть составлен пользователем, в полном объеме знакомым с той предметной областью, которая его интересует, а также с используемой ИПС. Но такому пользователю ИПС, очевидно, не нужна. Остальные же пользователи вынуждены довольствоваться или низкой точностью поиска, или низкой полнотой. Для повышения качества поиска можно использовать различные методы. Наиболее употребляемый из них – использование логических операторов AND, OR, NOT. Использование логических операторов – довольно простой способ повысить релевантность выдаваемых документов, но он имеет и свои недостатки. Главный из них – плохая масштабируемость. Применение оператора AND может сильно сузить выдачу, а оператора OR – сильно расширить.

Степень точности и полноты поиска зависит от того, насколько общие термины использовались при формулировке запроса. Может быть неверным использование как наиболее общих терминов (возрастает уровень информационного шума), так и слишком специфичных терминов (снижается полнота поиска). Использование слишком специфичных терминов может быть чревато еще и тем, что в словаре ИПС этого термина может не оказаться.

В общем, виде процедура поиска является процедурой итеративной, то есть за этапом выдачи результатов поиска следует коррекция запроса, поиск по этому запросу и т. д. Схематично такая процедура показана на рис.4.2.

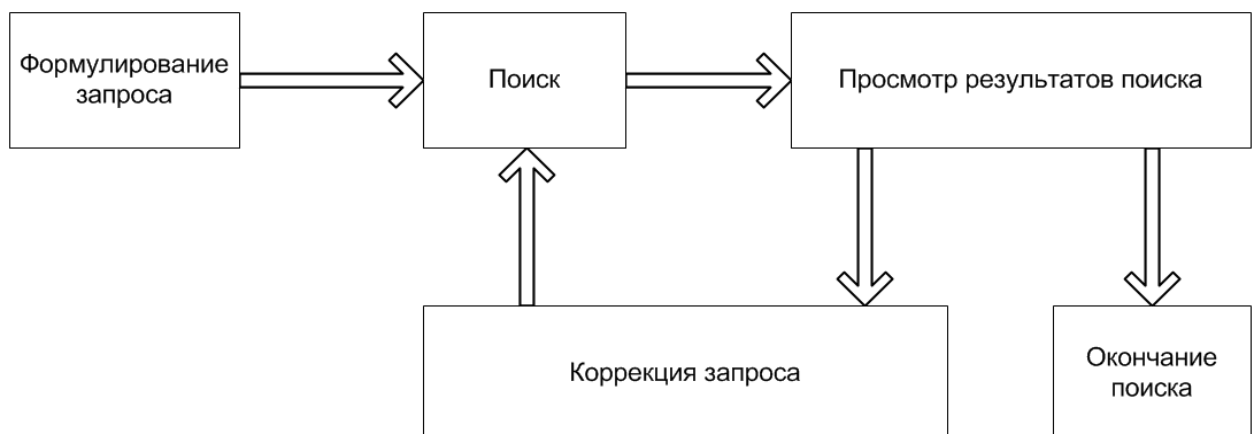


Рисунок 4.2 – Процедура поиска

В зависимости от соотношения полноты и точности, найденных документов пользователь может сузить или расширить область поиска, перейдя к более общим или, наоборот, более специфичным терминам, а также используя родственные понятия. В случае поиска по нескольким терминам такая коррекция области поиска может происходить по одному из нескольких терминов, что позволяет изменять эту область достаточно плавно. Не найдя их в списке найденных документов, область поиска надо расширить. Коррекция запроса системой информационного поиска происходит на основании анализа документов, помеченных пользователем как наиболее точно отвечающих его потребности. В таком случае при следующем поиске система ищет те документы, в которых, помимо заданных в первоначальном запросе, содержатся термины, встречающиеся в документах, отмеченных пользователем.

ИПС глобальных сетей. ИПС глобальной сети имеет отличия, обусловленные как характером сети, так и особенностями работы пользователей такой системы. Рассмотрим основные особенности использования ИПС в глобальной сети на примере сети Интернет (рис.4.3).

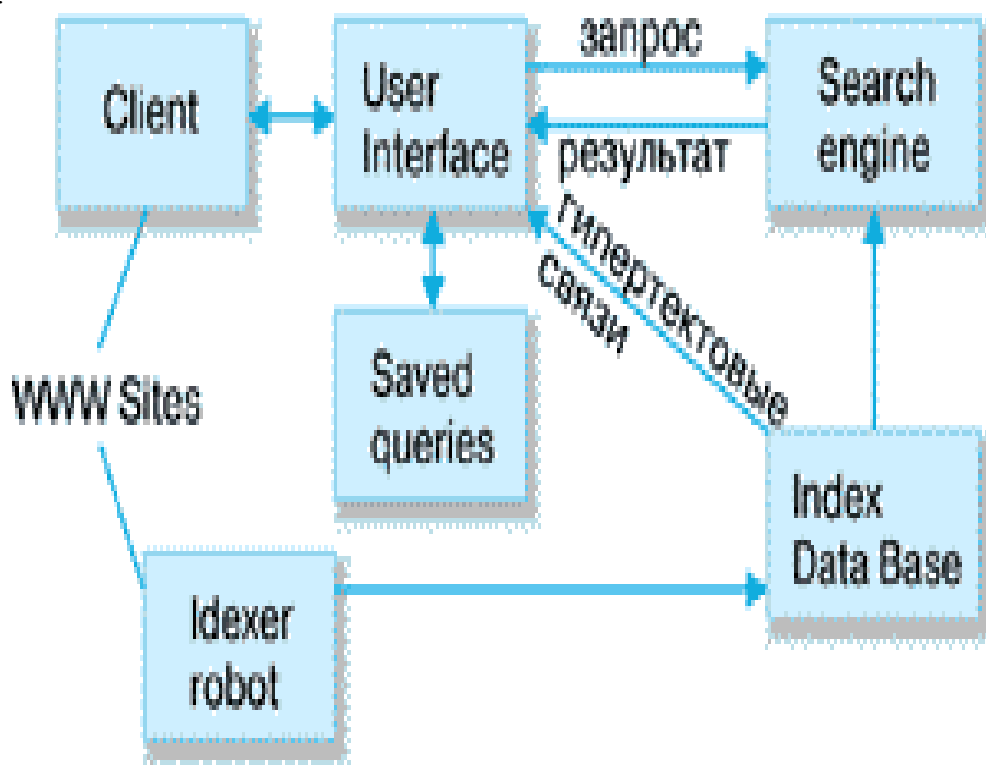


Рисунок 4.3 – ИПС для Интернета

Клиент на этой схеме – это программа просмотра конкретного информационного ресурса. Такая программа обеспечивает просмотр документов WWW, Gopher, Wais, FTP-архивов, почтовых списков рассылки и групп новостей. Все эти информационные ресурсы являются объектом поиска информационно-поисковой системы.

Интерфейс пользователя – способ общения пользователя с поисковым аппаратом: системой формирования запросов и просмотра результатов поиска.

Поисковая машина – служит для трансляции запроса на информационно-поисковом языке, в формальный запрос системы, поиска ссылок на информационные ресурсы Сети и выдачи результатов этого поиска пользователю.

Индекс базы данных – индекс, который является основным массивом данных ИПС и служит для поиска адреса информационного ресурса. Сохраненные запросы – сохраняются в личной базе данных пользователя. На отладку каждого запроса уходит достаточно много времени, и поэтому чрезвычайно важно запоминать запросы, на которые система дает хорошие ответы.

Робот индексирования – служит для просмотра данных в Интернете и поддержания базы данных индекса в актуальном состоянии. Эта программа является основным источником информации о состоянии информационных ресурсов сети.

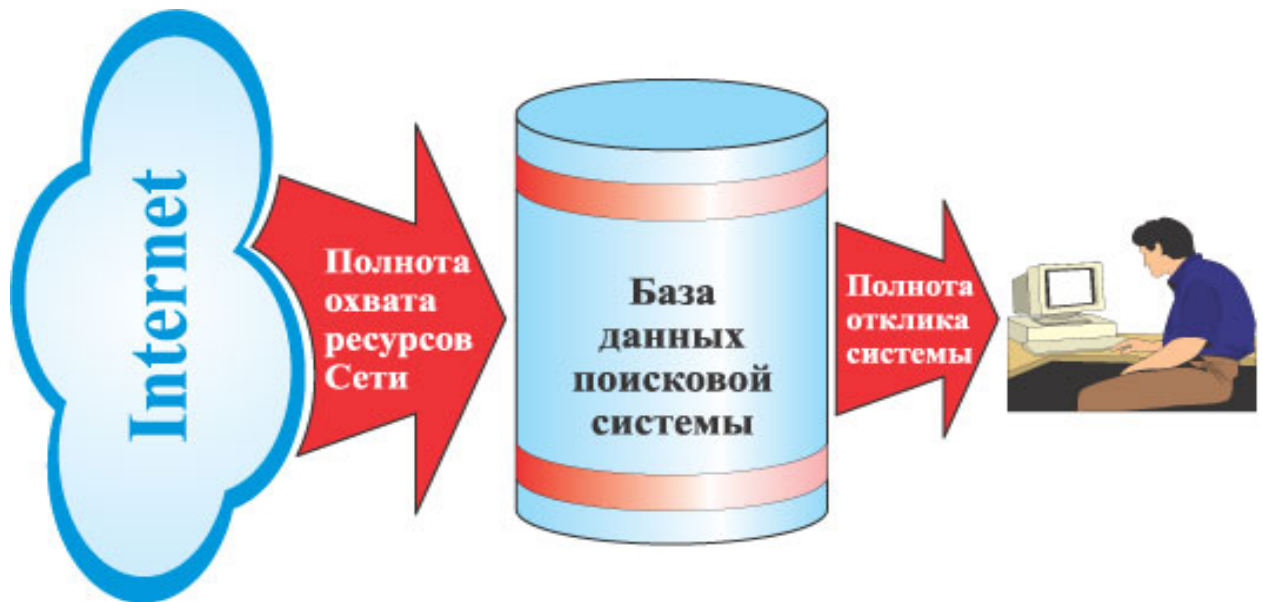
Ресурсы Интернет – это информационные ресурсы, просмотр которых обеспечивается программами просмотра. Таким образом, источником информации о состоянии информационных ресурсов сети является робот-индексировщик. Это программа, которая по определенному алгоритму «заходит» на различные страницы, «читает» их и индексирует. Индекс поисковых систем Интернета обновляется с периодичностью около недели. Отсюда видно, что в индекс поисковой системы не могут попасть материалы, например, периодических изданий, так как выходят они заведомо чаще, чем обновляется индекс. Еще одна проблема заключается в том, что не все документы хранятся в виде файлов HTML, с которыми роботу работать легче всего. Если информация хранится в другом формате, может сложиться ситуация, когда адрес страницы, выдаваемой пользователю, содержит параметры, которые робот не знает, и, следовательно, он не может эти данные проиндексировать.

Объем информации, опубликованной в Интернете, приводит также к ограничению количества терминов, которыми индексируется документ. Современные ИПС в Интернете используют порядка 100 терминов для индексирования документа. Выбор терминов, используемых для индексации, зависит от реализации конкретной системы. Чаще всего первым критерием является отношение частоты употребления этого термина во всех ранее проиндексированных документах. Наибольший вес присваивается тем терминам, которые наиболее часто встречаются в данном документе и наиболее редко – во всех остальных проиндексированных документах. Термины, которые в очень большом количестве документов, при индексировании не используются совсем.

При определении терминов индексирования, используемых для создания поискового образа, робот может использовать разметку индексируемой страницы. И в индексе присвоить наибольший вес термину, используемому, например, в заголовке. Этим пользуются авторы информационных ресурсов, чтобы повлиять на индексацию собственной страницы, указывая какие термины надо индексировать. Но многие поисковые системы отказались от использования описаний ресурсов, предоставленных авторами. Это сделано по причине недобросовестности некоторых авторов, которые использовали для описания своих страниц термины, наиболее часто встречающиеся в запросах. Результат запроса представляется в виде отсортированного списка. Наиболее часто используется сортировка по релевантности. Произвести точный поиск тем сложнее, чем шире круг потребностей пользователей системы. В глобальной сети эта проблема принимает глобальный характер. Многие отмечают постоянно растущий уровень шума (выдачи не релевантной информации) в результатах, выдаваемых в запрос.

Два аспекта полноты. Основопологающими характеристиками качества функционирования информационно-поисковых систем являются *полнота* и *релевантность*. Сегодня информации в Сети появляется больше, чем ее успевают проиндексировать поисковые системы. Вместе с тем, чем больше ресурсов соответствующего профиля включает база данных системы, тем выше должна быть ее полнота (рис.4.4).

Понятие полноты динамической базы данных из Internet тесно связано с оперативностью обновления информации. Сеть Internet представляет собой своеобразный "живой организм", – здесь постоянно добавляются новые ресурсы, удаляются устаревшие, некоторые документы меняют адреса, некоторые модифицируются. Созданная однажды база данных является "слепок" состояния информационных ресурсов Сети на конкретный момент. Если база данных ИПС не будет обновляться постоянно и оперативно, имеющиеся в ней ссылки на документы станут мертвыми, т.е. по адресам, представленным в этих ссылках, документы могут, либо не существовать, либо будут размещены с совершенно другим содержанием. Кроме того, отсутствие оперативности и обновления баз данных не позволит пользователю отслеживать последние изменения в его предметной области.



Два аспекта полноты

Рисунок 4.4 – Два аспекта полноты охвата ИПС ресурсов Internet

Второй аспект связан с полнотой информации, предъявляемой пользователю по его запросу к ИПС (рис.4.3). Если предположить, что по запросу пользователя Q в базе данных находится P документов, соответствующих этому запросу, а предъявлено для просмотра всего N документов, то полнота ИПС определяется по формуле:

$$П = (N / P) \times 100\%.$$

В случае, если $П$ оказывается больше 100%, очевидно, что пользователю выдано минимум $(N - P)$ документов, не соответствующих его запросу, т.е. нерелевантных.

Релевантность и пертинентность. Под релевантностью понимается формальное соответствие информации, выдаваемой системой, запросу. Если по запросу пользователя получено N документов, представляющих собой объединение двух множеств документов: соответствующих запросу (пусть их количество – $N1$) и не соответствующих (их количество – $N2$), т.е. $N = N1 + N2$, тогда релевантность, как степень соответствия, определяется по формуле

$$P = (N1/N) \times 100\%,$$

а шум – по формуле

$$S = (N2/N) \times 100\% = 100\% - P.$$

Если же обозначить количество соответствующих запросу документов в исходном массиве как R , то отношение

$$(N1/R) \times 100\%$$

будет определять полноту поиска.

Проиллюстрируем эти понятия. Допустим, исходный массив содержит 100 документов, из которых 50 соответствуют запросу.

Если в результате поиска будет выдан всего один документ, который при этом будет соответствовать запросу, релевантность выдачи будет равна 100%, шум – 0, а полнота – всего 2%.

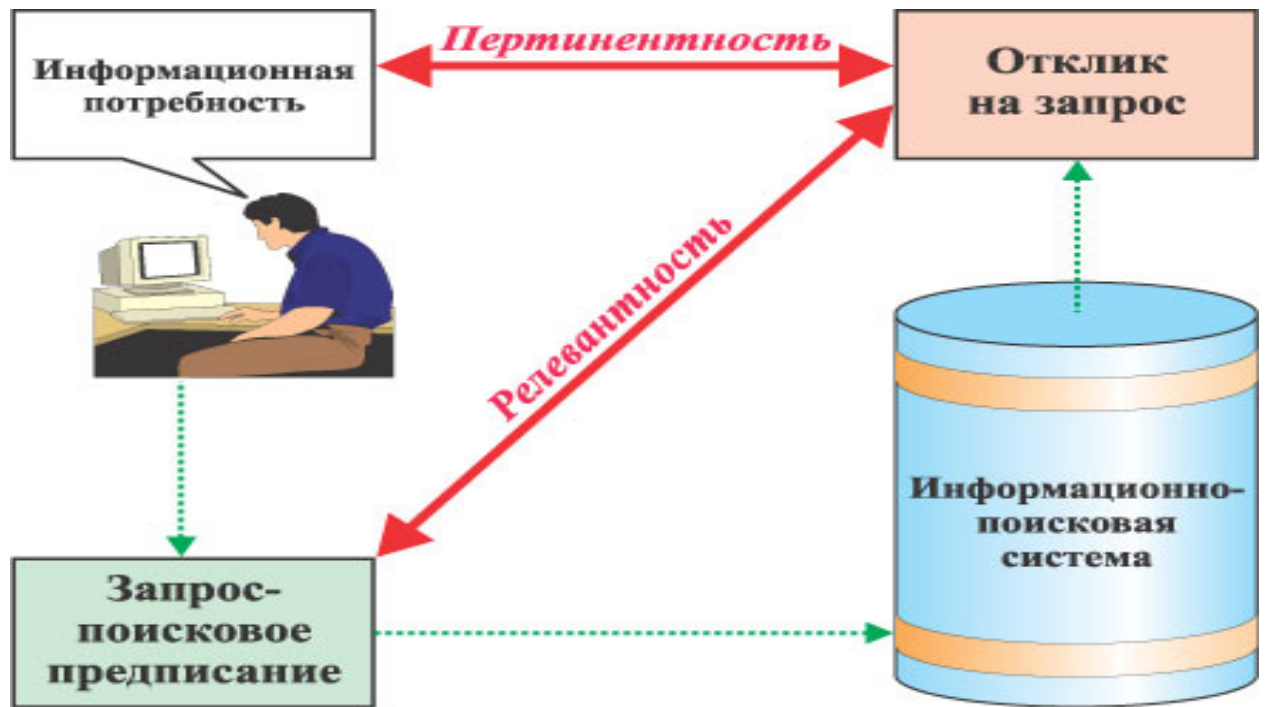
В другом, крайнем, случае, если будут выданы все 100 документов, релевантность результатов поиска составит 50%, шум – 50%, а полнота – все 100%.

Подобные рассуждения позволяют констатировать следующее: чем выше в системе релевантность, тем ниже полнота, и, соответственно, чем ниже релевантность, тем полнота выше.

Это определение характерно для формальной релевантности, однако, на практике используется другое, неформальное понятие – пертинентность. Для пользователя перти-

нентность, как соотношение объема полезной для него информации к общему объему полученной информации, имеет решающее значение (рис.4.5).

При этом следует учитывать, что формальный запрос к системе является предметом творческого осмысления информационной потребности и не всегда точно отражает последнюю. Неумение большинством пользователей правильно формулировать запросы и получать приемлемые объемы отклика породило в конце XX века мнение об Internet как об огромной информационной свалке.



Релевантность и пертинентность

Рисунок 4.5 – Релевантность и пертинентность

Средства повышения пертинентности в современных системах, помимо возможностей уточнения формулировки запросов, включает и весовые критерии, позволяющие ранжировать найденные документы и выдавать пользователю для просмотра наиболее весомые документы либо вообще ограничиваться выдачей не более заданного числа наиболее весомых документов. В последнем случае, естественно, страдает полнота выдачи, т. е. при этом полнота и релевантность являются антагонистическими характеристиками – чем выше релевантность, тем ниже полнота и наоборот.

Кроме характеристик полноты и релевантности, для пользователей ИПС большое значение имеют такие характеристики, как

- скорость обработки запросов,
- получения отклика от системы,
- достоверность отклика (например, оцениваемая по ее источникам), а также
- дополнительные сервисы (возможность нахождения документов, подобных уже имеющимся /like this/, подключения автоматических переводчиков и, конечно же, уточнения запроса непосредственно после выполнения процедуры поиска).

Статическая и динамическая составляющие Web – пространства. Все Internet-пространство можно с достаточной долей условности разделить на две составляющие – стабильную и динамическую. Стабильная составляющая сети содержит информацию «долговременного» плана, в то время как динамическая включает постоянно обновляемые ресурсы. Некоторая часть динамической составляющей со временем вливается в стабиль-

ную. Однако, большая ее часть «исчезает» из Сети или попадает в сегмент «скрытого» Web-пространства, не доступного пользователям с помощью информационно-поисковых систем.

В традиционной сетевой поисковой системе информационное пространство (состоящее из стабильной и новостной частей и индексируемое этой ИПС) меняет свое содержимое через N дней:

- некоторые новостные документы уходят в стабильную часть в виде архивов, а остальные
- исчезают.

В этом случае пользователь при обращении к ИПС получает соответствующие запросу ответы из стабильной части, устаревшие ссылки из новостной части и ничего из обновленной новостной части. Таким образом, ни одна из традиционных поисковых систем в достаточном объеме не помогает в поиске актуальной новостной информации, находящейся в динамической части сети.

Решение этой задачи требует создания своеобразного интеллектуального посредника между пользователем и Internet. Подобный посредник (или агент новостей) должен выполнять всю «черновую» работу по сбору и селекции информации и обеспечивать предпосылки для создания документальной базы данных. Принцип индексирования, используемый посредником, несколько отличается от аналогичного принципа традиционных поисковых систем: *индексируется не все пространство Internet, а только его новостная часть*. При этом, за счет относительно небольшого объема этих данных, частота индексирования выбирается достаточно малой – от нескольких минут до нескольких часов (в зависимости от источника).

В результате через N дней обрисовывается такая ситуация: пользователь получает необходимые ответы по новостной и «устаревшей» новостной части, подтвержденные документами из собственной архивной базы данных, но не получает полной выборки документов из стабильной части информационного наполнения сети. Таким образом, проблема получения полной информации из Сети в идеале может быть решена путем использования двух инструментов – традиционных ИПС и систем-агентов новостей.

Недостатки традиционного поиска. Количество источников новостной (динамической) информации в Сети постоянно возрастает, что еще более усложняет задачу поиска необходимых данных. Можно сказать, что извечная проблема поиска информации сегодня получила новое звучание: «поиск информации в неограниченной, неоднородной динамической информационной среде».

Традиционные поисковые системы предлагают лишь частичное решение этой проблемы. Периоды индексации у них составляют от недель до нескольких месяцев. И несмотря на то, что практически все известные поисковые порталы (Google, Yahoo!, AltaVista, Lycos и др.) имеют новостные разделы, они уже многих не устраивают. Традиционным подходам к организации поиска сетевой информации присущи такие недостатки, как низкая оперативность, зависимость от набора источников и ограниченность спектра этих источников, средние поисковые возможности, отсутствие средств уведомления о появлении новых данных.

Еще одна проблема нахождения информации в Сети обусловлена основным форматом, в котором представлена эта информация, – HTML. Этот формат был разработан, в первую очередь, для решения задач отображения содержания на каждом конкретном Web-ресурсе, поэтому он не всегда удобен для автоматической обработки информации, в том числе и организации поиска. В результате информация в Internet оказалась очень слабо приспособлена для автоматизированного обобщения, классификации и аналитической обработки. При импортировании в Web-ресурс информации с другого сайта (включении новостных сообщений и т.п.) возникает вопрос однотипного представления их содержания (контента). Если этот вопрос не решается, то изменение HTML-оформления сайта-источника приводит к необходимости одновременной модификации программного обес-

печения на всех сайтах, которые принимают от него информацию. Поэтому, объективно назрела необходимость использования некоего унифицированного формата представления данных.

4.2. Топология Web-пространства

Топологическая модель поверхностного Web. Сегодня каждый пользователь на своем опыте может почувствовать в действии один из самых больших парадоксов Internet – «полезной информации в WWW становится все больше, но найти что-то необходимое – все сложнее». Традиционные средства «учета» информационных ресурсов Сети – каталоги и информационно-поисковые системы – уже сегодня не справляются с задачей поиска информации, поставленной в общем виде. Web-пространство характеризуется большим количеством скрытых в нем неявных экспертных оценок, реализованных в виде гиперссылок. Именно гиперссылки оказались базой для построения модели Web-пространства. Для большего охвата информационных ресурсов средствами информационно-поисковых систем необходимо учитывать архитектуру всего Web-пространства, но именно этой информацией никто ранее не владел. Близкой к реальности математической модели не существовало до 1999 года.

В ноябре 1999 года компания IBM совместно AltaVista и Compaq совершили прорыв, математически описав карту ресурсов и гиперсвязей существующего пространства World Wide Web. Исследования опровергли расхожее мнение, будто Internet – это единое густое пространство. Проследив с помощью поискового механизма AltaVista свыше 200 млн Web-страниц и несколько миллиардов ссылок, размещенных на этих страницах, ученые пришли к следующим выводам о структуре Web-пространства. По их мнению, эта структура в действительности соответствует ориентированному графу с топологией «галстука-бабочки» (Bow Tie), в котором вершины соответствуют страницам, а ребра – соединяющим страницы гиперссылкам. Анализ структуры связей между отдельными Web-страницами, выполненный в рамках этой модели, позволил обнаружить следующее.

1. Центральное ядро (28% Web-страниц) – компоненты сильной связности (SCC) или узел галстука. Сюда относятся Web-страницы, связанные так тесно, что, просто следуя по гиперссылкам, из любой из них в конечном счете можно попасть на любую другую.

2. «Отправные» Web-страницы (IN) (22% Web-страниц) – они содержат гиперссылки, которые, в конечном счете, ведут к ядру, но из ядра к ним попасть нельзя.

3. «Оконечные» Web-страницы (OUT) (столько же - 22%) – к ним можно прийти по ссылкам из ядра, но нельзя вернуться назад в ядро.

4. «Отростки» (еще 22% Web-страниц) – полностью изолированы от центрального ядра: это либо «мысы», связанные гиперссылками со страницами любой другой категории, либо «перешейки», соединяющие две Web-страницы, не входящие в ядро.

Указанные четыре основных множества, в сумме составляющие более 90% исследованных Web-страниц, каждая из которых топологически относится одной компоненте связности, и обусловили название полученной модели рис.4.6) – Bow Tie («галстук-бабочка»). Помимо этого, в Web существуют «острова», которые вообще не пересекаются с остальными ресурсами Internet. Единственный способ обнаружить ресурсы этой группы – знать их адрес. Никакие поисковые машины не смогут найти эти острова, если они в прошлом каким-то образом не соединялись с другими частями Internet.

Исследователи обнаружили, что пропорции этих четырех категорий в течение нескольких месяцев оставались неизменными, несмотря на значительное увеличение общего объема Web-ресурсов.

Были исследованы такие параметры данной модели, как среднее количество сайтов, через которые связываются любые два сайта гиперссылками, а также распределение входящих и исходящих ссылок. Эксперимент выявил гораздо более детальную и сложную картину: значительная область WWW вообще отделена от других крупных частей («острова»), – говорится в отчете компаний. С большой степенью вероятности, случайно вы-

бренные Web-страницы окажутся никак не связаны. Если же путь все-таки существует, среднее количество щелчков, необходимых для переходов между ними, составляет 16. А если этот путь двусторонний, то среднее число промежуточных щелчков сокращается до семи. Топология и характеристики модели оказались примерно одинаковыми для различных подмножеств Web-пространства, подтверждая тем самым наблюдение о том, что Web – это фрактал, т. е. свойства структуры Bow Tie всего Web-пространства также верны и для его отдельных подмножеств. Таким образом, алгоритмы, использующие информацию о структуре Web-пространства, предположительно будут работать и на отдельных его подмножествах.

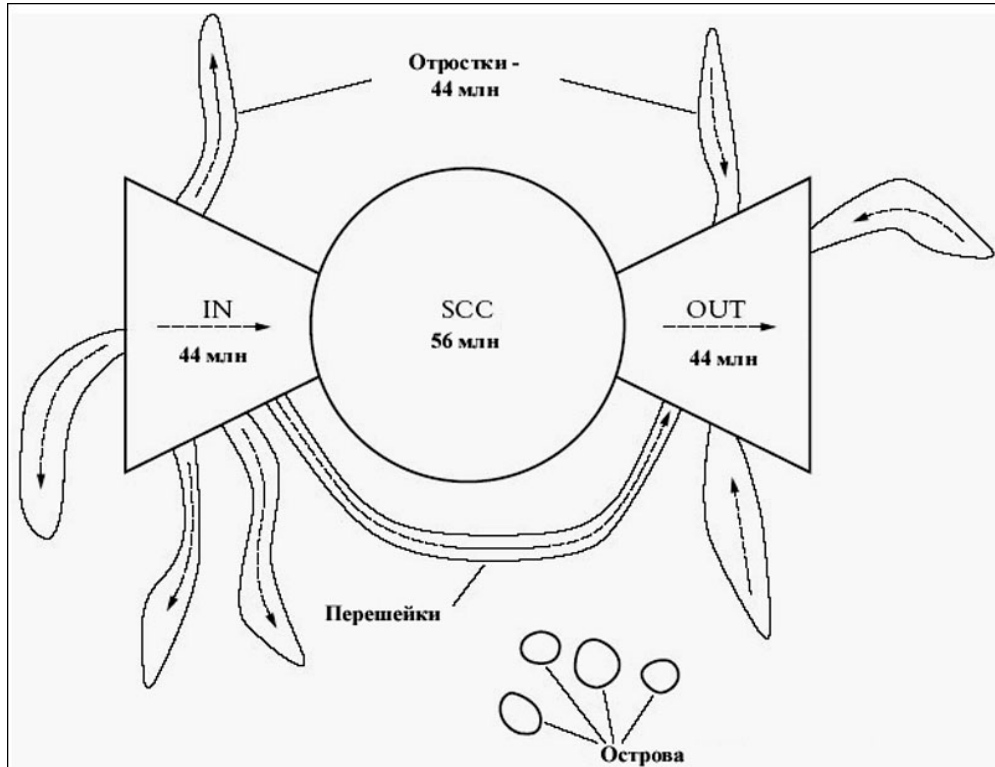


Рисунок 4.6 – Модель Bow Tie («галстук-бабочка»)

Скрытый Web. Кроме видимой для поисковых систем части Web-пространства, существует огромное количество страниц, которые ими не охватываются. При этом доступ пользователя к таким ресурсам возможен, хотя часто «прикрыт» паролями. Как правило, выйти на такие страницы трудно, а порой невозможно, если не знать точного адреса. Эти ресурсы уже десять лет как имеют собственное название – «скрытый WEB». Оно обозначает ресурсы, недоступные для обычных поисковых машин. Еще такие ресурсы называют невидимый – invisible Web. Они чаще всего охватывают динамически формируемые страницы, содержание которых хранится в базах данных и доступно лишь при запросах пользователей.

В 2000 году американская компания BrightPlanet (www.brightplanet.com) опубликовала доклад, в котором утверждается, что в Web-пространстве в сотни раз больше страниц, чем их удалось проиндексировать самыми популярными поисковыми системами. Эта же компания разработала программу LexiBot, которая позволяет сканировать некоторые динамические Web-страницы, формируемые из баз данных, и получила неожиданные данные. Выяснилось, что для традиционных поисковых систем огромная часть сети просто невидима.

Согласно построенной модели, получившей название Bow Tie, скрытые ресурсы – это «острова» в Web-пространстве, которые не пересекаются с остальными ресурсами сети. Единственный способ обнаружить ресурсы этой группы – точно знать их адрес. Поисковые машины в принципе не находят этих островов, если они в прошлом каким-то обра-

зом не соединялись с другими частями Internet. Но в модели Бредера исследовались в основном страницы открытого (поверхностного) Web, к тому же их отбор происходил, видимо, не совсем случайно. Поэтому, если процентное соотношение первых четырех составляющих «поверхностного» Web можно признать верным, то «острова» в реальности будут более объемными, чем в модели. Согласно исследованиям компании BrightPlanet, число скрытых (но не секретных) Web-страниц во много раз превышает количество видимых. Доступные сегодня посредством традиционных информационно-поисковых систем 10 млрд Web-страниц – это лишь видимая крупница. Непознанных, скрытых ресурсов сети в сотни (!) раз больше. Это, прежде всего, динамически генерируемые страницы, файлы нераспознаваемых поисковыми системами форматов, информация из многочисленных баз данных.

Типы скрытых ресурсов. Для того чтобы определить, какие из ресурсов невидимы для поисковых систем, следует рассмотреть принцип работы типового индексатора-робота таких систем. Эти программы-роботы, как правило, посещают Web-страницы по известным заранее адресам, анализируют их содержание и выделяют гиперссылки, идущие от них. Обычно, обработав текущую страницу, выделив ключевые слова и некоторые поля, робот переходит по адресам, найденным на ней, сканирует последующие страницы, выделяет новые адреса и т.д.

Обычно, если робот определяет, что в данный момент обращается к динамической странице, он останавливает свою работу. Эта тактика выбрана в предположении, что чаще всего для получения осмысленного ответа из баз данных требуется осмысленный запрос, а большинству из роботов чужды элементы интеллекта, даже искусственного. В результате содержимое онлайн-баз данных, доступных в сети попадает в «скрытый» Web в первую очередь. Динамической является и быстро обновляемая информация – новости, конференции, онлайн-журналы. Конечно, есть и явные «острова», на которые не указывают никакие гиперссылки и от которых никаких гиперссылок не исходит. Защищенные паролями коммерческие Web-сайты также попадают в категорию «скрытого» Web – о материалах этих сайтов большинство пользователей никогда не узнают с помощью поисковых систем. Однако относительное количество таких сайтов невелико. Например, среди крупнейших сайтов «скрытого» Web платными являются только 10% ресурсов, хотя именно они включают важнейшие ресурсы и базы данных.

В категорию скрытых попадают различные типы ресурсов, как традиционные базы данных (патенты, медицина и финансы), так и публичные ресурсы – объявления о поиске работы, чаты, библиотеки, справочники (www.leidenuniv.nl/ub/biv/specials.htm). К ним также могут быть отнесены специализированные поисковые системы, которые обслуживают определенные отрасли или рынки, базы данных которых не включаются в глобальные каталоги традиционных поисковых служб.

