**Тема 1. Введение в интернет-технологии и web-приложения**

**Содержание лекции** – Введение в интернет-технологии и web-приложения: знакомство с основными понятиями и принципами интернет-технологий и web-приложений в контексте экономики, история развития интернета, веб-стандарты, архитектуру веб-приложений и основные протоколы.

**Рекомендуемая литература**

1. Виктор Ромашев CMS Drupal: Система управления содержимым сайта. — Питер, 2010.
2. Хаген Граф Создание веб-сайтов с помощью Joomla! 1.5. — Издательский дом "Вильямс", 2009

**Основные понятия и термины**

Сайт (от англ. website: web — «паутина, сеть» и site — «место», буквально «место, сегмент, часть в сети») — совокупность электронных документов (файлов) частного лица или организации в компьютерной сети, объединённых под одним адресом (доменным именем или IP-адресом).

Все сайты в совокупности составляют Всемирную паутину, где коммуникация (паутина) объединяет сегменты информации мирового сообщества в единое целое — базу данных и коммуникации планетарного масштаба. Для прямого доступа клиентов к сайтам на серверах был специально разработан протокол HTTP.

Первый в мире сайт info.cern.ch появился в 1990 году. Его создатель, Тим Бернерс-Ли, опубликовал на нём описание новой технологии World Wide Web, основанной на протоколе передачи данных HTTP, системе адресации URI и языке гипертекстовой разметки HTML. Также на сайте были описаны принципы установки и работы серверов и браузеров. Сайт стал и первым в мире интернет-каталогом, так как позже Тим Бернерс-Ли разместил на нём список ссылок на другие сайты.

Все инструменты, необходимые для работы первого сайта, Бернерс-Ли подготовил ещё раньше — в конце 1990 года появились первый гипертекстовый браузер WorldWideWeb с функционалом веб-редактора, первый сервер на базе NeXTcube и первые веб-страницы.

«Отец» веба считал, что гипертекст может служить основой для сетей обмена данными, и ему удалось претворить свою идею в жизнь. Ещё в 1980 году Тим Бернерс-Ли создал гипертекстовое программное обеспечение Enquire, использующее для хранения данных случайные ассоциации. Затем, работая в Европейском центре ядерных исследований в Женеве (CERN), он предложил коллегам публиковать гипертекстовые документы, связанные между собой гиперссылками. Бернерс-Ли продемонстрировал возможность гипертекстового доступа к внутренним поисковику и документам, а также новостным ресурсам Интернета. В результате, в мае 1991 года в CERN был утверждён стандарт WWW.

Тим Бернерс-Ли является «отцом» основополагающих технологий веба — HTTP, URI/URL и HTML, хотя их теоретические основы были заложены ещё раньше. В 1940-х годах Ванневар Буш выдвинул идеи расширения памяти человека с помощью технических устройств, а также индексации накопленной человечеством информации для её быстрого поиска. Теодор Нельсон и Даг Энгельбарт предложили технологию гипертекста — «ветвящегося» текста, предоставляющего читателю разные варианты чтения. Xanadu, так и не законченная гипертекстовая система Нельсона, была предназначена для хранения и поиска текста, в который введены взаимосвязи и «окна». Нельсон мечтал связать перекрёстными ссылками все тексты, созданные человечеством.

В настоящее время Тим Бернерс-Ли возглавляет основанный им Консорциум Всемирной паутины (World Wide Web Consortium), который занимается разработкой и внедрением стандартов Интернета.

**По доступности сервисов**

Открытые — все сервисы полностью доступны для любых посетителей и пользователей.

Полуоткрытые — для доступа необходимо зарегистрироваться (обычно бесплатно).

Закрытые — полностью закрытые служебные сайты организаций (в том числе корпоративные сайты), личные сайты частных лиц. Такие сайты доступны для узкого круга пользователей. Доступ новым пользователям обычно даётся через т. н. инвайты (приглашения).

**По физическому расположению**

Общедоступные сайты сети Интернет.

Локальные сайты — доступны только в пределах локальной сети. Это могут быть как корпоративные сайты организаций, так и сайты частных лиц в локальной сети провайдера.

**По схеме представления информации**, её объёму и категории решаемых задач можно выделить следующие типы веб-ресурсов

Интернет-портал — многокомпонентная разветвлённая структура, скомпонованная из функционально самодостаточных сайтов самостоятельных организаций или подразделений корпоративной структуры.

**Информационные ресурсы:**

Тематический сайт — сайт, предоставляющий специфическую узкотематическую информацию по какой-либо теме.

