**15 неделя**

**Лекция 29** Защита информации в компьютерных сетях на основе Web-технологий. Сферы применения и перспективы развития Web-технологий.

## Web 1.0

Web 1.0 является ретронимом понятия, которое относится к статусу WWW и любому стилю дизайна вебсайта, используемому перед появлением термина Web 2.0. Это — общий термин, который был создан, чтобы описать Сеть перед явлением взрыв пузыря «.com» в 2001, который и был замечен многими как поворотный момент для Интернета. Самой простой формулировкой смысла Web 1.0 скорее всего следует считать «тот Web, который был до Web 2.0». Это обычная практика — сравнение сайтов по типу используемой технологии.

Основные тенденции Web 1.0 включали заботы о проблемах безопасности и приватности в одностороннем потоке информации, через вебсайты, содержащие материал 'только для чтения'. Характерным для Web 1.0 также являлись компьютерная неграмотность широких масс и распространённость медленных типов подключения к интернету, вдобавок к ограничениям самого Интернета.

### Элементы Web 1.0

Термин типичная употребляется здесь в обобщающем смысле — совершенно не обязательно, что для любой страницы Web 1.0 были присущи все элементы списка. Но, как правило, общая тенденция и большая часть элементов присутствовала.

* Статичные страницы, создаваемые и поддерживаемые разработчиком веб-сайта
* Бедная гипертекстовая разметка (большая часть контента де-факто являлась простым текстом, зачастую пренебрегавшим правилами HTML).
* Использование фреймов.
* Использование специфичных тегов HTML — следствие редактирования страниц в WYSIWYG редакторах, встроенных в конкретный браузер или сторонников конкретного браузера-участника «войны браузеров».
* гостевые книги, форумы или чаты — как попытка придания интерактивности.
* Кнопки формата GIF, обычно 88x31 пикселей, в качестве указания поддерживаемых браузеров.
* Использование информеров (погода, курс доллара и т. д.) вместо агрегации информации средствами CMS.
* Указание конкретного разрешения монитора, при котором дизайн сайта отображается корректно (не вылезает за пределы страницы, не разъезжается форматирование).
* Крайне редкое использование стилей CSS при оформлении страниц сайта.

## Web 2.0

Web 2.0 — совокупность веб-технологий, ориентированная на активное участие пользователей в создании контента веб-сайтов.

Появление названия Веб 2.0 принято связывать со статьёй Тима О’Рейли «What Is Web 2.0» от 30 сентября 2005 года

### Особенности web 2.0

* Привлечение «коллективного разума» для наполнения сайта;
* Взаимодействие между сайтами с использованием веб-сервисов;
* Обновление веб-страниц без перезагрузки;
* Агрегирование и синдикация информации;
* Объединение различных сервисов для получения нового функционала;
* Дизайн с применением стилевой разметки и акцентом на юзабилити.

### Основные элементы web 2.0

#### Веб-сервисы

Веб-сервисы (веб-службы) — это сетевые приложения, доступные по протоколу HTTP, в качестве протоколов взаимодействия использующие основанные на XML форматы данных (RPC, SOA и подобные). В результате программное обеспечение может использовать веб-службы вместо того, чтобы самостоятельно реализовывать требуемый функционал (например, проверить введенный в форме почтовый адрес). В отличие от обычных динамических библиотек, такой подход обладает рядом плюсов:

* Веб-служба находится на серверах компании, которая её создала. Поэтому в любой момент пользователю доступна самая свежая версия программы и ему не приходится заботиться об обновлениях и вычислительных мощностях, требуемых для выполнения операции.
* Инструменты для работы с HTTP и XML есть в любом современном языке программирования, поэтому веб-службы являются платформонезависимыми.

#### AJAX

Asynchronous JavaScript and XML — подход к построению пользовательских интерфейсов веб-приложений, при котором веб-страница, не перезагружаясь, асинхронно загружает нужные пользователю данные. Использование Ajax стало наиболее популярно после того, как Google начала активно использовать его при создании своих сайтов, таких как Gmail и Google Maps. Часто Ajax считают синонимом Веб 2.0, что совершенно не так. Веб 2.0 не привязан к какой-то одной технологии или набору технологий, с тем же успехом ещё в 1999 году возможность асинхронного обновления страницы уже предоставлял Flash 4.