К «скрытому» Web также относятся многочисленные системы интерактивного взаимодействия с пользователями – системы помощи, консультирования, обучения, требующие участия людей для формирования динамических ответов от серверов. К ним также можно отнести и закрытую (полностью или частично) информацию, доступную пользователям Сети только с определенных адресов, групп адресов, иногда городов или стран. Недавно появилась категория так называемых «серых» сайтов, функционирующих на основе динамических систем управления контентом (Dynamic Content Management Systems). В поисковых системах глубина индексирования таких сайтов обычно ограничивается во избежание возможного циклического просмотра одних и тех же страниц.

И, конечно же, «скрытыми» оказываются и Web-сайты, создатели которых не оповещают кого-либо о создании этих ресурсов. В настоящее время, все еще основной формат данных, с которым работают традиционные поисковые системы в Internet, – это HTML, причем статическая его часть. С другими форматами у многих поисковых систем имеются различные проблемы. К примеру, ресурсы, представленные в различных версиях формата PDF могут оказаться скрытыми для большой категории пользователей. Тем не

менее, некоторые современные поисковые системы уже вполне сносно индексируют документы в PDF-формате.

4.3. Средства и методы навигации в Internet

Традиционные средства навигации и поиска в Web-пространстве – это каталоги и поисковые системы. Причем первыми появились Web-каталоги, как психологически наиболее приближенные к образу мышления человека. Действительно, каталоги в принципе не требуют от пользователя ввода какой-либо информации с клавиатуры – достаточно воспользоваться гиперссылками, чтобы найти необходимую информацию. Трудно представить традиционный «бумажный» каталог, содержащий несколько миллионов ссылок.

Точно так же трудно ориентироваться в электронном Web-каталоге, не используя дополнительных возможностей, главной среди которых является возможность ввода «своего» запроса с клавиатуры.

Рост объема Web-ресурсов привел к появлению и бурному росту информационно-поисковых серверов в Сети. Сегодня наиболее развитые системы навигации в Internet обладают свойствами как Web-каталогов, так и информационно-поисковых серверов. Среди таких систем – мировые лидеры Google, Yahoo, AltaVista, Alltheweb. В России лидирующее положение занимают системы Яндекс, Rambler и Aport.

Web-каталоги и информационно-поисковые серверы стали прародителями нового типа Web-сервисов – порталов («ворота в Интернет»). Порталы как новые объекты WWW появились в 1998 году. Основная идея их создания заключалась в стремлении, наряду с возможностями навигации в Сети, предоставить пользователю максимальный уровень сервиса, сделать так, чтобы каждый свой сеанс работы в Internet он начинал именно с данного ресурса. С момента появления первых порталов основные функции «ворот в Internet» существенных изменений не претерпели: это средства реализации поиска данных, общения, новостная часть, торговля и службы приложений. Таким образом, можно дать следующее определение понятию «портал»: сайт (или совокупность сайтов), обеспечивающий удовлетворение основных потребностей пользователей путем реализации услуг (сервисов) в следующих областях: информация, бизнес, общение, а также предоставления инструментария, необходимого пользователю для продвижения собственного контента в рамках портала. В соответствии с данным определением портал должен включать четыре основных типа сервисов.

1. Информационный сервис – все, что помогает найти (при необходимости) и получить информацию.
2. Сервис реализации бизнес-функций – все то, что ориентировано непосредственно на продажу товаров/услуг.
3. Инструментарий пользователя – все, что помогает ему создавать и продвигать свой контент в сети, прежде всего бесплатный хостинг и бесплатный e-mail, рейтинги, баннеры, «анонсировщики» и др.
4. Сервис обеспечения общения (community) – все, что направлено на удовлетворение потребности в общении.

Среди поисковых систем, эволюционирующих в порталы, заслуживают внимания также Lycos (<http://www.lycos.com>) и Excite (<http://www.excite.com>). Другими путями пришли к «портальности» такие сайты, как Microsoft (<http://www.microsoft.com>) и AOL (<http://www.aol.com>).

История создания и назначение ИПС. В настоящее время информационные ресурсы только сети Internet составляют свыше десятка миллиардов документов (Web-страниц), к которым возможен свободный доступ любого пользователя. Естественно, чтобы найти необходимую информацию в этой крупнейшей распределенной полнотекстовой базе данных, необходимо использовать самые мощные ИПС. Такие системы существуют и конкурируют друг с другом на современном рынке информационных технологий.

В 1994 году университет Джорджии запустил пилотный проект "ГАЛИЛЕЙ" с использованием Site-Search – пакета программ Огайского центра в стандарте Z39.50. Стандарт Z39.50 положен в основу службы поиска распределенной информации в Internet – системы WAIS (Wide Area Information Service).

Сегодня миллионам пользователей Internet известны такие информационно-поисковые системы, как Google, Yahoo, AltaVista, AllTheWeb, каждая из которых охватывает свыше миллиарда Web-документов. За прошедшее десятилетие технология полнотекстового поиска стала повседневным инструментом миллионов пользователей. Для нахождения информации в Internet, чаще всего представленной в формате HTML. Часто возникает вопрос: как соотносятся системы, работающие с потоками документов в форматах гипертекстовой разметки, и популярные сегодня реляционные системы управления базами данных (СУБД)?

Современные реляционные СУБД поддерживают обработку запросов в стандарте SQL, позволяющем проводить поиск в рамках реляционной модели. Иными словами, стандартные средства этих систем обеспечивают эффективный поиск по совокупностям формализованных полей в рамках двумерной таблицы. Поэтому, несмотря на постоянное совершенствование информационно-поисковых систем, встроенных в СУБД, с сожалением, приходится констатировать их непригодность для решения задач глобального поиска информации в Internet-ресурсах.

В отличие от реляционных СУБД, у систем полнотекстового поиска не существует стандартизированного языка запросов. У каждой системы этого типа существует свой способ задания критериев поиска. Очень часто языки запросов поисковых систем приближены к SQL, однако каждой из них присущ ряд индивидуальных особенностей, связанных с такими моментами, как

- интерпретация операций, зависящих от порядка расположения слов в тексте (операций контекстной близости слов и др.);
- реализация вычисления близости, т.е. определения соответствия найденных документов запросам (релевантности) для представления результатов поиска;
- применение нестандартных функций, требующих, например, использования методов искусственного интеллекта (нахождение документов по принципу подобия, построение дайджестов из фрагментов документов и др.).

В различных полнотекстовых ИПС различаются архитектуры, структуры данных, алгоритмы их обработки, методологии организации поиска.

Лингвистическое обеспечение ИПС. На всех этапах развития полнотекстовых ИПС лингвистическое обеспечение играло важную роль. Именно средства лингвистики выступают интерфейсами между естественным языком и формальными поисковыми механизмами ИПС. Лингвистическое обеспечение включает такие основные элементы:

- языки представления данных в ИПС, которые определяют архитектуру, синтаксис и семантику представления информации в базах данных ИПС;
- информационно-поисковый язык, т. е. язык, на котором обращается пользователь к системе, чтобы получить интересующий его отклик.

Современные информационно-поисковые языки включают поддержку булевых операторов (И, ИЛИ, НЕ), операторов контекстной близости, средств управления приоритетами операторов, естественных языков и, наконец, языков разметки, на которых представлены документы-первоисточники. Большое значение в современных полнотекстовых ИПС уделяется морфологическому анализу, т.е. автоматическим средствам обработки отдельных слов, как в текстах исходных документов, так и в запросах пользователей.

Стоп-слова. Большинство естественных языков имеет так называемые вспомогательные слова типа артиклей и предлогов, которые входят в большинство документов и не влияют на процесс выявления документов, удовлетворяющих информационным потребностям пользователей, занимающихся поиском. Такие слова (например, а, an, the, on для английского языка) называются "стоп-словами".

Поисковые системы обычно не включают стоп-слова в свой индекс, однако учитывают при сквозной нумерации слов, что позволяет выполнять поиск фраз, содержащих «стоп-слова». Исключение стоп-слов из индекса ведет к его существенному сокращению и повышению эффективности работы. Однако некоторые запросы, состоящие только из стоп-слов (типа "to be or not to be"), в этих случаях уже не пройдут. Неудобство вызывают и некоторые случаи полисемии (многозначности слова в зависимости от контекста). Например, в одних случаях английское слово «sap» как вспомогательный глагол должно быть включено в список стоп-слов, однако как существительное оно часто несет большую содержательную нагрузку.

Морфемный анализ. При построении базы данных из массива документов (в случае сетевых ИПС такими документами выступают отдельные Web-страницы) формируется индекс из всех слов, входящих в эти документы, иногда за исключением так называемой «незначущей лексики» – предлогов, артиклей, частиц и т.д.

Файл незначущей лексики представляет собой стоп-словарь системы. Построенный словарный индекс системы во многих реализациях ИПС лемматизируется, т.е. все слова приводятся к каноническим формам, например существительные – к именительному падежу, глаголы – к инфинитивной форме и т.д. Это особенно характерно для славянских языков, для которых, в отличие, например, от английского, специфично достаточно много словоизменений. Построение индекса системы на основе лемматизированной лексики во многих случаях оправдано, но в системах, ориентированных на профессиональную работу, ориентация только на такой подход является спорной. В системах, работающих с учетом морфологии, лемматизации должны подвергаться и запросы пользователей, т.е. если в исходном документе присутствует словосочетание «белые ночи», то в индексе системы в этом случае имеются слова «белый» и «ночь». Если бы запрос пользователя «белые» и «ночи» был передан на вход поискового механизма без преобразования, то исходный документ не был бы найден, однако если данный запрос подвергнуть лемматизации, то он примет следующий вид: «белый» и «ночь», и исходный документ будет найден.

Для систем, рассчитанных на непрофессионалов (а таких большинство), лемматизация поискового индекса и запроса очень удобна. Например, представленное в запросе слово «люди» обеспечит поиск и по слову «человек».

Однако для профессионального поиска лемматизация не всегда пригодна, так как она может лишить поиск гибкости. Отсечение окончаний может увеличить количество документов, выдаваемых по запросу пользователей, однако может привести и к выдаче нерелевантных документов. Например, при использовании известного алгоритма Портера для отбора английских слов, одинаковая основа «univers» будет выделена в таких различных словах, как university и universal. Таким образом, на практике лемматизация далеко не всегда увеличивает число пертинентных документов. Механизм отсечения окончаний, полезный в некоторых случаях и неэффективный в других, в информационно-поисковых системах должен при необходимости подключаться или отключаться. Например, такая служба, как AltaVista, вообще не занимается морфологической обработкой текста. Все слова для нее – лишь последовательности символов.

Профессиональные запросы к традиционным системам. Традиционные системы пакетного поиска, обеспечивающие, например, рассылку результатов по электронной почте, не предполагают интерактивного взаимодействия с конечным пользователем, поэтому им присуща полнота, которая сродни избыточности. Так, профессиональный запрос к системе Интегрум по теме "Услуги связи" выглядит следующим образом: "услуги связи" или "междугородные переговоры" или "телефонные переговоры" или "мобильная связь" или "фиксированная связь" или "сотовая связь" или "сотовый оператор" или "средства связи" или "телефонная связь" или "спутниковая связь" или "космическая связь" или GPS или ростелеком или связьинвест или госкомсвязь или госкомтелеком или госсвязьнадзор или телекоммуникации или электросвязь или АТС или ПС или Минсвязи или "министерство связи" или "волоконно-оптическая линия связи" или ВОЛС. Очевидно, что для рабо-

ты в интерактивном режиме такие запросы неприемлемы. Пользователь желает ввести 1-2 слова и получить то, что ему необходимо. Тут на помощь могут прийти только интеллектуальные, семантические методы.

Тезаурус. Еще при появлении первых ИПС возникла дискуссия, предметом которой стало использование в качестве индексов систем автоматически формируемых словарей или подключение заранее подготовленных словарных массивов, снабженных рядом дополнительных атрибутов, – тезаурусов. В тезаурусах каждой лексической единице приписывается небольшой пояснительный текст – словарная статья и ссылки на другие слова этого словарного массива. Содержательно ссылки могут означать следующее: синонимию, противопоставление отдельных слов, подчиненность и т.д. Структура наполнения тезауруса регламентируется соответствующими стандартами – ISO 2788, ГОСТ 7.25-80 (для одноязычных тезаурусов) и ГОСТ 7.24-90 (для многоязычных тезаурусов).

Формирование поискового индекса во многих ИПС выполняется по правилам построения тезаурусов, в которые были включены такие типы лексических единиц:

- отдельные слова (существительные, прилагательные, глаголы, наречия);
- словосочетания;
- лексически весомые компоненты сложных слов;
- аббревиатуры;
- сокращения слов и словосочетаний.

В тезаурусах различные формы лексических единиц приводятся к каноническим формам. При формировании поискового индекса системы на основе тезауруса каждое слово из документов, входящих в базу данных ИПС, анализируется на вхождение в тезаурус. Особый смысл имеет использование тематических тезаурусов для специализированных баз данных, однако сегодня остается открытым вопрос построения политематического тезауруса и индекса ИПС на его основе. Хотя следует отметить, что при формировании баз данных на основе Web-сайтов Internet именно политематический тезаурус представляет самый большой интерес.

Другой подход, который чаще всего используется сегодня, основан на механизмах автоматического построения поискового индекса системы на основе входящих в документальный массив слов. Этот подход предполагает отказ от использования тезауруса или, по крайней мере, лишь минимальное его использование для второстепенных целей. В настоящее время этот подход считается более технологичным.

В самом начале истории ИПС противники использования тезаурусов приводили как один из основных аргументов то, что объемы тезаурусов не позволяют хранить их в памяти машины. В те времена объемы текстовых баз данных были относительно небольшими и автоматические индексы систем были на порядок менее объемными, чем соответствующие тезаурусы. Сегодня ситуация изменилась в корне – с одной стороны, объемы промышленных носителей информации позволяют хранить практически неограниченное количество тезаурусов, а с другой стороны, объемы текстовых баз данных настолько велики, что их индексы зачастую превышают объемы тезаурусов. Все это дает основание предполагать, что соотношение тезаурусных и бестезаурусных систем, сложившееся в результате инертности внедрения новых технологий, в недалеком будущем изменится.

Методологические основы поисковой процедуры

Семантические методы. В последнее время в технологии поиска все чаще стали внедряться элементы контент-анализа – методологии, возникшей в конце XIX-начале XX вв. Эта методология, изначально ориентированная на применение в психологии и социологии, сегодня все чаще используется в различных автоматизированных системах. Различают количественный и качественный контент-анализ. Если качественный контент-анализ базируется на глубоком лингвистическом и семантическом анализе отдельных предложений и всего текста, то основой количественного контент-анализа являются статистические подходы.

В последнее время получили развитие такие направления контент-анализа, как «Data Mining» и «Text Mining», которые предполагают автоматическое выявление из текстовых массивов нового смысла, новых данных, феноменов, фактов-знаний. Все чаще возникают попытки привлечения методов контент-анализа, а точнее Text Mining, в реальные поисковые системы. И эти попытки не умозрительны – они обусловлены объемами и темпами роста Сети. Во многие современные сетевые поисковые системы внедрены такие компоненты, как:

- автоматическая группировка документов по определенному заранее классификатору;
- автоматическое определение новых, не заданных заранее классов на основе неструктурированных или слабо структурированных документов;
- ранжирование документов по смысловой релевантности;
- выявление семантически подобных документов – поиск подобных документов на основе эталона;
- автоматический анализ и смысловое преобразование запросов пользователей.

Группировка результатов поиска. Большинство современных интеллектуальных систем обеспечивает группировку своих откликов по заранее определенным классификаторам. Но создатели поисковых служб реализовали технологию выявления «смысла» слов путем построения обучаемой внутренней семантической сети. Например, технология, предлагаемая службой AltaVista (<http://www.av.com>), обеспечивает реализацию режима уточнения поиска (Refine Your Search), автоматическое определение классов и последующую группировку (кластеризацию) откликов ИПС в соответствии с ними. В этой системе активизация соответствующего класса приводит к уточнению первоначального запроса.

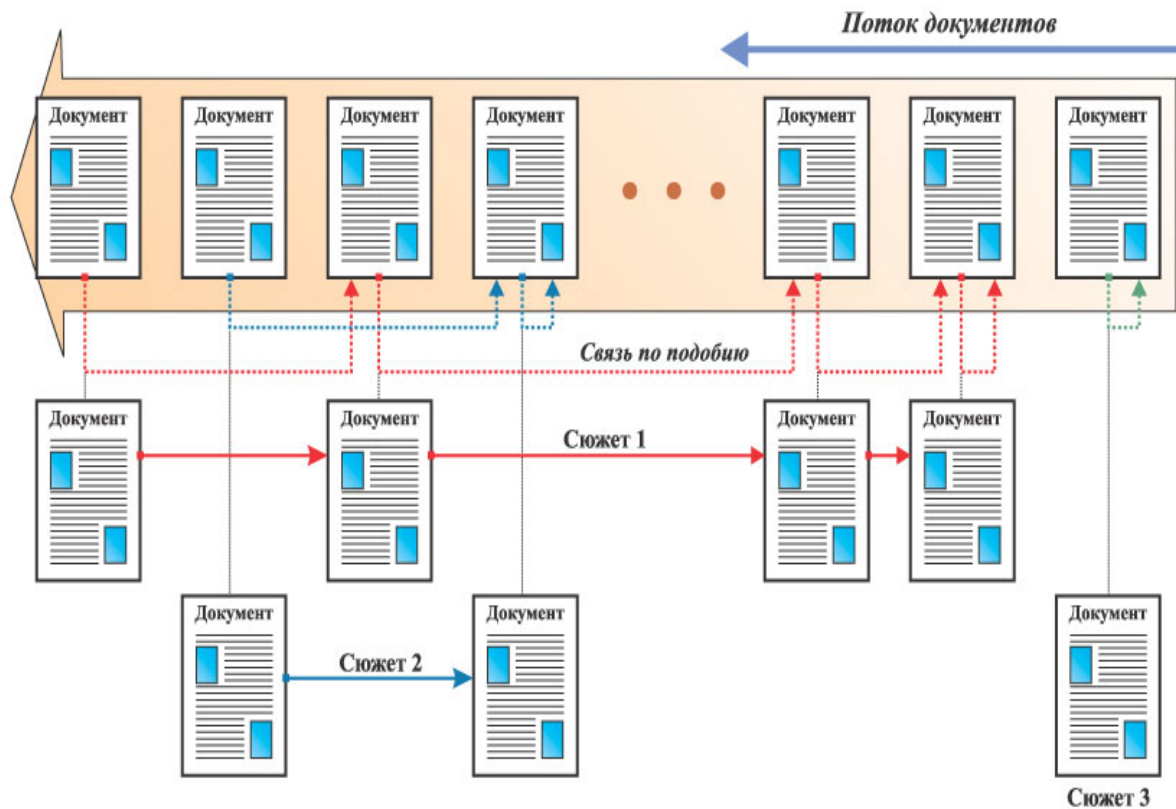
Сюжетный подход. При поиске новостной информации всегда возникает задача нахождения и объединения в сюжетные темы документов, описывающих одни и те же события и ранжирования сюжетов по некоторым признакам, что должно обеспечить не только выявление самой важной темы, но и «вверное», многоаспектное освещение всех наиболее значимых событий. Эта задача решается во многих системах, но с использованием различных подходов и алгоритмов. При этом неизменной остается технологическая цепочка: построение семантической сети из документов, кластеризация – автоматическое выявление наиболее взаимосвязанных групп (т.е. сюжетов), «взвешивание» этих сюжетов и наглядная визуализация самых важных из них. При выделении сюжетных цепочек для определения попарной близости текстов, как правило, используются алгоритмы выявления похожих документов, ставшие уже традиционными в поисковых системах.

Матрица попарной близости документов обрабатывается алгоритмами кластеризации. Выделенные классы документов и представляют собой сюжетные цепочки (рис.4.7).

Для предъявления пользователям сюжеты должны быть ранжированы. Основные факторы, влияющие на ранжирование по важности, – оперативность информации и размер сюжетной цепочки. Под оперативностью понимается некоторая функция от времени публикации всех сообщений в сюжете, а размер сюжета отражает общий интерес к конкретной теме. Во всех этих подходах центральная задача состоит в отождествлении сообщений, относящихся к одному сюжету, и выявлении «непересекающихся» сюжетов. Для результирующего отображения каждого отдельно взятого сюжета используются отобранные по содержательной близости документы из различных источников, отсортированные в хронологическом порядке.

При этом сюжеты могут представлять собой дайджесты, интегрирующие общие места документов по теме, а также уникальную информацию, содержащуюся в отдельных документах. Реферирование сюжета в этом случае сводится не к свертыванию информации, а к построению расширенной версии, по сравнению с любым документом из сюжетной цепочки.

Например, в системе Yandex. Новости (<http://news.yandex.ru>) для этого строится матрица попарной близости документов, которая обрабатывается алгоритмом кластеризации с эмпирически подобранными параметрами (в частности, радиусом метрики близости).



Сюжеты - результат построения семантических цепочек

Рисунок 4.7 – Построение сюжетных цепочек

Для того чтобы увеличить связность крупных сюжетов, в системе Yandex. Новости дополнительно используется кластеризация второго уровня, обеспечивающая сбор атомарных кластеров в более крупные. В результате внедрения этой системы, все сообщения в результатах поиска на сайте Yandex. Новости сгруппированы по сюжетам, при этом ранжирование построено на стандартных для Yandex принципах ранжирования сгруппированной выдачи. Оно основано на числе и ранге новостей внутри новостных сюжетов, при этом ранг отдельной новости определяется как ее свежесть с учетом приоритетов текстуального совпадения. Можно сказать, что задача автоматического построения качественных тематических сюжетов на основе потоков сетевой новостной информации сегодня практически решена.

Полностью автоматические средства многих систем, обрабатывают поток новостной информации, превышающий десятки тысяч документов в сутки и обеспечивают полноту свыше 80% и точность около 95%.

Структуризация поисковой процедуры. Целью создания ИПС является предоставление пользователю возможности поиска информации по интересующей его тематике, выраженной специальными запросами. Различные ИПС имеют собственные языки запросов или, как их еще называют, информационно-поисковые языки (ИПЯ), позволяющие в той или иной мере описывать предметные области пользователей. Очевидно, что составление запросов должно базироваться на этих ИПЯ, однако сам процесс составления запросов допускает многовариантность и является своего рода искусством. При поиске в Internet следует четко определить информационные потребности, необходимую ретроспективу информации, круг поисковых серверов, специализирующихся на индексирова-

нии подобной информации, и даже предусмотреть заранее возможный результат, подобрав несколько известных документов сходной тематики. Пользователь Internet должен определить необходимые ему поисковые серверы и каталоги. Выбрав необходимые поисковые ресурсы, пользователь Internet составляет поисковое предписание, соответствующее интересующей его тематике. Только при этом он осознанно формирует запрос на ИПЯ.

Пользователь Internet, анализируя отклик ИПС, выбирает ссылки на документы, которые, по его мнению, действительно соответствуют его информационным потребностям. Далее он выходит непосредственно на первоисточники, анализирует их и копирует себе только ту информацию, которая является наиболее полезной для него.

Как и любой поиск, поиск в Internet является искусством, и ему, как и многим видам искусства, присуща многовариантность и творческий подход. Поиск в Internet можно рассматривать и с точки зрения его этапности.

Этапы поисковой процедуры. Процедура поиска имеет вполне определенную этапность – от определения информационной потребности и области поиска до анализа результатов и выбора пертинентных объектов. Информационные потребности пользователя могут относиться к разным областям, которые могут быть как узкоспециализированными, так и достаточно типовыми.

На практике основная часть информационных потребностей приходится именно на типовые области применения: поиск отдельных Web-страниц; поиск новостей; поиск людей и организаций; поиск литературных произведений; поиск программного обеспечения; поиск музыкальных произведений; поиск графических изображений; поиск видеоинформации; поиск коммерческой информации.

Оперативная, часть поисковой процедуры предполагает многовариантность подходов и решений при формализации запросов в процессе их отработки. В этом случае аналитик-профессионал приходит к необходимости использования весьма ограниченного числа поисковых серверов, каталогов и отдельных web-ресурсов для решения своей задачи. Основной задачей второго этапа является формирование эффективных запросов к ИПС. Наибольшую проблему при формировании запросов представляет то, что на каждом поисковом сервере используется свой информационно-поисковый язык (ИПЯ), несмотря на то, что у различных языков этого типа много общего, – например, схожий набор булевых операций. В настоящее время не существует единого стандарта, подобного стандарту языка SQL для СУБД, хотя на протяжении многих лет ведутся попытки такой стандартизации.

Третий этап поиска в сети Internet является определяющим, – от его реализации зависит, будет ли найденное решение пертинентно. На этом этапе пользователь работает с конечными документами, полученными в виде отклика ИПС. От правильного выбора набора документов-первоисточников зависит результат работы всех трех этапов поисковой процедуры.

Полученные в результате обработки запросов отклики ИПС требуют, с одной стороны, скрупулезной работы пользователей-аналитиков и, с другой стороны, развитых средств автоматизации аналитической работы, обеспечивающих:

- итеративное уточнение запросов;
- поиск по подобию;
- ранжирование выдаваемых документов;
- построение графических отчетов, визуализацию.

Запросы пользователей. Казалось бы, с развитием технологических возможностей современные поисковые системы должны обеспечить гарантированное нахождение информации, однако «ленивые» пользователи все же очень часто недовольны качеством их работы. Основная масса пользователей не хочет прикладывать особых интеллектуальных усилий при формировании критериев поиска. Удивительно низким оказывается процент использования запросов, усложненных хотя бы одним логическим или контекстным опе-

ратором. Если и используются операторы, то это, в основном, булевы AND и OR. Доля использования операторов контекстной близости и логического отрицания (NOT) не превышает 1-2%. В то же время реализация отработки сложных запросов (которых пока не более 20%) и определяет эффективность использования времени, проводимого пользователем в Internet.

Согласно исследованию, проведенному OneStat.com в 2004 году (табл. 4.1), большинство поисковых запросов в Сети состоят из двух слов – 32,58% от общего количества. Из трех слов состоит 25,61% запросов и 19,02% запросов состоит из одного слова.

Таблица 4.1 – Распределение запросов по количеству слов, полученное аналитической службой OneStat

Количество слов в запросе	Количество запросов в процентах
1	19,02
2	32,58
3	25,61
4	12,83
5	5,64
6	2,32
7	0,98
8 и более	1,02

Эти результаты нельзя признать утешительными, так как они наглядно свидетельствуют о том, что человек в Internet не слишком проявляет признаки разумности. Но все же среди поисковых запросов год от года преобладают все более сложные конструкции – чтобы найти что-то конкретное, пользователям приходится прибегать к все более сложным поисковым запросам. Для сравнения еще в апреле 2003 года процент поисковых запросов из одного слова составлял 24,76%.

Кроме того, очень большое значение имеет ранжирование результатов поиска, т. е. порядок следования документов, предъявляемых пользователю. Так, исследователи из IST, проанализировав характер свыше 450 тыс. запросов, выданных за сутки поисковой системе AlltheWeb.com, обнаружили, что пользователи чаще всего просматривают первые три ссылки, полученные по запросу, очень быстро оценивают найденные сайты и еще быстрее разочаровываются в результатах. Другие исследования показали, что 75% пользователей удовлетворяются первыми 10-15 результатами поиска. И только 20% просматривают результаты на второй странице и менее 5% добиваются до третьей и последующей страниц с результатами поиска.

Для ввода сложных запросов требуется использование булевых и контекстных операторов, скобок, указание полей и тому подобное, что недоступно для среднестатистического пользователя. Поисковые службы обычно создают два интерфейса – простой (по умолчанию) и расширенный (называемый в разных системах детальным, мощным или профессиональным), однако главная задача коммерческих поисковых служб как раз и заключается в удовлетворении информационных потребностей среднестатистического пользователя. Создается впечатление, что при поиске очень редкого слова любая ИПС даст хороший результат. Действительно, если слово или словосочетание редкое, то документов выдается, как правило, немного, они все быстро просматриваются без особых затрат на аналитическую работу. Правда, все усложняется, если система не находит ни одного документа. В этом случае следует обратиться к другим системам или изменить критерии поиска. При получении результатов запроса возможны несколько случаев. На, казалось бы, логично сформулированный запрос выдается тысяча документов, имеющих слабое

отношение к информационным потребностям. В этом случае рекомендуется применить два метода: первый – кардинальный – полностью переформулировать запрос, изменив представление о возможном поисковом образе, второй – уточнить запрос с помощью добавления еще одного условия с применением операции конъюнкции (оператора логического "И"). Второй путь реализуется в большинстве систем опцией «искать в найденном». В этом случае, не изменяя логики предыдущего запроса, а лишь уточняя его, можно добиться удовлетворительных результатов.

Второй случай, когда в результате поиска по запросу найдено избыточное количество документов, но при просмотре первых страниц результатов поиска найдено несколько пертинентных документов. Естественно, у пользователя возникает желание найти еще документы (или ссылки на них), сходные с ними по содержанию, не затрачивая интеллектуальных усилий на анализ и составление запроса. В результате многие ИПС реализовали опции «найти подобное», «find similar», «like this». Однако этот режим не всегда ведет к удовлетворительным результатам при целевом поиске, но иногда приводит к получению полезных документов, имеющих косвенное отношение к теме первичного запроса. Что означает «подобный документ» и по каким критериям это определяется, зачастую остается загадкой для пользователя. Один из подходов к ее решению может быть таким: каждое значимое, по мнению системы, слово ранжируется по какому-то критерию, из наиболее весомых слов автоматически формируется запрос, рассматриваемый как новый критерий поиска. Такой режим реализован во многих современных ИПС, например, на серверах Excite, Google и Yandex.

Ранжирование откликов. Ранжирование выдаваемых документов, в отличие от предыдущей опции, имеет большое значение в работе современных ИПС. Инструменты повышения пертинентности в современных системах, помимо возможностей уточнения формулировки запросов, предусматривают использование весовых критериев, что позволяет ранжировать найденные документы и выдавать пользователю для просмотра наиболее весомые документы либо вообще ограничиваться выдачей не более заданного числа наиболее весомых документов. Следует отметить, что в современных системах проблеме релевантности, а особенно пертинентности, уделяется все большее внимание. Ранжирование выдаваемых документов может выполняться по дате создания/обновления документа, по степени важности (многие системы оценивают важность документов по весовым критериям или по количеству ссылок на них, т.е. по цитированию). Ранжирование по дате имеет особое значение при поиске новостных сообщений средств массовой информации и информационных агентств. Ранжирование по индексу цитирования, аналогичное оценке значимости научных публикаций в традиционной научной среде, впервые ввела Google, продемонстрировавшая эффективность такого подхода для Web-пространства.

Поиск по словам и словоформам. Все поисковые системы обеспечивают поиск хотя бы по одному слову. Средства навигации в Internet, не обеспечивающие такого поиска, называются каталогами, коллекциями ссылок и т.п. Иначе дело обстоит с усечениями слов. Некоторые системы рассматривают все слова запроса как правые усечения. У других известных систем возможность поиска по усечениям попросту не реализована (Google, AUtheweb, Рамблер). Однако в большинстве систем для маскирования правого усечения слова достаточно поставить символ "*" (AltaVista, Yandex). Некоторые системы не чувствительны к регистрам букв в словах запросов. К таким системам относится AUtheweb, Google и UAport. Однако в большинстве приведенных выше систем «чувствительность» к регистрам включается при употреблении хотя бы одной прописной буквы в слове запроса.

Поиск по словоформам является результатом серьезного лингвистического анализа и реализован в русскоязычных системах Апорт, Яндекс и Рамблер, а также в украинской системе МЕТА. К примеру, в системе Апорт, независимо от того, в какой грамматической форме указано слово в запросе, оно находится в базе данных во всех своих формах. В этой системе запрос «ребенок шел» эквивалентен запросу «дети идут».

В системах Yandex и Рамблер, если слово участвует в запросе, учитываются также все его формы. Для поиска по конкретному слову, а не всем словоформам, перед ним ставится символ "!" (Yandex) или оно берется в кавычки (Рамблер).

Логические операторы. Для ввода сложных запросов требуется использование булевых и контекстных операторов, скобок, указание полей и т.п. Хотя для большинства случаев (по статистике 70% запросов состоят из одного слова) этого не требуется. Поэтому поисковые службы обычно создают два интерфейса – простой (по умолчанию) и расширенный (называемый в разных системах детальным, мощным или профессиональным). Но есть и такие системы, которые с помощью одного и того же механизма позволяют вводить, а затем обрабатывают простые и сложные запросы, обеспечивая пользователей руководствами различного уровня сложности.

Во всех современных системах реализованы булевы операторы AND, OR и NOT, а также работа со скобками. Однако в двух из них – AltaVista и Excite – оператор NOT записывается в виде AND NOT, что подчеркивает его бинарность (в математической логике оператор NOT в чистом виде является унарным). В режимах простого поиска булевы операторы реализуются не всегда указанием их в явном виде. Например, во многих поисковых системах пробел между словами запроса по умолчанию воспринимается как оператор AND (Allthenews, Google, META и UAport). В то же время при указании опций типа any of the words, пробел в таких системах воспринимается как OR. Кроме того, в Alltheweb допускается использование перед словами операторов «+» и «-» фактически как синонимов операторов AND и NOT соответственно. Точно так же используются эти операторы в AltaVista, Excite, Lycos и Апорт. Можно отметить, что у самой популярной сегодня системы Google – самый лаконичный набор логических операторов: «+», «OR» и «-».

