

ТЕХНОЛОГИЯ ПОСТРОЕНИЯ РАСПРЕДЕЛЁННЫХ СЕТЕЙ САЙТОВ

Резниченко Антон Дмитриевич, аспирант кафедры АСОИиУ
ГОУВПО «Мордовский государственный университет им. Н. П. Огарева», г. Саранск
Тел. 8 (937) 515-88-67. E-mail: altervision13@gmail.com

Аннотация. В докладе рассматриваются принципы построения сетей сайтов с единым информационным пространством. Основной упор сделан на выделение оптимальных методов единой аутентификации пользователей и информационного обмена между различными компонентами сети.

Ключевые слова: единое информационное пространство, сеть сайтов, методы аутентификации, сайт, репозиторий.

Постановка задачи

В настоящее время сеть Интернет стала наиболее важным и актуальным источником информации. Крупные сообщества и организации нуждаются в предоставлении своей информации, разбитой по различным отраслям деятельности. В качестве критериев этого разделения может выступать территориальное расположение (филиалы организации в различных городах, регионах, странах), сфера деятельности различных подразделений организации, отдельные персоналии и подразделения, составляющие организацию или сообщество, и подобные. Существует задача построить информационное пространство организации, логически выделив отдельные структурные единицы.

В данный момент для реализации этой задачи используются некоторые стандартные решения. Наиболее распространённым является *усложнение структуры* сайта организации, выделение отдельных иерархических веток страниц и разделов, описывающих ту или иную отрасль. Сайт организации значительно усложняется, ухудшается поиск по сайту и доступ к вложенным разделам. Второй вариант решения поставленной задачи состоит в выделении независимых крупных разделов сайта, разделяемых на основе *независимых поддоменов*. При углублении в это решение, возможно также выделение для каждой из структурных единиц *отдельного независимого сайта*. В данном случае каждая структурная единица становится обособленной и осуществляет свою деятельность независимо от остальных, информационный обмен между ними затруднён и осуществляется вручную.

В настоящей статье рассматривается методика организации информационного пространства для организаций и сообществ, учитывающая минусы описанных выше способов. Предлагается организация информационного пространства на основе распределённой сети сайтов.

Описание принципов построения сети

Распределённая сеть сайтов – совокупность веб-сайтов, объединённых общими информационным пространством, базами данных основных структурных единиц (пользователи, подразделения и пр), расположенных на различных хостинг-платформах, осуществляющих взаимосвязь на основе определённых протоколов. В данном случае организация (сообщество) рассматривается как совокупность некоторых структурных единиц (ячеек), объединённых с общим управляющим центром.

Рассмотрим основные принципы построения данных сетей сайтов.

1. Распределение составляющих частей.

Основным принципом построения сети является разделение информационного пространства на ряд сайтов по определённому признаку. Критерии выделения составляющих элементов описаны выше: территориальное и отраслевое деление, выделение персоналий и аналогичные варианты, зависящие от рассматриваемой организации или сообщества.

В указанном информационном пространстве выделяется две крупных базовых составляющих – центральный сайт и множественные сайты-сателлиты.

Центральный сайт – интернет-ресурс, отвечающий за осуществление функций контроля и мониторинга сети сайтов, хранит в себе базу данных пользователей, сайтов и других необходимых структурных единиц системы. На основе центрального сайта организуется техническая поддержка и сопровождение сайтов-сателлитов системы, в том числе их установка, развёртывание и обновление. Совместно с центральным сайтом целесообразно создание нестандартных ресурсов, реализующих специфические функции для всех сайтов-сателлитов. Это может быть общий центральный форум проекта, интернет-магазин, объединяющий в себе товары подразделений, и другие ресурсы, основной задачей которых является агрегация данных сайтов-сателлитов. Совместно с центральным сайтом они составляют ядро системы.