#### Веб-синдикация

Одновременное распространение информации в том числе аудио- и видео- на различные страницы или web-сайты, как правило, с использованием технологий RSS или Atom. Принцип заключается в распространении заголовков материалов и ссылки на них (например, последние сообщения форумов, и т. п.). Первоначально эта технология использовалась на новостных ресурсах и в блогах, но постепенно сфера применения расширилась.

#### Mash-up

Веб mash-up (дословный перевод — «смешение») — сервис, который полностью или частично использует в качестве источников информации другие сервисы, предоставляя пользователю новую функциональность для работы. В результате такой сервис может становиться также новым источником информации для других веб mash-up сервисов. Таким образом образуется сеть зависимых друг от друга сервисов, интегрированных друг с другом.

Например, сайт транспортной фирмы может использовать карты сервиса Google Maps для отслеживания местонахождения перевозимого груза.

#### Метки (теги)

Ключевые слова, описывающие рассматриваемый объект, либо относящие его к какой-либо категории. Это своего рода метки, которые присваиваются объекту, чтобы определить его место среди других объектов. С понятием меток тесно связано понятие фолксономии — термина, о котором широко заговорили именно в связи с ростом сервисов Веб 2.0, таких как Flickr, del.icio.us, и, в дальнейшем, Wink.

Появление и быстрое распространение блогов тоже вписывается в концепцию Веб 2.0, создавая так называемую «редактируемую Паутину» (writable web).

Возможность пометить документ ключевыми словами существует и в языке HTML (англ. keywords), однако этот способ был полностью скомпрометирован широким его использованием в целях поискового спама.

#### Социализация

Использование разработок, которые позволяют создавать сообщества пользователей.

* В понятие социализация сайта можно также включить возможность индивидуальных настроек сайта и создание личной зоны (личные файлы, изображения, видео, блоги) для пользователя, чтобы пользователь чувствовал свою уникальность.
* Поощрение, поддержка и доверие «коллективному разуму».
* При формировании сообщества большое значение имеет соревновательный элемент, Репутация или Карма, которые позволяют сообществу саморегулироваться и ставить пользователям дополнительные цели присутствия на сайте.

#### Дизайн

Понятие Веб 2.0 также отразилось и в дизайне. Предпочтительными стали округлость, имитация выпуклых поверхностей, имитация отражений на манер глянцевого пластика современных hi-end устройств (к примеру, плееры). В целом, восприятие внешнего вида на глаз кажется более приятным. Графика таких сайтов занимает больший объём, нежели при использовании аскетичного дизайна. Отчасти эта тенденция связана с совпавшим по времени выходом новых версий операционных систем использующих вышеупомянутые идеи.

Однако однообразие таких сайтов явно и в последнее время считается, графический облик классического дизайна веб 2.0, устаревшим и не креативным. Особенно это отражается в современной тенденции создания информативных сайтов, где главную роль играет простота, изящество, графичность и юзабилити. В дизайне не должно быть ограничений, но веб 2.0 их прививает.

### Недостатки Веб 2.0

При использовании технологий web 2.0 вы становитесь арендатором сервиса и/или дискового пространства у какой-то сторонней компании. Возникающая при этом зависимость формирует ряд недостатков новых сервисов:

* зависимость сайтов от решений сторонних компаний, зависимость качества работы сервиса от качества работы многих других компаний;
* слабая приспособленность нынешней инфраструктуры к выполнению сложных вычислительных задач в браузере;
* уязвимость конфиденциальных данных, хранимых на сторонних серверах, для злоумышленников (известны случаи хищения личных данных пользователей, массовых взломов учётных записей блогов).

Фактически сайт эпохи Веб 2.0 на первый взгляд интерактивен и дружелюбен, позволяет себя легко настраивать. Однако сбор статистики о пользователях, их предпочтениях и интересах, личной жизни, карьере, круге друзей могут помочь владельцу сайта манипулировать сообществом. По самым пессимистичным прогнозам многочисленные сайты Веб 2.0 вкупе с другими современными технологиями дают прообраз тоталитарной системы «Большого брата».