Операторы контекстной близости. Большинство профессиональных поисковых систем обеспечивает выполнение операций контекстной близости, одна из реализаций которой – поиск выражений в кавычках. Например, в системе Google реализована только возможность поиска по фразам в кавычках, в AltaVista реализован оператор NEAR (~), обеспечивающий нахождение документов, у которых два слова находятся на расстоянии не более 10 слов. В системе Lycos функции контекстной близости получили наибольшее развитие и реализованы с помощью четырех операторов: ADJ, NEAR, FAR и BEFORE. Оператор ADJ обеспечивает близость двух слов в тексте в любом порядке, а оператор NEAR позволяет находить документы, в которых слова-операнды удалены не более, чем на 25 слов. FAR – оператор, противоположный по смыслу оператору NEAR, т.е. он исключает близость терминов запроса в пределах 25 слов текста документа, а оператор BEFORE похож на оператор ADJ, только с учетом порядка встречаемости терминов в тексте. Оригинально решен вопрос контекстной близости в системе Рамблер. Значение ограничения контекста в этой системе можно изменять конструкцией (число, запрос), где число – любое положительное число, а запрос – любой корректный запрос, состоящий более чем из одного слова. Таким образом, по запросу (2, красная роза) будут найдены только те документы, в которых между словами «красная» и «роза» хотя бы раз не стоит ни одного слова. В системе Yandex режим контекстного поиска называется «поиском с расстоянием». В общем виде ограничение по расстоянию задается в строке данных выражением вида

$$I(n\ m),$$

где n – минимальное, а

m – максимальное допустимое расстояние.

В системе Апорт существует два вида ограничения по расстоянию:

- в словах $wN(\dots)$,

где N – число слов, и

- в предложениях $sN(\dots)$,

где N – число предложений.

В этой системе также подвергаются интеллектуальной обработке выражения в кавычках. Например, запрос "яблоки на снегу":

- эквивалентен запросам "яблоки и снег",
- "яблоки под снегом",
- "яблоко снег".

Большинство из названных систем способно реализовать контекстный поиск: заключенной в кавычки фразы (Google, Alltheweb, AltaVista, Lycos и др.). Такая способность – это реализация неявно указанных с помощью кавычек операторов контекстной близости.

Поиск по параметрам. Отдельного рассмотрения заслуживает возможность поиска по параметрам документов, которая позволяет ограничивать диапазон поиска значениями URL, датами, заглавиями и т.п. Чаще всего получить такую возможность можно из режима расширенного поиска, в котором для ввода значений отдельных параметров предлагается весь диапазон возможностей Web-интерфейса.

Например, в системе Alltheweb в запросах можно указать параметры, обеспечивающие поиск по таким элементам:

- URL – *url:* (например, по запросу *url:energ* будут найдены документы, в URL которых имеется строка «energ»);
- ссылки на страницы сайтов – *link::*;
- доменные имена – *site:* (например, *site:ua* обеспечит нахождение документов из украинского домена);
- заголовки – *title:*.

В этой системе допустим поиск, кроме всех вариантов текстовых файлов, еще трех типов файлов – PDF, MS Word, Flash.

В системе AltaVista имеются все приведенные для Alltheweb возможности; кроме того, в режиме расширенного поиска обеспечивается поиск по датам (с явным указанием диапазона поиска либо с указанием типа «искать за последние *n* месяцев»). Этот режим в системе традиционно называется «Web-археологией». В Google обеспечивается поиск по сайту (site), определение ссылок на сайт (admission site-), поиск по ценам, странам, датам, доменам и т.д. В поле ввода запроса можно вводить и арифметические выражения, используя интерфейс Google как калькулятор, что, конечно же, подчеркивает своеобразность данной системы (например, по запросу 4^2 будет выведен результат 16).

4.4. XML – язык разметки и модель данных

Для эффективного представления и последующего поиска в Internet возникла необходимость связать воедино информационные ресурсы, позволить программным приложениям устанавливать связи между объектами и обмениваться информацией на одном языке. В связи с этим в 1996 году консорциумом World Wide Web (<http://www.w3.org>) была предпринята попытка приступить к проектированию расширяемого языка разметки, который сочетал бы в себе гибкость и мощность промышленного стандарта для издательских приложений – языка SGML (Standard Generalized Markup Language) и был совместим с уже получившим повсеместное распространение языком HTML. Новый язык получил название XML (eXtensible Markup Language) в феврале 1998 года его версия XML 1.0 была принята как рекомендация W3C.

На языке HTML составлена большая часть Web-сайтов и отдельных Web-страниц в Internet. Тем не менее, этот язык не лишен многих ограничений по отображению символов, не входящих в стандартные таблицы кодировок. Например, в нем отсутствуют возможности для отображения музыкальных нот или математических формул. Кроме того, изначально HTML не предполагал серьезного интерактивного взаимодействия. Возможности подключения CGI-форм или JavaScript-приложений лишь частично решают эту проблему. Именно для решения проблем внешнего и смыслового описания документов и была в свое время начата разработка нового стандарта хранения и обеспечения доступа к информации (прежде всего, сетевой), который получил наименование XML – «расширяемый язык разметки». Благодаря своей простоте и расширяемости, XML практически сразу же получил широкое распространение и послужил основой для создания других специализированных языков разметки – CML, MathML, WML и десятков других. Даже традици-

онный HTML можно рассматривать как дочерний язык XML, поскольку его описание также укладывается в стандарт XML.

Создатели языка XML убрали изобразительные «излишества» HTML, в результате чего пришли к простому, эффективному и наглядному метаязыку. В отличие от большинства подобных языков, XML одинаково понятен и человеку, и программам обработки – парсерам.

При создании XML-документа нет ограничений, связанных с фиксированным набором элементов разметки, – в этом языке не существует заранее определенного набора тегов. Все необходимые для описания документа теги разработчик может определять самостоятельно. При этом теги XML не предлагают решения вопроса, как информация должна выглядеть на экране. При публикации можно применять множества правил, собранных в листы стилей для автоматического форматирования документов. Для этого был разработан специальный расширяемый язык стилей (eXtensible Stylesheet Language, или XSL). XML-документ представляет собой дерево вложенных открывающих и закрывающих тегов, каждый из которых может включать в себя несколько пар «атрибут-значение».

При этом не существует фиксированного словаря тегов и набора их допустимых комбинаций. В XML 1.0 описание правил построения документа выполняется с помощью специального языка определения типа документа DTD (document type definition), накладывающего ограничения на используемые теги и указывающего, каким образом должна быть организована их вложенность внутри документа. В DTD определяется грамматика, т.е. допустимые комбинации и вложенность имен тегов и атрибутов.

В XML применяется два вида указания на определение DTD, использованное для конкретного документа.

- Декларации внутренних подмножеств DTD-определений, помещаемых непосредственно в XML-документ. При этом команда-определение DTD заключается в квадратные скобки, например `<!DOCTYPE rootElement [declarations] >`.

- Ссылки на внешние DTD-определения, например `<!DOCTYPE wml PUBLIC "-//WAPFORUM/DTD WML 1.1//EN" "http://www.wapforum.org/DTD/wml_1.1.xml">`.

Указание PUBLIC во втором случае свидетельствует о том, что DTD является общедоступным стандартом, в частности в данном примере это стандарт языка WML. Язык DTD позволяет определить логическую структуру документа, т.е.

- задать порядок следования элементов;
- определить вложенность элементов;
- установить количество возможных элементов;
- установить типы атрибутов;
- определить сущности и нотации.

Однако языку DTD присущи серьезные недостатки, а именно ограниченность описания типов данных и синтаксис, отличный от XML. Поэтому в настоящее время W3C активно работает над тем, чтобы заменить язык DTD на новый стандарт – использование XML-схем (XML Schema). Техническая рекомендация «XML-схема» была принята в 2001 году (<http://www.w3.org/TR/xmlschema-formal>). Ведущие производители программного обеспечения во всем мире приняли концепцию XML-схем и внедрили ее в своих продуктах. Так, корпорация Microsoft на условиях бесплатного лицензирования предоставила разработчикам доступ к XML-схемам, используемым в Microsoft Office 2003. В настоящее время по адресу www.microsoft.com/office/xml/default.msp доступны следующие XML-схемы:

- SpreadsheetML для Microsoft Office Excel 2003;
- FormTemplate Schemes для Microsoft Office InfoPath 2003;
- WordprocessingML для Microsoft Office Word 2003;
- DataDiagramingML для Microsoft Office Visio 2003.

По мнению представителей корпорации, доступность схем значительно облегчит реализацию поддержки возможности обмена данными между разрабатываемыми программами и приложениями офисного пакета.

Безусловное преимущество XML – это использование им современного стандарта кодировки символов Unicode, который позволяет в одном документе комбинировать тексты, написанные на основных языках мира (в том числе имеется поддержка кириллицы). Тем самым XML дает возможность с легкостью обмениваться информацией вне национальных границ.

XML как модель данных. XML – это удобный формат обмена данными, язык описания документов. При этом оказалось, что спецификацию XML можно рассматривать еще и как реализацию иерархической модели данных. На основе этой модели можно строить системы управления XML-базами данных, причем первые же попытки показали высокую эффективность таких систем. В свое время аналитик Yankee Group Роб Перри отметил: «XML-документы могут содержать мегабайты текстовых строк и тегов XML. Такие документы имеют иерархическую структуру. И лучше оставить эту иерархию в неизменном виде, ведь поиск в этом случае окажется гораздо более эффективным». Даже отдельный XML-файл сам по себе – уже готовая иерархическая база данных. Дело остается лишь за языком запросов, основой для которого тоже может служить сам XML.

Представление данных в виде XML-документов является естественным отражением структуры реальных документов. Представлять данные как XML-документы значительно легче, чем помещать их в реляционные таблицы, которые проще трактовать лишь как фрагменты документов. При этом манипулировать данными с помощью присущей XML-технологии гиперсвязности пользователю удобнее, чем ключами, используемыми в реляционной модели. Поэтому сегодня стоит вопрос не о том, возможно ли создание XML-СУБД, а о том, можно ли заменить такими системами традиционные реляционные СУБД.

При создании СУБД нового поколения для XML-модели данных (их часто называют native – естественными) разработчики исходят из необходимости отражения операций, которые совершаются с документами в реальной жизни. Такие базы данных в качестве модели используют XML-модель данных. С одной стороны, именно такие базы данных намного лучше подходят для хранения XML-контента, чем широко распространенные сейчас реляционные СУБД. В то же время такие компании, как Oracle, IBM и Microsoft, разрабатывают технологии, улучшающие возможности работы их реляционных СУБД с XML-документами. При этом, конечно, следует отличать XML-ориентированные базы данных от реляционных и смешанных, которые также основаны на реляционной модели, но поддерживают обмен данными на языке XML (такие БД также называют постреляционными).

Под языком XML подразумевается совокупность трех тесно связанных стандартов. Сюда входит сам XML как средство описания структуры документов; XSL – как средство преобразования XML-документов для отображения; а также XLL-расширяемый язык связывания документов. Последний позволяет устанавливать многонаправленные ссылки и ссылаться не на весь документ, а на его конкретные элементы. Кроме того, в XML структурированные данные документа отделены от описания способа их логико-графического представления в виде списков, параграфов, таблиц, диаграмм и так далее, а определение логического представления отделено от задания конкретного внешнего вида (стиля). Для разработчиков приложений существует спецификация программного интерфейса XML OM (например, компания Microsoft использовала его в виде Document Object Model – DOM).

Работа с XML-данными заранее неизвестной структуры – это принципиальная особенность XML-ориентированных СУБД, выгодно отличающая их от реляционных СУБД. XML-ориентированные СУБД обеспечивают значительно большую скорость выполнения транзакций, что объясняется, с одной стороны, меньшими затратами на выполнение преобразований данных, а с другой стороны, способом управления памятью – по алгоритму

двоичного дерева. В XML-СУБД данные могут быть записаны без предварительного ввода описаний, при этом они сразу же становятся доступными пользователям за счет унификации обработки XML-тегов. XML-СУБД также характеризуются простотой разработки приложений, что обусловлено естественностью и простотой их архитектуры. Кроме того, технология XML-СУБД полностью соответствует концепции корпоративных хранилищ данных. Обычно такие комплексы физически представляют собой совокупность нескольких, связанных между собой баз данных, содержащих, помимо прочего, аналитическую информацию, необходимую для принятия решений. В XML-базах данных каждый документ может быть интерпретирован как аналитический, т.е. представляющий собой ответ на запрос, непосредственно аналитический расчет, входную или выходную форму. Возможность преобразования XML-документов в реляционное представление и обратно позволяет подсоединить к XML-СУБД реляционные базы данных из уже имеющихся корпоративных хранилищ. Кроме того, технология XML предоставляет возможность создания децентрализованных баз данных, логически объединенных через Internet/intranet.

Отсутствие до последнего времени унифицированных стандартов для XML-запросов и достаточного числа примеров масштабного развертывания XML-баз данных пока еще сдерживает массовый переход к технологии XML – большинство корпоративных пользователей по-прежнему предпочитают хранить данные в реляционных СУБД. Однако внедрение этих технологий дает немалые шансы на существенный выигрыш в оперативности обработки информации и естественности ее представления. Для ускорения подобного перехода многие производители реляционных СУБД добавляют поддержку XML к своим продуктам.

Так, в СУБД Oracle версии 9i (www.oracle.com) появилось множество новых возможностей, включая поддержку методов Data Mining (глубинный анализ данных), персонализацию и работу с XML. Теперь сервер Oracle является не только объектнореляционным, но и позволяет хранить и обрабатывать XML-данные. В свою очередь, компания IBM (www.ibm.com) создала расширение своей СУБД DB2 Universal Database для представления и обработки слабоструктурированных данных DB2 XML Extender, предназначенное, прежде всего, для работы с XML-данными. Кроме того, IBM объявила о создании новой технологии Xperanto, призванной облегчить поиск информации в реляционных базах данных, в документах формата XML, в плоских (текстовых) файлах, электронных таблицах и прочих источниках данных, как если бы они были организованы в единую базу данных. В своей работе Xperanto использует язык XQuery для поиска информации в XML-документах. За исключением языка XQuery, все технологии, применяемые в Xperanto, уже применялись в последней версии СУБД IBM DB2.

Контрольные вопросы к лекции 4.

- 4.1. Каковы общие принципы поиска информационных ресурсов?
- 4.2. Что собой представляет топология Web-пространства?
- 4.3. Назовите средства и методы навигации в Internet
- 4.4. Для чего нужен XML – язык разметки и модель данных?

Лекция 5. МИРОВОЙ РЫНОК ИНФОРМАЦИОННЫХ УСЛУГ

Вопросы:

- 5.1. Характеристика мирового рынка информационных услуг
- 5.2. Основные секторы рынка информационных продуктов и услуг
- 5.3. Профессиональные базы данных и ресурсы Internet

5.1. Характеристика мирового рынка информационных услуг

Для того чтобы использовать информационные ресурсы, необходимо знать, где они находятся или как их можно найти, сколько они стоят, кто ими владеет, насколько они доступны и т. п. Для этого существует рынок информационных продуктов и услуг. Рынок информационных продуктов и услуг – это система экономических, правовых и организационных отношений в сфере торговли продуктами интеллектуального труда. Как любой другой, информационный рынок имеет свои особенности – это специфичный товар со своими механизмами и условиями продажи. В качестве товара здесь выступают информационные системы, технологии, патенты и лицензии, ноу-хау, товарные знаки, различного рода информация и прочие виды информационных ресурсов.

Основными источниками ресурсов для информационного обслуживания в современном обществе являются различные базы данных. Они интегрируют в себе поставщиков и потребителей информационных услуг, связи и отношения между ними, порядок и условия продажи и покупки информационных услуг.

Стремительному увеличению объемов продажи информационной продукции способствовали широкое внедрение в информационную деятельность вычислительной техники и возможность удаленного доступа к БД по национальным и международным каналам связи и информационным сетям. Развитие информационных технологий привело к появлению двух фундаментальных возможностей, обеспечивающих конкурентные преимущества фирмам, которые их используют:

1. доступ к удаленным информационным ресурсам;
2. распределенная обработка информации;

В настоящее время рынок информационных услуг представляет собой совокупность экономических, правовых и организационных отношений по продаже и покупке информационных продуктов и услуг, складывающихся между их поставщиками и потребителями.

Поставщиками информационных продуктов и услуг могут выступать:

- центры, где создаются и хранятся базы данных;
- центры, распределяющие информацию на основе баз данных;
- службы телекоммуникации и передачи данных;
- специальные службы, куда стекается информация по конкретной сфере деятельности для ее анализа, обобщения, прогнозирования, например консалтинговые фирмы, банки, биржи;
- коммерческие фирмы;
- информационные брокеры;

Потребителями информационных продуктов и услуг могут быть различные юридические и физические лица, решающие определенные задачи.

5.2. Основные секторы рынка информационных продуктов и услуг

Основными компонентами рынка информационных услуг являются:

- *техническая составляющая* – информационные технологии;
- *нормативно-правовая составляющая* – юридические документы, обеспечивающие правовые нормы на информационном рынке.
- *информационная составляющая* – информационно-поисковые системы, помогающие ориентироваться в огромных массивах информации;

• *организационная составляющая* – методы государственного регулирования взаимодействия производителей и распространителей информационных продуктов и услуг.

Инфраструктура информационного рынка – это совокупность секторов, каждый из которых объединяет группу людей или организаций, предлагающих однородные информационные продукты и услуги.

В последние годы сформировалось следующее представление инфраструктуры информационного рынка, которую можно представить в виде пяти больших секторов:

- деловая информация;
- информация для специалистов;
- потребительская информация;
- услуги образования;
- обеспечивающие информационные системы и средства.

Деловая информация. Данный сектор состоит из следующих частей:

- биржевая и финансовая информация – котировки ценных бумаг, валютные курсы, учетные ставки, рынок товаров и капиталов, инвестиции, цены. Поставщиками являются специальные службы биржевой и финансовой информации, брокерские компании, банки;
- статистическая информация – ряды фактических данных и прогнозы социально-экономического и демографического характера. Поставщиками являются государственные службы, компании, консалтинговые фирмы;
- коммерческая информация по компаниям, фирмам, корпорациям, направлениям работы и их продукции, ценам; о финансовом состоянии, связях, сделках, руководителях, деловых новостях в области экономики и бизнеса. Поставщиками являются специальные информационные службы.

Информация для специалистов. Содержит следующие части:

- профессиональная информация – специальные данные и информация для юристов, врачей, фармацевтов, преподавателей, инженеров, геологов, метеорологов и т. д.;
- научно-техническая информация – документальная, библиографическая, реферативная, справочная информация в области естественных, технических, общественных наук, по отраслям производства и сферам человеческой деятельности;
- доступ к первоисточникам – организация доступа к источникам информации через библиотеки и специальные службы, возможности приобретения первоисточников, их получения по межбиблиотечному абонементу в различных формах.

Потребительская информация. Сектор потребительской информации включает разделы:

- новости и литература – информация служб новостей и агентств прессы, электронные журналы, справочники, энциклопедии;
- обслуживающая потребительская информация – расписания транспорта, резервирование билетов и мест в гостиницах, заказ товаров и услуг, банковские услуги и т. п.;
- развлекательная информация – игры, телетекст, видеотекст.

Услуги образования. Данный сектор включает все формы и ступени образования:

- дошкольное, школьное,
- специальное,
- среднее профессиональное,
- высшее,
- повышение квалификации и
- переподготовку.

Информационная продукция может быть представлена в компьютерном или некомпьютерном виде: учебники, методические разработки, практикумы, развивающие компьютерные игры, компьютерные обучающие и контролирующие системы, методики обучения и пр.

Обеспечивающие информационные системы и средства. Сектор обеспечивающих информационных систем содержит части:

- программные продукты – программные комплексы с разной ориентацией – от профессионала до неопытного пользователя компьютера: системное программное обеспечение, программы общей ориентации, прикладное программное обеспечение по реализации функций в конкретной области принадлежности, по решению задач типовыми математическими методами и др.;

- технические средства – компьютеры, телекоммуникационное оборудование, оргтехника, сопутствующие материалы и комплектующие;

- разработка и сопровождение информационных систем и технологий – обследование организации в целях выявления информационных потоков, разработка концептуальных информационных моделей, разработка структуры программного комплекса, создание и сопровождение баз данных;

- консультирование по различным аспектам информационной индустрии, – какую информационную технику приобретать, какое программное обеспечение необходимо для реализации профессиональной деятельности, какая нужна информационная система, на базе какой информационной технологии лучше организовать свою деятельность;

- подготовка источников информации – создание баз данных по заданной теме, области, явлению и т.п.

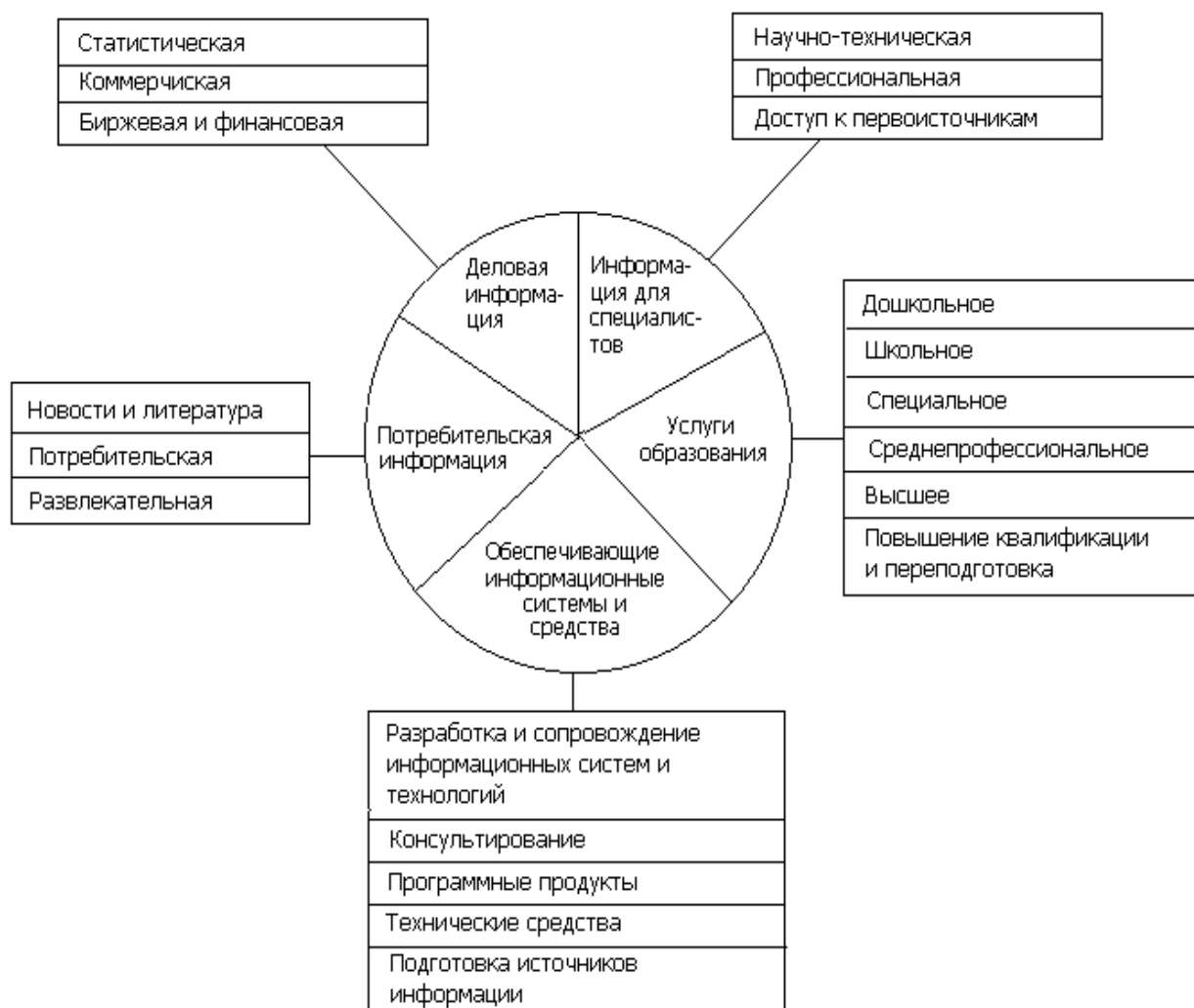


Рисунок 5.1 – Секторы информационного рынка

Информационный рынок, несмотря на разные концепции и мнения относительно его инфраструктуры, существует и развивается, а значит, можно говорить о бизнесе информационных продуктов, услуг, под которым понимается не только торговля и посредничество, но и производство.

5.3. Профессиональные базы данных и ресурсы Internet

Производители информационных ресурсов. Среди фирм, занимающихся производством и торговлей информационными продуктами, могут быть названы:

- Lexis-Nexis;
- Bell & Howell Information and Learning; Dow Johns и
- др.

Например, основные характеристики фирмы Lexis-Nexis таковы:

- объем базы данных около 23 Тбайт – это более одного миллиарда документов;
- 18 600 информационных источников. Свыше 13 600 полнотекстовых источников информации, включая крупнейшие журналы и газеты (New York Times, Washington Post, Newsweek, Financial Times и др.);
- более 140 информационных агентств (CNN, Agence France Presse, Associated Press, Bloomberg, Business Wire, ITAR-TASS, UPI, Economist Intelligence Unit, Sports Wires);
- данные о десятках миллионов компаний во всем мире – базы Dun&Bradstreet, Standard&Poor, Disclosure, Hoppenstedt, Credireform и др.;
- финансовые отчеты компаний и информация о рынках во всем мире – Investext, Market Research Reports, Predicast, ABI/Inform, ICC, UMI, Country Forecasts, Market Insight и др.;
- законодательство, судебные дела, договоры, юридическая информация по различным странам и регионам (США, страны ЕС, Канада, Китай, Россия и пр.).

Распределение баз данных по тематике:

- 10% — потребительский рынок;
- 12% — законодательство;
- 19% — наука и техника;
- 26% — другие;
- 33% — бизнес.

Фирмы UMI и Bell & Howell Information and Learning. Более 50 лет UMI (University of Microfilm International) обеспечивает библиотеки многих стран мира материалами. Применяя новейшие информационные технологии, она удовлетворяет информационные запросы различных организаций, предлагая широкий спектр традиционных и электронных продуктов и услуг:

- реферативные и полнотекстовые базы данных на CD-ROM;
- диссертации. UMI дает возможность обеспечить тезисами и диссертациями, выносившимися на обсуждение практически во всех университетах Северной Америки. Любую из диссертаций можно получить, приобретя ксерокопию, микроафишу или фильм;
- сериальные издания на микрофильмах. UMI начала использовать микрофильмы в 1938

году, собрав с тех пор 18 000 названий периодических изданий и 7000 газет, включая многие такие, которые недоступны из других источников;

- репринтные издания;
- собрания исследований.

Реферативные и полнотекстовые базы данных на CD-ROM включают в себя:

- базы данных, содержащие указатели и рефераты статей;
- ABI/INFORM Ondisc — основные периодические издания по бизнесу и проблемам управления;
- INSPEC Ondisc — реферативная база данных по физике, электронике и компьютерным технологиям;
- Periodical Abstracts Ondisc — охватывает основные реферативные издания по всем направлениям, включая социальные и гуманитарные науки, бизнес, общеобразовательные дисциплины, спорт;

- полнотекстовые базы данных содержащие не только указатели и рефераты статей, но и полный их текст:

- Businesslink – полнотекстовая база данных основных 100 журналов по бизнесу и промышленности;

- Newspapers – тексты наиболее популярных американских газет;

- базы данных, содержащие графику, включают указатели и рефераты статей, а также полное изображение журнальной статьи вместе с иллюстративными материалами:

- Business Periodicals Ondisk – обширный спектр журналов по бизнесу и проблемам управления, содержащие точные копии журнальных статей;

- Social Sciences Ondisk – электронные копии статей из 200 наименований, по разделу «Социальные науки»;

- General Periodicals Ondisk – электронные копии большого числа реферативных журналов по 5 основным направлениям: социальные и гуманитарные науки, бизнес и общеобразовательные дисциплины.

UMI располагает коллекцией из более 500 собраний исследований, каждое из которых содержит исторические справочные материалы о знаменитых ученых и событиях почти по каждому направлению академических исследований. UMI обеспечивает доставку материалов. Этим занимается отдел компании UMI InfoStore. Коллекция периодических изданий UMI InfoStore охватывает все направления науки, техники, бизнеса и содержит большое количество журналов другой тематики. Любой документ из этого хранилища может быть извлечен за 24 часа и доставлен. К настоящему моменту UMI объединилась с информационной компанией Bell & Howell.

Сетевые информационно-поисковые службы. Безусловно, для обеспечения полного поиска необходимо знать степень охвата информационных ресурсов Internet поисковыми системами. Сегодня ведущими по охвату информационных ресурсов Internet являются поисковые системы Google и Alltheweb. Вместе с тем, даже эти системы охватывают всего лишь третью часть существующих Web-страниц. Количество поисковых серверов, охватывающих Internet, а не отдельные его части, ограничено несколькими десятками. Лидерами здесь являются такие поисковые машины, как:

- <http://www.google.com>
- <http://search.yahoo.com>
- <http://www.ask.com>
- <http://www.alltheweb.com>
- <http://www.altavista.com>
- <http://www.lycos.com>

Среди российских поисковых серверов особого внимания заслуживают три – это Yandex (<http://www.yandex.ru>), Рамблер (<http://www.rambler.ru>) и Апорт (<http://www.aport.ru>). В Украине две лидирующие поисковые системы: МЕТА (<http://meta.ua>) – по стабильной части украинского сегмента Сети, и UAпорт (<http://uaport.net>) – по новостной части.

Google. В январе 1996 года будущие основатели Google, студенты Сергей Брин и Лари Пейдж, начали совместную работу над поисковой системой под названием BackRub. В сентябре 1998 года ими была основана компания Google. К 2000 году служба Google заняла лидирующее положение на рынке сетевых поисковых систем; трафик к ней непрерывно растет в течение шести лет. В 2002 году Google на короткое время отдала первенство по объему поискового индекса системе Alltheweb, но в настоящее время вновь заняла устойчивое первое место, охватывая свыше 4 млрд документов, и осуществляет более 200 млн поисковых операций в день. Поисковая машина Google позволяет искать как без учетов специфики алфавитов и языков, так и с учетом особенностей свыше 97 языков.

Компания является лидером поискового рынка во всем мире. В США ее предпочитают 34,7% пользователей, тогда как в мире доля Google на рынке англоязычного поиска достигает 43,3%. Большинство пользователей службы находятся за пределами США. Са-

мым близким преследователем Google является компания Yahoo!, до недавнего времени также применявшая поисковую технологию Google, но в начале 2004 года сменившая ее на собственную систему.

Поисковые устройства этой компании используют в своей работе армия США, администрация калифорнийского города Сан-Диего, фармацевтический гигант Pfizer, корпорации Boeing, Procter & Gamble, Cisco Systems и др. Поисковый механизм комплекса обеспечивает работу более чем с двумястами типами файлов (естественно, включая .html, .pdf, .doc). При этом осуществляется учет синонимов при полнотекстовом поиске по запросам и возможна работа более чем с пятьюдесятью естественными языками. Google Search Appliance поддерживают функции поиска защищенной информации, находящейся на закрытых серверах. При этом пользователь может обратиться к защищенному документу лишь при наличии у него соответствующих полномочий доступа. //Далее **Yahoo! Search, Alltheweb, AltaVista, Lycos** и службы поиска в Российском сегменте: **Yandex, Rambler, Апорт.**

Программные средства для поиска информации

RetrievalWare. Информационно-поисковая система RetrievalWare (www.convera.com) представляет собой средство полнотекстового и атрибутивного поиска. К документам, с которыми способна работать система RetrievalWare, относятся тексты в различных форматах и кодировках, электронные таблицы, базы данных, почтовые сообщения и т.п. – всего более двухсот форматов. Система обладает дополнительным инструментарием, позволяющим настроиться на поддержку документов специфических форматов. Объем архива при необходимости может измеряться терабайтами.

Архитектура системы RetrievalWare позволяет ей работать как через локальную корпоративную сеть, так и через Internet. Серверная часть системы поддерживает все распространенные серверные платформы, а клиентским местом может быть любой компьютер, имеющий графический Web-браузер. Система обладает возможностью работы в различных многопроцессорных и распределенных многосерверных конфигурациях. Попытки анализа больших объемов неструктурированных или слабо структурированных данных очень часто усложняют процесс принятия решений. Если широкий спектр поисковых систем достаточно легко справляется с «простым» полнотекстовым поиском, то для подобного анализа нужны технологии совсем другого типа, представленные системами извлечения знаний (Knowledge Mining). Стоимость внедрения таких систем составляет сотни тысяч долларов.

Итак, основная задача – выявление знаний в массивах неструктурированных данных с целью их использования в процессе принятия решений. Чтобы добиться этого, необходимо сделать информацию доступной для анализа, выявить классы понятий и сопоставить их с документами.

Как правило, информационные массивы преобразуются такими системами в хранилища данных (Data Warehouse) или корпоративные порталы знаний – интегрированные информационные репозитории, доступные для оперативного обобщения и анализа. Часто такие хранилища являются самообучаемыми за счет использования статистических байесовских алгоритмов. Последние обеспечивают адаптацию критериев группирования документов. Большую роль играют и «отклики» реальных пользователей.