Одной из основных функций центрального сайта является предоставление пользователю краткой информации о сайтах-сателлитах. Обязательным условием является наличие полного списка сайтов-сателлитов системы. Рекомендуется создание информационных страниц для каждого из сайтов системы, содержащих краткое описание сайта, его региональную или персональную принадлежность, заголовки последних новостей и публикаций сайта и подобную информацию в сокращённом виде (дайджест). Состав данной информации определяется администратором сайта-сателлита.

Сайт-сателлит – это элементарная структурная единица сети. Данные сайты строятся на основе одинаковых логических принципов, единой системы управления контентом и общих стандартов представления информации в данной сети. Каждый из сайтов отвечает за обслуживание одной структурной ячейки системы: региона, города, подразделения, персоны.

Предполагается, что сайты-сателлиты и центральный сайт распложены на различных хостинг-платформах и серверах, за обслуживание каждого из которых отвечает владеющая сайтом структурная единица. Таким образом, достигается ряд преимуществ:

- Система является устойчивой к выходу из строя одного или нескольких элементов, в том числе не только сайтов-сателлитов, но и различных составляющих центрального сайта. Основной функционал сайтов-сателлитов не зависит от работоспособности центрального сайта системы, равно как и центральный сайт системы способен поддерживать свою функциональность в случае отказа некоторых сайтов-сателлитов, предоставляя пользователю краткую информацию (дайджест) с этих сайтов.

- Значительно снижаются технические требования к аппаратной платформе для сайтов-сателлитов, по причине распределения нагрузки и посещаемости по множеству сайтов в рамках сети.

- Уменьшаются технические требования к аппаратной платформе центрального сайта за счёт оттока посетителей на сайты-сателлиты.

Распределение информационных ресурсов сети по различным сайтам позволяет значительно сократить затраты на обеспечение аппаратной технической базы для её функционирования.

2. Единая аутентификация.

Распределение ресурсов и информации по различным сайтам требует унификации доступа ко всем информационным ресурсам. Необходимым условием в данном случае является введение единой для всех сайтов базы пользователей системы с выделениями им некоторых базовых прав доступа.

Для хранения учётных записей пользователей используется центральный сайт. Его средствами реализуется регистрация пользователей, открытая либо закрытая, и управление

их профайлами. Таким образом, на центральном сайте системы располагается полная объёмная база данных пользователей сети и их основных прав доступа к ресурсам. Каждый из сайтов-сателлитов содержит локальную копию указанной базы данных в сокращённом виде. Использование сокращённой базы данных позволяет значительно уменьшить требования к аппаратной платформе за счёт облегчения выполнения запросов к базе данных. При этом способе в локальной базе данных сайта-сателлита сохраняется информация только о тех пользователях, кто хотя бы один раз авторизовался на сайте.

Авторизация пользователей на сайте может осуществляться по двум различным алгоритмам: OAuth (OpenID) и прямая авторизация с вводом логина и пароля.

OAuth - это открытый протокол авторизации, который позволяет предоставить третьей стороне доступ к защищенным ресурсам пользователя, без необходимости передавать ей (третьей стороне) логин и пароль. OAuth позволяет скрипту веб-приложения получить ограниченный доступ к данным провайдера (центрального сайта системы), если пользователь разрешает [4]. **OpenID** — протокол для ускоренной регистрации. OpenID позволяет пользователю без ввода пароля получить аккаунт на каком-либо сервисе, если он уже зарегистрирован где-то еще в интернете [5].

Упрощённая схема взаимодействия сайта-сателлита с OAuth-сервером центрального сайта системы представлена на рисунке 1 (на примере API Яндекс.Деньги).