## Web 3.0

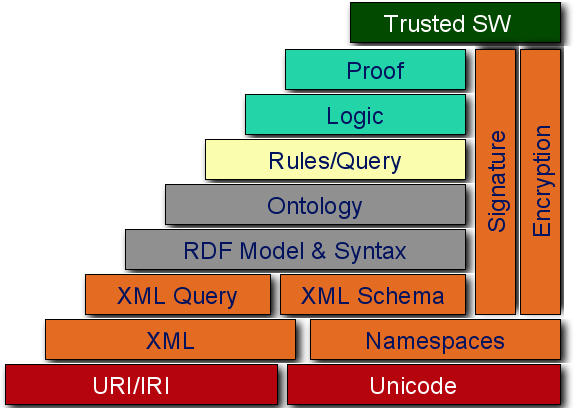
Web 3.0 — это принципиально новый подход к обработке информации, представленной во Всемирной паутине. Web 3.0 в первую очередь подразумевает под собой иной подход к обработке информации сообществом пользователей. Если Web 1.0 предполагает веб-мастера в качестве поставщика контента, а Web 2.0 — сообщество равноправных пользователей, генерирующих контент в рамках тематического проекта, то Web 3.0 уже позволяет этим самым равноправным пользователям "выбирать" экспертов в заданной области (или в нескольких областях) и "наделять его властью". Такие общепризнанные эксперты-выдвиженцы постепенно расширяют свое влияние на сообщество, могут выступать его модераторами, управлять сообществом при помощи дополнительных прав и расширенных возможностей в рамках интернет-проекта. Это исключает возможность управления сообществом некомпетентных и малозначимых его участников, что является очень важным вещью — достаточно вспомнить «вебдванольное» равноправие на проекте Wikipedia, из-за которого ее создатель должен был более десятка раз исправлять свою собственную (!) биографию – пользователи считали, что в ней есть неточности и честно ее исправляли.

Также термином Web 3.0 часто называют концепцию семантической паутины (Semantic Web).

## Semantic Web

**Семантическая паутина** (англ. Semantic Web) — часть глобальной концепции развития сети Интернет, целью которой является реализация возможности машинной обработки информации, доступной во Всемирной паутине. Основной акцент концепции делается на работе с метаданными, однозначно характеризующими свойства и содержание ресурсов Всемирной паутины, вместо используемого в настоящее время текстового анализа документов. Термин впервые введён Тимом Бернерсом-Ли в мае 2001 года в журнале «Scientific American», и называется им «следующим шагом в развитии Всемирной паутины». В семантической паутине предполагается повсеместное использование, во-первых, унифицированных идентификаторов ресурсов (URI), а во-вторых — онтологий и языков описания метаданных.

Эта концепция была принята и продвигается Консорциумом W3 ([Рис. 2](http://www.4stud.info/web-programming/?lecture=9#semantic-web#semantic-web)). Для её внедрения предполагается создание сети документов, содержащих метаданные о ресурсах Всемирной паутины и существующей параллельно с ними. Тогда как сами ресурсы предназначены для восприятия человеком, метаданные используются машинами (поисковыми роботами и другими интеллектуальными агентами) для проведения однозначных логических заключений о свойствах этих ресурсов.

*[](http://www.4stud.info/web-programming/img/semantic_web_stack)*

*Рис. 2. Стек протоколов Semantic Web*

### Основная идея

Семантическая паутина — это надстройка над существующей Всемирной паутиной, которая призвана сделать размещённую в ней информацию более понятной для компьютеров. Машинная обработка возможна в семантической паутине благодаря двум её важнейшим характеристикам:

* Повсеместное использование унифицированных идентификаторов ресурсов (URI). Традиционная схема использования таких идентификаторов в современном Интернете сводится к установке ссылок, ведущих на объект, им адресуемый. Очевидным свойством такой ссылки является возможность «загрузки» объекта, на который она указывает. Таким объектом может быть веб-страница, файл произвольного содержания, фрагмент веб-страницы, а также неявное указание на обращение к реально существующему физическому ресурсу по протоколу, отличному от HTTP (например, ссылки mailto:). Концепция семантической паутины расширяет это понятие, включая в него ресурсы, недоступные для скачивания. Адресуемыми с помощью URI ресурсами могут быть, например, отдельные люди, города и другие географические сущности, художественные артефакты и т. д. К идентификатору предъявляются несколько простых требований: он должен быть уникальной строкой определённого формата, адресующей реально существующий объект.
* Повсеместное использование онтологий и языков описания метаданных. Современные методы автоматической обработки данных, доступных в Интернете, как правило, основаны на частотном и лексическом анализе текстового содержимого, которое прежде всего предназначено для восприятия человеком. В семантической паутине предлагается использовать форматы описания, доступные для машинной обработки (например, семейство форматов, часто упоминаемое в литературе как «Semantic Web family»: RDF, RDF Schema или RDF-S, и OWL), в свою очередь, использующие URI для адресации описываемых и описывающих объектов, а также онтологии и дескрипционные логики в качестве базовых математических формализмов.

### Критика

Несмотря на все преимущества, предоставляемые семантической паутиной в случае её внедрения, существуют сомнения в возможности её полной реализации.

#### Практическая нереализуемость

Разные комментаторы высказывают различные причины, которые могут быть препятствием к этому, начиная с человеческого фактора (люди склонны избегать работы по поддержке документов с метаданными, открытыми остаются проблемы истинности метаданных, и т. д.), и заканчивая сложностью определения онтологии верхнего уровня (корня иерархии), критической для семантической паутины.

#### Дублирование информации

Необходимость описания метаданных так или иначе приводит к дублированию информации. Каждый документ должен быть создан в двух экземплярах: размеченным для чтения людьми, а также в машинно-ориентированном формате.

#### Невозможность получения коммерческой выгоды

Известно, что основное финансирование современных интернет-ресурсов (за исключением строго некоммерческих проектов) обеспечивают рекламодатели. Главный критерий, от которого зависит стоимость рекламного места — посещаемость сайта. Однако в случае реализации семантических поисковых систем, которые будут сами отбирать и сразу предоставлять нужную пользователю информацию, отпадает необходимость посещать сайт — источник материала, а значит пользователь не увидит рекламу, и как следствие прекратится финансирование интернет-проектов.

### Реализация

#### Языки описания

Техническую часть семантической паутины составляет семейство стандартов на языки описания, включающее XML, XML Schema, RDF, RDF Schema, OWL, а также некоторые другие. Располагая их в порядке повышения уровня абстракции, реализуемого тем или иным языком, получаем:

* XML предоставляет синтаксис для определения структуры документа, подлежащего машинной обработке. Синтаксис XML не несёт семантической нагрузки.
* XML Schema определяет ограничения на структуру XML-документа. Стандартный синтаксический анализатор языка XML в состоянии проверить произвольный XML-документ на соответствие его структуры так называемой схеме документа, описанной в XML Schema.
* RDF представляет собой простой способ описания экземплярных данных в формате субъект-отношение-объект, в котором в качестве любого элемента этой тройки используются только идентификаторы ресурсов. Существует стандартизованное отображение этих троек на XML-документы предопределённой структуры (т. е. консорциумом W3 определена схема XML-документов, содержащих RDF-описания), а также на другие форматы представления (например, в нотацию N3).
* RDF Schema описывает набор атрибутов (здесь их точнее назвать отношениями), таких, как rdfs:Class, для определения новых типов RDF-данных. Языком поддерживается также отношение наследования типов rdfs:subClassOf.
* OWL расширяет возможности по описанию новых типов (в частности, добавлением перечислений), а также позволяет описывать новые типы данных RDF Schema в терминах уже существующих (например, определять тип, являющийся пересечением или объединением двух существующих).

#### Логический вывод

Форматы описания метаданных в семантической паутине предполагают проведение логического вывода на этих метаданных, и разрабатывались с оглядкой на существующие математические формализмы в этой области. Математическое обоснование тех или иных конструкций языка описания необходимо для проведения заключений о свойствах программ, обрабатывающих данные в этом формате.