За счет предварительной обработки информации, проводимой на этапе формирования хранилищ данных, значительно повышается эффективность таких процессов, как интеллектуальный анализ данных, глубинный анализ текстов и обнаружение новых знаний в текстах. Как неожиданную производную этих процессов можно назвать появление средств, упрощающих поиск для пользователя, таких как реализация *нечеткой логики запросов* (нечеткого поиска), средств построения функциональных информационных портретов, визуализации семантических связей и т.д. В свою очередь, эти возможности напрямую связаны с распознаванием образов, поиском мультимедийных данных, анализом речевого ввода.

Yandex.Server. Yandex.Server Standard 3.2 (<http://company.yandex.ru/technology/products/yandex-server.xml>) представляет собой системный сервис для организации полного текстового поиска информации в заданной коллекции документов. Он предназначен для работы с текстами как в локальной, так и в глобальной сети. Система не содержит лицензионных ограничений на число индексируемых документов, их размер или суммарный размер индекса, позволяет индексировать документы как через HTTP-соединение, так и чтением локальной файловой системы. Yandex.Server Standard представляет результаты поиска во встроенном дизайне. Yandex.Server 3.0 состоит из двух основных логических частей: индексатора и поискового сервера. Индексатор анализирует документы, среди которых должен проводиться поиск, и сохраняет информацию о них в специальных индексных файлах. Обычно используется режим работы, при котором не создаются заново индексные файлы, а отрабатывается информация только по изменившимся, новым и удаленным документам. Поисковый сервер после запуска находится в постоянном ожидании запросов, которые могут быть представлены на естественном языке. Поиск может осуществляться с учетом морфологии языка, в одной или нескольких коллекциях документов. Yandex.Server 3.2 поддерживает форматы .html, .xml, .rtf, .pdf, .doc, .mp3 и многие другие. Содержимое индексируемых документов также может быть получено при обращении к реляционной базе данных, в частности MySQL и MS SQL Server. Система предоставляет возможность кластеризации результатов поиска (группирует найденные документы в соответствии с внешними атрибутами), а также ранжирует результаты (сортирует документы по степени соответствия запросу).

Порталы знаний. По данным недавно проведенного исследования, сотрудники компаний могут тратить до трех часов в день на поиски информации, которые зачастую оказываются безрезультатными, вследствие чего тысяча крупнейших фирм США ежегодно теряет 2,5 млрд долларов.

Именно для решения этой проблемы созданы и продолжают создаваться корпоративные поисковые системы и порталы знаний как среды для эффективного поиска знаний и обмена ими, инструменты, которые представляют собой совокупность технологических решений для выявления, хранения, классификации, обработки и распространения знаний. В настоящее время широко используется система IBM Lotus Discovery Server – программный продукт, предназначенный для управления знаниями в корпоративных порталах, для нахождения экспертов, идентификации связей и общего управления интеллектуальным капиталом. Lotus Discovery Server является логическим продолжением ранее популярного программного продукта Lotus Raven – системы построения корпоративных порталов знаний. Благодаря возможности анализа информации, хранящейся в организации, Lotus Discovery Server в состоянии указывать области экспертных знаний и подразумеваемые знания сотрудников, находя и организуя динамические связи между информацией, людьми и их деятельностью.

Современные порталы знаний обеспечивают решение целого комплекса задач, среди которых – сбор информации об объектах, определение связи между объектами, выявление тенденций. Функциональные возможности таких систем позволят проводить многофакторные динамические исследования, выполнять диагностику и прогнозирование развития ситуации. В дополнение к возможностям глубинного анализа данных и текста, в порталах знаний широко используется человеческий опыт – знания экспертов в процессах выявления, сохранения и эффективного использования знаний.

Около пяти лет назад по заказу группы аналитиков Гарвардского университета российские разработчики из «Инфорус» создали систему Avalanche, которая в процессе поиска формирует модель предметной области в виде набора «умных папок», каждая из которых знает, что в нее должно попадать. Наполнением папок занимается специализированный робот, который запускается с компьютера «хозяина» и приносит только то, что у него просили. Это одно из первых эффективных решений на базе современной технологии глубинного анализа текстов.

Очень близка по идеологии и технология компании Vivisimo, в рамках которой результаты Internet-поиска распределяются по папкам-категориям, которые система создает автоматически. Достигается это за счет лексического сопоставления запросов и результатов поиска. Естественно, свое применение Vivisimo сразу же нашла в корпоративных сетях и порталах знаний.

Специализированные ресурсы поиска скрытого контента. "Скрытый" Web представляет собой гигантский репозиторий документов, звуков, изображений, фильмов и т.п. Безусловно, если большая часть этой информации не доступна традиционным поисковым системам, то существует потребность в специальных инструментах поиска «скрытого» контента. Эти инструменты включают каталоги, метапоисковые сайты, доступные через Web базы данных, а также большое количество глобальных, региональных и специальных поисковых систем.

Для поиска в скрытой Сети, а именно в том ее сегменте, который составляют базы данных, сегодня уже существуют некоторые специализированные ресурсы. Среди них, например, системы BigHub (www.bighub.com) и Invisible Web (www.invisible-web.net) компании IntelliSeek.

Сайт Invisible Web включает в себя каталог баз данных, большинство из которых не проиндексированы известными поисковыми машинами. При введении запроса этот сайт выдает ссылки на ресурсы, с помощью которых поиск необходимой информации станет наиболее оптимальным. На этом сайте собраны коллекции ссылок на различные базы данных, среди которых содержится немало уникальных ресурсов. Программный пакет BullsEye компании IntelliSeek осуществляет поиск более чем в 800 сетевых ресурсах. Лидером среди навигаторов в скрытом Web является сайт CompletePlanet (www.completeplanet.com) компании BrightPlanet. Этот сайт является крупнейшим каталогом, насчитывающим свыше 100 тыс. ссылок. Компания BrightPlanet также создала персональную утилиту для поиска в он-лайн-базах данных – LexiBot, которая может обеспечивать поиск в нескольких тысячах поисковых систем скрытого Web. Метапоисковый пакет DeepQueryManager (DQM) этой же компании обеспечивает поиск по 55 тыс. скрытых Web-ресурсов.

Сайт Direct Search (<http://www.freepoint.com/gary/direct.htm>), созданный Гари Прайсом, также обеспечивает поиск в базах данных скрытого Web. На сайте содержится ссылка на лучшие ресурсы ценовой информации (MySimon.com), финансовой информации (FinancialFind.com), а также ссылки на информацию из научно-популярных журналов и научных баз данных по биотехнологиям (Biolinks.com). В Internet есть и другие сайты-навигаторы, а также специализированные программы поиска.

Особенность большинства скрытых ресурсов – в их узкой специализации. Для поиска в них используются те же механизмы, что и для поверхностного Web, однако чаще всего роботы поисковых систем для скрытого Web включают уникальные для каждого такого ресурса модули доступа к данным.

Системы поиска в скрытом Web. Многие поисковые системы, как глобальные, так и локальные, описаны на сайтах Search Engine Watch (<http://www.searchenginewatch.com>) и Search Engine Showdown (<http://www.searchengineshowdown.com>). На этих сайтах приведены, среди прочих, и поисковые системы скрытого Web.

- Singingfish (<http://www.singingfish.com>) – эта поисковая система обеспечивает поиск аудио- и видеофайлов, представленных на Web-сайтах.

- Scirus (<http://www.scirus.com>) – поисковая система по представленным в Internet научным материалам, включая статьи из журналов и отчеты. Со страницы расширенного поиска (Advanced Search) доступны многочисленные тексты из баз данных EBSCO и ProQuest.

- UFOSeek (<http://www.ufoseek.com>) – поисковая система по материалам о паранормальных явлениях и НЛО.

Качественный и полноценный поиск информации в скрытом Web возможен и с использованием таких специализированных коммерческих баз данных, как Dialog, ProQuest, Web of Science. Но эти базы данных, ввиду своей платности, сами являются объектами скрытого Web.

Базы данных «скрытой» Сети. Пожалуй, самыми большими из известных ресурсов «скрытого» Web являются базы данных служб Dialog и LexisNexis. Одной из крупнейших мировых служб информационного поиска является американская компания Dialog (<http://www.dialog.com>), созданная при поддержке NASA и до 1988 года принадлежавшая аэрокосмической фирме Lockheed. Сегодня Dialog принадлежит корпорации Thomson (США) – одному из всемирных лидеров в области предоставления интегрированных информационных решений. Корпорация Thomson имеет свыше 20 миллионов пользователей в 130 странах мира. Образованная в 1965 году как первая в мире он-лайновая информационно-поисковая служба, Dialog фактически определила современные стандарты управления информацией. На сегодняшний день она включает такие продукты и сервисы, как Dialog®, Dialog Profound®, Dialog DataStar®, Dialog NewsEdge® и Dialog Intelliscope, которые обеспечивают доступ к более 1,4 млрд. документов через Internet или сети intranet. При этом в компании Dialog определяют свои ресурсы как часть «скрытого» Web (Deep Web), заявляя, что содержат полезной, не дублирующейся информации в 500 (!) раз больше, чем доступно с помощью традиционных информационно-поисковых систем. Коллекция баз данных службы Dialog содержит 900 баз данных, доступных 700 000 пользователям, которые только за один час прочитывают свыше 17 млн. документов из этих баз данных.

Основанная в 1973 году, крупнейшая в мире он-лайновая служба LexisNexis (<http://www.lexisnexis.com>) предоставляет своим пользователям юридическую, политическую, коммерческую, новостную, регистрационную и другую информацию. С 1979 года система баз данных LexisNexis – первая в мире служба полнотекстового поиска. В настоящее время эта служба охватывает свыше 35 000 источников информации, содержащих в совокупности более 4,6 млрд документов с глубиной ретроспективы до 200 лет. Каждый час в базы данных LexisNexis добавляется 57 500 документов. LexisNexis представлена сегодня в 20 странах, пользователи сервиса находятся в более чем 100 странах. К коммерческим базам данных «скрытого» Web можно отнести и информационные ресурсы крупнейших мировых информационных агентств, уже много лет работающих на рынке финансовой информации, таких как Reuters, Tenfore, Dow Jones Telerate, Bloomberg. С другой стороны, в «скрытом» Web существует множество альтернатив коммерческим базам данных. Среди них, например, сайт www.10kwizard.com, предлагающий доступ к полным текстам корпоративных документов, хранящихся в Комиссии США по ценным бумагам и биржам. Существуют тысячи баз данных «скрытого» Web, свободно доступные для пользователей, но чаще всего не охватываемые традиционными поисковыми системами.

Приведем еще несколько примеров.

- Educator's Reference Desk (<http://www.askeric.org>) – этот ресурс содержит свыше двух тысяч учебных планов, несколько тысяч ссылок на образовательные документы, а также ссылки, представляющие собой запросы к архиву. С этого сайта обеспечивается доступ к базе данных ERIC – крупнейшему источнику информации по проблемам образования, а также к полнотекстовым дайджестам, составляемым экспертами.

- Nuclear Explosions Database (<http://www.ga.gov.au>) – австралийская база данных по географии. Для работы с системой достаточно перейти в режим «Online Tools», после чего будет представлен список баз данных и карт.

- PubMed (<http://www.ncbi.nlm.nih.gov>) – с ресурса обеспечивается доступ к свыше 14 млн ссылок системы MEDLINE, включая ссылки на полные тексты статей и информационные ресурсы. Имеется возможность перехода к службе PubMed Central (PMC), к свободно доступному архиву статей (свыше 90 тысяч) из научных журналов. Обеспечивается

также доступ к глобальной поисковой системе NCBI, охватывающей базы данных по естествознанию.

- LookSmart's FindArticles (<http://www.findarticles.com>) – база данных FindArticles – доступный через Web-интерфейс архив, содержащий 2,8 млн. статей из более чем 500 источников, накапливаемый с 1998 года.

Скрытый Web в каталогах. Каталоги, как глобальные, так и специальные, могут содержать ссылки на скрытые ресурсы, прежде всего базы данных. Например.

- Портал WebData.com на первый взгляд ничем не отличается от других подобных ресурсов, однако содержит гиперссылку «Add Your Database» (добавить Вашу базу данных), говорящую о том, что на данном портале можно зарегистрировать базу данных – часть скрытого Web.

- Librarians Index to the Internet (<http://lii.org>) – каталог, содержащий свыше 14 000 Internet-ресурсов. ЛII также включает ссылки на скрытые в Web-пространстве базы данных. У владельцев таких баз данных есть возможность поместить соответствующую гиперссылку в этом каталоге на свой ресурс (в ЛII есть ссылка «and databases» (добавить базу данных)).

- FindLaw (<http://www.findlaw.com>) – один из наиболее популярных в мире юридических Web-сайтов, представляющий собой огромный каталог правовых ресурсов, содержащий аннотированный список свободно доступных баз данных нормативно-правовых документов, для которых данный ресурс является «точкой входа».

- InfoMine (<http://infomine.ucr.edu>) – ресурс, содержащий ссылки на 120 000 документов, представленных в 9 аннотированных базах данных. Этот каталог позиционирует себя как «виртуальную библиотеку Internet-ресурсов», ориентированную на студентов и исследователей-профессионалов.

- About.com (<http://www.about.com>) – портал, охватывающий тысячи снабженных комментариями ссылок на Web-ресурсы, в том числе и на ресурсы скрытого Web (имеется ссылка Invisible Web). На портале предоставляется возможность поиска в каталоге. Ресурс также включает несколько статей по проблематике «невидимого» Web.

К разряду каталогов можно также отнести следующие коллекции ссылок, поисковые системы и «скрытые» базы данных.

- Direct Search (<http://www.freepint.com/gary/direct.htm>) – ресурс, содержащий ссылки на ресурсы скрытого Web. Например, присутствует ссылка на сайт ResourceShelf (<http://www.resourceshelf.com>), обеспечивающий поиск в блогах (сетевых журналах) и новостных сообщениях.

- The Invisible Web Directory (<http://www.invisible-web.net>) – Web-сайт Шермана и Прайса (Chris Sherman & Gary Price), соавторов термина Invisible Web.

- CompletePlanet (<http://www.completeplanet.com>) – сайт корпорации BrightPlanet Corporation, который охватывает свыше 70 000 поисковых баз данных и специальных поисковых систем.

Контрольные вопросы к лекции 5.

1. Какова основная характеристика мирового рынка информационных услуг? Что она отражает?
2. Назовите ведущие секторы рынка информационных продуктов и услуг
3. Для чего нужны профессиональные базы данных и ресурсы Internet? Каковы их основные функции?

Лекция 6. ГОСУДАРСТВЕННЫЕ ИНФОРМАЦИОННЫЕ РЕСУРСЫ

Вопросы:

- 6.1. Общая характеристика государственных информационных ресурсов
- 6.2. Органы и организации, ответственные за формирование и использование государственных информационных ресурсов
- 6.3. Основные правовые акты, определяющие порядок формирования и использования информационных ресурсов
- 6.4. Информационные ресурсы министерств и ведомств
- 6.5. Информационные ресурсы региональных и муниципальных органов власти
- 6.6. Статистическая информация
- 6.7. Научная и техническая информация
- 6.8. Информационные ресурсы библиотечной сети России
- 6.9. Информационные ресурсы социальной сферы
- 6.10. Информация о природных ресурсах, явлениях, процессах

6.1. Общая характеристика государственных информационных ресурсов

Понятие «информационные ресурсы» определено в Законе «Об информации, информатизации и защите информации». В соответствии с ним к государственным информационным ресурсам относятся находящиеся в собственности Российской Федерации и субъектов Российской Федерации отдельные документы и отдельные массивы документов, документы и массивы документов в информационных системах (библиотеках, архивах, фондах, банках данных, других информационных системах), созданные, приобретенные, накопленные за счет средств федерального бюджета, бюджетов субъектов Российской Федерации. Согласно переписи, проведенной Госкомстатом России в 1996 г., в ведении только 30 тыс. российских предприятий находится около 800 тыс. баз данных общей стоимостью не менее 320 млрд. долл. Не менее 30% национальных информационных ресурсов России являются государственными, т. е. созданными за счет средств налогоплательщиков в процессе деятельности государственных органов власти и управления всех уровней, а также при выполнении предприятиями всех форм собственности государственных заказов.

6.2. Органы и организации, ответственные за формирование и использование государственных информационных ресурсов

Функции контрольного и нормативно-методического характера на общероссийском уровне по отношению ко всем государственным информационным ресурсам осуществляют:

- Министерство Российской Федерации по связи и информатизации – контроль за созданием информационных ресурсов в органах и организациях, их регистрацией, доступностью и порядком использования, а также контроль систем навигации и общая координация работ по формированию и ведению государственных информационных ресурсов.
- Федеральное агентство правительственной связи и информации при Президенте Российской Федерации и Государственная техническая комиссия при Президенте Российской Федерации – контроль за защитой государственных информационных ресурсов от незаконного использования и разрушения.
- Министерство имущественных отношений Российской Федерации – учет государственных информационных ресурсов как имущества, порядка их закрепления в оперативном управлении и хозяйственном ведении.
- Российское агентство по патентам и товарным знакам – учет информационных ресурсов как интеллектуальной собственности.

- Министерство финансов Российской Федерации – порядок финансирования и финансовой отчетности деятельности по формированию и применению информационных ресурсов с использованием бюджетных средств, а также оказания платных услуг на основе государственных информационных ресурсов.

В 1999 г. межведомственной группой специалистов под руководством Госкомсвязи России, правопреемником которого является Минсвязи России, подготовлен доклад «Информационные ресурсы России», в котором подробно рассмотрены состояние, тенденции развития государственных информационных ресурсов и выделены основные категории информационных ресурсов России.

6.3. Основные правовые акты, определяющие порядок формирования и использования информационных ресурсов

Одним из важных направлений государственной политики по созданию и эффективному использованию информационных ресурсов является организация работ под руководством Минсвязи России по возможно более полному и достоверному учету этих ресурсов и публикации сведений о них. В результате в России создана система государственной регистрации электронных информационных ресурсов. Рассмотрим основные положения, определяющие порядок функционирования этой системы. Нормативно-правовые основы государственной регистрации электронных ресурсов определены рядом федеральных законов и постановлений Правительства РФ, в том числе:

- Федеральным законом «Об информации, информатизации и защите информации»;
- Федеральным законом «Об обязательном экземпляре документов»;
- Постановлением Правительства РФ № 226 от 28 февраля 1996 г. «Целями организации государственной регистрации электронных информационных ресурсов (баз данных, электронных изданий и т.п.)» в соответствии с упомянутым выше Постановлением Правительства РФ являются:
 - информирование граждан и организаций о содержащихся в базах и банках данных информационных ресурсах, а также о порядке доступа к ним;
 - организация информационного обеспечения органов государственной власти РФ.

Государственной регистрации подлежат базы и банки данных любой тематической направленности, структуры и назначения. Регистрация является бесплатной и осуществляется в двух формах – обязательной и добровольной. Обязательная форма распространяется на государственные (созданные, приобретенные или накапливаемые за счет или с привлечением средств федерального бюджета) базы и банки данных. Добровольная форма регистрации применяется для электронных информационных ресурсов, созданных или накапливаемых негосударственными организациями. Работы по учету и регистрации баз и банков данных и электронных изданий проводятся подведомственной Минсвязи России организацией – Научно-техническим центром (НТЦ) «Информрегистр». В соответствии с возложенными на него задачами НТЦ «Информрегистр» ведет Государственный регистр баз и банков данных и Государственный депозитарий электронных изданий. Материалы государственной регистрации электронных информационных ресурсов публикуются в следующих печатных изданиях НТЦ «Информрегистр»:

- Каталог «Базы данных России», имеющий статус официального издания Государственного регистра баз и банков данных. Содержит информацию (5000 записей) о российских базах данных по всем отраслям науки, техники и производства и об их владельцах. Снабжен предметным указателем. Поиск может осуществляться по ключевым словам, по рубрикам базы данных, а также по сочетанию этих характеристик.
- Каталог «Российские электронные издания». Содержит информацию об электронных изданиях, включенных в Государственный депозитарий, в том числе библиографическое описание, аннотацию, системные требования и адресные данные организации разработчика, номер государственной регистрации. Электронные издания упорядочены по раз-

делам. Внутри разделов электронные издания расположены в алфавитном порядке заглавий. Здесь так же, как и в предыдущем случае, наименование организации представляет собой ссылку на дополнительную информацию о ней. Электронные версии этих каталогов доступны в Интернет на сервере НТЦ «Информрегистр» www.inforeg.org.ru.

Сервер содержит еще целый ряд полезных сведений по информационным продуктам, разработанным НТЦ «Информрегистр». Это тексты законодательных и организационно-методических документов, а также тексты отчетов и аналитических справок о результатах учета и государственной регистрации электронных информационных ресурсов.

6.4. Информационные ресурсы министерств и ведомств

Представительство Минфина России в Интернет (www.minfin.ru) открывается страницей со списком тематических разделов, по которым сгруппирована имеющаяся информация. Здесь представлены утвержденные федеральные бюджеты РФ, отчетные материалы по исполнению федерального бюджета, материалы о выпуске ценных бумаг и внешних облигационных займах России, материалы по финансовым взаимоотношениям центра и регионов. С этой страницы по ссылке «Нормативные и методические материалы Министерства финансов» можно перейти на страницу поиска этих материалов в базе данных системы «Гарант». По ссылке «Официальная информация» можно выйти на страницу с перечнем официальных информационных материалов министерства и, указав название документа, просмотреть его текст.

Представительство Банка России расположено в Интернет по адресу www.cbr.ru. На страницах этого сервера размещена как общая информация о Банке России (исторический очерк, правовой статус, функции, организационная структура и др.), так и информация о его текущей деятельности. В частности, в разделе «Ежедневная информация» можно познакомиться с последними новостями. В разделе «Денежная политика» содержатся сведения о динамике изменения денежной массы, золотовалютных резервах, ставках рефинансирования и процентных ставках. В разделе «Банковская система» приводится обзор банковского законодательства и материалы о лицензировании кредитных операций. Раздел «Финансовые рынки» содержит сведения о межбанковском кредитном рынке, валютном рынке (включая курсы валют) и др. Дополнительную официальную финансово-экономическую информацию можно получить в разделах «Информационные материалы» и «Публикации и доклады». На этих страницах можно просмотреть тексты документов, а также электронные версии выпусков «Бюллетень банковской статистики». Справочно-информационный сервер Государственного таможенного комитета Российской Федерации расположен в Интернет по адресу www.gtk.ru. Информация на сервере предназначена для различных категорий участников внешнеэкономической деятельности и сгруппирована по тематическим разделам. В разделе «Новости», например, публикуются таможенные новости, новости таможенного законодательства и другие материалы, представляющие интерес для участников внешнеэкономической деятельности. На этом сервере можно просмотреть иерархический список таможенных органов России, получить доступ к базам данных адресных сведений участников внешнеэкономической деятельности, получить целый ряд справочных данных по таможенному законодательству, котировкам валют, денежным единицам различных стран мира и др.

По ссылкам, приведенным на этом сервере, можно попасть на сайты подведомственных ГТК России организаций. Например, по ссылке «ГНИВЦ ГТК России» можно попасть на начальную страницу сервера Главного научно-информационного вычислительного центра ГТК России. В состав доступной здесь информации входят, в частности, различные справочники, например по товарной номенклатуре – «Разделы ТН ВЭД», «Группы ТН ВЭД», «Товарные позиции ТН ВЭД» и целый ряд других.

6.5. Информационные ресурсы региональных и муниципальных органов власти

В органах государственной власти субъектов Российской Федерации и муниципальных органах за последнее время создано большое количество различных собственных ин-

формационных ресурсов в виде массивов документов, баз данных и информационных массивов, которые можно разделить на следующие категории.

- Информация о земельных ресурсах (земельные кадастры и др.).
- Информация об объектах недвижимости (здания и сооружения, инженерные сети, транспортные сети и др.).
- Информация о юридических лицах (предприятия, организации и другие физические лица).
- Информация о физических лицах (основные паспортные данные жителей региона).
- Документы научно-технической информации.
- Нормативно-правовые документы.
- Социально-экономические и финансовые показатели административно-территориальных единиц и хозяйствующих субъектов.

Свои представительства в Интернет имеет достаточно большое число региональных органов власти. Виды информационных ресурсов органов, к которым можно получить доступ в Интернет – это новости, акты, общие сведения, транспорт, сельское хозяйство, промышленность, инвестиции, культура, структура власти и др. Подобный перечень информации характерен для большинства серверов, на которых представлены региональные органы власти. Списки адресов Интернет-представительств федеральных и региональных органов власти России можно найти, например, в каталогах ресурсов Интернет.

6.6. Статистическая информация

Структура Государственной системы статистики включает районный, областной и федеральный уровни (дополнительно выделяются Москва и Санкт-Петербург), объединяя 89 территориальных комитетов и 2,2 тыс. районных отделов. На начальной странице сайте Госкомстата России (www.gks.ru) расположена ссылка на каталог изданий Госкомстата (строка «Каталог» в правой части страницы). Перейдя по этой ссылке, можно получить на экране перечни и краткие характеристики периодических изданий, статистических сборников и статистических бюллетеней Госкомстата. На этой же странице расположена ссылка «Организационная структура Госкомстата России», по которой можно получить информацию о центральном аппарате Госкомстата, подведомственных ему организациях и его территориальных отделениях. Например, указав интересующий регион на карте, можно получить список территориальных комитетов, расположенных в этом регионе. Указав на наименование территориального комитета, – получить на экране страницу с его координатами. Для территориальных комитетов, имеющих свои представительства в Интернет, приводятся ссылки на соответствующие адреса. Поэтому можно сразу перейти к сайту определенного территориального комитета и ознакомиться с содержащейся на нем информацией.

С первой страницы представительства Госкомстата можно получить доступ к ряду общих статистических сведений Российской Федерации по ссылке «Основные экономические показатели России». За дополнительной информацией следует обращаться по ссылкам раздела «Распространение информации», расположенного в правом нижнем углу страницы. Здесь размещены ссылки на сайты Информационного центра Госкомстата России и Главного межрегионального центра обработки и распространения статистической информации (ГМЦ).

Информационный центр Госкомстата России является организацией, которая создана для удовлетворения запросов российских и зарубежных клиентов в статистической информации о положении в экономике России, внешнеторговой деятельности, населении, его занятости и уровне жизни. Информационному центру предоставлено эксклюзивное право на издание и распространение официальных статистических публикаций Госкомстата. На сервере Информационного центра предоставляются услуги по доступу к электронным версиям этих публикаций непосредственно в Интернет, а также по подписке на книжные издания и электронные издания на CD-ROM или дискетах. На начальной стра-

нице сервера Информационного центра Госкомстата России расположены текущая статистическая информация и ссылки на вновь поступившие документы. Список разделов сервера размещен в верхней части страницы. Чтобы получить доступ к электронным версиям статистических публикаций Госкомстата, нужно перейти по ссылке на раздел «Web-доступ».

Доступ к подавляющей части статистических публикаций является платным. С правилами оплаты можно ознакомиться на этом же сервере. Главный межрегиональный центр обработки и распространения информации Госкомстата России (ГМЦ) – головная организация в системе государственной статистики по сбору, обработке, анализу и распространению официальной статистической информации в России и зарубежных странах, обеспечивающая ведение и актуализацию государственных регистров и баз данных. В состав информационного фонда ГМЦ входит: общеэкономическая, отраслевая и региональная статистическая информация, объединенная в блоки.

- *Промышленность.* Динамика обобщающих стоимостных показателей развития промышленности и производства промышленной продукции в натуральном выражении. Сведения, характеризующие количество промышленных предприятий, объем выпуска промышленной продукции, численность занятых, стоимость основных производственных фондов, балансовую прибыль. Оперативная (ежемесячная, квартальная) информация по развитию промышленности России.

- *Сельское хозяйство.* Информация по развитию сельского хозяйства России, его отраслей по регионам страны. Сведения, характеризующие динамику обобщающих стоимостных показателей, размер и структуру посевных площадей, производство сельхозпродукции в натуральном выражении, численность поголовья скота и птицы.

- *Наука и инновации.* Основные показатели деятельности организаций, выполняющих научные исследования и разработки. Освоение, сертификация и снятие с производства промышленной продукции. Созданные впервые в России образцы новых типов машин, оборудования, аппаратов и приборов. Технологические инновации.

- *Уровень жизни и доходы населения.* Рынок труда, социальные вопросы, информация о переписи населения, ежемесячные данные о ценах (тарифах) и индексе цен.

- *Статистика внешнеэкономической деятельности.* Информация по сводным показателям хозяйственной деятельности совместных 1 предприятий, движению иностранной валюты, иностранным инвестициям в экономику, экспорт (импорт) товаров по России, странам, группам стран, государствам-участникам СНГ, товарам.

- *Паспорт территорий Российской Федерации.* Содержит информацию по основным показателям развития экономики регионов России. Включает сведения, характеризующие численность и состав населения, развитие секторов экономики, производство промышленной и сельскохозяйственной продукции, развитие инфраструктуры регионов, социальной сферы, структуру занятости и уровень безработицы.

- *Регистр городов России.* Содержит информацию по городам России; включает сведения, характеризующие численность, прирост, состав и занятость населения, уровень безработицы, развитие секторов экономики, финансовые показатели, производство промышленной продукции, развитие инфраструктуры, развитие социальной сферы, площадь городских земель.

- *Единый государственный регистр предприятий и организаций (ЕГРПО).* Включает информацию о 2 700 000 юридических лиц и их обособленных подразделениях, прошедших государственную регистрацию в соответствии с законодательством РФ. Содержит сведения о наименовании, местонахождении, виде деятельности, форме собственности, учредителях, их доле акций, государственной регистрации.

- *Бухгалтерская отчетность.* Информация по показателям, характеризующим финансово-хозяйственную деятельность юридических лиц и их обособленных подразделений. По каждому предприятию имеется информация в составе показателей годового и квартального бухгалтерского баланса и его приложений, в частности, информация об ос-

новных средствах, запасах и затратах, денежных средствах, источниках собственных средств, финансовых результатах деятельности, движении денежных средств дебиторской и кредиторской задолженностях и др.

- *Регистр промышленных предприятий*. Информация о 28 000 предприятий России. По каждому предприятию приводятся сведения об объеме продукции в стоимостном выражении, стоимости основного капитала, численности занятых, прибыли, затратах на рубль продукции, размерах оплаты труда, номенклатуре выпускаемой продукции, а также адрес, телефон, факс.

- *Регистр сельскохозяйственных предприятий*. Представляет информацию о развитии сельскохозяйственных предприятий России. По каждому предприятию приведена информация об объеме продукции в стоимостном выражении, стоимости основного капитала, численности занятых, финансовом положении, посевных площадях, выработке сельхозпродукции, а также адрес, телефон, факс.

Запрос на получение необходимой информации можно сформировать непосредственно на Интернет-сервере ГМЦ (www.gks.ra/gmc). О принятии запроса пользователь уведомляется в течение 3 часов с момента его отправки. В зависимости от вида запроса и типа запрашиваемой информации срок выполнения запроса может составлять от нескольких часов до нескольких дней.

6.7. Научная и техническая информация

К этой информации в настоящее время принято относить документированную информацию, возникающую в результате экономической, научной и научно-технической деятельности. Примерами подобной информации являются книги, периодические издания, патенты, описания изобретений, стандарты, технические условия, классификаторы, отчеты о научно-исследовательских и опытно-конструкторских работах, диссертации, промышленные каталоги, описания экономических и научно-технических достижений и т. п.

Государственная система научно-технической информации (ГСНТИ) представляет собой совокупность научно-технических библиотек и организаций – юридических лиц независимо от формы собственности и ведомственной принадлежности, специализирующихся на сборе и обработке научно-технической информации и взаимодействующих между собой с учетом принятых на себя системных обязательств.

К федеральным органам НТИ и научно-техническим библиотекам, обеспечивающим формирование, ведение и организацию использования федеральных информационных фондов, баз и банков данных по различным видам источников НТИ и направлениям науки и техники относятся более 30 российских организаций информационного профиля. В их информационных фондах накапливаются, обрабатываются и предоставляются пользователям десятки миллионов информационных источников в области науки и техники.

Одним из важных элементов ГСНТИ является объединение Росинформресурс Минпромнауки России. Оно было образовано в 1966 г. как общероссийская территориально распределенная специализированная информационная сеть. Региональные центры Росинформресурса (69 центров в различных регионах страны) осуществляют формирование, ведение и организацию использования региональных информационных фондов, баз и банков данных НТИ. Представительство Росинформресурса расположено в Internet по адресу www.rosinf.ru и содержит ссылки на представительства региональных центров. На этом сайте представлены, в частности, следующие базы данных:

- *Результаты научно-технической деятельности (РНТД)*. Информация о новых технических решениях в промышленности, строительстве, сельском хозяйстве, торговле и организации производства в России и СНГ. Объем 110 тыс. документов.

- *Технологии и наукоемкая продукция России (ТНЕПР)*. Полнотекстовая двуязычная база данных о разработках по приоритетным направлениям науки и техники в регионах России.

- *Рынок промышленной продукции* (РПП). Сведения о продукции, товарах и услугах 57 тыс. предприятий и организаций России, которые регулярно обновляются и пополняются.

- *Товары народного потребления* (ТНП). Информация о разработанной и выпускаемой в России так называемой непрофильной продукции предприятий. Около 35 тыс. наименований с систематическим обновлением.

Сводные базы данных НТП. Результат регулярной обработки машиночитаемых массивов ведущих российских информационных центров - ВИНТИ, ВНИИЦ, ВИМИ, ГПНТБ и др. На сайтах региональных центров НТИ Росинформресурса представлены сведения о региональных фондах НТИ. Например, на сайте Ярославского центра НТИ www.csti.yar.ru наряду с электронными версиями каталогов различных документов представлена база данных информационных листов, содержащих сведения по научно-техническим разработкам и передовому опыту предприятий Ярославского региона. Задавая те или иные условия поиска, можно получить на экране список названий информационных листов, соответствующих запросу. При этом каждое название представляет собой ссылку на содержание соответствующего информационного листа.