Тематический портал — это очень большой веб-ресурс, который предоставляет исчерпывающую информацию по определённой тематике. Порталы похожи на тематические сайты, но дополнительно содержат средства взаимодействия с пользователями и позволяют пользователям общаться в рамках портала (форумы, чаты) — это среда существования пользователя.

Интернет-представительства владельцев бизнеса (торговля и услуги, не всегда связанные напрямую с Интернетом):

Сайт-визитка — содержит самые общие данные о владельце сайта (организация или индивидуальный предприниматель). Вид деятельности, история, прайс-лист, контактные данные, реквизиты, схема проезда. Специалисты размещают своё резюме. То есть подробная визитная карточка.

Представительский сайт — так иногда называют сайт-визитку с расширенной функциональностью: подробное описание услуг, портфолио, отзывы, форма обратной связи и т. д.

Корпоративный сайт — содержит полную информацию о компании-владельце, услугах/продукции, событиях в жизни компании. Отличается от сайта-визитки и представительского сайта полнотой представленной информации, зачастую содержит различные функциональные инструменты для работы с контентом (поиск и фильтры, календари событий, фотогалереи, корпоративные блоги, форумы). Может быть интегрирован с внутренними информационными системами компании-владельца (КИС, CRM, бухгалтерскими системами). Может содержать закрытые разделы для тех или иных групп пользователей — сотрудников, дилеров, контрагентов и пр.

Каталог продукции — в каталоге присутствует подробное описание товаров/услуг, сертификаты, технические и потребительские данные, отзывы экспертов и т. д. На таких сайтах размещается информация о товарах/услугах, которую невозможно поместить в прайс-лист.

Интернет-магазин — сайт с каталогом продукции, с помощью которого клиент может заказать нужные ему товары. Используются различные системы расчётов: от пересылки товаров наложенным платежом или автоматической пересылки счета по факсу до расчётов с помощью пластиковых карт.

Промо-сайт — сайт о конкретной торговой марке или продукте, на таких сайтах размещается исчерпывающая информация о бренде, различных рекламных акциях (конкурсы, викторины, игры и т. п.).

Сайт-квест — Интернет-ресурс, на котором организовано соревнование по разгадыванию последовательности взаимосвязанных логических загадок.

Веб-сервис — сайт, созданный для выполнения каких-либо задач или предоставления услуг в рамках сети WWW:

Доска объявлений представляет собой ресурс, на котором есть возможность размещения публичного объявления о продаже или покупке товаров и услуг, также возможно оставить какую-либо информацию краткого содержания.

Каталог сайтов — это ресурс, на котором размещаются сайты и блоги, например, Open Directory Project. Каталоги бывают платные и бесплатные. Также каталоги могут способствовать продвижению ресурса, который размещается в каталоге сайтов.

* Поисковые сервисы — например, Yahoo!, Google.
* Почтовый сервис.
* Веб-форумы.
* Блоговый сервис.
* Файлообменный пиринговый сервис — например, Bittorrent.
* Облачное хранилище данных — например, Skydrive.
* Сервис редактирования данных — например, Google Docs.
* Фотохостинг — например, Picnik, ImageShack, Panoramio, Photobucket.
* Видеохостинг — например, YouTube, Dailymotion.
* Социальные медиа.

Комбинированные веб-сервисы (Социальные сети) — например, Facebook, Twitter.

Комбинированные веб-сервисы (Специализированные социальные сети) — например, MySpace, Flickr.

Страницы сайтов — это набор текстовых файлов, размеченных на языке HTML. Эти файлы, будучи загруженными посетителем на его компьютер, понимаются и обрабатываются браузером и выводятся на средство отображения пользователя (монитор, экран КПК, принтер или синтезатор речи). Язык HTML позволяет форматировать текст, различать в нём функциональные элементы, создавать гипертекстовые ссылки (гиперссылки) и вставлять в отображаемую страницу изображения, звукозаписи и другие мультимедийные элементы. Отображение страницы можно изменить добавлением стилей на языке CSS, что позволяет централизовать в определенном файле все элементы форматирования (размер и цвет заглавных букв 2-го уровня, размер и вид блока вставки и другое) или сценариев на языке JavaScript, с помощью которого имеется возможность просматривать страницы с событиями или действиями.

Страницы сайтов могут быть простым статичным набором файлов или создаваться специальной компьютерной программой на сервере. Она может быть либо сделана на заказ для отдельного сайта, либо быть готовым продуктом, рассчитанным на некоторый класс сайтов. Некоторые из них могут обеспечить владельцу сайта возможность гибкой настройки структурирования и вывода информации на веб-сайте. Такие управляющие программы называются системами управления содержимым (CMS).