Особенно сильно это относится к языку OWL. Базовым формализмом для него являются дескрипционные логики, а сам язык разбит на три вложенных подмножества (в порядке вложенности): OWL Lite, OWL DL и OWL Full. Доказано, что логический вывод на метаданных с выразительностью OWL Lite выполняется за полиномиальное время (другими словами, задача вывода принадлежит к классу P). OWL DL описывает максимальное разрешимое в данный момент подмножество дескрипционных логик, но некоторые запросы по таким данным могут требовать экспоненциального времени выполнения. OWL Full реализует все существующие конструкторы дескрипционных логик, но не каждый запрос в этом подмножестве языка может быть разрешён (слово «разрешён» здесь употребляется в значении, основанном на корне «решать»).

Простая структура предикатов языка RDF, в свою очередь, позволяет использовать при его обработке опыт из теорий логических баз данных, логики предикатов, и т. д.

### Проекты

#### Дублинское ядро

Одним из первых серьёзных и популярных проектов, основанным на принципах семантической паутины, стал проект «Дублинское ядро» (англ. Dublin Core), реализуемый инициативной организацией Dublin Core Metadata Initiative (DCMI). Это открытый проект, цель которого — разработать стандарты метаданных, которые были бы независимы от платформ и подходили бы для широкого спектра задач. Конкретнее, DCMI занимается разработкой словарей метаданных общего назначения, стандартизирующих описания ресурсов в формате RDF.

#### RSS (версий 0.90 и 1.0)

Версии 0.90 и 1.0 формата RSS основаны на RDF. Информация в нём представляется как и в RDF, тройками субъект-отношение-объект. Необходимо отметить, что несмотря на то, что ему присущи многие недостатки семантической паутины (дублирование информации, например), этот простейший формат быстро стал чрезвычайно популярным за счёт узкой категоризации подмножества используемых метаданных. Отличие RSS от RDF состоит в том, что субъектом тройки всегда является сайт-источник RSS-файла, а в качестве отношений используются самые очевидные свойства документов, имеющие отношение к часто обновляющимся источникам информации: дата написания, автор, постоянная ссылка, и т. д. Другими словами, RSS — узкоспециализированное подмножество RDF.

Помимо недостатков, RSS унаследовал и все достоинства форматов из семейства семантической паутины: гибкость RSS позволяет использовать его не только для проверки на наличие новой информации на регулярно обновляющихся сайтах, но и для подкастов, и торренттрекеров.

Заметим, что формат RSS версии 2.0, хотя и не является форматом, основанным на RDF, позволяет внедрение произвольного XML-содержимого, находящегося в собственных пространствах имён XML. Это позволяет использовать RDF-описания также и в нём (используя пространство имён rdf).

#### FOAF

Проект «Friend of a Friend» («Друг друга») позволяет описывать отношение знакомства с помощью RDF. Любой его участник может идентифицировать себя уникальным образом с помощью URI (например, mailto-адресом электронной почты, адресом блога, и т. п.), создать свой профиль, используя предопределённые для FOAF отношения на языке RDF, и перечислить идентификаторы людей, которых этот участник знает. Это описание может обрабатываться автоматически; на его основе можно строить сети доверия, анализировать структуру социальных групп, и т. д.

### Семантические веб-сервисы

В то время как совокупность ресурсов и их метаданных можно считать статической частью семантической паутины, её динамическую часть представляют т. н. семантические веб-сервисы — законченные элементы программной логики с однозначно описанной семантикой, доступные через Интернет и пригодные для поиска, композиции и выполнения.

Технически, семантический веб-сервис отличается от обычного веб-сервиса наличием не только описания интерфейса (обычно на языке WSDL) в терминах типов данных, передаваемых сервису, возвращаемых значений и генерируемых ошибок, но и семантического описания всех его характеристик. Заметим, что дублирования данных, упомянутого в числе недостатков семантической паутины, здесь не происходит: WSDL-описания изначально были предназначены для машинной обработки.

Потенциальная выгода от использования семантических веб-сервисов заключается в возможности автоматического поиска (а также композиции) программными агентами подходящих сервисов для решения поставленных задач. Тем не менее, сложность этой задачи в её общей формулировке пока позволяет добиваться некоторых положительных результатов только в узкоспециализированных отраслях, явным образом выигрывающих от внедрения сервисно-ориентированной архитектуры, например в интеграции корпоративных приложений.