6.8. Информационные ресурсы библиотечной сети России

Библиотечная сеть России представляет собой объединение ряда ведомственных библиотечных сетей и насчитывает в настоящее время около 150 тыс. библиотек. Несмотря на ряд серьезных проблем развития библиотек, в первую очередь связанных с недостаточными объемами бюджетного финансирования, библиотечная сеть России продолжает функционировать.

К настоящему времени автоматизированные библиотечно-информационные технологии функционируют в более чем 2,5 тыс. научных и публичных библиотек, из которых 1250 – муниципального уровня. Почти во всех центральных универсальных научных библиотеках субъектов РФ созданы локальные вычислительные сети, объединяющие от 30 до 50 компьютеров. Основным результатом внедрения информационных технологий в практику работы библиотек является создание электронных каталогов и других библиографических и реферативных баз данных. Самые значительные по объему базы данных сформированы в библиотеке Института информации по общественным наукам, Российской национальной библиотеке, Российской государственной библиотеке, Государственной публичной научно-технической библиотеке России и других центральных библиотеках.

Основные направления развития библиотечных информационных ресурсов России и, в частности, задачи расширения их представления в сети Интернет, сформулированы в программе «Создание общероссийской информационно-библиотечной компьютерной сети – ЛИБНЕТ», реализуемой под эгидой Министерства культуры Российской Федерации. Свои представительства в сети Интернет имеет сегодня уже целый ряд центральных федеральных и региональных библиотек. Списки адресов этих представительств можно найти на сайтах центральных библиотек, например, на сайте Государственной публичной научно-технической библиотеки России (ГПНТБ) www.gpntb.ru. Основными видами электронных информационных ресурсов российских библиотек, к которым можно получить доступ в Интернет, являются электронные каталоги, различного рода базы данных (библиографические и др.), а также разнообразная справочная информация. Например, на сайте ГПНТБ можно получить свободный доступ к электронному каталогу, в котором содержится библиографическая информация о всех видах литературы, поступающей в фонд России – авторефераты диссертаций (поступления с 01.10.91 г.), неопубликованные переводы (поступления с 01.01.92 г.), зарубежные книги (поступления с 01.01.92 г.), зарубежная периодика (поступления отдельных номеров с 01.01.93 г.), отечественные книги (поступления с 01.01.93 г.), отечественные журналы (поступления с 01.01.94 г.). При обращении к электронному каталогу можно задать различные варианты поиска интересующих изданий (по автору, названию, тематике), году издания, ключевым словам и т.п.). В ответ на запрос

будет выведен список литературы, имеющейся в фондах и соответствующий заданным в запросе критериям.

На сайте ГПНТБ можно также получить доступ к электронной версии Российского сводного каталога по научно-технической литературе, который содержит сведения о зарубежных и отечественных книгах и зарубежных периодических изданиях по естественным наукам, технике, сельскому хозяйству и медицине, поступившие в организации-участницы Автоматизированной системы Российского сводного каталога. Этот Каталог отражает фонды более чем 400 библиотек страны и содержит около 480 тыс. библиографических записей. Его ежегодное пополнение составляет около 30 тыс. записей по всем видам изданий.

6.9. Информационные ресурсы социальной сферы

Социальная сфера представляет собой совокупность целого ряда отраслей, из которых в состав наиболее значимых входят:

- здравоохранение;
- образование;
- занятость и социальное обеспечение;
- пенсионное обеспечение;
- миграционная служба;
- физическая культура и туризм.

К настоящему времени наиболее развитыми системами государственных информационных ресурсов обладают отрасли здравоохранения и образования. Доступ к информационным ресурсам в области здравоохранения обеспечивается рядом региональных органов власти. В качестве примера здесь можно привести сервер Центра фармацевтической информации Комитета здравоохранения Правительства Москвы, расположенный по адресу www.pharm.mos.ru. На этом сервере учреждениям, занимающимся фармацевтической деятельностью, предоставляется целый ряд информационных услуг. В частности, имеется информация о наличии лекарственных препаратов в оптовой и розничной продаже (для Москвы).

6.10. Информация о природных ресурсах, явлениях, процессах

Информация о природных ресурсах, явлениях и процессах сосредоточена в нескольких отраслевых системах и секторах информационной сферы. Наиболее крупной из этих систем является создаваемая в Министерстве природных ресурсов Российской Федерации (МПР России) Единая информационная система недропользования (ЕИСН) в составе:

- всероссийских геологических фондов — Российский федеральный геологический фонд (Росгеофонд), 5 специализированных и 62 территориальных геологических фондов субъектов Российской Федерации;
- Государственного банка цифровой геологической информации (ГБЦГИ) и информации о недропользовании в России;
- банка данных государственного мониторинга геологической среды (ГМГС) – государственный, региональные и территориальные центры ГМГС;
- музейно-библиотечных и коллекционных фондов, фондов эталонов минерального сырья и керна материала.

Важнейшим информационным ресурсом в области гидрометеорологии является Российский государственный фонд данных о состоянии окружающей природной среды.

Здесь, в частности, на машиночитаемых носителях представлены результаты наблюдений за параметрами окружающей среды с начала инструментальных наблюдений (1725 г.). Информация по другим тематическим направлениям, связанным с природными ресурсами и явлениями (например, экологическая информация), в значительной степени сосредоточена по организациям различных отраслей, органам управления и научным учреждениям, занимающимся исследованиями разного рода природных объектов, явлений, процессов, сбором и анализом данных о них. Особенностью данных работ являются высокие

требования к полноте и точности сведений о природных объектах и процессах, а также высокая трудоемкость и соответственно высокая стоимость сбора этих сведений. Информация о природных объектах, ресурсах, явлениях и процессах весьма разнообразна и часто трудно сопоставима. Одним из путей преодоления проблемы сопоставимости разнородной информации данного типа является ее пространственная привязка на основе географических информационных систем (ГИС), которые представляют собой специализированные базы данных, интегрирующие картографическую и другие виды информации.

В России работы по созданию ГИС возглавляет Федеральная служба геодезии и картографии России (Роскартография), которая руководит всеми топографо-геодезическими и картографическими работами в стране. ГИС разного назначения и территориального охвата создаются в исследовательских институтах Академии наук, отраслевых научно-производственных организациях и учреждениях и российских университетах. За последние 5 лет появилось довольно много коммерческих фирм, занимающихся цифровым картографированием и ГИС в целом. Сложился небольшой, но быстро развивающийся российский рынок ГИС-продуктов. В 1995 г. для объединения специалистов, занятых в области разработки и применения геоинформационных технологий, создана Межрегиональная общественная организация содействия развитию рынка геоинформационных технологий и услуг (ГИС-Ассоциация).

Информацию по земельной политике, землеустройству и смежным областям можно получить на сервере Федерального кадастрового центра «Земля» (ФКЦ «Земля») и Федеральной службы земельного кадастра России — www.fccland.ru. Большой объем информации по ГИС и геоинформационным технологиям представлен на сервере ГИС-Ассоциации — www.gisa.ru.

Контрольные вопросы к лекции 6.

1. Какова общая характеристика государственных информационных ресурсов? Что она отражает?
2. Какова суть деятельности органы и организаций, ответственных за формирование и использование государственных информационных ресурсов? Как они организованы?
3. Перечислите основные правовые акты, определяющие порядок формирования и использования информационных ресурсов? Как они регулируют правоотношения в области мировых информационных ресурсов?
4. Для чего нужны информационные ресурсы министерств и ведомств?
5. Каковы функции информационные ресурсы региональных и муниципальных органов власти?
6. В чем суть статистической информации?
7. Каково назначение научной и технической информация?
8. Как формируются и используются информационные ресурсы библиотечной сети России?
9. Для чего нужны информационные ресурсы социальной сферы?
10. Что отражает информация о природных ресурсах, явлениях, процессах? Для чего она нужна?

Лекция 7. ПРАВОВАЯ ИНФОРМАЦИЯ

Вопросы:

- 7.1. Источники правовой информации
- 7.2. Основные агентства, обеспечивающие потребителя правовой информацией
- 7.3. Информационные справочные правовые системы (СПС)
- 7.4. Технологии доступа к правовым базам данных

7.1. Источники правовой информации

Правовая информация включает тексты и другие материалы, содержащие сведения о законодательстве, праве и правоприменительной практике, а также иные данные, которые требуются для соблюдения норм права.

В настоящее время основу государственной системы правовой информации составляют информационные ресурсы организаций Федерального агентства правительственной связи и информации при Президенте Российской Федерации (ФАПСИ) и Министерства юстиции Российской Федерации (Минюст России). В рамках ФАПСИ работы в области формирования и организации использования ресурсов правовой информации проводятся Научно-техническим центром правовой информации «Система» (НТЦ «Система»). Фонд правовой информации НТЦ «Система» представлен следующими основными информационно-справочными системами:

- Официальная электронная версия бюллетеня «Собрание законодательства Российской Федерации». Содержание: полные тексты документов федерального уровня – законы, правовые акты Государственной Думы и Совета Федерации Федерального Собрания РФ, Президента РФ, Правительства РФ и Конституционного Суда РФ.
- Официальная электронная версия «Бюллетень нормативных актов федеральных органов исполнительной власти Российской Федерации» и Банк правовых актов высших органов государственной власти Российской Федерации. Содержание: полные тексты ведомственных документов федерального уровня, зарегистрированных в Минюсте России.
- Электронная версия «Бюллетеня Верховного Суда Российской Федерации». Содержание: полные тексты документов Верховного Суда РФ и субъектов РФ, комментарии и обзоры судебной практики.
- Электронная версия «Вестник Высшего Арбитражного Суда Российской Федерации». Содержание: полные тексты документов Высшего Арбитражного Суда РФ, комментарии и обзоры судебной практики.
- Электронная версия «Бюллетень международных договоров Российской Федерации». Содержание: полные тексты многосторонних и двусторонних договоров, федеральных законов.

Центральным узлом информационно-вычислительной системы Минюста России является Научный центр правовой информации при Минюсте России (НЦПИ), созданный в 1975 г. В ряде субъектов Российской Федерации созданы учреждения Минюста России – центры правовой информатизации. Центры правовой информатизации Минюста России (ЦПИ) — государственные учреждения Министерства юстиции Российской Федерации, которые создаются для обеспечения судов общей юрисдикции, органов и учреждений юстиции, находящихся на территории субъектов РФ, пользуются правовой информацией (базами данных и печатными изданиями), средствами вычислительной и оргтехники. К числу основных информационных ресурсов Минюста России относятся:

- Комплекс баз данных правовой информации, объединенных в программно-технологический комплекс «Фонд», разработанный специалистами НЦПИ и представляющий собой архив правовых актов СССР и РФ. Фонд содержит более 340 000 документов с 1917 г. в их поворсионной истории развития. В базах данных «Фонд» содержатся: право-

вые акты СССР и РФ, судебная, нотариальная и арбитражная практика, международные договоры и соглашения, акты субъектов РФ, прошедшие экспертизу в Минюсте России, и другие документы.

- База данных действующего российского законодательства «Эталон». Это полнотекстовая база данных по действующему российскому законодательству, разработанная в НЦПИ. «Эталон» содержит около 50 000 текстов действующих нормативных актов. Среди них — федеральные законы, акты Президента РФ и Правительства РФ, а также приказы и инструкции министерств и ведомств, судебная, нотариальная, арбитражная практика.

- Государственный реестр общественных объединений и религиозных организаций. Ведется НЦПИ на основе данных, предоставляемых Минюстом России и региональными органами юстиции. Он содержит информацию о зарегистрированных общественных и религиозных организациях и объединениях. В реестре указаны: учредитель, название объединения, регистрационный номер, дата регистрации, адрес, контактные телефоны, а также цели и задачи данного объединения (организации). Реестр включает в себя сведения более чем о 3000 общественных организаций и объединений; этот список постоянно дополняется новыми данными.

7.2. Основные агентства, обеспечивающие потребителя правовой информацией

Кроме перечисленных систем правовой информации, разработанных государственными организациями, в России существует и активно развивается рынок коммерческих правовых компьютерных систем.

Рассматривая коммерческие правовые системы, важно отметить два обстоятельства, определяющих их связь с государственной системой правовой информации. Во-первых, в основе создаваемых в этих системах баз данных лежат государственные ресурсы правовой информации, которые используются коммерческими фирмами в рамках оформленных тем или иным образом соглашений. Во-вторых, в настоящее время идет активный процесс интеграции коммерческих правовых систем в государственные структуры, поскольку сегодня государственные органы всех ветвей власти наряду с системами Минюста России и ФАПСИ используют также и коммерческие системы.

В настоящее время на коммерческом рынке правовых систем России работают и конкурируют между собой более десятка фирм-производителей. Ведущие фирмы-производители коммерческих правовых систем создали разветвленные сети распространения систем, охватывающие практически всю страну. Они обеспечивают достаточно высокий уровень обслуживания пользователей. К явным лидерам этого рынка можно отнести системы: «Гарант» (Научное производственное предприятие «Гарант-Сервис»), «Кодекс» (Информационно-правовой консорциум «Кодекс») и «Консультант Плюс» (АО «Консультант Плюс»).

Практически все ведущие государственные и коммерческие организации-разработчики информационных правовых систем имеют в настоящее время свои представительства (серверы, сайты) в Интернет. Информацию, размещаемую на этих серверах, можно разделить на категории:

- информация о фирме-разработчике,
- информация о предлагаемых фирмой правовых системах (характеристики, варианты поставки, варианты обновления, прайс-листы и др.);
- собственно правовые системы с доступом через Интернет (демонстрационные или рабочие версии).

Говоря о государственной системе правовой информации, нельзя не упомянуть сервер Министерства юстиции РФ www.sch.ru. На сервере можно найти самые разнообразные сведения о правовых информационных ресурсах, в том числе описания баз данных НЦПИ, сведения о работах НЦПИ и ЦПИ, познакомиться с новыми документами и др.

Коммерческие организации-разработчики правовых систем также весьма полно представлены в Интернет:

- НПП «Гарант-сервис» — <http://www.garant.ru>;
- Консорциум «Кодекс» — www.kodeks.net;
- АО «Консультант Плюс» — www.consultant.ru.

На этих серверах в свободном доступе имеется информация о новых нормативных актах и текущие обзоры законодательства по различным направлениям (мониторинг законодательства). На ряде серверов можно подписаться на данные по мониторингу законодательства и бесплатно получать их по электронной почте, а также просматривать содержание баз данных по законодательству и сформировать запрос на поиск нужного документа.

В результате просмотра можно получить не только название документа, содержащегося в базе, но и доступ к текстам документов в основных базах данных, как правило, платный.

7.3. Информационные справочные правовые системы

Понятие, назначение и преимущества. Принятие обоснованных решений в сфере, как экономики, так и политики невозможно без обладания достаточным объемом правовой информации. Особенно остро эта необходимость ощущается в период реформирования экономического и политического устройства. Задачу удовлетворения потребности в своевременном предоставлении необходимого объема правовой информации решает различные средства массовой информации (СМИ). В этой области конкурируют как традиционные СМИ, так и справочно-правовые системы (СПС). По-настоящему эффективная СПС может быть создана только с применением современных информационных технологий. Созданная таким образом СПС называется компьютерной. Компьютерная справочно-правовая система – это программный комплекс, включающий в себя массив правовой информации и инструменты для работы с ним. Эти инструменты могут позволять производить поиск документов, формировать подборки документов, выводить документы или их фрагменты на печать. Преимущества компьютерных СПС очевидны. Это и доступность информации, и удобство работы с ней. Проблема же, присущая таким системам – недостаточная оперативность – может быть решена с помощью сети Интернета.

Обзор рынка СПС в России. На рынке справочно-правовых систем в России работает большое количество фирм как разрабатывающих собственные программные комплексы, так и обслуживающих существующие. Наиболее известны следующие продукты таких фирм (по данным АО «Консультант Плюс»):

- «Консультант Плюс» (АО «Консультант Плюс»);
- «ГАРАНТ» (НПП «Гарант-Сервис»);
- «Кодекс» (Центр компьютерных разработок).

Системы, созданные государственными предприятиями для обеспечения потребностей в правовой информации государственных ведомств:

- «Эталон» (НЦПИ при Министерстве юстиции РФ);
- «Система» (НТЦ «Система» при ФАПСИ).

Кроме того, на российском рынке представлены такие системы, как:

- «ЮСИС» (фирма «Инталекс»);
- «Референт» (ЗАО «Референт-Сервис»);
- «Юридический мир» (издательство «Дело и право»);
- «Ваше право» и «Юрисконсульт» (фирма «Информационные системы и технологии»);
- «1С: Кодекс», «1С: Гарант», «1С: Эталон» (компания «1С»);
- «Законодательство России» (Ассоциация развития банковских технологий) и некоторые другие.

Различные продукты могут существенно различаться не только по задачам, решаемым с их помощью, но и по качеству. Качество СПС зависит как от качества предоставляемой информации, так и от качества инструментов, используемых для работы с ней.

Применение самых последних компьютерных технологий не поможет, если в СПС не содержится полной правовой информации или если информация обновляется с недостаточной периодичностью. И наоборот, СПС, содержащая даже самую полную и оперативно обновляющуюся информацию, не будет достаточно эффективна, если не предоставлены качественные инструменты для обработки этой информации.

Таким образом, основными параметрами, позволяющим определить качество содержания информационной базы, являются:

- полнота информации;
- достоверность информации;
- оперативность обновления информации.

Параметры, характеризующие качество программной оболочки:

- поисковые возможности системы;
- средства актуализации информации;
- дополнительные сервисные функции.

Основные свойства информационных банков СПС. Оценка полноты, достоверности и оперативности обновления информации основывается на количественных показателях. Оценка же качества юридической обработки поступающих в информационный банк документов достаточно субъективна. Без юридической обработки СПС является всего лишь электронным аналогом бумажных изданий. Ее цель – систематизация документов для повышения эффективности их дальнейшего использования.

Юридическая обработка обычно состоит из следующих основных этапов:

- классификации документов;
- выявления взаимосвязей между различными документами;
- составления примечаний к документу.

Классификация документов предназначена для последующего их поиска по некоторым признакам. Это могут быть как формальные признаки, такие как выходные данные, так и неформальные – темы, которые рассматриваются в этих документах. Классификация производится на основании классификатора данной системы. В общем случае классификатор – это иерархическая структура, содержащая все понятия, используемые для описания документов, входящих в информационную базу. Качество последующей юридической обработки документа, а также эффективность его поиска напрямую зависят от корректности классификации документа в системе.

Выявление взаимосвязей документа позволяет создать список документов, его дополняющих. Таким образом, пользователю предоставляется возможность максимально полно ознакомиться с интересующей его темой, начав работу всего лишь с одним документом.

В качестве примечаний к документу обычно используются ссылки на связанные с ним информационные ресурсы, данные о текущем статусе документа, комментарии юриста.

7.4. Технологии доступа к правовым базам данных

Хранение и обработка больших объемов правовой информации. Необходимость хранения и регулярного обновления больших объемов информации накладывает на СПС ряд требований. Эти требования связаны с потребностями, как пользователей системы, так и разработчиков, производящих обновление и юридическую обработку массива правовой информации. Объем ежемесячно вводимой и обновляемой информации может достигать нескольких тысяч документов, что вынуждает разработчиков с самого начала создания системы приспособлять ее к таким условиям поддержки актуальности информационного банка СПС.

Основные поисковые и сервисные возможности. Поиск документов можно условно разделить на два вида:

- поиск по реквизитам документов;
- поиск по тексту документов.

Поиск по реквизитам документа позволяет найти документ с определенными

- датой,
- номером,
- выпущенный определенным органом и т.д.

Для повышения эффективности поиска условия в таком виде поиска можно сочетать, задавать интервалы номеров или дат.

Поиск по тексту документов заключается в простом *переборе массива документов и нахождении среди них тех, которые содержат слово, заданное пользователем*. Таким образом, такой поиск позволяет *найти документы по какой-либо тематике*. Недостатком поиска по тексту является большой шум, то есть попадание в список найденных документов большого количества документов, удовлетворяющих запросу, но пользователю в данный момент не нужных. Уменьшить уровень такого шума может, лишь опыт составления запросов к системе.

Многие СПС позволяют использовать одновременно поиск и по реквизитам, и по тексту документов, что может существенно повысить эффективность поиска.

Дополнительные сервисные возможности. Некоторые СПС обладают дополнительными возможностями, позволяющими сделать работу с системой более простой и эффективной. Из таких возможностей наибольшее значение для пользователя имеют

- *улучшенная навигация* по информационной базе,
- *возможность сохранять личные предпочтения* и подборки документов,
- *возможности редактирования* документов и вывода их на печать.

Для обеспечения удобной навигации по информационной базе обычно используют гипертекстовые ссылки между документами, а также перемещение по списку ранее просмотренных документов и использование закладок. Возможность сохранять подборки документов позволяет пользователю приспособить систему к своим собственным нуждам. Средства редактирования текста документов необходимы для их обработки пользователем, особенно в том случае, когда информационная база содержит типовые формы документов, предназначенные для заполнения. В СПС может быть не включен собственный текстовый редактор, если предусмотрены функции экспорта документов в форматы популярных текстовых редакторов.

Технологии передачи информации. Доступ пользователя к информации, хранящейся в СПС, может осуществляться двумя способами, каждый из которых имеет свои достоинства и недостатки:

- работа с удаленной базой;
- работа с локальной базой.

При работе с удаленной базой пользователю нет необходимости хранить на своем компьютере данные системы, они хранятся на сервере разработчика и доступны через сеть. Большинство СПС имеют версии, доступные через глобальную сеть Интернет. Основное преимущество работы с такими версиями заключается в том, что пользователь всегда имеет доступ к самым последним данным. Также при таком подходе упрощается процесс предоставления пользователю бесплатной демонстрационной версии системы, которая имеет ограниченный набор функций и документов, но может дать представление о системе.

Очевидным недостатком СПС, расположенной на сервере разработчика, является невозможность работы с ней при отсутствии связи с сервером, что может быть вызвано как перебоями в работе сервера, например из-за большого числа одновременных обращений, так и неполадками в линиях связи. Еще одним минусом СПС с удаленным доступом является то, что не все сервисные функции системы в настоящий момент могут быть представлены в такой версии системы.

При работе с локальной версией этих проблем не существует, но возникает необходимость постоянной поддержки информационной базы в актуальном состоянии. Обновле-

ние информации локальной базы может происходить как с помощью Интернета, так и с помощью традиционных носителей информации – дискет, компакт-дисков и т.д.

Контрольные вопросы к лекции 7.

1. Назовите источники правовой информации?
2. Для чего нужна правовая информация?
3. Каковы технологии доступа к правовым базам данных?
4. Функции основных агентств, обеспечивающих потребителя правовой информацией? Как эти функции реализуются?
5. Перечислите информационные справочные правовые системы (СПС)
6. Назовите основные функции СПС

Лекция 8. БИРЖЕВАЯ И ФИНАНСОВАЯ ИНФОРМАЦИЯ

Вопросы:

- 8.1. Российский рынок биржевой и финансовой информации
- 8.2. Информационные агентства биржевой и финансовой информации, предоставляемые ими услуги
- 8.3. Виды биржевой и финансовой информации, ее анализ
- 8.4. Программные средства анализа

Финансовая информация. Характеризует текущее и перспективное положение фирм, сложившуюся конъюнктуру на рынке капиталов, инвестиции, эмиссии ценных бумаг, формируется в результате проведения собственных или заказных исследований, а также получаемая из независимых источников (например, консалтинговых, инвестиционных и аудиторских фирм, специализированных агентств, баз данных, периодических изданий и т.д.).

Биржевая информация – это информация о котировках ценных бумаг, валютных курсах, учетных и процентных ставках, ценах, индикаторах, предоставляемая банками, биржами, брокерскими фирмами, а также специальными агентствами и службами. Следует отметить, что в связи с глобализацией хозяйственной деятельности часто бывает трудно провести четкую границу между выделенными видами информации.

8.1. Российский рынок биржевой и финансовой информации

Формирование рынка деловой информации в России началось в середине 1990-х годов. Основными поставщиками информационных продуктов и услуг здесь являются:

- службы обмена информацией между финансовыми институтами;
- специализированные агентства и службы, ориентированные на профессиональных инвесторов;
- агентства и службы, ориентированные на потребительский рынок.

Предоставляемые ими информационные продукты и услуги доступны в следующих формах:

- в реальном масштабе времени;
- в виде баз данных на магнитных носителях или с доступом on-line;
- как специализированные обзоры в электронном или печатном виде;
- материалов периодических деловых изданий.

Особую роль здесь играет сеть Интернет, которая используется для оперативного доступа к деловой информации. Следует отметить, что роль Интернет в качестве источника доступа к информации в реальном времени очень высока. Целый ряд фирм, например «Финмаркет», RBC, «Сфинкс», предлагают подобные услуги отечественным потребителям.

На рынке финансовой информации представлены базы данных коммерческой информации по предприятиям, как адресные, так и характеризующие их финансовое состояние, а также специальные базы по сделкам с предприятиями (приобретениям и продажам предприятий, слияниям и разделением фирм и т.п.) и коммерческим предложениям по продаже товаров и услуг.

Специализированные обзоры, как правило, содержат аналитическую информацию, посвященную основным тенденциям развития рынка, конкретному активу или предприятию. Наиболее популярными из них являются обзоры агентств МАФИ, «Скейт-Пресс», АК&М, RBC, «Прайм», «Пал Информ», публикуемые в печатной форме, а также распространяемые по каналам электронной почты, в режиме диалога или в глобальной сети Интернет. Таким образом, отечественный рынок информационных продуктов и услуг в финансово-кредитной сфере в целом соответствует уровню развитых стран. Он имеет определенную специфику, связанную с особенностями развития нашего общества на этапе пе-

реходного периода. Специфика отечественного рынка заключается в ограниченных возможностях получения и предоставления достоверной информации, что отрицательно влияет на качество информационных продуктов и услуг, а также на информационную прозрачность хозяйственной деятельности в целом.

Серьезной проблемой является отсутствие на сегодняшний день общенациональных открытых БД, содержащих текущую и прогнозную информацию о состоянии экономики. Парадоксально, но доступными источниками подобной информации по России являются БД Всемирного банка, Международного валютного фонда, а также ведущих зарубежных поставщиков.

Другой серьезной проблемой является отсутствие стандартизации форматов представления информационных сообщений и несогласованность технической политики ведущих отечественных поставщиков информации.

Не менее значительным препятствием на пути становления отечественного рынка деловой информации служит неразвитость телекоммуникационной инфраструктуры и высокие цены на качественные услуги связи.

В совокупности все это делает недоступным получение достоверной информации для многих средних и мелких предприятий, а также индивидуальных предпринимателей.

8.2. Информационные агентства биржевой и финансовой информации, предоставляемые ими услуги

Деловая информация в реальном масштабе времени по российским и международным рынкам доступна только профессиональным участникам и предоставляется в основном зарубежными поставщиками (Reuters, Dow Jones Telerate, Tenfore, Bloomberg и др.). В финансово-кредитной сфере информационные услуги ведущих мировых поставщиков сочетаются с возможностью проведения электронных сделок.

В настоящее время по уровню сервиса и оперативности к ним существенно приблизились ряд известных отечественных компаний (МФД, АК&М, «Прайм», «Интерфакс»), а также служб институтов финансового рынка (ММББ, РТС, МФБ).

Ведущими производителями и операторами БД международной деловой информации, доступных в режиме online или на CD ROM, являются такие известные фирмы, как Zacks Investment Research Inc. (Corporate Earnings Estimator, Zacks Fundamentals), S&P (Compustat, Corporation Records Online, S&P Marketscope), Telerate Systems Inc. (Telstat, Telerate Financial Information Network), Value Line Inc. (Value Line Data Base II), Data Resource Inc. (DRI-FACS, DRI-SEC, Security Industry Data Bank), Merrill Lynch Inc. (Merrill Lynch Research Service) и др.

Информация с финансовых рынков достаточно широко представлена в системе Dow Jones Telerate (DJT). К числу услуг, предоставляемых системой, относятся: электронный дилинг; передача сообщений; электронная почта; специальные тематические выпуски; доступ к базам данных; услуги по предоставлению новостей; программное обеспечение биржевых операций.

Отечественных генераторов БД деловой информации можно условно разделить на две группы. Первую составляют специализированные службы финансовых институтов, предоставляющие информацию в режиме on-line и преимущественно для своих клиентов.

Доступна также ретроспективная информация. Однако выпуск БД на CD ROM или других носителях информации этими поставщиками пока не практикуется. Вторую группу составляют различные информационные агентства (РИА), которые помимо доступа к БД предлагают другие услуги – тематические обзоры, бюллетени, консультации и т.д. В настоящее время число подобных агентств достаточно велико. Примером может служить РИА RBC (RosBusinessConsulting), предоставляющее тематические БД (итоги торгов, котировки, ставки, финансовые показатели предприятий и т.д.) в формате ППП Excel на условиях подписки.

Следует также выделить одну из самых известных российских компаний – K&M. Основной продукт фирмы – база данных АК&М-List. Структура базы разбита на шесть относительно самостоятельных блоков: брокеры, эмитенты, текущие котировки, информация о выпуске ценных бумаг, финансовая информация, новости. Оперативная информация собирается у различных производителей. БД содержит сведения о более чем 2000 эмитентов, доступны их текущие и ретроспективные балансы. Блок новостей формируется на основе сообщений центральной и региональной прессы. База содержит также различные аналитические обзоры и расчеты популярных отечественных фондовых индексов.

Интересным примером отечественных баз данных выступает Stock Navigator фирмы «ДИАСОФТ», являющейся информационно-справочной системой поддержки операций с ценными бумагами и хранения в систематизированном виде подробных сведений о субъектах и объектах фондового рынка. В качестве другого примера можно назвать базу данных «Ценные бумаги и приватизация», подготовленную и поддерживаемую Ассоциацией развития банковских технологий (АРБТ). Одним из немногих источников адресной информации по институциональным инвесторам выступают базы данных «Коммерческие структуры» (АО «Мосвнешинформ») и «Финансы» (ЦНИИ экономики).

8.3. Виды биржевой и финансовой информации, ее анализ

Биржевая информация является по своему содержанию текущей, изменяется ежеминутно, отражая события, происходящие на рынке, и включает сведения о видах и котировках ценных бумаг (товаров), реализуемых через биржи и на внебиржевом рынке, процентных и учетных ставках, курсах валют, объемах торгов, сводных индексах и т.д. Биржевая информация в большинстве случаев требуется потребителю в реальном масштабе времени, т. е. по возможности одновременно с совершением сделки на рынке или немедленной ее регистрации. Счет здесь может идти на доли секунды, а минутная задержка уже рассматривается как значительная. Сектор биржевой информации, пожалуй, единственный, где особенности предоставляемой информации (высокая цена ошибок, повышенные требования к ее своевременности и качеству) обусловили необходимость и оправданность многократного дублирования и запараллеливания каналов связи.

На каждом этапе принятия решений используемая информация может быть представлена одновременно в нескольких формах. Исторически (с момента изобретения в 1867 году тикера – специального автоматического печатающего аппарата) основным стандартом представления биржевой информации является тикерная строка. Это – бегущая строка на электронном табло и мониторах брокеров в биржевом зале. Она начинается с тикерного символа компании – эмитента ценной бумаги, содержит группы цифр, отделенных друг от друга разделителями, которые показывают: цену товара или курс ценной бумаги на моменты открытия и закрытия биржи, среднюю цену или котировку за прошлый день продаж, максимальную, минимальную и среднюю котировки в день продажи, котировки и цены на момент открытия и закрытия рынка, данные об объемах продаж товаров или ценных бумаг и т. д.

Но воспринимать биржевую информацию посредством тикерной строки не просто. В настоящее время основные подходы к представлению биржевой информации базируются на фильтрации и визуализации данных. Фильтрация позволяет пользователю задавать определенные ограничения на данные (например, на пороговые изменения котировок) и получать на свой терминал только информацию, отвечающую этим критериям, причем в виде, пригодном для дальнейшего анализа и моделирования.

Визуализация обеспечивает пользователю многооконное представление данных. Поступающая информация различается цветом, шрифтовым оформлением и т.п. Будучи обработанной специальными аналитическими программами, информация представляется также в графической форме, что значительно облегчает ее восприятие. Достаточно перспективным является представление биржевой информации в форматах популярных пакетов прикладных программ. Такой подход уже используется целым рядом ведущих зарубежных и отечественных поставщиков. В качестве примера можно привести специальный

продукт агентства Reuters, позволяющий получать биржевую информацию в виде таблицы Microsoft Excel, изменяющейся в реальном масштабе времени. При этом пользователь может вносить свои данные, а также моделировать требуемые изменения.

Примерами из отечественной практики могут служить ММВБ и РТС, информация о ходе торгов на которых доступна пользователям в формате другого офисного пакета – СУБД Microsoft Access.

В отличие от биржевой финансовая информация гораздо шире как по содержанию, так и по форме представления. Будучи гораздо менее критичной к фактору времени, она является, как правило, аналитической, а не текущей. Спектр охватываемых ею сведений достаточно широк – от ежедневных обзоров текущей конъюнктуры, складывающейся на рынке тех или иных активов (товаров), до перспективных прогнозов будущего состояния, как отдельных фирм-эмитентов, так и конкретных секторов финансового рынка в целом. Формы представления финансовой информации достаточно разнообразны – от привычных таблиц, до трехмерных графиков, снабженных динамическими пояснениями (комментариями), гипертекстовыми ссылками, аудио- и видеофрагментами и т.п.