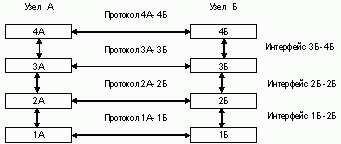
Определённый класс сайтов иначе называют интернет-представительством человека или организации. Как комментарий к ссылке может быть страничка-визитка на полнофункциональном сайте (портале). Когда говорят «своя страничка в Интернете», то подразумевают целый сайт или личную страницу в составе чужого сайта (портал). Кроме сайтов (порталов), в сети Интернет также доступны WAP-сайты для мобильных телефонов.

Изначально сайты представляли собой совокупность статичных документов, например — сайт-визитка. По мере развития коммуникаций, количество внутренних и внешних ссылок увеличивалось. Сайт стал выполнять не только роль справки, аннотации, но и функционального офиса, новостного или медийного центра. В настоящее время большинству из них свойственна динамичность и интерактивность. Для таких случаев специалисты используют термин веб-приложение — готовый программный комплекс для решения задач сайта. Веб-приложение входит в состав сайта, но веб-приложение без данных сайтом является только технически. Оболочку (форму, шаблон) нужно наполнить и активизировать. Продвижение сайтов стало ёмкой индустрией в сети.

1. Изготовление сайтов как работающих целостных информационных ресурсов и систем — составной процесс, вовлекающий труд различных специалистов. Этот вид деятельности называется веб-разработка. Владельцы будущего сайта (частные лица или организации) разрабатывают сайты своими силами, либо обращаются к специализированным разработчикам (фрилансерам, студиям, бюро, конторам и т. п.). Отношения между заказчиком и исполнителем регулируется с помощью договоров, технических заданий, специальных систем (различных сайтов, выступающих посредником между заказчиком и фрилансерами), либо устной договорённостью. Заказанная работа может представлять собой как полный комплекс создания сайта, вплоть до придумывания названия и регистрации домена, так и расширение сайта, техническую оптимизацию и редизайн. Всё больше разработка и сопровождение сайта (портала) становится мощным сегментом активов предприятий (организаций). Поэтому разработчиков предпочитают штатных или поручают проект вести одному из директоров аппарата управления (коммерческий директор, директор департамента по связям или непосредственно руководителю проекта с группой штатных специалистов и/или совместителей). Особую роль выполняют «тестеры» конечного продукта. Это ответственная роль в продвижении и оценке проекта, так как стадия разработки для динамического большого проекта никогда не прекращается. Если вы видите сайт невредимым 2-3 года без изменений, то он возможно никому не нужен, то ли пользуется спросом на базисную информацию. Но сопровождение проекта становится не менее ответственным делом.

Как следует из всего вышеизложенного, Интернет является очень сложной сетью, и соответственно такой же сложной является задача организации взаимодействия между устройствами сети. Для решения такого рода задач используется декомпозиция, т.е. разбиение сложной задачи на несколько более простых задач-модулей. Одной из концепций, реализующих декомпозицию, является многоуровневый подход. Такой подход дает возможность проводить разработку, тестирование и модификацию каждого отдельного уровня независимо от других уровней. Иерархическая декомпозиция позволяет, перемещаясь в направлении от более низких к более высоким уровням переходить к более простому представлению решаемой задачи.

Специфика многоуровневого представления сетевого взаимодействия состоит в том, что в процессе обмена сообщениями участвуют как минимум две стороны, для которых необходимо обеспечить согласованную работу двух иерархий аппаратно-программных средств. Каждый из уровней должен поддерживать интерфейс с выше- и нижележащими уровнями собственной иерархии средств и интерфейс со средствами взаимодействия другой стороны на том же уровне иерархии. Данный тип интерфейса называется протоколом (см. рисунок [Организация взаимодействия между уровнями иерархии при иерархической декомпозиции в сети Интернет](http://www.kgau.ru/istiki/umk/ipr/ch01s02.html#r1_2)).

**Рисунок 1. Организация взаимодействия между уровнями иерархии при иерархической декомпозиции в сети Интернет**

Иерархически организованный набор протоколов, достаточный для организации взаимодействия узлов в сети, называется стеком протоколов.

В начале 80-х годов международные организации по стандартизации ISO (International Organization for Standardization), ITU (International Telecommunications Union) и другие разработали стандартную модель взаимодействия открытых систем OSI (Open System Interconnection). Назначение данной модели состоит в обобщенном представлении средств сетевого взаимодействия. Ее также можно рассматривать в качестве универсального языка сетевых специалистов ( справочной модели ).