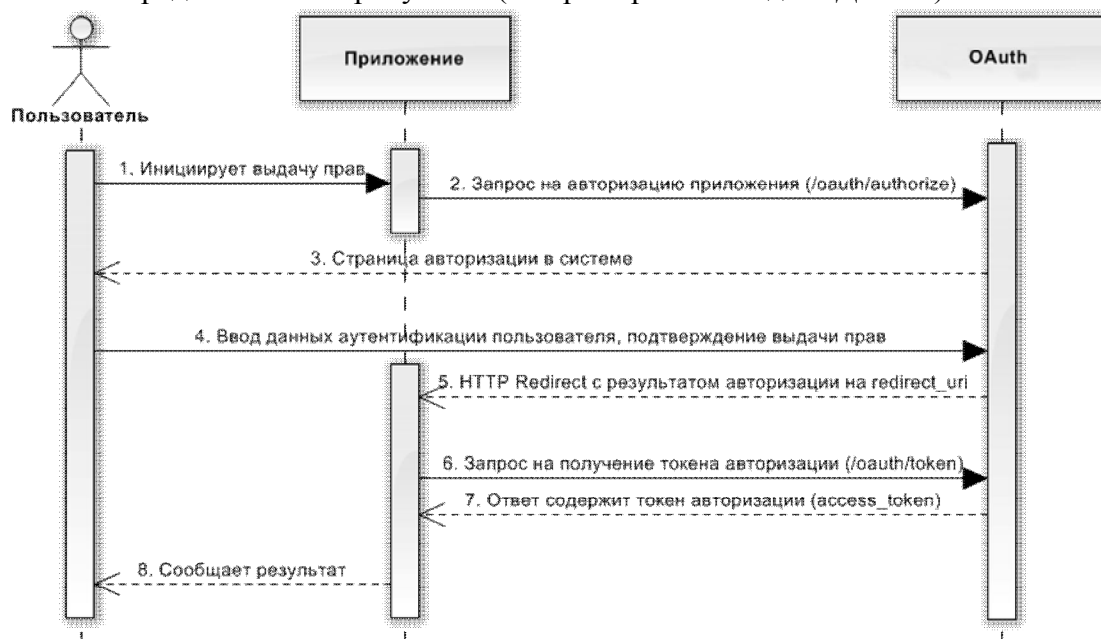


Рис. 1 – Упрощенная схема аутентификации с помощью OAuth.

Указанные протоколы являются наиболее распространёнными в сети Интернет и используются многими крупными веб-проектами. Основным достоинством данных протоколов является безопасность данных пользователя и сложность перехвата его личной информации. Из минусов можно выделить необходимость перехода пользователя на внешние сайты OpenID/OAuth-провайдера (подключение внешних скриптов с этих сайтов) и использование дополнительных специализированных библиотек авторизации.

Прямая авторизация является альтернативным вариантом решения поставленной задачи. При данном способе аутентификации пользователь вводит свои логин и пароль системы на любом сайте-сателлите и получает доступ к его ресурсам. При этом сам механизм аутентификации является скрытым от пользователя и происходит на стороне сервера.

Авторизация происходит по следующему алгоритму:

1. Пользователь передаёт на сайт-сателлит свои логин и пароль.
2. Сайт-сателлит невидимо для пользователя осуществляет запрос на сервер авторизации центрального сайта системы.

3. Сервер авторизации обрабатывает запрос от сайта-сателлита, и предоставляет в качестве результата в случае неудачи – код ошибки, в случае успеха - массив данных пользователя, в который обязательно входят идентификатор пользователя в системе и уникальный для данного сайта-сателлита и пользователя код, который используется для идентификации сессии пользователя на сайте.

4. Сайт-сателлит обрабатывает ответ сервера. При необходимости, в локальной базе данных сайта создаётся новый пользователь, или обновляются данные существующего пользователя.

5. Пользователь осуществляет вход на сайт или получает сообщение об ошибке входа.

Таким образом, аутентификация и создание локального профиля происходит незаметно для пользователя, без внешнего использования дополнительных серверов аутентификации. При каждом входе пользователя в систему осуществляется проверка его данных и обновление устаревшей информации в локальной базе сайта. Таким образом, данные пользователя, такие как его отображаемое имя, логин, контактная информация, всегда являются актуальными. Схема аутентификации представлена на рисунке 2.

