УДК 004.4

І.В. Кобзев, К.Е. Петров

Харківський національний університет внутрішніх справ, Харків

ПРИНЦИПИ ПОБУДОВИ РОЗПОДІЛЕНИХ WEB-СИСТЕМ

В статті розглядаються принципи побудови мереж сайтів з єдиним інформаційним простором на основі сайтів-сателітів. Основну увагу приділено виділенню оптимальних методів єдиної аутентифікації користувачів та інформаційного обміну поміж різними компонентами мережі.

Ключові слова: *єдиний інформаційний простір, мережа сайтів, методи аутентифікації, сайт, сателіт.*

Вступ

В теперішній час мережа Інтернет стала найбільш важливим і актуальним джерелом інформації. Крупні співтовариства і організації потребують надання своєї інформації, яка розбивається по різних галузях діяльності. Відомо, що розподілена інформаційна система складається з сукупності програмних компонентів, які взаємодіють один з одним. Кожен з таких компонентів є програмним модулем, що виконується в рамках окремого процесу. Критерієм цього розділення може виступати територіальне розташування (філії організації в різних містах, регіонах, країнах), сфера діяльності різних підрозділів організації, окремі персоналії і підрозділи, організації, співтовариства.

Постановка завдання. Актуальним є завдання побудови інформаційного простору організації з виділенням окремих структурних одиниць на основі сучасних Інтернет-орієнтованих технологій.

В теперішній час для реалізації цього завдання можуть бути використані різні стандартні рішення. Найбільш поширеним є ускладнення структури сайту організації, виділення окремих ієрархічних гілок сторінок і розділів, що описують ту або іншу ділянку діяльності організації. При цьому WEB-сайт значно ускладнюється, погіршується пошук по сайту і доступ до вкладених розділів. Другий варіант рішення поставленої задачі полягає у виділенні незалежних крупних розділів сайту. Ці розділи базуються на основі незалежних піддоменів. Також є можливість виділення для кожної з структурних одиниць окремого незалежного сайту (сателіта). В даному випадку кожна структурна одиниця стає відособленою і здійснює свою діяльність незалежно від інших. Інформаційний обмін між структурами в даному випадку здійснюється через адміністратора сайту.

В даній статті розглядається методика організації інформаційного WEB-простору для організацій і співтовариств, що враховує недоліки описаних вище способів. Пропонується організація інформаційного простору на основі розподіленої мережі сайтів-сателітів.

Опис принципів побудови мережі

Розподілена мережа сайтів – сукупність WEB-сайтів, об'єднаних загальними інформаційним простором, базами даних основних структурних одиниць (користувачі, підрозділи і т.і.), які розташовані на різних хостинг-платформах, що здійснюють взаємозв'язок на основі відповідних протоколів. В даному випадку організація (співтовариство) розглядається як сукупність деяких структурних одиниць. Управління структурними одиницями здійснює єдиний центр керування. Розглянемо основні принципи побудови мереж сайтів.

Розподіл складових частин

Основним принципом побудови мережі є розділення інформаційного простору на ряд сайтів за певною ознакою. Критерії виділення складових елементів наступні: територіальне і галузеве ділення, виділення персоналій і аналогічні варіанти, які залежать від даної організації або співтовариства.

У вказаному інформаційному просторі виділяється дві крупних базових складових – центральний сайт і чисельні сайти-сателіти.

Центральний сайт – інтернет-ресурс, що відповідає за здійснення функцій контролю і моніторингу мережі сайтів, зберігає в собі базу даних користувачів, сайтів і інших необхідних структурних одиниць системи. На основі центрального сайту організується технічна підтримка і супровід сайтів-сателітів системи, у тому числі їх встановлення, розгортання і оновлення. Спільно з центральним сайтом доцільне створення нестандартних ресурсів, які реалізовують специфічні функції для всіх сайтів-сателітів. Це може бути загальний центральний форум проекту, інтернет-магазин, який об'єднує в собі товари підрозділів, і інші ресурси, основним завданням яких є агрегація даних сайтів-сателітів. Спільно з центральним сайтом вони складають ядро системи.

Однією з основних функцій центрального сайту є надання користувачеві короткої інформації про сайти-сателіти. Обов'язковою умовою є наявність повного списку сайтів-сателітів системи. Рекомендується створення інформаційних сторінок для кожного з сайтів системи, що містять короткий опис сайту, його регіональну або персональну приналежність, заголовки останніх новин і публікацій сайту та подібну інформацію в скороченому вигляді (дайджест). Зміст цієї інформації визначається адміністр-

© І.В. Кобзев, К.Е. Петров

ратором сайту-сателіту. Сайт-сателіт – це елементарна структурна одиниця мережі. Дані сайти будуються на основі однакових логічних принципів, єдиної системи управління контентом і загальних стандартів представлення інформації в такій мережі. Сателіт в Інтернеті – це невеликий сайт, який входить в систему якого-небудь більш значущого сайту. Звичайно, в одного добре просунутого сайту (згідно рейтингу) є не один, а декілька сателітів. Кожен сателіт містить в собі частину інформації з основного сайту, але більш вузькоспеціалізовану або містять інформацію (посилання) про головну сторінку основного сайту. Сайти-сателіти призначені для ширшого кола користувачів, яких цікавить в цілому область запропонованого товару або послуги. Основною їх функцією є просування сайту в пошукових системах на вищі позиції [1].