Поскольку сеть – это соединение разнородного оборудования, актуальной является проблема совместимости, что в свою очередь, требует согласования всеми производителями общепринятых стандартов. Открытой является система, построенная в соответствии с открытыми спецификациями.

Спецификация представляет собой формализованное описание аппаратных (программных) компонентов, способов их функционирования, взаимодействия с другими компонентами, условий эксплуатации, особых характеристик. Под открытыми спецификациями понимаются опубликованные, общедоступные спецификации, соответствующие стандартам и принятые в результате достижения согласия после всестороннего обсуждения всеми заинтересованными сторонами. Использование открытых спецификаций при разработке систем позволяет третьим сторонам разрабатывать для этих систем аппаратно-программные средства расширения и модификации, а также создавать программно-аппаратные комплексы из продуктов разных производителей.

Если две сети построены с соблюдением принципов открытости, это дает следующие преимущества:

* Возможность построения сети из аппаратных и программных средств различных производителей, придерживающихся стандарта;
* Безболезненная замена отдельных компонентов сети другими, более совершенными;
* Легкость сопряжения одной сети с другой.

В рамках модели OSI средства взаимодействия делятся на семь уровней: прикладной, представления, сеансовый, транспортный, сетевой, канальный и физический. В распоряжение программистов предоставляется прикладной программный интерфейс, позволяющий обращаться с запросами к самому верхнему уровню, а именно, - уровню приложений.

Сеть Интернет строилась в полном соответствии с принципами открытых систем. В разработке стандартов этой сети принимали участие тысячи специалистов-пользователей сети из вузов, научных организаций и компаний. Результат работы по стандартизации воплощается в документах RFC.

RFC (англ. Request for Comments) — документ из серии пронумерованных информационных документов Интернета, содержащих технические спецификации и Стандарты, широко применяемые во Всемирной сети. В настоящее время первичной публикацией документов RFC занимается IETF под эгидой открытой организации Общество Интернета ( ISOC ). Правами на RFC обладает именно Общество Интернет. Формат RFC появился в 1969 г. при обсуждении проекта ARPANET. Первые RFC распространялись в печатном виде на бумаге в виде обычных писем, но уже с декабря 1969 г., когда заработали первые сегменты ARPANET, документы начали распространяться в электронном виде. В таблице 1.2 приведены некоторые из наиболее известных документов RFC.

Таблица 1.2. Примеры популярных RFC-документов.

|  |
| --- |
| Номер RFC Тема |
| RFC 768 UDP |
| RFC 791 IP |
| RFC 793 TCP |
| RFC 822 Формат электронной почты, заменен RFC 2822 |
| RFC 959 FTP |
| RFC 1034 DNS — концепция |
| RFC 1035 DNS — внедрение |
| RFC 1591 Структура доменных имен |
| RFC 1738 URL |
| RFC 1939 Протокол POP версии 3 (POP3) |
| RFC 2026 Процесс стандартизации в Интернете |
| RFC 2045 MIME |
| RFC 2231 Кодировка символов |
| RFC 2616 HTTP |
| RFC 2822 Формат электронной почты |
| RFC 3501 IMAP версии 4 издание 1 (IMAP4rev1) |

Основным организационным подразделением, координирующим работу по стандартизации Интернет, является ISOC (Internet Society), объединяющее порядка 100 тысяч участников, которые занимаются различными аспектами развития данной сети. ISOC курирует работу IAB (Internet Architecture Board), включающую две группы:

* IRTF (Internet Research Task Force). Координирует долгосрочные исследовательские проекты, относящиеся к TCP/IP;
* IETF (Internet Engineering Task Force). Инженерная группа, определяющая спецификации для последующих стандартов Интернет.

Разработкой стандартов для сети Веб, начиная с 1994 года, занимается Консорциум W3C (World Wide Web Consortium), основанный и до сих пор возглавляемый Тимом Бернерсом-Ли.

Консорциум W3C — организация, разрабатывающая и внедряющая технологические стандарты для Интернета и WWW. Миссия W3C формулируется следующим образом: "Полностью раскрыть потенциал Всемирной паутины путем создания протоколов и принципов, гарантирующих долгосрочное развитие Сети". Две другие важнейшие задачи Консорциума — обеспечить полную "интернационализацию Сети" и сделать ее доступной для людей с ограниченными возможностями.