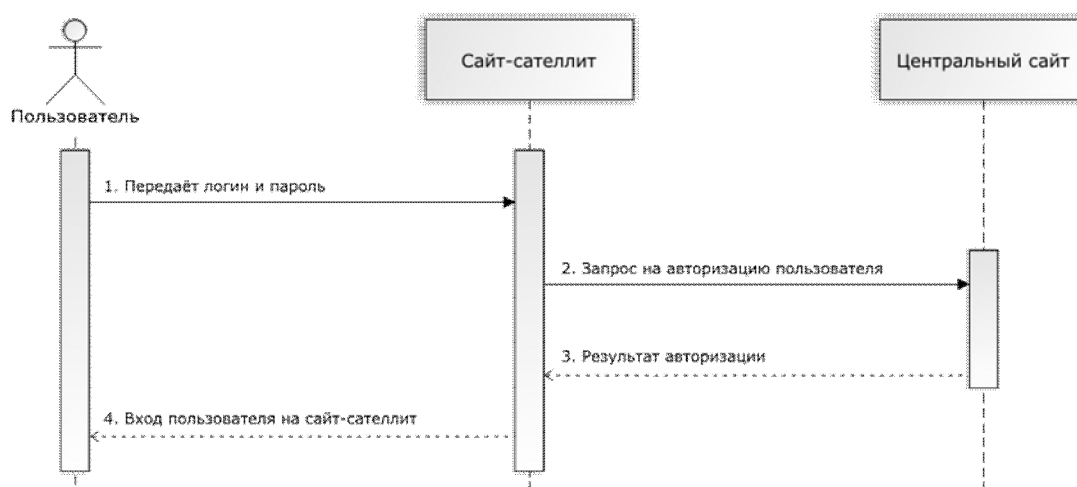


Рис. 2 – Схема прямой авторизации пользователей.

Плюсами прямой авторизации пользователей являются видимая простота реализации с точки зрения пользователя, использование привычного механизма логина и пароля, отсутствие дополнительных громоздких библиотек авторизации. К минусам подобного алгоритма стоит отнести проблемы с безопасностью использования логина и пароля на сторонних сайтах и подтверждение подлинности сайта и принадлежности его к системе до ввода пользователем своих персональных данных.

3. Центральный репозиторий.

Для организации множества сайтов-сателлитов и поддержки их в работоспособном состоянии, требуется центральный репозиторий и система обновления используемого на сайтах-сателлитах программного обеспечения.

Предполагается, что для всех сайтов-сателлитов в системе используется единая система управления контентом (CMS). CMS - информационная система или компьютерная программа, используемая для обеспечения и организации совместного процесса создания, редактирования и управления контентом (то есть содержимым) сайта. На основе единой CMS строится унифицированное взаимодействие сайтов с центральным сервером и пользователями системы. Допускается, что центральный сайт может также базироваться на этой CMS и отличаться от сайтов-сателлитов настройкой метода аутентификации пользователей.

Обязательным требованием к системе является модульность её построения и возможность сборки требуемого программного комплекса для сайта на основе готовых существующих

щих модулей из центрального репозитория. Таким образом, достигается гибкость системы и возможность оптимизации отдельных сайтов путём отключения тех или иных модулей.

В качестве используемой CMS могут выступать как системы общего назначения, такие как WordPress [6], phpBB [3] и Joomla [2], со специализированными модулями аутентификации и поддержки связи между сайтами, так и специализированная система AlterVision Core CMS.

Обязательным требованием к системе является автоматическое обновление CMS всех сайтов в её составе для поддержания актуальной версии и быстрого внедрения нового функционала. В случае с WordPress, автоматическое обновление и установка модулей (плагинов) может осуществляться средствами самой CMS и её центрального репозитория WordPress.org с возможностью инициации обновления системы и её модулей ответственным пользователем сайта-сателлита [6]. Система AlterVision Core CMS предлагает использование центрального сервера обновлений, отвечающего за автоматическую установку новых версий ядра и модулей системы на всех сайтах сети. Данный способ позволяет детально проконтролировать процесс обновления и снизить нагрузку на сервер обновлений за счёт контроля потока обновляющихся сайтов.