8.4. Программные средства анализа

Базы данных деловой информации, распространяемые на магнитных носителях (преимущественно на CD ROM), пока еще не получили должного распространения в России. Например, первые БД биржевой информации на CD ROM появились только в конце 1997 года – «Котировки РТС за последний год» («Финмаркет») и «История российского рынка» (МФД) на 7 компакт-дисках. Широкую популярность приобрели БД коммерческой информации, такие, как «Товары и цены» и др.

В настоящее время ряд производителей приступили к созданию коммерческих БД, ориентированных на информационную поддержку своих программных продуктов. В качестве примера можно привести программный продукт «Альт-Эксперт» («Альт», Санкт-Петербург), выполненный в виде надстройки к Microsoft Excel и обеспечивающий проведение сравнительного анализа основных финансовых коэффициентов компаний со среднеотраслевыми данными. Используемая при этом БД содержит средние значения финансовых показателей, полученные на основе обработки отчетности отечественных предприятий из регионов РФ. Необходимая информация для создания БД была предоставлена информационным агентством Прайм-Тасс. По определяемой пользователем выборке программа представляет набор среднеотраслевых финансовых показателей и сопоставляет его с показателями анализируемого предприятия. Система совместима с программным продуктом этой же фирмы «Альт-Финансы» версии 1.5, что позволяет осуществлять прямой обмен данными. Существует также возможность непосредственного ввода данных через специальную форму пользователя.

Использование БД на магнитных носителях значительно увеличивает эффективность финансовых решений, обеспечивая специалисту уникальную возможность дешевого и эффективного доступа к практически неограниченным массивам как оперативной, так и ретроспективной информации непосредственно с рабочего места. Особая ценность таких БД заключается также в том, что информация в них, как правило, представлена уже в готовом для использования виде (в виде таблиц, графиков, прогнозов, форматов данных популярных пакетов прикладных программ и т.д.).

Основным поставщиком БД международной и отечественной деловой информации является агентство Reuters. Базы данных Reuters содержат информацию по ценным бумагам, валюте, товарным рынкам, а также общеполитические и деловые новости, материалы других агентств новостей и прессы, международных и правительственных организаций, предоставляемые в режиме реального времени.

Информационные услуги Reuters сгруппированы в несколько служб-комплексов.

- Базы данных Reuter Money, Reuter Treasury, Reuter Commodities, Reuter Equities, Reuter News, охватывающие весь спектр инструментов финансового рынка.

- Система осуществления транзакций Dealing, работающая в сетях с протоколом X.25, благодаря которой абонент может получать и передавать котировки, заключать сделки и обмениваться информацией в режиме двусторонней телексной связи с партнерами.

- Программы технического анализа в реальном времени Reuters Technical Analysis и Reuters Graphics 3.5 Professional.

- Рабочая станция ATW (Advanced Trader Workstation), работающая в среде UNIX и включающая систему регистрации сделок, ведения позиций, комплексного анализа и управления финансовыми рисками Kondor+ (новое поколение рабочих станций представлено системой Kobra).

- Система аналоговой коммутации потоков данных Prism+, позволяющая эффективно использовать многотерминальные конфигурации в рамках единого рабочего места и др.

Еще одним информационным продуктом Reuters является финансовое телевидение – качественно новый вид услуг, с которым компания выходит на российский рынок мультимедиа. Подписчики имеют возможность смотреть прямые репортажи с мировых финансовых рынков, обзоры новостей, трансляции пресс-конференций, выступлений, а также комментарии финансовых экспертов.

Другим примером систем удаленного доступа к базам данных международной и отечественной деловой информации является служба Tenfore, предлагаемая компанией «Tenfore-Russia». Эта спутниковая информационная система ориентирована на широкий круг подписчиков – от предприятий до частных лиц, имея весьма конкурентное соотношение цена/качество и перекрывает основные рынки, получая информацию от всех ведущих участников. Базовый продукт – информационная система Tenfore Workstation – устанавливается на компьютер пользователя. Она реализована в среде Windows, имеет удобный русифицированный интерфейс, средства формирования рабочей среды, накопления и обработки информации в формате электронных таблиц Excel, поддерживает протокол DDE, предоставляет средства фильтрации и поиска новостей по ключевым словам.

Контрольные вопросы к лекции 8.

1. Какие функции выполняет Российский рынок биржевой и финансовой информации?
2. Перечислите информационные агентства биржевой и финансовой информации. Какие услуги они предоставляют?
3. Каковы различия видов биржевой и финансовой информации?
4. Как осуществляется анализ биржевой и финансовой информации?
4. Каковы основные программные средства анализа?

V ПРАКТИЧЕСКИЕ ЗАДАНИЯ

ПЗ-1. Знакомство с поисковыми машинами

Поисковыми машинами называются программы, предназначенные для поиска информации, размещенной на сайтах сети Internet. С их помощью можно найти текстовую, графическую, мультимедийную информацию, файлы. Однако, поиск при этом проводится лишь среди текстовой информации, т.е. в теле станицы должны находиться ссылки на нетекстовую информацию в виде файла, чтобы ее можно было найти.

Для поиска документов на русском языке чаще всего используются поисковые машины универсального назначения: "Yandex", "Rambler", "Апорт". Мировой лидер - поисковая система "Google".

Наиболее популярные в России поисковики имеют и варианты с компактным интерфейсом: ya.ru - "Yandex", r0.ru - "Rambler", au.ru - "Апорт". Интерфейс "Google" изначально достаточно компактен.

Главная страница поисковой машины содержит поле для ввода ключевых слов и условий запроса, кнопку запуска поиска. Найденные результаты выводятся ниже в виде списка. Предваряет этот список указание на число найденных страниц, содержащих искомый текст запроса или отдельные слова, его составляющие.

Как правило, результат поиска содержит:

- · найденный текст,
- · часть документа, окружающую найденный текст (контекст),
- · информацию о размещении найденного текста,
- · тип документа,
- · дата создания/обновления документа.

Обычно, полученные результаты упорядочены по релевантности, но их можно расположить на поисковой странице и по дате создания - от новейших документов к самым старым.

Среди результатов поиска можно провести новый поиск, уточнив или изменив условия запроса. В ответ на запрос, содержащий несколько слов, например, Утро Стрелецкой Казни будут найдены документы как содержащие всю фразу; так и содержащие все три слова, но по отдельности; и документы, где находится хотя бы одно из этих слов, в разных падежах, склонениях, числах. Если необходимо найти лишь документы с целой фразой, то запрос следует записать так: "Утро Стрелецкой Казни". Грамматическая форма слов как-вычками не фиксируется.

Внимание! Часть результатов поиска содержит рекламную информацию и включена в список результатов по согласованию с фирмой-владельцем поисковой системы.

Запрос можно также адресовать к сайтам, содержащим изображения для поиска рисунков и фотографий ("картинки"); музыку (alltheweb.com); новости; информацию о товарах и ценах; файловым архивам.

Указания к оформлению отчета

Отчет оформляется в виде текстового файла и должен содержать:

- · фамилию, имя, группу и вариант студента;
- · номер задания, ответ на задание (если требуется);
- · комментарий к выполнению задания (если требуется);
- · ссылку или набор ссылок на найденные сайты согласно заданию.

Практические задания

Вариант 1. Откройте с помощью программы Internet Explorer поисковые системы "Yandex", "Rambler", "Апорт", "Google" или "Yahoo", а также варианты сокращенного ин-

терфейса: ya.ru, r0.ru. Задайте запрос: мировые информационные ресурсы. Запишите результаты поиска - количество найденных сайтов и страниц (документов) в каждой из четырех поисковых машин. Укажите на дополнительную информацию по запросу, предоставляемую каждой из поисковых систем. Выпишите адрес первой найденной страницы в каждом случае. Измените условие сортировки на "по дате" и снова выпишите адреса первых результатов поиска.

Измените условие запроса на "мировые информационные ресурсы" и снова запишите количества найденных сайтов и страниц во всех четырех поисковых системах.

Дополните условие запроса с помощью опции поиск среди результатов на слово Краснодар. Запишите результат поиска: количество найденных страниц, сайтов, адреса первых ссылок.

Найдите с помощью любой из поисковых машин ответ на следующие вопросы:

1. Укажите дату начала выпуска процессора Pentium 4.
2. Какие химические элементы входят в состав изумруда?
3. Что такое "охлупень"?
4. Перечислите области, окружающие Свердловскую.
5. Назовите марку пива, "бренд года - 2010".
6. Чем смартфон отличается от коммуникатора?
7. Найти ссылку или скачать на книгу К. Маркса «Капитал»

Вариант 2. Откройте с помощью программы Internet Explorer поисковые системы "Yandex", "Rambler", "Апорт", "Google", а также варианты сокращенного интерфейса: ya.ru, r0.ru. Задайте запрос: факультет прикладной информатики. Запишите результаты поиска - количество найденных сайтов и страниц (документов) в каждой из четырех поисковых машин. Укажите на дополнительную информацию по запросу, предоставляемую каждой из поисковых систем. Выпишите адрес первой найденной страницы в каждом случае. Измените условие сортировки на "по дате" и снова выпишите адреса первых результатов поиска.

Измените условие запроса на "факультет прикладной информатики" и снова запишите количества найденных сайтов и страниц во всех четырех поисковых системах.

Дополните условие запроса с помощью опции поиск среди результатов на слово Краснодар. Запишите результат поиска: количество найденных страниц, сайтов, адреса первых ссылок.

Найдите с помощью любой из поисковых машин ответ на следующие вопросы:

1. Сколько продано персональных компьютеров в России в первом квартале 2009 года?
2. В каком году началось производство шампанского в Абрау-Дюрсо?
3. Что такое "хабилитированный доктор наук"?
4. К какому острову у берегов Америки впервые приплыл Колумб (современное название)?
5. Кто является обладателем смежных прав по российскому законодательству?
6. Каков должен быть состав икебаны с пожеланием долголетия?
7. Найти ссылку или скачать на книгу В.Я. Ищейнов «Защита конфиденциальной информации», М.–2009

Вариант 3. Откройте с помощью программы Internet Explorer поисковые системы "Yandex", "Rambler", "Апорт", "Google", а также варианты сокращенного интерфейса: ya.ru, r0.ru. Задайте запрос: количество пользователей интернет. Запишите результаты поиска - количество найденных сайтов и страниц (документов) в каждой из четырех поисковых машин. Укажите на дополнительную информацию по запросу, предоставляемую каждой из поисковых систем. Выпишите адрес первой найденной страницы в каждом случае. Из-

мените, условие сортировки на "по дате" и снова выпишите адреса первых результатов поиска.

Измените, условие запроса на "количество пользователей интернет" и снова запишите количества найденных сайтов и страниц во всех четырех поисковых системах.

Дополните условие запроса с помощью опции поиск среди результатов на слово Краснодар. Запишите результат поиска: количество найденных страниц, сайтов, адреса первых ссылок.

Найдите с помощью любой из поисковых машин ответ на следующие вопросы:

1. Сколько городов в Краснодарском крае с населением более 50 000 жителей.
2. Сколько service pack'ов было выпущено для Windows NT?
3. Что такое "балясина"?
4. Где похоронен солист группы "Beatles" Джон Леннон?
5. Найти пять бесплатных программ для Pocket PC.
6. Какова грузоподъемность самолета "Boing 737"?
7. Найти ссылку или скачать на книгу В.Н. Волкова и А.А. Денисов «Системный анализ...»

Вариант 4. Откройте с помощью программы Internet Explorer поисковые системы "Yandex", "Rambler", "Апорт", "Google", а также варианты сокращенного интерфейса: ya.ru, r0.ru. Задайте запрос: компьютерные технологии и системы. Запишите результаты поиска - количество найденных сайтов и страниц (документов) в каждой из четырех поисковых машин. Укажите на дополнительную информацию по запросу, предоставляемую каждой из поисковых систем. Выпишите адрес первой найденной страницы в каждом случае. Измените условие сортировки на "по дате" и снова выпишите адреса первых результатов поиска.

Измените условие запроса на "компьютерные технологии и системы" и снова запишите количества найденных сайтов и страниц во всех четырех поисковых системах.

Дополните условие запроса с помощью опции поиск среди результатов на слово Краснодар. Запишите результат поиска: количество найденных страниц, сайтов, адреса первых ссылок.

Найдите с помощью любой из поисковых машин ответ на следующие вопросы:

1. Столицей, какого государства был город Туапсе?
2. Какой автомобильной компании принадлежит марка "Траккер"?
3. Каково значение слова "серпень"?
4. Сколько пуль попало в президента Линкольна?
5. Как называется ответ на сообщение, посланное командой ping?
6. Каков должен был быть доход или образовательный уровень избирателя в Родезии?
7. Найти ссылку или скачать на книгу Ч. Петцольд (Charles Petzold) «Аннотация исторической статьи через вычисления на машине Алана Тьюринга». (The Annotated Turing A Guided Tour through Alan Turing's History Paper on Computability and the Turing Machine (Wiley, 2008)).

ПЗ-2. Расширенный язык запроса поисковых машин

Расширенным языком запроса называется набор команд и служебных символов, сопровождающих запрос с целью определения области поиска или ее уточнения для получения более релевантных результатов (пример: "Апорт"). Условия расширенного поиска могут быть заданы и заполнением полей форм, выбора пункта списка (пример: "Rambler").

Поиск изображений

Поисковые системы не индексируют сами изображения, поэтому поиск осуществляется по названию рисунка и по подписи к изображению, найдем изображение флага Того (Togo) в системе "Яндекс".

Поиск по подписи к изображению. В строке запроса введем служебное слово #hint=, затем, в скобках, текст запроса (togo). Общий вид запроса: #hint=(togo). По одному из результатов поиска найдем изображение флага Того.

Поиск среди изображений. Под строкой поиска "Яндекс" выбрать категорию поиска картинки. Ввести в поисковую строку ключевое слово togo. По одному из результатов поиска найдем изображение флага Того.

Поиск сайтов, ссылающихся на данный сайт

Введем в строку запроса поисковой системы "Google" служебное слово link:, затем, через пробел, адрес сайта;

Запрос будет иметь вид, например: link:www.kubsau.ru;

После запуска поиска, подсчитаем количество сайтов, путь с которых ведет на заданный.

Поиск информации на сайте

Найдем ответ на вопрос: Как называются питоны, которые размножаются в неволе только в Московском зоопарке?

Введем в строку поиска "Rambler" слова: "Московский зоопарк";

Скопируем найденный адрес сайта Московского зоопарка;

Откроем окно расширенного поиска "Rambler";

Введем в строку поиска ключевые слова питон размножается, а в поле формы "Искать документы только на следующих сайтах:" вставим адрес Московского зоопарка;

Среди небольшого числа найденных документов найдем нужный и выпишем ответ и ссылку.

Указания к оформлению отчета

Отчет оформляется в виде текстового файла и должен содержать:

- · фамилию, имя, группу и вариант студента;
- · номер задания, ответ на задание (если требуется);
- · комментарий к выполнению задания (если требуется);
- · текст запроса;
- · ссылку или набор ссылок на найденные сайты согласно заданию.

Вариант 1. Составьте таблицу соответствий команд языка запроса поисковых систем "Яндекс" и "Rambler", если запрос представляет собой фразу из нескольких слов.

Зная адрес Московского зоопарка ответьте, при каком кружке в московском зоопарке работает пони-клуб? Использовать технику поиск на сайте с помощью Яндекс.

С помощью поиска в заголовках найти при помощи "Апорта" значение термина башибузуки.

С помощью "Google" найти количество сайтов, ссылающихся на адрес www.pupkin.ru.

В поисковой системе "Яндекс" запросом с использованием знаков: +, -, ~, () найти ответ на вопрос: сколько американских долларов купюрами по 10\$ поместятся в багажнике ИЖ-2717?

Найдите с помощью любой из поисковых машин ответ на следующие вопросы: в какие разъемы устанавливались процессоры i486?

Вариант 2. Составьте таблицу соответствий команд языка запроса поисковых систем "Яндекс" и "Апорт", если запрос представляет собой фразу из нескольких слов.

Зная адрес Московского зоопарка ответьте, в чем особенность рациона фламинго? Использовать технику поиск на сайте с помощью Google (форма запроса такова: запрос site:www.name_site.ru).

С помощью поиска в заголовках найти при помощи "Rambler" значение термина герлыга.

С помощью "Яндекс" найти все ссылки, ведущие на сайт chgk.org.ru.

В поисковой системе "Rambler" запросом с использованием знаков: and, or, not, () ограничения расстояния между словами запроса найти ответ на вопрос: с автомобилем какого знака, по гороскопу, происходит большинство аварий?

Найдите с помощью любой из поисковых машин ответ на следующие вопросы: какие фирмы выпускали клоны процессоров i486?

Вариант 3. Составьте таблицу соответствий команд языка запроса поисковых систем "Rambler" и "Апорт", если запрос представляет собой фразу из нескольких слов.

С помощью техники поиск на сайте в поисковой системе Яндекс найти на сайте www.sobesednik.ru: в каком году и где состоялась впервые встреча "большой восьмерки"?

С помощью поиска в заголовках найти при помощи "Google" значение термина ликвидамбар.

С помощью "A9.com" найти количество сайтов, ссылающихся на адрес izh-auto.ru.

В поисковой системе "Апорт" запросом с использованием знаков: and, or, not, () и операторами расстояния между словами запроса найти ответ на вопрос: какое советское оружие времён ВОВ называли "зверобоем" (поиск без Ф.Купера)?

Найдите с помощью любой из поисковых машин ответ на следующие вопросы: какая максимальная частота процессора, который можно было установить в разъем Socet7?

ПЗ-3. Почтовые системы (электронная почта)

Услуги почтового сервиса в Internet представлены во множестве. Для выполнения этого практического задания необходимо иметь почтовый ящик. Если вы не являетесь зарегистрированным пользователем почтовой службы, можно воспользоваться почтовыми услугами какой-либо из поисковых систем. Внимательно прочтите инструкции и следуйте указанием мастера создания и настройки почтового ящика.

Практические задания

1. Зарегистрировать почтовый ящик на одной из бесплатных почтовых служб.
2. Настроить главную страницу почтового ящика.
3. Познакомиться с возможностями фильтров против «спама». Настроить фильтр, настроить защиту от вирусов в электронной почте.
4. Создать папку для писем, получаемых от студентов своей группы, по полю «от кого» (пример: «ПИЭ 31»).
5. Послать письма двум своим сокурсникам. Получить от них ответ.
6. Изучить возможность получения информации с помощью рассылок. Подписаться на любую рассылку. Отказаться от подписки.
7. Переслать отчет о проделанной работе на адрес: mirinf@mail.ru с темой: МИР-3?, где? – номер группы от 1 до 5.

Указания к оформлению отчета:

- Фамилию, имя, группу студента;
- Почтовую службу, услугами которой воспользовались. С описанием возможностей, предоставляемых клиенту.

- Описание личных настроек главной страницы.
- Структура папок почтового ящика.
- Описание возможностей почтовой системы по защите от спама и вирусов.
- Ссылку на подписку, услугами которой воспользовались.
- Ссылку на страницу, где можно отказаться от указанной рассылки.

ПЗ-4. Поиск материалов для самостоятельной работы

Ниже приведены темы индивидуальных заданий, по которым необходимо приготовить реферат. Содержимое реферата должно включать материалы как минимум пяти электронных документов (ссылки обязательны). Текст набирается и редактируется самостоятельно. Оформление свободное. Срок сдачи: не позднее четырех недель со дня получения задания.

Тема задания регистрируется преподавателем на текущем занятии, в конце занятия необходимо предоставить найденный предварительный материал для выполнения самостоятельной работы.

1. Политика РФ в области информационных ресурсов.
2. Правовое обеспечение информационной области в Российской Федерации.
3. Сравнение законодательства РФ и США/ЕС в области обращения информационных ресурсов.
4. Научно-техническая информация в РФ: виды, способы обращения и доступа.
5. Меры по обеспечению информационной безопасности государства.
6. Политика Китая в области информационных ресурсов государства.
7. Корпорация назначения имен и номеров в Internet (ICANN) и ее роль в обращении мировых информационных ресурсов
8. Развитие информационных сетей в СССР и РФ.
9. Традиционные информационные ресурсы и ресурсы Internet.
10. Сравнение поисковых систем в рунете.
11. Дополнительные возможности поисковых систем.
12. Организация электронных информационных архивов.
13. Способы организации подключения и работы в Internet компьютерных устройств, не относящихся к ПК.
14. Критерии оценки информационного бизнеса
15. Электронная цифровая подпись. Понятие, виды и практика их применения (использования).
16. Сравнение законодательств разных стран об электронной цифровой подписи.
17. Информационные ресурсы как объекты рыночных отношений.
18. Рынок информационных услуг в РФ.
19. Способы организации торговли программным обеспечением.
20. Сравнение систем электронных платежей.
21. Электронная коммерция в Краснодарском крае.
22. Кризис в электронной коммерции начала XXI века его причины и последствия.
23. Влияние новых информационных технологий на развитие сетевой торговли.
24. Обзор виртуальных продуктов, предметов электронной коммерции. Сравнение виртуального товара с традиционным.
25. Обзор платных услуг, предоставляемых в сети Internet.
26. Виды рекламы в сети Internet.
27. Объекты рекламы в Internet, потребитель сетевой рекламы, ее эффективность.
28. Использование Internet и мировых информационных ресурсов в системах обучения и образования.
29. Дистанционное обучение. Понятие, обзор возможностей.

30. Роль Интернет и мировых информационных ресурсов в развитии образования в России.
31. Роль Интернет и мировых информационных ресурсов в развитии образования в развитых капиталистических странах.
32. Виды предоставления пользователю прав на программный продукт.
33. Проблема защиты авторских прав в цифровую эпоху.
34. Информационное пиратство в РФ.
35. Средства защиты от информационного пиратства.
36. Обеспечение безопасности и сохранности электронной информации.
37. Сетевые вирусы, трояны, spyware.
38. Программные средства и меры защиты информационных ресурсов.
39. Коммуникационные возможности Microsoft Windows в глобальных информационных сетях.
40. Работа с удаленными информационными ресурсами средствами офисных программ.
41. Платформа Microsoft .NET. Обзор.
42. Концепция интеллектуального жилища.
43. Концепция сетевого офиса.
44. Обзор развития сенсорных сетей.
45. Оценка эффективности использования информационных ресурсов.
46. Корпорация как создатель и пользователь информационных ресурсов Интернет.
47. Основные цели (побудительные мотивы) размещения информационных ресурсов в глобальной сети Интернет.
48. Структура затрат на доступ к информационным ресурсам.
49. Системы управления проектами. Обзор.

ПЗ-5. Знакомство с предприятиями электронной торговли

Электронная торговля – предоставление продуктов и услуг с помощью сети Internet.

Предприятия электронной торговли представлены как организациями, для которых электронная торговля является основным бизнесом, так и теми, что в основном торгуют offline.

В первом случае, предприятие экономит на торговой площади, складах, обслуживающем персонале, но должен озаботится своевременной доставкой товара, оплатой, сервисным обслуживанием.

Во втором случае, электронный магазин лишь витрина, и способ передачи клиенту прайса.

Электронная торговля - лишь побочная сторона работы, но риски минимизируются основным бизнесом.

В настоящее время в России, наиболее востребованной является вторая модель.

Помимо удобства работы с сайтом электронного магазина, для клиента важно удобство оплаты, своевременность доставки, возврат брака и сервисное обслуживание приобретенного товара.

Оплата за товар возможна либо наличными при получении товара, либо с помощью кредитной карты, либо электронными деньгами, иные способы менее распространены.

Доставка товара осуществляется как правило курьером. Это может быть собственная служба доставки магазина, или специализированная курьерская служба. В отдельных случаях товар можно получить по почте.

Описанные выше способы доставки товара свойственны покупке обычного товара. Если же вы приобретаете товар в электронной форме (программное обеспечение, мультимедийный контент, и т.п.), возможна непосредственная доставка товара по сети. В таком

случае товар вы можете скачать самостоятельно, получить по подписке, по почте, получить пароль доступа к информационному ресурсу, получить код активации, и т.д.

Особое место занимает виртуальный товар, т.е. товар которым вы можете пользоваться лишь в электронной сети, и не можете придать ему какую-либо иную объектную форму.

Это может быть удаленное дисковое пространство; услуга, привязанная к вашему ip-адресу; программный код, реализуемый в виртуальном пространстве, исполнение услуги по запросу и т.п.

Последняя категория товаров в меньшей степени представлена на рынке товаров и услуг, но будет приобретать все большее значение по мере дальнейшей информатизации общества.

Отдельной стороной электронной торговли является предоставление услуг. Также как и товары, эти услуги могут быть вполне обычными (Internet используется для организации заказа), так и специфичными для информационных сетей.

Рассмотренные аспекты электронной торговли более свойственны сегменту бизнес - клиент, в то же время сегмент бизнес - бизнес; имеет свои особенности и отличия.

Реклама в сети Internet использует формы как близкие обычной наружной рекламе (баннеры, всплывающие окна, рекламные объявления), так и специфические. Например: ссылки, ведущие непосредственно на сайт рекламодателя, контекстная реклама и реклама в поисковых системах, скрытая реклама на форумах и чатах, целевая реклама, зависящая от региона, программного обеспечения и поведения посетителя.

Указания к оформлению отчета

Отчет оформляется в виде текстового файла и должен содержать:

- · фамилию, имя, группу студента;
- · номер задания, ответ на задание;
- · комментарий к выполнению задания (если требуется);
- · ссылку или набор ссылок на найденные сайты согласно заданию.

Практические задания

1. Найдите предприятие электронной торговли, которое предлагает товар, начинающийся на первую букву Вашей фамилии.

2. Изучите процесс покупки товара в электронном магазине:

Найдите 5 магазинов, действующих на территории Краснодарского края. Укажите, какие из этих магазинов являются самостоятельными организациями, а какие - филиалы (подразделения) обычных (не электронных) предприятий торговли.

Укажите 5 способов доставки товара (с указанием магазина, предлагающего то или иное решение).

3. Найдите 3 магазина, торгующие программным обеспечением. Укажите различие в организации оплаты и доставки приобретенного товара.

4. Найдите виртуальный товар и укажите способ его приобретения - продавец, способ оплаты, доставка/регистрация.

5. Укажите 3 организации, предоставляющие платные электронные услуги частному пользователю, корпоративным клиентам.

6. Приведите пример 5 платежных средств, с помощью которых можно оплатить покупку в интернет-магазине.

7. Укажите различия в этих платежных средствах (с точки зрения покупателя).

8. Приведите примеры 5 видов рекламы в сети Internet.

VI ТЕСТОВЫЕ ЗАДАНИЯ

1. ИНФОРМАЦИОННЫЕ РЕСУРСЫ И ИХ РОЛЬ В ЭКОНОМИКЕ

1. В теории информации под информацией понимают...	<input checked="" type="checkbox"/> сведения, устраняющие или уменьшающие неопределённость; <input type="checkbox"/> характеристику объекта, выраженную в числовых величинах; <input type="checkbox"/> сигналы от органов чувств человека; <input type="checkbox"/> повтор ранее принятых сообщений.
2. Важнейшие свойства экономической и управленческой информации: - достоверность и полнота; - -	<input checked="" type="checkbox"/> ясность и понятность; <input type="checkbox"/> своевременность и точность; <input checked="" type="checkbox"/> ценность и актуальность; <input type="checkbox"/> доступность и правильность.
3. Для экономической информации характерны: - большие объёмы; - значительный вес логических операций при ее обработке; - -	<input type="checkbox"/> алгоритмическая сложность в обработке данных; <input checked="" type="checkbox"/> многократное повторение циклов ее получения и преобразования в установленные периоды (месяц, квартал, год, и т.д.); <input checked="" type="checkbox"/> многообразие источников и потребителей; <input type="checkbox"/> ведение двойной бухгалтерии.
4. <i>Мировые информационные ресурсы</i> – это документы и массивы информации, находящиеся в глобальных компьютерных сетях,.....	<input type="checkbox"/> библиотеках, архивах, фондах, банках данных, других информационных системах; <input type="checkbox"/> подготовленные людьми и зафиксированные на материальном носителе; <input type="checkbox"/> в них в разных формах представлены знания, которыми обладали люди, создававшие их; <input checked="" type="checkbox"/> доступ к которым не имеет территориальных ограничений и контролируется только источником информации.
5. Сколько видов бизнес-информации Вы знаете?	<input checked="" type="checkbox"/> 2; <input type="checkbox"/> 5; <input type="checkbox"/> 7; <input type="checkbox"/> 12.
6. Классификация внешней информации: - информация о регулировании и налогообложении; - макроэкономическая и геополитическая информация; - внешняя финансовая информация; - - -	<input checked="" type="checkbox"/> информация о поставщиках; <input checked="" type="checkbox"/> информация о конкурентах; <input type="checkbox"/> информация о трудовых ресурсах; <input type="checkbox"/> информация о производстве; <input checked="" type="checkbox"/> рыночная информация.
7. Классификация источников бизнес-информации по формальной стратегии поиска:	<input type="checkbox"/> поставщики; <input checked="" type="checkbox"/> коммерческие учреждения;

<ul style="list-style-type: none"> - торговые ассоциации; - правительственные учреждения; - библиотеки; - поставщики коммерческой информации; - действующая государственная система, банки, статистические органы; - - 	<ul style="list-style-type: none"> <input type="checkbox"/> конференции; <input type="checkbox"/> добровольно представленная литература и рассылка писем; <input checked="" type="checkbox"/> инструкции, информационные бюллетени, каталоги торговых организаций; <input type="checkbox"/> базы данных в режиме он-лайн.
<p>8. Существует представление, что вся работа управленческого персонала в современном офисе делится на шесть основных частей:</p> <ul style="list-style-type: none"> - ввод, сбор необработанных данных; - сохранение обработанной информации; - поиск – процесс доступа к хранящимся данным; - - - 	<ul style="list-style-type: none"> <input checked="" type="checkbox"/> коммуникации, перемещение информации от одного источника к другому; <input type="checkbox"/> снижение риска и уменьшение неопределенности; <input checked="" type="checkbox"/> обработка, преобразование информации из одной формы в другую; <input type="checkbox"/> получение власти и средств воздействия на других; <input checked="" type="checkbox"/> вывод, формирование данных или информации в виде, удобном для пользователя; <input type="checkbox"/> контроль и оценка производительности и эффективности фирмы.
<p>9. Имя ученого, внесшего большой вклад в исследования последствий информатизации на общества.</p>	<ul style="list-style-type: none"> <input type="checkbox"/> Флойд; <input checked="" type="checkbox"/> Хессиг; <input type="checkbox"/> Шеннон; <input type="checkbox"/> Гурвиц; <input type="checkbox"/> Вальрас.

2. ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

<p>I Росту индустрии информации и знаний способствовали следующие факторы:</p> <ol style="list-style-type: none"> 1. Перемещение экономики от промышленного производства к производству услуг. 2. Глобализация рыночной экономики 3. Создание экономических союзов и заключение глобальных торговых соглашений 4. Вышеупомянутые факторы и высокая потребность в усовершенствованных технологиях и системах стимулировали спрос на информационные системы 5. Развитие систем искусственного интеллекта, экспертных систем и нейронных сетей, содержащих инструментальные средства и методы для создания интеллектуальных систем для принятия решений и обработки информации. 6. Повышение уровня знаний и навыков работы с компьютером новых поколений коммерческих и индивидуальных потребителей информации и знаний по сравнению с предыдущими поколениями. 	<ul style="list-style-type: none"> <input checked="" type="checkbox"/> Идея об информационной супермагистралах <input checked="" type="checkbox"/> Удаление организационных культур от централизованного, иерархического и бюрократического подхода к распределенному, единообразному и предпринимательскому стилю управления. <input checked="" type="checkbox"/> Возросшая потребность в высокопрофессиональных знаниях и навыках привели к тому, что специалистов стало очень дорого (а иногда и невозможно) заменить, а неопытные работники стали менее полезными. <input type="checkbox"/> Разработка новых видов топологии сети. <input type="checkbox"/> Стимулирование спроса и предложения в мировых информационных ресурсах.
--	---

<p>7. Логистическая координация стала одним из способов получения конкурентного преимущества на рынке.</p> <p>8.....</p> <p>9.....</p> <p>10.....</p>	
<p>II. Структура информационных систем основана на следующих компонентах:</p> <ul style="list-style-type: none"> • Данные, информация и знания. • Аппаратные средства. • Языки программирования. • Средства анализа и проектирования. • Системные разработчики и обслуживающий персонал. • • 	<p><input checked="" type="checkbox"/> Прикладное программное обеспечение.</p> <p><input type="checkbox"/> Системное программное обеспечение</p> <p><input type="checkbox"/> САПР</p> <p><input checked="" type="checkbox"/> Коммуникации.</p>
<p>III. Основные принципы и организационная структура Интернет:</p> <ul style="list-style-type: none"> • принцип открытой топологии; • принцип маршрутизации; • • 	<p><input type="checkbox"/> принцип создание общей инфраструктуры;</p> <p><input type="checkbox"/> принцип эволюционирования новые приложения;</p> <p><input checked="" type="checkbox"/> принцип децентрализованного управления;</p> <p><input checked="" type="checkbox"/> принцип интегрированной доставки информации.</p>
<p>IV. Сколько уровней в эталонная модель OSI</p>	<p><input type="checkbox"/> 3;</p> <p><input checked="" type="checkbox"/> 7;</p> <p><input type="checkbox"/> 9;</p> <p><input type="checkbox"/> 10;</p> <p><input type="checkbox"/> 12.</p>
<p>V. Что такое «trailer» в модуле данных PDU?</p>	<p><input type="checkbox"/> небольшой видеоролик;</p> <p><input type="checkbox"/> трассировщик;</p> <p><input checked="" type="checkbox"/> концевик;</p> <p><input type="checkbox"/> устройство для передачи данных.</p>
<p>VI. Сколько уровней в сетевой модели TCP/IP?</p>	<p><input type="checkbox"/> 2;</p> <p><input checked="" type="checkbox"/> 4;</p> <p><input type="checkbox"/> 6;</p> <p><input type="checkbox"/> 7;</p> <p><input type="checkbox"/> 8;</p> <p><input type="checkbox"/> 9.</p>
<p>VII. Что такое протокол сети Интернет?</p>	<p><input checked="" type="checkbox"/> единый набор правил, определяющих способ взаимодействия узлов сети;</p> <p><input type="checkbox"/> устройство, содержащее запись всего, что было описано, сделано и решено при компиляции;</p> <p><input type="checkbox"/> задаёт набор правил взаимодействия функциональных блоков, расположенных на одном уровне;</p> <p><input type="checkbox"/> описывает: синтаксис сообщения, имена элементов данных, операции управления и состояния;</p> <p><input type="checkbox"/> регистрация событий в хронологическом порядке.</p>

VIII. Перечислите важнейшие задачи для межсетевого уровня TCP/IP	<input type="checkbox"/> Коммутация IGMP; <input checked="" type="checkbox"/> Фрагментация MTU; <input checked="" type="checkbox"/> Адресация; <input type="checkbox"/> Репакинг данных в контейнеры меньшим размером; <input checked="" type="checkbox"/> Маршрутизация; <input type="checkbox"/> Идентификация провайдера.
IX. Что такое DNS?	<input checked="" type="checkbox"/> система доменных имён; <input type="checkbox"/> система цифрового нормирования; <input type="checkbox"/> система передачи данных; <input type="checkbox"/> динамическая система имён; <input type="checkbox"/> система загрузочных сетей.
X. Что такое мультимплексирование?	<input type="checkbox"/> идентификация отдельных протоколов; <input type="checkbox"/> дробление входящего потока данных на отдельные блоки, доставляемые соответствующим приложениям; <input type="checkbox"/> эмуляция терминала; <input type="checkbox"/> работа электронной почты; <input checked="" type="checkbox"/> объединение различных источников исходящих данных в единый исходящий поток; <input type="checkbox"/> передача файлов.
XI. Виды адресации TCP/IP?	<input checked="" type="checkbox"/> Физическая числовая <input checked="" type="checkbox"/> Логическая числовая; <input type="checkbox"/> Континуальная; <input type="checkbox"/> Дискретная; <input type="checkbox"/> Экспоненциальная; <input checked="" type="checkbox"/> Символическая.
XII. Сколько типов данных можно передавать по стандарту телеписьма?	<input type="checkbox"/> 3; <input type="checkbox"/> 5; <input type="checkbox"/> 6; <input checked="" type="checkbox"/> 7; <input type="checkbox"/> 8; <input type="checkbox"/> 9.
XIII. Что такое транзакция?	<input checked="" type="checkbox"/> Группа последовательных операций, которая представляет собой логическую единицу работы с данными; <input type="checkbox"/> Двухфазный протокол фиксации данных в сети; <input type="checkbox"/> Минимальная логически осмысленная операция, которая имеет смысл и может быть совершена только полностью (или полностью отменена); <input type="checkbox"/> Обналичивание денежных средств по средствам терминалов (банкоматов).

3. ДОСТУП К ИНФОРМАЦИОННЫМ РЕСУРСАМ

I. Кем были разработаны принципы информационного поиска	<input type="checkbox"/> Минковским; <input type="checkbox"/> Ляпунов; <input type="checkbox"/> Шеннон; <input checked="" type="checkbox"/> Баттенон;
--	--

	<input type="checkbox"/> Винер; <input type="checkbox"/> Харкевич.
II. Что входит в состав информационно-поискового языка входят: 1. Кодовый словарь; 2. Словарь входов; 3. Вспомогательные средства языка индексирования;	<input checked="" type="checkbox"/> Словарь индексационных терминов; <input type="checkbox"/> Язык запросов; <input type="checkbox"/> Семантика; <input type="checkbox"/> Синтаксис; <input type="checkbox"/> Тезаурус; <input checked="" type="checkbox"/> Правила использования языка индексирования.
III. Сколько этапов имеет процедура поиска?	<input checked="" type="checkbox"/> 5; <input type="checkbox"/> 7; <input type="checkbox"/> 8; <input type="checkbox"/> 10; <input type="checkbox"/> 13; <input type="checkbox"/> 21.
IV. Что такое пертинентность?	<input type="checkbox"/> Обозначение количество соответствующих запросу документов в исходном массиве; <input checked="" type="checkbox"/> Соотношение объема полезной для него информации к общему объему полученной информации; <input type="checkbox"/> Формальное соответствие информации, выдаваемой системой, запросу; <input type="checkbox"/> Творческое осмысление информационной потребности пользователя.
V. Дайте краткое определение статической составляющей Web – пространства.	
VI. Дополните структуру топология Web-пространства: 1. Центральное ядро; 2. «Отправные» Web-страницы (IN);	<input checked="" type="checkbox"/> «Оконечные» Web-страницы (OUT); <input checked="" type="checkbox"/> «Отростки»; <input type="checkbox"/> Иерархическая структура; <input type="checkbox"/> Фрактал; <input type="checkbox"/> «Острова»; <input type="checkbox"/> «Полученные» Web-страницы (OUT).
VII. Что такое invisible web?	<input type="checkbox"/> Невероятный; <input type="checkbox"/> Недоступный; <input type="checkbox"/> Неформальный; <input type="checkbox"/> Платная сеть; <input checked="" type="checkbox"/> Скрытый.
VIII. Что такое контент?	<input type="checkbox"/> Размещенный материал на веб-сайте с разрешения правообладателя, являющийся результатом интеллектуального труда и охраняющийся законом об авторском праве; <input type="checkbox"/> Доступное содержание веб-сайта; <input checked="" type="checkbox"/> Любое информационно значимое наполнение информационного ресурса тексты, графика, мультимедиа - вся информация, которую пользователь может скопировать (загрузить) на диск компьютера с соблюдением соответствующих законностей, как правило, только для личного пользования; <input type="checkbox"/> Это агентства копирайтинга, копирайтеры.

	ры-фрилансеры, биржи текстового формата.
IX. Какие типы лексических единиц предусмотрены при построении тезаурусов: <ul style="list-style-type: none"> отдельные слова (существительные, прилагательные, глаголы, наречия); аббревиатуры; сокращения слов и словосочетаний; 	<input type="checkbox"/> стоп-слова; <input checked="" type="checkbox"/> словосочетания; <input type="checkbox"/> силлогизмы; <input checked="" type="checkbox"/> лексически весомые компоненты сложных слов.
X. Что такое сюжетная цепочка?	
XI. Операторы контекстной близости	<input type="checkbox"/> AND; <input checked="" type="checkbox"/> BEFORE; <input type="checkbox"/> NOT; <input checked="" type="checkbox"/> FAR; <input type="checkbox"/> OR; <input checked="" type="checkbox"/> ADJ; <input type="checkbox"/> TRUE/FALSE; <input checked="" type="checkbox"/> NEAR.

4. МИРОВОЙ РЫНОК ИНФОРМАЦИОННЫХ УСЛУГ

I. Развитие информационных технологий привело к появлению двух фундаментальных возможностей, обеспечивающих конкурентные преимущества фирмам, которые их используют:	<input type="checkbox"/> санкционированный промышленный шпионаж; <input type="checkbox"/> передача данных по новым сетевым протоколам; <input checked="" type="checkbox"/> доступ к удаленным информационным ресурсам; <input type="checkbox"/> применение телеконференций; <input checked="" type="checkbox"/> распределенная обработка информации.
II. Поставщиками информационных продуктов и услуг могут выступать: <ul style="list-style-type: none"> центры, где создаются и хранятся базы данных; службы телекоммуникации и передачи данных; коммерческие фирмы; информационные брокеры; 	<input checked="" type="checkbox"/> центры, распределяющие информацию на основе баз данных; <input type="checkbox"/> специалисты по аутсорсингу бизнес-процессов; <input type="checkbox"/> интернет-трейдеры; <input checked="" type="checkbox"/> специальные службы, куда стекается информация по конкретной сфере деятельности для ее анализа, обобщения, прогнозирования, например консалтинговые фирмы, банки, биржи.
III. Основными компонентами рынка информационных услуг являются: <ul style="list-style-type: none"> нормативно-правовая составляющая; организационная составляющая; 	<input type="checkbox"/> деловая составляющая; <input type="checkbox"/> потребительская составляющая; <input type="checkbox"/> сервисная составляющая; <input checked="" type="checkbox"/> техническая составляющая; <input checked="" type="checkbox"/> информационная составляющая.
IV. Основы сектора информационного рынка.	<input type="checkbox"/> 3; <input type="checkbox"/> 4; <input checked="" type="checkbox"/> 5; <input type="checkbox"/> 7; <input type="checkbox"/> 9.

5. ГОСУДАРСТВЕННЫЕ ИНФОРМАЦИОННЫЕ РЕСУРСЫ

<p>I. Какие организации осуществляют контроль за информационными ресурсами на общероссийском (государственном) уровне:</p>	<p><input checked="" type="checkbox"/> Министерство РФ по связи и массовых коммуникаций; <input type="checkbox"/> Управление «К» МВД России; <input checked="" type="checkbox"/> Федеральное агентство правительственной связи и информации при Президенте РФ и Государственная техническая комиссия при Президенте РФ; <input type="checkbox"/> Министерство образования и науки РФ; <input type="checkbox"/> Следственный комитет при прокуратуре РФ; <input checked="" type="checkbox"/> Российское агентство по патентам и товарным знакам; <input type="checkbox"/> Министерство Обороны РФ; <input checked="" type="checkbox"/> Министерство имущественных отношений РФ; <input type="checkbox"/> Общественный Совет при ФСБ России; <input type="checkbox"/> Национальный антитеррористический комитет России.</p>
<p>II. В состав информационного фонда ГМЦ входит: общеэкономическая, отраслевая и региональная статистическая информация, объединенная в блоки</p> <ul style="list-style-type: none"> - Наука и инновации. - Сельское хозяйство. - Промышленность. - Бухгалтерская отчетность. - Уровень жизни и доходы населения. 	<p><input checked="" type="checkbox"/> <i>Регистр промышленных предприятий;</i> <input type="checkbox"/> <i>Государственная система научно-технической информации;</i> <input checked="" type="checkbox"/> <i>Регистр сельскохозяйственных предприятий;</i> <input checked="" type="checkbox"/> <i>Регистр городов России;</i> <input type="checkbox"/> <i>Пенсионный фонд;</i> <input checked="" type="checkbox"/> <i>Паспорт территорий Российской Федерации;</i> <input type="checkbox"/> <i>Здравоохранение;</i> <input checked="" type="checkbox"/> <i>Единый государственный регистр предприятий и организаций;</i> <input checked="" type="checkbox"/> <i>Статистика внешнеэкономической деятельности;</i> <input type="checkbox"/> <i>Научно-технические библиотеки.</i></p>

6. ПРАВОВАЯ ИНФОРМАЦИЯ

<p>I. Что является источником правовой информации в России (согласно ФАПСИ и Минюста России)?</p>	<p><input checked="" type="checkbox"/> Государственный реестр общественных объединений и религиозных организаций; <input checked="" type="checkbox"/> База данных действующего российского законодательства «Эталон»; <input checked="" type="checkbox"/> Комплекс баз данных правовой информации, объединенных в программно-технологический комплекс «Фонд»; <input type="checkbox"/> Международное публичное право (МПП); <input checked="" type="checkbox"/> Официальная электронная версия бюллетеня «Собрание законодательства</p>
--	---

	<p>Российской Федерации»;</p> <p><input type="checkbox"/> Информационно-правовая система «Консультант»;</p> <p><input checked="" type="checkbox"/> Официальная электронная версия «Бюллетень нормативных актов федеральных органов исполнительной власти Российской Федерации»;</p> <p><input checked="" type="checkbox"/> Электронная версия «Бюллетеня Верховного Суда Российской Федерации»;</p> <p><input checked="" type="checkbox"/> Электронная версия «Вестник Высшего Арбитражного Суда Российской Федерации»;</p> <p><input checked="" type="checkbox"/> Электронная версия «Бюллетень международных договоров Российской Федерации»;</p> <p><input type="checkbox"/> Информационно-правовая система «Гарант».</p>
<p>II. Перечислите основными параметрами, позволяющие определить качество содержания информационной базы:</p> <p>- полнота информации;</p> <p>.....</p> <p>.....</p>	<p><input type="checkbox"/> средства актуализации информации;</p> <p><input type="checkbox"/> дополнительные сервисные функции;</p> <p><input checked="" type="checkbox"/> оперативность обновления информации;</p> <p><input checked="" type="checkbox"/> достоверность информации;</p> <p><input type="checkbox"/> поисковые возможности.</p>
<p>III. Перечислите основные поисковые и сервисные возможности.</p>	<p><input type="checkbox"/> возможность сохранять личные предпочтения и подборки документов;</p> <p><input type="checkbox"/> поиск по сервисным функциям;</p> <p><input type="checkbox"/> улучшенная навигация по информационной базе;</p> <p><input checked="" type="checkbox"/> поиск по тексту документов;</p> <p><input checked="" type="checkbox"/> поиск по реквизитам документов.</p>

VII ВОПРОСЫ ДЛЯ ПОДГОТОВКИ К ЗАЧЕТУ

по дисциплине "МИРОВЫЕ ИНФОРМАЦИОННЫЕ РЕСУРСЫ"
по направлению подготовки 230700 – "Прикладная информатика"
специализации – прикладная информатика в экономике,
3 курс (5 семестр)

1. Свойства экономической информации. Определение информационного ресурса.
2. Виды бизнес-информации. Классификация источников и роль бизнес-информации.
3. Компоненты информационных систем.
4. Технологии передачи и обмена информационными ресурсами.
5. Интеграционные процессы на основе современных информационных технологий.
6. Основные принципы и организационная структура Интернета.
7. Инфраструктура сети Интернет.
8. Эталонная модель OSI.
9. Сравнение уровней эталонной модели OSI и сетевой модели TCP/IP.
10. Уровень доступа к сети модели TCP/IP.
11. Межсетевой уровень модели TCP/IP.
12. Транспортный уровень модели TCP/IP.
13. Прикладной уровень модели TCP/IP.
14. Номера протоколов и портов TCP/IP, сокет, инкапсуляция данных TCP/IP.
15. Разновидности схем адресации в сети Интернет.
16. Классы IP-адресов.
17. IP-сети, подсети и маски.
18. Расчет маски подсети постоянной длины.
19. Расчет суперсетей.
20. Частные и общедоступные IP-адреса.
21. Структура баз данных DNS, работа серверов доменных имен.
22. Электронная почта: структура адреса, формат сообщения, схема передачи почтовых сообщений по протоколу SMTP.
23. Система World Wide Web: концепция WWW (схема), основные технологии, составляющие основу WWW.
24. Универсальная форма адресации информационных ресурсов, схема взаимодействия элементов при передаче данных по протоколу HTTP.
25. Служба FTP: ключевые компоненты FTP-сообщения, процесс взаимодействия клиента и сервера по протоколу FTP.
26. Основы информационного поиска, предметное индексирование и механизм поиска, типовая схема ИПС.
27. Стратегии поиска, концептуальная схема процедуры поиска, ИПС глобальных сетей – концептуальная схема.
28. Два аспекта полноты, определяющие качество ИПС: релевантность и пертинентность. Недостатки традиционного поиска. Статическая и динамическая составляющие Web-пространства.
29. Топология Web-пространства: топологическая модель поверхностного Web. Скрытый Web, типы скрытых ресурсов.
30. Средства навигации в сети Интернет: каталоги, информационно-поисковые серверы, порталы. Основные типы сервисов, включаемых в портал.
31. Лингвистическое обеспечение информационно-поисковых серверов, стоп-слова, морфемный анализ, тезаурус. Отличие профессиональных запросов от запросов пользователей.

32. Методологические основы поисковой процедуры: семантические методы, группировка результатов, сюжетный подход (схема построения сюжетных цепочек).
33. Этапы поисковой процедуры. Запросы пользователей: ранжирование откликов, поиск по словам и словоформам, логические операторы, операторы контекстной близости, поиск по параметрам.
34. XML – история и цель создания. Преимущества по сравнению с HTML. XML как модель данных.
35. Характеристика мирового рынка информационных услуг. Основные секторы рынка информационных продуктов и услуг (схема).
36. Производители мировых информационных ресурсов. Сетевые информационно-поисковые службы.
37. Программные средства для поиска информации: перечень, основные особенности, производители.
38. Государственные информационные ресурсы: общая характеристика. Правовые акты, определяющие порядок их формирования и использования. Организации, ответственные за формирование и использование.
39. Информационные ресурсы министерств и ведомств. Информационные ресурсы региональных и муниципальных органов власти.
40. Статистическая информация. Научно-техническая информация. Информация о природных ресурсах и процессах.
41. Источники правовой информации. Основные агентства-производители на рынке правовой информации.
42. Информационные справочные правовые системы: назначение, основные свойства информационных банков СПС, обзор рынка СПС в России.
43. Технологии доступа к правовым базам данных: поисковые и сервисные возможности, технологии передачи информации.
44. Виды биржевой и финансовой информации. Программные средства анализа биржевой и финансовой информации.

VIII КОНТРОЛЬНЫЕ РАБОТЫ

Контрольная работа выполняется в виде письменного ответа на указанные в индивидуальном задании вопросы.

Изложение материала необходимо обосновывать положениями законодательных актов, нормативных и инструментальных материалов, регулирующих деятельность субъектов в области мировых информационных ресурсов.

Объем контрольной работы не должен превышать 16-18 страниц рукописного текста.

Перечень вопросов для выполнения контрольной работы

(порядок определения варианта определяется студентом самостоятельно)

Раздел I. КОНЦЕПЦИЯ ДИСЦИПЛИНЫ

А. Информации, ее определение, виды и свойства. Информационные процессы. Информационные революции.

Б. Информатизация общества. Информатизация как совокупность процессов и явлений, связанных с целенаправленной обработкой информации, связанной с применением средств вычислительной техники и связи, информационных технологий и программного обеспечения. Окинавская Хартия Глобального информационного общества.

В. Мировые информационные ресурсы: определение, классификация и характеристика основных структур (баз данных, сетей) по различным признакам. Определение информационных ресурсов (ИР), представление информации в виде ИР.

Г. Формы, виды, классификация, измерение ИР. Представление ИР в глобальных информационных сетях. ИР России. Мировые информационные ресурсы. Объект, предмет, цель и задачи дисциплины.

Раздел II. ЗАКОНОДАТЕЛЬСТВО, РЕГУЛИРУЮЩЕЕ ОТНОШЕНИЯ В ОБЛАСТИ МИР

Д. Государственная политика в области ИР. Понятие. Средства обеспечения государственной политики. Информационная политика РФ. Федеральная целевая программа «ЭЛЕКТРОННАЯ РОССИЯ на 2002–2010 годы». Национальный доклад «Информационные ресурсы России». Информационное неравенство. Информационная безопасность государства. Меры по обеспечению информационной безопасности. Доктрина информационной безопасности Российской Федерации.

Е. Законодательство РФ в области ИР. Предпосылки государственного регулирования в области информационных технологий. Федеральные законы: «О праве на информацию», «Об информации, информатизации и защите информации», «Об электронной цифровой подписи», «О правовой охране программ для ЭВМ и баз данных», «Об обязательном экземпляре документов».

Ж. Законодательство ЕС и США в области ИР. Особенности правового регулирования мировых ИР. Влияние международного законодательства на развитие сферы обращения ИР в РФ и мире. Национальное и международное законодательство.

З. Конфликты юрисдикции, связанные с характером мировых ИР.

И. Проблемы правового регулирования. Авторские и смежные права. Законы США об авторских правах в цифровую эпоху. Влияние международных и национальных коммерческих организаций на обращение ИР. Проблемы налогообложения сетевых коммерческих организаций. Информационное пиратство.

Раздел III. МИРОВЫЕ ИНФОРМАЦИОННЫЕ СЕТИ: ИХ СТРУКТУРА, ПРАВИЛА ПОИСКА

Й. Глобальные информационные сети. Причина возникновения и история развития. Классификация информационных сетей. Корпоративные сети. Образовательные сети. Общедоступные и коммерческие сети. Обзор технологий размещения и доступа к информации в

информационной сети. Сети BBS. FIDO. Bitnet. SprintNet. AOL. Internet 2. Next Generation Internet.

К. Семь Internet. История образования. Принципы организации. Преимущества и недостатки. Способы размещения информации в Internet. Файловые архивы. Сайты. Конференции. Рассылки. Способы доступа к информации в Internet. Организация информационного трафика. Открытая и закрытая информация в Internet. Обращение к специализированным базам данных и информационным сетям. Получение информации по запросу, подписка на рассылки. Настройка динамических страниц. Обмен информацией в сети. Peer-to-peer.

Л. Поиск информации в сети Internet. Каталогизация информации в сети Internet. Адресация в сети Internet. Проблема поиска информации и сортировки найденного по релевантности. Особенности поиска текстовой информации. Кодировки, форматы записи. Поиск файлов, визуальной, аудио и мультимедийной информации. Каталоги IP. Рейтинг IP.

М. Поисковые системы (ПС). Принципы работы ПС. Понятие запроса к ПС. Вид запроса в ПС. Обработка, каталогизация, индексирование, сортировка, анализ информационных ресурсов, выдача информации по запросу, выделение главного и отсеечение второстепенного, повторяющегося (резюмирование) первичная аналитическая обработка запроса, уточнение требований к выборке информации. Возможности ПС. Классификации ПС. Сравнение ПС, действующих в русскоязычной части Internet. Международные ПС. Меры по оптимизации поиска.

Н. Электронная почта. Принцип работы. Виды предоставляемых услуг. Получение информации при помощи почтовых служб. Уязвимости работы с почтой: получение вирусов, spam.

Раздел IV. ИНФОРМАЦИЯ И БИЗНЕС

О. Торговля в сети Internet. Торговля традиционными товарами. Преимущества и недостатки сетевого (Internet) магазина. Особенности товарного предложения. Способы организации

торговли. Доставка товара. Система оплаты. Клиент сетевого магазина. История развития сетевой торговли. Сетевая торговля в РФ. Кризис DOT-COMов.

П. Реклама в сети Internet. Виды рекламы. Особенности Internet рекламы. Эффективность Internet рекламы.

Р. Оплата в сети Internet. Способы оплаты в сетевых магазинах. Платежные системы в Internet, электронные деньги. WebMoney, PayCash, Yandex. Деньги. Особенности денежного обращения в сети Internet.

С. Рынки информационных ресурсов: особенности спроса, предложения, рыночного равновесия. Особенности торговли IP: информация, музыка, документы, программные продукты.

Т. Виды предоставления пользователю прав на программный продукт. Платный продукт, бесплатный продукт (freeware), пробный продукт (demo, shareware), с рекламной нагрузкой (bannerware), со сбором информации о покупателе (spyware).

Ф. Ограничение покупателя в правах пользования приобретенным виртуальным товаром: по времени, по функциональности, по технической поддержке, по числу использований, по передаче другому пользователю, по изменению и совершенствованию, по географическому, национальному признаку. Товары для личного и коммерческого использования.

У. Виртуальные товары и услуги. Особенности виртуальных товаров и услуг. Виртуальная собственность.

Х. Авторское право и коммерция в сети Internet. Вопросы отчуждения собственности на информационный продукт. Виды лицензий. GNU.

Раздел V. ОБЕСПЕЧЕНИЕ БЕЗОПАСНОСТИ В РАБОТЕ С IP.

У. Несанкционированное использование программных продуктов и IP (информационное пиратство) в РФ. Понятие информационного пиратства. Причины и следствия пи-

ратства. Государственная политика по борьбе с пиратством. Борьба корпорации Microsoft с пиратством.

Ч. Несанкционированный доступ к ИР и информационному трафику. Определение несанкционированного доступа. Меры по защите информации. Аппаратные и программные средства защиты ИР.

Ш. Электронная цифровая подпись (ЭЦП). Анонимность в сети Internet. Потребность в авторизации потребителя сетевых услуг. Понятие ЭЦП. Сфера обращения ЭЦП. Международное и российское законодательство об ЭЦП. Примеры использования ЭЦП в РФ.

Щ. Защита от сетевых вирусов. Особенности компьютерных вирусов в информационных сетях. Классификация сетевых вирусов. Сетевые черви, трояны, шпионские модули.

Ъ. Меры безопасности при работе с ИР. Меры безопасности при работе с электронной почтой. Программные средства защиты информационных ресурсов.

Раздел VI. ИСПОЛЬЗОВАНИЕ ИР В ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ.

Ы. Технология и практика эффективного взаимодействия индивидуального и коллективного пользователя с ИР через специализированные сетевые структуры.

Б. Особенности работы индивидуального и коллективного пользователя с ИР. Использование специализированных, корпоративных ИР. Учет использования ИР в профессиональной деятельности.

Э. Оценка затрат на использование ИР на конечную стоимость выполнения работ. Комплексная оценка эффективности использования мировых ИР.

Ю. Планирование использования ИР. Средства управления проектами. Обзор. Назначение. Сфера применения. Оценка экономической эффективности.

Я. Технология microsoft.net. Определение. Предпосылки возникновения. Особенности работы на платформе microsoft.net. Альтернативные разработки фирмы Sun.

Выбор индивидуального задания по контрольной работе

Контрольная работа выполняется по приведенным вопросам и включает 10 вопросов, которые формируется по буквам фамилии, имени, отчества в именительном падеже слева направо без повторения. Например, ИВАНОВ ОЛЕГ ПЕТРОВИЧ выполняет контрольную работу по вопросам **И, В, А, Н, О, Л, Е, Г, П, Т.**

При выполнении контрольных работ используется литература, приведенная в разделе данной рабочей программы.

ГЛОССАРИЙ

А

Авторизация (клиента) - совершение определенных действий (например, введения логина и пароля) для получения расширенных полномочий при работе с сайтом - например, отправка сообщений, заполнение форм заказа и др.

Административная система - management system - система, обеспечивающая управление веб-узлом либо его частью. Осуществляет также контроль доступа и защиту от несанкционированного вторжения.

Анимация - придание графическому изображению динамических свойств (картинки чередуются в определенном порядке, "бабочка машет крыльями", "кузнец стучит молотком", вращение, перемещение графического объекта и др.). Как правило, рекомендуется не использовать анимацию на web-сайте без крайней необходимости, т.к. она отвлекает пользователя от основной информации, утомляет глаза, да и просто может раздражать.

Аудитория (веб-сайта) - совокупность посетителей веб-узла. Объем аудитории, вопреки распространенному мнению, не является однозначной характеристикой успеха Интернет-проекта. Web-разработчики шутят - "если вы хотите иметь гигантскую аудиторию, поместите на ваш сайт фотографии обнаженных женщин и ссылки на порноресурсы". Но ведь Вам не нужны такие посетители?

Аудит сайта - комплексный анализ сайта, проводимый с целью выявления недостатков сайта и выработке рекомендаций по их устранению.

Б

База данных - совокупность данных, организованных по строгим правилам. Создание баз данных и управление ими относится к области web-программирования и необходимо в тех случаях, где посетителю сайта предлагается выбрать что-то (товар, услугу, информацию) из общей совокупности. Вообще, БД - оптимальная система хранения любой информации на сайтах выше определенного уровня сложности. Каталог товаров, регистрационные данные пользователей, архив сообщений гостевой книги и форума - удобное хранилище любой информации, которая может понадобиться пользователю или разработчику сайта.

Back-office - интеграция информационно-учетной системы компании с интернет, информация об онлайн-заказах на закупку/продажу автоматически поступает для обработки во внутреннюю систему. Проще говоря, сотрудники компании (менеджеры, бухгалтеры, администраторы и пр.), которые стоят за веб-интерфейсом её сайта, т.е. делают всю реальную работу по выполнению заказа.

Баннер - banner - изображение или текстовый блок рекламного характера, представляющий собой гиперссылку на страницы с подробным описанием продукта или услуги. Баннер может быть статическим или анимированным. Необходимо учитывать, что чересчур навязчивые баннеры (особенно всплывающие) зачастую негативно воспринимаются посетителями web-страниц, причем это негативное отношение может быть перенесено на сам ресурс, разместивший у себя подобный баннер.

Баннерная сеть - объединение сайтов, на страницах которых на определенных условиях размещаются баннеры участников, а также баннеры компаний, оплативших размещение своего баннера владельцу баннерной сети.

"Бизнес для бизнеса" - business-to-business, B2B - сектор рынка, ориентированный на организацию работы с контрагентами и партнерами в процессе производства и продажи товаров или услуг. В данный сектор входят все торговые отношения между различными

фирмами, организация поставок, продаж, согласование контрактов и планов. Под системами B2B понимают: 1) Procurement Systems - системы связи потребителей продукции с производителями, покупателей с продавцами. Как правило, в этих системах покупателями и продавцами выступают юридические лица и, как правило, существуют в той или иной мере развитые бэк-офисные системы. 2) Supply Chain Management - обеспечивает предприятию бесперебойную информацию о поставках и логистику всех контрагентов для основного производства. Обычно предприятия начинают свою деятельность в секторе B2B с создания собственного вертикального портала.

"Бизнес для потребителя" - business-to-customer, B2C - сектор рынка, ориентированный на работу с потребителями товаров или услуг. В сектор B2C входят электронные магазины, организации, торгующие продукцией через Интернет непосредственно для потребителей и др.

Брандмауэр - firewall - средство защиты компьютера (локальной сети) от несанкционированного доступа (в т.ч. от проникновения вирусов) по Интернет-каналам, иначе - барьер, запрещающий доступ к защищаемым узлам по всем протоколам, кроме разрешенных.

Браузер (броузер) - программа навигации по сети Интернет и просмотра веб-ресурсов, позволяет запрашивать и просматривать файлы в Интернет. В современных версиях ОС Windows используется встроенный браузер Internet Explorer (IE), однако пользователь может установить на свой компьютер любой другой браузер, если это покажется ему более удобным. По последним данным, большинство пользователей используют при работе с Интернетом IE, на втором месте по популярности - Mozilla Firefox, на третьем - Opera.

В

WWW - world wide web, Web - сервис сети Интернет, позволяющий получать доступ к массивам информации, размещенным в глобальной сети.

Веб - то же, что WWW

Веб-броузер - то же, что Броузер

Веб-дизайн - особая разновидность художественно-проектной деятельности, направленная на создание и обеспечение удобства использования веб-ресурсов с учетом особенностей восприятия пользователей в сети Интернет.

Веб-кольцо - сайты сходной тематики, объединенные взаимными ссылками. Эффективный инструмент для поиска информации и продвижения контент-ресурсов.

Веб-сайт - web site - совокупность веб-страниц, объединенных по смыслу, имеющих общую структуру и навигацию.

Веб-сервер (сервер WWW) - сервер, предназначенный для представления взаимосвязанной мультимедийной информации и содержимого баз данных в Интернет. Место, где физически расположены файлы веб-страниц, базы данных и пр., а также серверные программы, взаимодействующие с обращающимся к ним браузером и организующие, по соответствующему запросу браузера, компоновку требуемой веб-страницы.

Веб-страница - page - статический, или динамически сформированный HTML-файл, который можно просмотреть с помощью браузера.

Видеоконференция - video conferencing - методология проведения совещаний и дискуссий между группами удаленных пользователей при помощи технологии высокоскоростной передачи видео в среде Интернет.

Виртуальный офис - virtual office - интернет-ресурс, или его часть, позволяющие географически разбросанным сотрудникам компании, взаимодействовать посредством единой системы для обмена, хранения, обработки и передачи информации и управляющих воздействий.

Владелец информации - owner of information - субъект, в непосредственном ведении которого в соответствии с законом находится информация.

"Всемирная паутина" - то же, что WWW

WYKIWYL феномен - "что вам знакомо, то вам и нравится". Психологическая особенность человека, заключающаяся в том, что он с большей лояльностью относится к уже знакомым вещам или тому, что напоминает ему уже знакомую вещь. Обычно учитывается при разработке пользовательского интерфейса программ (в частности веб-приложений).

Webmoney Transfer - система электронных расчетов, использующая специальные электронные деньги - Webmoney (WM). 1 WM может быть эквивалентна 1 рублю (WMR), 1 доллару США (WMZ), 1 евро (WME).

Г

Гиперсреда - hypermedia - технология представления любых видов информации в виде относительно небольших блоков, ассоциативно связанных друг с другом посредством гипертекстовых технологий.

Гиперссылка - hyperlink - объект веб-страницы (текст или изображение), устанавливающий связь с другим объектом в сети Интернет. Гиперссылки могут связывать между собой объекты, находящиеся на одной странице, на одном сайте, а также объекты, находящиеся в самых различных частях сети Интернет.

Главная страница - первая страница, "обложка" сайта. Как правило, несет основную презентационную и навигационную нагрузку.

Д

Домашняя страница - home page - первая страница, которая загружается браузером и появляется на экране при открытии браузера. Если в настройках вашего браузера в качестве домашней страницы выбрана опция "about blank", т.е. "с чистого листа", при открытии браузера будет загружаться чистая страница без текста и изображений, а вам необходимо будет ввести в адресную строку адрес того ресурса, на который вы хотите попасть. Домашней страницей (страничкой) так же иногда именуют небольшие сайты, сделанные для частных лиц и несущие информацию частного характера.