W3C разрабатывает для WWW единые принципы и стандарты, называемые " Рекомендациями ", которые затем внедряются разработчиками программ и оборудования. Благодаря Рекомендациям достигается совместимость между программными продуктами и оборудованием различных компаний, что делает сеть WWW более совершенной, универсальной и удобной в использовании.

Все Рекомендации W3C открыты, то есть, не защищены патентами и могут внедряться любым человеком без каких-либо финансовых отчислений Консорциуму.

Для удобства пользователей Консорциумом созданы специальные программы-валидаторы (англ. Online Validation Service), которые доступны по сети и могут за несколько секунд проверить документы на соответствие популярным Рекомендациям W3C. Консорциумом также созданы многие другие утилиты для облегчения работы веб-мастеров и программистов. Большинство утилит — это программы с открытым исходным кодом, все они бесплатные. В последнее время, повинуясь мировым тенденциям, Консорциум, в целом, гораздо больше внимания уделяет проектам с открытым исходным кодом.

В российском сегменте Интернета имеется своя организация - Российский НИИ Развития Общественных Сетей РОСНИИРОС (Russian Institute for Public Networks, RIPN). РОСНИИРОС занимается координацией российских исследований и разработок в Интернете.

Прежде чем перейти к описанию структуры, принципов работы и основных протоколов сети Веб, рассмотрим основной стек протоколов сети Интернет - стек TCP/IP.

***Стек протоколов TCP/IP***

Эти протоколы изначально ориентированы на глобальные сети, в которых качество соединительных каналов не идеально. Он позволяет создавать глобальные сети, компьютеры в которых соединены друг с другом самыми разными способами от высокоскоростных оптоволоконных кабелей и спутниковых каналов до коммутируемых телефонных линий. TCP/IP соответствует модели OSI достаточно условно и содержит 4 уровня. Прикладной уровень стека соответствует трем верхним уровням модели OSI: прикладному, представления и сеансовому.

В сети данные всегда передаются блоками относительно небольшого размера. Каждый блок имеет префиксную часть (заголовок), описывающую содержимое блока, и суффиксную, содержащую, например, информацию для контроля целостности передаваемого блока данных.

Название стека протоколов TCP/IP состоит из названий двух разных протоколов. Протокол IP (Internet Protocol) представляет собой протокол нижнего (сетевого) уровня и отвечает за передачу пакетов данных в сети. Он относится к так называемым протоколам датаграмм и работает без подтверждений. Последнее означает, что при его использовании доставка пакетов данных не гарантируется и не подтверждается. Не гарантируется также и то, что пакеты достигнут пункта назначения в той последовательности, в которой они были отправлены.

К протоколам сетевого уровня относится также протокол межсетевых управляющих сообщений ICMP (Internet Control Message Protocol), предназначенный для передачи маршрутизатором источнику информации об ошибках при передаче пакета.

Очевидно, что намного удобнее передавать данные по каналу, который работает корректно, доставляя все пакеты по порядку. Поэтому над протоколом IP работает протокол передачи данных более высокого (транспортного) уровня — TCP (Transmission Control Protocol). Посылая и принимая пакеты через протокол IP, протокол TCP гарантирует доставку всех переданных пакетов данных в правильной последовательности.

Следует отметить, что при использовании протокола IP обеспечивается более быстрая передача данных, так как не тратится время на подтверждение приема каждого пакета. Есть и другие преимущества. Одно из них заключается в том, что он позволяет рассылать пакеты данных в широковещательном режиме, при котором они достигают всех компьютеров физической сети. Что же касается протокола TCP, то для передачи данных с его помощью необходимо создать канал связи между компьютерами. Он и создается с использованием протокола IP.

Для идентификации сетевых интерфейсов используются 3 типа адресов:

* аппаратные адреса (или MAC-адреса);
* сетевые адреса (IP-адреса);
* символьные (доменные) имена.

В рамках IP протокола для создания глобальной системы адресации, не зависящей от способов адресации узлов в отдельных сетях, используется пара идентификаторов, состоящая из номера сети и номера узла. При этом IP-адрес идентифицирует не отдельный компьютер или маршрутизатор, а одно сетевое соединение в составе сети, в которую он входит; то есть конечный узел может входить в несколько IP-сетей.

**Вопросы для самостоятельного изучения**

1. Что означает термин «web-ресурс»?

2. Как соотносятся новые медиа и web-системы?

3. Опишите веб-сервисы по доступности.

4. Что представляет собой протокол IP?

5. В чем заключается функцииконсорциума W3C?