Рекомендуемым требованием к центральному репозиторию является система автоматического запуска сайтов. Это понятие подразумевает наличие инструментов, позволяющих осуществлять подготовку сервера, установку и запуск системы управления контентом, предварительную настройку сайта-сателлита, в автоматическом режиме без участия технических специалистов. При указанном способе расширение сети сайтов может происходить автоматически, путём подачи заявки на регистрацию сайта структурной единицей организации или сообщества, и последующим одобрением заявки администратором.

4. Информационный обмен между сайтами.

Центральный сайт системы и сайты-сателлиты являются связанными в единое информационное пространство. Для реализации этой связи, необходимо осуществление быстрого и удобного обмена информацией между сайтами в автоматическом режиме. В данном случае применяется механизм запросов, называемый Data Ping.

Data Ping – механизм реализации запросов между сайтами системы, позволяющий обмениваться массивами данных необходимого формата. Он является упрощённой реализацией стандарта XML-RPC. Технически Data Ping реализуется на базе стандартного GET- или POST-запроса к HTTP-серверу сайта. Обязательными полями в запросе являются его тип (идентификатор) и формат возвращаемого результата. Идентификатор запроса определяет функцию, которая отвечает за его обработку. В качестве формата представления возвращаемого результата, могут использоваться следующие варианты:

– **Сериализованные данные PHP.** При обмене информацией между скриптами, написанными на языке PHP, наиболее удобным способом передачи данных является сериализация. Она позволяет описать любые данные в качестве представления их в виде строки символов, с возможностью последующего развёртывания в аналогичную структуру данных на стороне сервера. Данный способ является наиболее удобным для внутреннего обмена информацией, так как не требует использования дополнительных инструментов, реализуется встроенными средствами PHP (стандартные функции `serialize` и `unserialize`) и является наиболее компактным с точки зрения объёма передаваемых данных.

– **Формат JSON.** Для взаимодействия сайтов с инструментами, описанными на JavaScript, оптимальным является использование стандарта описания данных JSON. JSON (англ. JavaScript Object Notation) — текстовый формат обмена данными, основанный на JavaScript и обычно используемый именно с этим языком. Несмотря на происхождение от JavaScript, формат считается языконезависимым и может использоваться практически с любым языком программирования. Для многих языков существует готовый код для создания и обработки данных в формате JSON [1].

– **Экспорт переменных в JavaScript.** Альтернативным вариантом для взаимодействия с JavaScript является использование данных, представленных напрямую на языке

JS, для последующего их «выполнения» с помощью стандартной команды *eval*. Данный способ не требует использования дополнительных библиотек поддержки формата JSON и позволяет получить данные на любом внешнем сайте.

– **XML**. Взаимодействие с любыми сторонними сервисами, работающими на любых языках и платформах, может быть реализовано на основе стандартизированного описания данных в виде XML. Средства обработки XML-документов в настоящее время реализованы для всех распространённых языков программирования и позволяют получать информацию с любых источников данных. Минусом использования формата XML является применение дополнительных библиотек обработки данных этого стандарта и значительное увеличение объёма передаваемых данных.

– **XML-RPC**. Расширением и уточнением использования формата XML является реализация стандарта взаимодействия на основе интерфейса XML-RPC. Плюсом данного стандарта является полная его поддержка любыми платформами. К минусам стоит отнести значительное увеличение объёма передаваемых данных по сравнению с XML [7].

При реализации механизма Data Ping подразумевается возможность использования любого из описанных выше стандартов представления данных в зависимости от запроса и свободного конвертирования внутреннего представления данных в системе в любой из указанных форматов без потери информации.