Данных база - информация, представленная в специальном формализованном виде, пригодном для автоматизированной обработки.

DNS - Domain Name Service - распределенная иерархическая база данных, содержащая информацию об именах серверов Интернета и позволяющая по имени системы определить ее IP-адрес. Иерархическая организация DNS наглядно проявляется в структуре доменного имени. Каждый из разделенных точками компонентов адреса соответствует определенному домену (или зоне).

Домен - группа компьютеров или других устройств в Сети, управляемых как единое целое в рамках общих правил и процедур. В Интернет домены определяются IP-адресами. Считается, что все устройства, "делящие между собой" общую часть IP-адреса, находятся в одном и том же домене. Домены бывают трех уровней. Домены верхнего уровня существуют практически для всех признанных ООН государств и территорий. Все они состоят из двух букв, а их поддержкой в каждой стране по поручению ICANN занимается специальная организация. Кроме доменов для стран, существует ряд глобальных доменов верхнего уровня, таких, как .com, .net,

.biz и ряд других. Домен второго уровня - это домен типа www.adres.ru, домен третьего уровня - домен типа www.adres.domen2.ru. При этом в нашем примере .ru - это домен верхнего уровня.

Доменное имя - адрес ресурса в сети Интернет. Представляет собой буквы или буквенные сочетания, которые разделены точками на отдельные информационные блоки. На последнем месте в нем стоит домен первого уровня. Имя домена первого уровня может отражать географическое положение сервера либо тематику предоставляемой информации. Домен второго уровня выбирается организацией - хозяином web-ресурса. Регистрация домена второго уровня является платной. Домен третьего уровня так же выбирается организацией - хозяином web-ресурса, однако в структуре доменного имени обязательно будет присутствовать название домена второго уровня (www.adres.narod.ru). Как правило, регистрация домена третьего уровня является бесплатной, однако при этом хозяин домена второго уровня оставляет за собой право размещать на вашем сайте баннеры произвольного содержания (не ставя вас в известность и не спрашивая вашего мнения на этот счет), а также изменять условия предоставления вам места на своем сервере. В доменном имени не имеет значения регистр написания букв; использование заглавных букв вместо строчных не меняет имени домена. Не имеет значения, наберете ли вы в адресной строке "W-sight.com" или "w-sight.com" - вы все равно попадете к нам.

Е

Е-mail - электронная почта. Передача сообщений (в виде текстового или html-файла) по сети при помощи специализированных почтовых программ или онлайн-сервисов. Сообщение, передаваемое по электронной почте, может содержать тексты, таблицы, графики, а также присоединенные файлы любого формата. Распространено также присоединение (злоумышленниками) к файлу сообщения или приложения вирусов или других вредоносных микропрограмм, проникающих таким образом на компьютеры пользователей.

Ж

Живой Журнал - <http://www.lifejournal.com> - (ой! оцепятка... ;))) конечно же, <http://www.lifejournal.com>) - популярный международный сервис для ведения онлайн-дневников (блогов) и сетевого общения.

И

Идентификатор - уникальное сочетание имени и пароля пользователя для обеспечения процесса его идентификации.

Идентификация пользователя - user identification - опознавание пользователей (по имени и паролю) для определения его полномочий - права на доступ к данным и выбора режима их использования.

Индексирование (документов) - indexing - процесс описания содержания документов и поисковых запросов в терминах информационно-поискового языка; назначение документу набора ключевых слов, отражающих его смысловое содержание. Применяется в библиотечно-информационных и электронных технологиях описания и поиска документов.

Интерактивность в Web (диалог с пользователем) - характеристика протекания процесса коммуникации в Интернет. Примерами интерактивности может служить ссылка на web-сайте, нажимая на которую, вы можете отправить письмо хозяевам сайта или отправить сообщение, которое появится на страницах сайта и на которое администрация сайта может дать свой ответ, разместив его в специально отведенном месте, доступном для вас и иногда для других посетителей.

Интернет - Internet - открытая мировая коммуникационная инфраструктура, состоящая из взаимосвязанных компьютерных сетей, обеспечивающая доступ к удаленной информации и обмен информацией между компьютерами. В глобальную сеть входят правительственные, академические, коммерческие, военные и корпоративные сети всего мира. Первоначально Internet был разработан для использования в американской армии и только затем стал сетью, которая широко используется академическими и коммерческими организациями. Пользователи, работающие в Internet, могут читать и загружать данные по любой теме практически со всего света. Более формально это зафиксировано в определении Federal Networking Council USA от 24. 10. 1995: "Интернет - глобальная информационная система, части которой логически взаимосвязаны друг с другом посредством уникального адресного пространства, основанного на протоколе IP или его последующих расширениях, способная поддерживать связь с использованием комплекса протоколов TCP/IP, их последующих расширений или других IP-совместимых протоколов, и которая обеспечивает, использует или делает доступным, публично или частным образом, коммуникационный сервис высокого уровня". См. также Сетевое информационное пространство.

Интернет-вещание - динамическое изменение информации, передаваемой по каналам Интернета: новостные ленты, видео, радио и т.д.

Интернет-магазин - то же, что Электронный магазин

Интернет-провайдер - Internet service provider, ISP - компания, предоставляющая пользователям доступ к Интернет.

Интернет-сообщество - сложившаяся в процессе общения и совместной деятельности относительно устойчивая система связей и отношений между пользователями сетевого информационного пространства. В отличие от межличностных контактов, агенты интернет-сообщества часто взаимодействуют на основе прямого обмена информацией по типу "все со всеми". Члены интернет-сообщества, как правило, связаны общими интересами, областью деятельности, сферой бизнеса и т.п.

Интернет-услуги - Internet services - услуги пользователям по обеспечению доступа в сеть, разработке, организационному и информационному сопровождению интернет-ресурсов, разработке и размещению рекламы в сети.

Интерфейс - interface - комплекс условий, обеспечивающих взаимодействие объектов в окружающем мире, в т.ч. информационном пространстве. Элементы диагностики и управления системой, представленные в удобном для взаимодействия с системой формате.

Интерфейс пользователя - то же, что Пользовательский интерфейс

Интранет -Intranet - внутренняя частная сеть организации, Intranet-сети защищены от общего доступа и предназначены для хранения информации и эффективного взаимодействия пользователей внутри одной организации; важно, что вся информация в них хранится в том же формате и организована по тем же принципам, что и в World Wide Web.

Информационная безопасность - information security - состояние защищенности информационной среды, обеспечивающее ее формирование, использование и развитие в интересах граждан, организаций, государства.

Информационная поддержка сайта - комплекс мероприятий по добавлению, удалению и редактированию новостей и другой информации на сайте.

Информационно-поисковая система - information retrieval system - система, предназначенная для поиска документов в информационных массивах, базах данных и всей совокупности информационных ресурсов.

Информационный ресурс - information resources - отдельные документы и массивы документов в информационных системах (библиотеках, архивах, фондах, банках данных, других видах информационных систем).

IP-адрес - уникальный адрес компьютера в Сети. Цифровые IP-адреса состоят из четырех целых чисел (от 0 до 255), разделенных точками, например 195.36.221.141. Числовая форма адреса используется компьютерами, для человека числовой адрес неудобен, поэтому его заменяет доменное имя.

К

Канал - channel - средство или путь, по которому передаются сигналы или данные.

Каталог - catalog - перечень однородных объектов, составленный в порядке, облегчающем их нахождение. В информационных комплексах - один из вспомогательных ресурсов, облегчающий поиск.

Качество информации - information quality - совокупность свойств, отражающих степень пригодности конкретной информации об объектах и их взаимосвязях для достижения целей, стоящих перед пользователем, при реализации тех или иных видов деятельности. В состав наиболее общих параметров входят: достоверность, своевременность, новизна, ценность, полезность, доступность.

Клиент-сервер - Client/server - сетевая архитектура, в которой все устройства являются либо клиентами, либо серверами. Клиентом (front end) является запрашивающая машина (обычно ПК пользователя), сервером (back end) - машина, которая отвечает на запрос пользователя. Оба термина (клиент и сервер) могут быть применены как к физическим устройствам, так и к программному обеспечению.

Ключ - key - конкретное секретное состояние некоторых параметров алгоритма криптографического преобразования данных, обеспечивающее выбор одного преобразования из совокупности возможных.

Кодирование - coding - процесс представления данных последовательностью символов.

Контент - content - любое информационно значимое наполнение сервера, информационного комплекса - тексты, графика, мультимедиа. Проще говоря, информационное содержание веб-страниц. Существенными параметрами контента является его объем, актуальность и релевантность.

Криптография - cryptography - способ изменения данных с целью сделать их непонятными для непосвященных лиц. Является важным компонентом системы безопасности данных. Её сущность заключается в превращении данных перед их передачей в бессмысленный набор символов либо сигналов и восстановлении данных в первоначальной форме после их приёма адресатом.

Л

Ламер - человек, не разбирающийся в тонкостях компьютерного дела, новичок, компьютерный "чайник"

LOL - принятое в Интернет-среде сокращение, текстовый смайлик, обозначающий что-то донельзя смешное. То же, что и :)))

М

Мобильный Интернет - технология беспроводного доступа в Интернет на основе протокола WAP.

Модем - modem - внешнее или внутреннее устройство, подключаемое к компьютеру для передачи и приема сигналов по разным линиям связи. Сокращение от "модулятор - демодулятор", что указывает на первоначальный принцип работы этого устройства: преобразование цифрового сигнала, полученного от компьютера, в аналоговую форму для передачи по проводным сетям и обратное преобразование принятого сигнала из аналоговой формы в цифровую. С развитием технологий понимается прежде всего, как устройство для формирования входящих и исходящих сигналов от компьютера.

Мониторинг - monitoring - форма организации исследований, обеспечивающая непрерывное поступление информации о том или ином объекте с целью анализа ожидаемого и конечного результатов.

Магистраль - backbone - основная линия связи, к которой подключена компьютерная сеть, а также основной участок передачи данных внутри локальной сети.

Мультимедиа - multimedia - визуальные и аудио эффекты под управлением интерактивного программного обеспечения. Обычно означает сочетание текста, звука и графики, анимации и видео.

О

Онлайновые технологии - on line - совокупность соответствующего программного обеспечения, человека-оператора и физических устройств (компьютеров и сетей), обеспечивающих синхронный обмен информацией в сетевом информационном пространстве в реальном времени.

Он-лайн обслуживание - консультационное обслуживание клиентов, происходящее в режиме реального времени (запрос-ответ), либо продажи посредством сети Интернет (например, клиент через Интернет делает заказ, оплачивает его при помощи телебанкинга или электронных платежных систем и получает товар по сети (программное обеспечение, информационные продукты - тексты, музыка, видео)).

Операционная система - operating system - комплекс программ, обеспечивающий в системе (компьютере) выполнение других программ, распределение ресурсов, планирование, ввод-вывод и управление данными.

Отклик баннера - отношение числа "кликов" на баннер к числу его показов.

П

Пакет - packet - автоматически формируемая производственная единица информации, передаваемая по сети или по каналу связи и содержащая собственно передаваемые данные и информацию об отправителе и адресате. Размер пакета определяется используемым протоколом передачи данных.

Параметры информации - parameter of information - характеристики, с помощью которых оцениваются информационные ресурсы. К основным параметрам относятся: содержание, охват, время, источник, качество, соответствие потребностям, способ фиксации, язык, стоимость.

Пароль - password - набор символов, предъявляемый пользователем системе для получения доступа к данным и программам. Является средством их защиты от несанкционированного доступа.

Платформа - platform - основа, на которой строится и работает компьютер. В зависимости от контекста термин может относиться к аппаратуре, в частности к типу процессора, либо к комбинации аппаратуры и операционной системы.

Поисковые системы (поисковики) - автоматизированные системы обработки web-страниц, позволяющие организовать поиск информации в сети Интернет по определенным ключевым словам.

Пользовательский интерфейс - user interface - правила взаимодействия пользователя с приложением (или операционной средой), а также средства и методы, с помощью которых эти правила реализованы. Основные требования к характеристикам пользовательского интерфейса таковы:

- Естественность интерфейса. Естественный интерфейс не вынуждает пользователя существенно менять привычные способы решения задачи. Использование знакомых пользователю понятий и образов (метафор) обеспечивает интуитивно понятный интерфейс.

- Согласованность интерфейса. Согласованный интерфейс позволяет пользователям переносить имеющиеся знания на новые задания.

- Дружественность интерфейса. Дружественный интерфейс должен предотвращает ситуации, которые могут закончиться ошибками, предупреждает пользователя о возможных ошибках и способах их предупреждения, предлагает отмену ошибочных действий.

- Принцип "обратной связи". Каждое действие, произведенное пользователем должно получать подтверждение (визуальное или звуковое). Полезно также предоставление информации о состоянии процесса.

- Простота интерфейса. "Делайте простые вещи простыми, а сложные возможными". Простота интерфейса означает обеспечение простоты работы в сочетании с реализацией доступа к широким функциональным возможностям. Один из возможных путей поддержания простоты ПИ - представление на экране информации, минимально необходимой для выполнения пользователем очередного шага задания. Другой путь - размещение и представление элементов управления на экране с учетом их смыслового значения и логической взаимосвязи. Этот подход ориентируется на ассоциативное мышление пользователя.

Портал - portal - сайт, организованный как системное многоуровневое объединение разных ресурсов и сервисов. В их числе могут быть информационные ресурсы, поисковые системы, электронный шоппинг, бесплатная электронная почта, торговая реклама, электронная рассылка сообщений, интернет-аукционы, чаты. Порталы обладают возможностью привлекать большое число пользователей и собирать информацию о них. Порталы бывают горизонтальные или вертикальные. Вертикальный портал имеет узкую тематическую направленность и предоставляет в ее рамках различные сервисы, в отличие от горизонтального, объединяющего в рамках одного ресурса несколько тематически разных, но равноправных проектов.

Посетитель - visitor - пользователь конкретного ресурса сети, обладающий рядом технических и психологических характеристик - ip-адрес, маршрут следования, время пребывания на сайте, историю пребывания и т.д.

"Потребитель для потребителя" - customer to customer, C2C - сектор рынка, где коммерческие отношения строятся на общении потребителей друг с другом. Пример веб-ресурсов этого сектора - интернет-аукционы.

Программа-репликант - копия типовой программы, предназначенная для создания нового интернет-ресурса с другим контентом.

Программное обеспечение (ПО) - software - комплекс программ, обеспечивающий обработку или передачу данных.

Программное приложение - application - совокупность программ, реализующих обработку данных в определенной области применения (в частности, веб-приложение - в среде Веб).

Протокол передачи данных - protocol - набор правил, которым следуют компьютеры и программы при обмене информацией. Протоколы определяют формат, временной по-

рядок, контроль и последовательность передачи данных по сети. Существует масса различных протоколов, которые управляют всеми аспектами связи и передачи данных - от аппаратного до прикладного уровня, но все они сходны в том, что задают правила, делающие связь возможной. Протокол передачи гипертекстовой информации - hyper text transfer protocol, HTTP - транспортный протокол, обеспечивающий доступ к документам на веб-узлах. В этом качестве он фактически выполняет все запросы к веб-узлам.

Протокол управления передачей / межсетевой протокол, TCP/IP (в Интернет) - transmission control protocol / Internet protocol, TCP/IP - набор протоколов, разработанный для Интернета и ставший его основой. TCP гарантирует, что каждый посланный байт дойдет до получателя без потерь. IP присваивает локальные IP-адреса физическим сетевым адресам, обеспечивая тем самым адресное пространство, с которым работают маршрутизаторы. В семейство TCP/IP входят и протокол Telnet, который позволяет удаленным терминалам подключаться к удаленным узлам (компьютерам), система доменной адресации DNS, дающая возможность пользователям адресоваться к узлам сети по символьному доменному имени вместо цифрового IP-адреса, протокол передачи файлов FTP, который определяет механизм хранения и передачи файлов, а также протокол передачи гипертекста HTTP.

Р

Раздел сайта - одна или несколько web-страниц, объединенных единой или логически связанной тематикой.

Рейтинг - процент аудитории целевой группы, охваченной каким-либо интернет-ресурсом (аналогично и телепрограммой, печатным изданием, радиостанцией) в данный момент времени. Один пункт рейтинга соответствует одному проценту.

Рекламная поддержка сайта - комплекс мероприятий по продвижению сайта в сети Интернет.

Релевантность - relevance - мера соответствия получаемого результата желаемому. В терминах поиска - это мера соответствия результатов поиска задаче поставленной в поисковом запросе. Определяет, насколько полно тот или иной документ отвечает критериям, указанным в запросе пользователя. Необходимо учитывать, что в каждой поисковой системе работает собственная программа (спайдер), индексирующая веб-страницы, каждая система индексирует страницы своим особым способом и приоритеты при поиске по индексам тоже различны. Поэтому запрос по одним и тем же ключевым словам в каждой из поисковой систем порождает разные результаты.

Репликация (базы данных) - replication - дублирование базы данных на нескольких серверах. Повышает эффективность доступа к информации, но чтобы быть действительно полезной, должна гарантировать регулярное обновление и синхронизацию копий.

Рунет - Runet - российская часть Интернета: совокупность ресурсов, находящихся в доменной зоне России .ru, а также прочих, преимущественно русскоязычных, ресурсов.

С

Санкционированный доступ - authorized access - доступ к программам и данным пользователей, имеющих право (полномочия) на ознакомление или работу с ними.

Сгорание баннера - увеличение вероятности того, что по мере показа баннера в определенной баннерной системе или на определенном интернет-ресурсе он будет показан одному и тому же посетителю несколько раз, что приводит к падению отклика баннера. Время, за которое баннер "сгорит", зависит от интенсивности его показа и от того, насколько широкой аудитории он демонстрируется

Сервер - server - объект, предоставляющий сервис другим объектам по их запросам. В Интернете - компьютер, подключенный к сети, или выполняющаяся на нем программа,

предоставляющие клиентам доступ к общим ресурсам и управляющие этими ресурсами. Наиболее важными типами серверов являются:

1. серверы WWW, предназначенные для представления взаимосвязанной мультимедийной информации и содержимого баз данных;
2. серверы электронной почты;
3. серверы FTP, предназначенные для обмена файлами;
4. серверы общения в реальном времени (чаты);
5. серверы, обеспечивающие работу интернет-телефонии;
6. системы трансляции радио и видео через Интернет.

Сервис - service - совокупность средств для обслуживания пользователей; набор функций одного из уровней программной структуры сети, обеспечивающих доступ к объектам вышележащего уровня через интерфейс между этими уровнями.

Сервисы Интернет - Internet servise - процессы обслуживания объектов Интернет. Сервисы предоставляются пользователям, программам, системам, уровням, функциональным блокам. Наиболее распространенными видами являются: хранение данных, передача сообщений и блоков данных, электронная и речевая почта, организация и управление диалогом партнеров, предоставление соединений, проведение сеансов, видео-сервис. Сервис осуществляют сетевые службы.

Сетевая операционная система - network operating system - комплекс программ, обеспечивающих в сети обработку, хранение и передачу данных. Определяет взаимосвязанную группу протоколов верхних уровней, обеспечивающих основные функции сети. К ним в первую очередь относятся: адресация объектов, функционирование служб, обеспечение безопасности данных, управление сетью.

Сетевое информационное пространство - information networks space - неотъемлемая часть глобального информационного пространства, ограниченная рамками коммуникационных сетей. Учитывая конвергенционные тенденции, оно практически является синонимом понятия "Интернет".

Система - system - любой объект, который одновременно рассматривается и как единое целое, и как совокупность разнородных объектов, объединенных для достижения определенного результата.

Система обеспечения безопасности - security system - стандартные защитные меры, такие, как: криптографическое кодирование, присваивание пароля или идентификатора, электронная цифровая подпись и т. д.

Система управления контентом (Content Management System, CMS) - система программного обеспечения, позволяющая вносить определенные изменения в контент сайта. Существует очень большое количество CMS, начиная с набора коротких скриптов, позволяющих добавлять новости или пресс-релизы на одной странице сайта, кончая полноценными системами паблишинга, поддерживающих процессы workflow (цепочки Автор-Корректор-Редактор), обеспечивающих различные схемы разграничения доступа, автоматически создающих "связанные" документы и т.п. Выбор наиболее подходящей CMS для конкретного проекта определяется бюджетом проекта, предполагаемой сложностью сайта и техническим обеспечением.

Система электронных платежей - electronic payments system - комплекс аппаратных и программных средств, позволяющих авторизованному пользователю произвести оплату товаров и услуг в реальном времени без обращения к наличным деньгам или банковским структурам. Отличается от безналичного расчета значительно большей простотой, оперативностью и автономностью платежей.

Скрипт - микропрограмма, обеспечивающая выполнение определенных действий с элементами web-страницы (реализацию определенного сценария), как под руководством пользователя, так и вне зависимости от его действий.

Спам - spam - рассылка какого-либо сообщения (чаще всего - рекламного или коммерческого содержания) множеству адресатов, для которых данное сообщение нежелательно, или во множество списков и групп новостей, тематика которых не соответствует содержанию сообщения. Справедливо считается тяжелым нарушением этикета и правил применения компьютерных сетей.

Списки рассылки - maillists - простой сервис Интернет, не имеющий собственного протокола передачи данных и работающий исключительно через электронную почту. Осуществляет коммуникацию сообщений всем подписчикам с одного специального адреса.

Старение информации - ageing of information - свойство информации утрачивать со временем свою практическую ценность, обусловленное изменением состояния отображаемой ею предметной области.

CSS (Cascade Style Sheets) - каскадные таблицы стилей - технология компоновки и оформления web-страниц. Разработаны в 1996 г. в виде дополнения к стандарту HTML 3.2.

Т

Таблицы стилей (каскадные таблицы стилей) или CSS (Cascade Style Sheets) - технология компоновки и оформления web-страниц. Разработаны в 1996 г. в виде дополнения к стандарту HTML 3.2.

Таргетинг - targeting - точный охват целевой аудитории, осуществляемый по тематическим сайтам, по географии и по времени.

Телеконференция - telecoferencing - метод проведения дискуссий между удаленными группами пользователей. Она осуществляется в режиме реального времени или путем просмотра документов.

Тестирование сайта в различных браузерах - проверка адекватности отображения всех элементов web-страницы различными браузерами.

Техническое задание (ТЗ) - описание требований к системе и самой системы, документ, в соответствии с которым Заказчик оценивает готовую систему. Согласно ГОСТ 34.602-89 ТЗ является основным документом, определяющим требования и порядок создания (развития или модернизации) информационной системы, в соответствии с которым проводится ее разработка и приемка при вводе в действие. Согласно указанному ГОСТу, ТЗ должно содержать следующие основные разделы:

1. общие сведения
2. назначение и цели создания (развития) системы
3. требования к системе
4. характеристика объектов
5. состав и содержание работ по созданию системы
6. порядок контроля и приемки системы
7. требования к составу и содержанию работ по подготовке объекта разработки к вводу в действие
8. требования к документированию
9. источники разработки

Техническая поддержка сайта - комплекс мероприятий, обеспечивающих стабильную работоспособность сайта в сети Интернет, предотвращение и устранение возможных проблем, связанных с хостингом и функционированием удаленного сервера.

Тракинг - tracking - совершенствование сайта, продукта и маркетинга на основе мониторинга и анализа поведения посетителей на сайте.

Транзакция - transaction - короткий по времени цикл взаимодействия объектов, включающий запрос - выполнение задания - ответ. Обычно осуществляется в режиме диалога.

Трафик - traffic - совокупный объем передаваемой информации за единицу времени, выраженный в единицах измерения компьютерной памяти (бит/с).

У

Удаленный доступ - remote access - технология взаимодействия абонентских систем с локальными сетями через территориальные коммуникационные сети.

Узел - node - компьютер, терминал или любое другое устройство, подключенное к сети. Каждому узлу сети присвоен уникальный адрес, позволяющий другим компьютерам сети связываться с ним.

Управляющие элементы пользовательского интерфейса - важная составляющая пользовательского интерфейса (ПИ). Графические управляющие элементы позволяют выбирать пункты и свойства, а также инициировать выполнение действий. Некоторые из них предоставляют информацию в виде списков и вариантов (радиокнопки, выпадающие списки и т.п.)

Х

Хост - host - установленный в узлах сети компьютер (сервер), решающий вопросы коммуникации и доступа к сетевым ресурсам: модемам, факс-модемам, большим компьютерам и др.; главный, ведущий, центральный компьютер.

Хостинг - услуга по размещению файлов на удаленном сервере, содержащем программы, взаимодействующие с браузерами пользователей. Иными словами, услуга, обеспечивающая размещение сайта в сети Интернет.

Хостер - компания, предоставляющая услуги хостинга.

HTML - то же, что Язык гипертекстовой разметки HTML.

HTML-версия сайта - комплект web-документов, оформленных с применением языка гипертекстовой разметки HTML.

Ц

Целевая аудитория (веб-узла) - обозначение потенциальных посетителей веб-узла, на которых в первую очередь ориентирован данный ресурс. От специфики целевой аудитории зависит стиль представления ресурса и предлагаемые сервисы.

Цифровые деньги - digital cash - электронный аналог наличных денег. В числе наиболее популярных в России платежных систем, использующих виртуальные деньги, сегодня входят Яндекс-деньги и Web-money.

Ш

Шифрование информации - encryption - взаимно однозначное математическое (криптографическое) преобразование, зависящее от ключа (секретный параметр преобразования), которое ставит в соответствие блоку открытой информации, представленной в некоторой цифровой кодировке, блок шифрованной информации, также представленной в цифровой кодировке.

Шлюз - gateway - программа, предназначенная для соединения двух сетей, использующих различные протоколы, благодаря которой осуществляется обмен данными между ними. Перед передачей данных из одной сети в другую программа их преобразует, обеспечивая совместимость протоколов.

Электронная биржа - electronic exchange - биржа, ведущая торги с использованием информационной сети. Операции купли-продажи осуществляются пользователями с помощью абонентских систем, включенных в информационную сеть. Брокеры - посредники при заключении сделки могут находиться в различных географических пунктах, странах.

Электронная доска объявлений - bulletin board system - частный случай телеконференции, специальная база данных, на которой "вывешиваются" различные объявления и сообщения с целью их обмена.

Электронная книга - electronic book - книга, представленная на электронном носителе информации.

Электронная коммерция - electronic commerce - специальным образом построенная модель реального бизнеса, в основе которого лежит использование информационных технологий. Является одним из направлений электронного бизнеса. Означает представление коммерческих предложений и проведение коммерческих операций с использованием технологий и сервисов Интернет. Привлекательность е-коммерции для потенциального предпринимателя заключается в том, что каждый желающий может сделать предложение о продаже товара (или услуги) через Интернет. Первоначальные затраты при этом сравнительно невысоки. Потенциальный покупатель, в свою очередь, получает следующие преимущества - удобство, оперативность, полноту информации, широкий выбор для сравнения, анонимность. Однако, при ведении бизнеса в сети Интернет необходимо учитывать ряд специфических особенностей как данного информационного пространства, так и специфических законов поведения потребителя в нём.

Электронная почта - electronic mail - средства передачи сообщений или документов между пользователями без применения бумажного носителя. Один из сервисов Интернет.

Электронная цифровая подпись (ЭЦП) - electronic digital signature - аналог собственноручной подписи физического лица, представленный как последовательность символов, полученная в результате криптографического преобразования электронных данных с использованием закрытого ключа ЭЦП, позволяющая пользователю открытого ключа установить целостность и неизменность этой информации, а также владельца закрытого ключа.

Электронный денежный перевод - electronic funds transfer - перевод средств, иницируемый с терминала, телефона или магнитоносителя, посредством передачи инструкций или полномочий, касающихся дебетования или кредитования счета финансовому учреждению.

Электронный документооборот - electronic data interchange, EDI - создание, хранение и обмен деловыми документами в цифровой, электронной форме.

Электронный консалтинг - electronic consulting - профессиональные консультации клиентов по электронной почте.

Электронный магазин - веб-узел, где осуществляются прямые продажи товаров потребителю (юридическому или физическому лицу), включая доставку. При этом потребительская информация, заказ товара и сделка осуществляются на сайте е-магазина. С точки зрения технологий электронный магазин представляет собой веб-приложение, работающее с базой данных (электронным Каталогом товаров).

Электронный маркетинг - electronic marketing - маркетинг, осуществляемый с помощью информационных систем и сетей.

Элементы Веб - совокупность аппаратных средств, программ и протоколов, образующих Веб.

Ю

Юзабилити - "Удобство применения - это эффективность, рентабельность и удовлетворение, с которым пользователи могут выполнить те или иные задачи в заданной среде". Область деятельности, направленная на создание функционально удобных с точки зрения пользователя веб-ресурсов.

Я

Язык гипертекстовой разметки HTML - hypertext markup language - язык, используемый для создания документов, содержащих специальные команды форматирования и гипертекстовые ссылки и предназначенных для размещения в WWW и других службах и сетях.

Язык гипертекстовой разметки WML - wireless markup language - язык, специально созданный для отображения информации на маленьком экране мобильного телефона в соответствии с протоколом WAP. Эквивалент HTML на базе XML.

Язык гипертекстовой разметки XML - eXtensible markup language - расширяемый язык гипертекстовой разметки, используемой для создания и размещения документов в среде WWW. В отличие от языка HTML, его можно использовать не только для передачи инструкций по представлению данных, но и для описания содержимого файлов практически любых типов. Позволяет автоматизировать обмен данными, не прибегая для этого к существенному объему программирования.

Язык Java - language Java - интерпретируемый язык с синтаксисом C++, специально рассчитанный на работу в открытой сетевой среде. Разработан компанией Sun Microsystems (США). Появление Java произвело переворот в области создания, применения и распространения программ и мультимедийной информации. Программы, написанные на этом языке, можно переносить на разные платформы (типы компьютеров), выполнять в мультипроцессорных системах, загружать по сети и выполнять программы в машине пользователя, а также одновременно по сети всем пользователям.

Язык программирования - programming language - алфавит, грамматика и синтаксис, используемые для построения набора инструкций, заставляющих компьютер выполнять те или иные действия. Инструкции, написанные на языке программирования, называют исходным кодом. Перед тем, как исходный код будет реально выполнен компьютером, его нужно либо интерпретировать, либо компилировать в машинный код.

Яндекс-деньги - система электронных расчетов между пользователями Рунета.

СПИСОК ЛИТЕРАТУРЫ

Основная:

1. Хорошилов А. Н., Селетков С. Н. Мировые информационные ресурсы. – М., СПб.: Питер, 2004. – 170 с.
2. Симонович С. В. Интернет. – М.: АСТ-Пресс, 2003. – 432 с.
3. Перфильев Ю. Ю. Российское интернет-пространство: развитие и структура. – М.: Гардарики, 2003. – 271 с.
4. Норенков И. П., Трудоношин В. А. Телекоммуникационные технологии и сети. – М.: МГТУ им. Баумана, 2000. – 248 с.
5. Семёнова И. Интернет и российское общество. – М.: Гендальф, 2002. – 278 с.
6. Семёнов Ю. А. WWW ГНЦ ИТЭФ. Проверено 8 февраля 2007 г.
7. Семёнов Ю. А. Гипертекстовый протокол HTTP ГНЦ ИТЭФ. Проверено 8 февраля 2007 г.
8. Храмцов П. Лабиринт Internet. – Электронинформ, 1996.
9. Описание протокола HTTP Проверено 8 февраля 2007 г.
10. HTTP – CGI Проверено 8 февраля 2007 г.
11. Храмцов, Павел Администрирование сети и сервисов Internet Проверено 8 февраля 2007 г.
12. Вереvченко А. П. Информационные ресурсы Проверено 8 февраля 2007 г.
13. Семёнов Ю. А. Принципы построения сетей. Протоколы Интернет. Проверено 8 февраля 2007 г.
14. Семёнов Ю. А. Телекоммуникационные технологии (сайт). Проверено 8 февраля 2007 г.
15. Мамаев, Максим Телекоммуникационные технологии (Сети TCP/IP) Проверено 8 февраля 2007 г.

Дополнительная:

16. Технологии интернет Проверено 8 февраля 2007 г.
17. Адресация в IP-сетях Проверено 8 февраля 2007 г.
18. Введение в IP-сети Проверено 8 февраля 2007 г.
19. Браун Марк Р. Использование HTML 4 / Марк Р. Браун, Джерри Хоникатт. – М.; СПб.; К.: Издательский дом «Вильямс», 1999. – 784 с.
20. Гончаров А. Самоучитель HTML. – СПб.: Питер, 2000. – 360 с.
21. Захаркина В. В. Основы создания Web-страниц. – СПб.: Институт «Открытое общество», 2000. – 23 с.
22. Нечаев В.И. Основы теории защиты информации. – М.: Наука, 2001. – 212с.
23. Ярочкин В.И. Информационная безопасность: Учебник для студентов вузов. – М.: Академический проект; Гаудеамус, 2004, - 544с.

Учебное издание

Лойко Валерий Иванович
Заслуженный деятель науки РФ,
заведующий кафедрой компьютерных технологий
и систем, профессор, д.т.н.

Жмурко Даниил Юрьевич
доцент кафедры компьютерных технологий
и систем, к.э.н.

МИРОВЫЕ ИНФОРМАЦИОННЫЕ РЕСУРСЫ

Учебно-методическое пособие

В авторской редакции

Подписано в печать 16.12.2012 г. Формат $60 \times 84 \frac{1}{16}$.
Тираж 100 экз. Усл. печ. л. – 18. Уч.-изд. л. – 13.
Заказ № 141.

Редакционный отдел и типография
Кубанского государственного аграрного университета,
350044, г. Краснодар, ул. Калинина, 13