Использование данного механизма позволяет запрашивать у сайтов различные данные, такие как информацию о последних новостях, ленту обновлений; инициировать различные процессы на самом сайте, как автоматическое обновление ядра и модулей, обновление данных пользователя; осуществлять аутентификацию пользователей по механизму прямой авторизации.

Дополнительной возможностью информационного обмена является использование наиболее распространённых стандартов агрегации информации. В этих целях целесообразна реализация различных протоколов альтернативного отображения данных.

– **RSS/Atom** протокол позволяет отображать информацию сайтов-сателлитов в качестве ленты обновлений, для использования в любых сервисах автоматического сбора обновлений с сайтов. Применение протоколов RSS/Atom для экспорта информации позволяет организовать оперативное отображение необходимой актуальной информации сайтов-сателлитов в соответствующих разделах центрального сайта.

– **YML-описание** данных и аналогичные форматы позволяют осуществлять экспорт данных товаров и услуг, предлагаемых центральным сайтом и сайтами-сателлитами. YML (Yandex Market Language) — это стандарт, разработанный Яндексом для принятия и размещения информации в базе данных Яндекс.Маркета. YML основан на стандарте XML (XML DTD). Единый формат представления данных позволяет быстро и качественно обрабатывать предоставляемую магазинами информацию о товарах и услугах [8].

– **Sitemap** представление сайта необходимо для большинства поисковых сервисов и систем, в целях удобного построения карты сайта и качественной индексации его поисковыми системами. Возможно применение данных указанного формата для построения иерархической карты сайта-сателлита в рамках центрального сайта.

Возможно использование дополнительных форматов для передачи специфических данных поисковым системам и агрегаторам контента. В этот перечень могут входить данные словарей и энциклопедий, кулинарные рецепты, информация о вкладах и кредитах, вакансиях и различных объявлениях и прочая информация.

Вывод

Обозначенные выше принципы положены в основу построения распределённых сетей сайтов, позволяющих наиболее оптимальным образом организовать информационное пространство средних и крупных организаций и сообществ. На основе данных принципов воз-

можно минимизация затрат на организацию программной и аппаратной платформы для указанных сетей и построение удобной иерархической структуры составляющих сеть элементов.

Библиографические ссылки

1. JSON // Wikipedia [Электронный ресурс]. Режим доступа: <http://ru.wikipedia.org/wiki/JSON> (дата обращения: 12.05.11)
2. phpBB – официальный сайт [Электронный ресурс]. Режим доступа: <http://www.phpbb.com> (дата обращения: 12.05.11)
3. Joomla! – официальный сайт [Электронный ресурс]. Режим доступа: <http://www.joomla.org> (дата обращения: 12.05.11)
4. OAuth // Wikipedia [Электронный ресурс]. Режим доступа: <http://ru.wikipedia.org/wiki/OAuth> (дата обращения: 08.05.11)
5. OpenID // Wikipedia [Электронный ресурс]. Режим доступа: <http://ru.wikipedia.org/wiki/OpenID> (дата обращения: 08.05.11)
6. WordPress – официальный сайт [Электронный ресурс]. Режим доступа: <http://www.wordpress.org> (дата обращения: 12.05.11)
7. XML-RPC // Wikipedia [Электронный ресурс]. Режим доступа: <http://ru.wikipedia.org/wiki/XML-RPC> (дата обращения: 08.05.11)
8. Яндекс.Вебмастер - передача данных о содержимом сайта [Электронный ресурс]. Режим доступа: <http://content.webmaster.yandex.ru> (дата обращения: 10.05.11)

DISTRIBUTED WEBSITE NETWORKS CONSTRUCTION TECHNOLOGY

Reznichenko D. Anton, postgraduate student of ASOIU chair
Mordovia State University named after N.P. Ogariov, Saransk
Tel.: +7 (937) 515-88-67. E-mail: altrvision13@gmail.com

Summary. The article examines principles of construction on the distributed site website networks with unified information space. It is focused on selection of the optimal principles of unified user authentication and data between the various components of the network.

Keywords: unified information space, website network, authentication method, website, repository.