Передбачається, що сайти-сателіти і центральний сайт розмішені на різних хостинг-платформах та серверах, за обслуговування кожного з яких відповідає структурна одиниця, що володіє сайтом. Таким чином, досягається ряд переваг:

- система є стійкою до виходу з ладу одного або декількох елементів, у тому числі не лише сайтів-сателітів, але і різних складових центрального сайту. Основний функціонал сайтів-сателітів не залежить від працездатності центрального сайту системи, так само як і центральний сайт системи здатен підтримувати свою функціональність в разі відмови деяких сайтів-сателітів, надаючи користувачу коротку інформацію (дайджест) з цих сайтів;
- значно знижуються технічні вимоги до апаратної платформи для сайтів-сателітів, внаслідок розподілу навантаження і відвідуваності по множині сайтів в рамках мережі;
- зменшуються технічні вимоги до апаратної платформи центрального сайту за рахунок відтоку відвідувачів на сайти-сателіти.

Розподіл інформаційних ресурсів мережі по різних сайтах дозволяє значно скоротити витрати на забезпечення функціонування апаратної технічної бази.

Єдина аутентифікація

Розподіл ресурсів і інформації по різних сайтах вимагає уніфікації доступу до всіх інформаційних ресурсів. Необхідною умовою в даному випадку є введення єдиної для всіх сайтів бази користувачів системи з виділеннями ним деяких базових прав доступу.

Для зберігання облікових записів користувачів використовується центральний сайт. Його засобами реалізується реєстрація користувачів, відкрита або закрита, і управління їх профайлами. Таким чином, на центральному сайті системи розташовується повна об'ємна база даних користувачів мережі і їх основних прав доступу до ресурсів. Кожен з сайтів-сателітів містить локальну копію вказаної бази даних в скороченому вигляді. Використання скороченої бази даних дозволяє значно зменшити вимоги до апаратної платформи за рахунок полегшення виконання запитів до бази даних. При цьому способі в локальній базі даних сайту-сателіта зберігається інформація лише про тих користувачів, хто хоч би один раз авторизувався на сайті.

Авторизація користувачів на сайті може здійснюватися за двома різними алгоритмами: OAuth (OpenID) і пряма авторизація з введенням логіну і паролю.

OAuth (скорочення від англ. Open Authentication) – це відкритий стандарт аутентифікації, який дозволяє користувачам відкривати доступ до своїх приватних даних (фотографії, відео, списки контактів), що зберігаються на одному сайті, іншому сайту, без необхідності вводу імені користувача та паролю.

OAuth дозволяє користувачам роздавати сайтам маркери доступу до даних, що розміщуються на сайтах-сервісах. Кожен маркер доступу надає доступ конкретному сайту (наприклад, сайту редагування відео) до конкретних ресурсів (наприклад, тільки відео від конкретного альбому) та на визначений термін (наприклад, на наступні 2 години). Це дозволяє користувачам надавати доступ третім сайтам до їх інформації, що зберігається на інших сайтах-постачальниках послуг, не передаючи повною мірою самих даних та без застосування імені/паролю [2].

OpenID – протокол для прискореної реєстрації. OpenID – це децентралізована система єдиного входу, яка дозволяє використовувати один обліковий запис (ім'я користувача і пароль) на великій кількості сайтів. На сайтах з підтримкою OpenID, користувачам не доводиться запам'ятовувати дані для аутентифікації. Натомість, їм достатньо бути зареєстрованими на сайті «провайдера ідентифікації» OpenID (який надає ідентифікатор). Оскільки технологія OpenID децентралізована, то будь-який сайт може використовувати програмне забезпечення OpenID в якості засобу аутентифікації користувачів. OpenID розв'язує проблему не покладаючись на централізований сайт для підтвердження істинності користувача.

Вказані протоколи є найбільш поширеними в мережі Інтернет і використовуються багатьма крупними веб-проектами. Основною перевагою даних протоколів є безпека даних користувача і складність перехоплення його особистої інформації. До недоліків можна віднести необхідність переходу користувача на зовнішні сайти OpenID/OAuth-провайдера (підключення зовнішніх скриптів з цих сайтів) і використання додаткових спеціалізованих бібліотек авторизації.

Пряма авторизація є альтернативним варіантом вирішення поставленої задачі. При даному способі аутентифікації користувач вводить свої логін і пароль системи на будь-якому сайті-сателіті і отримує доступ до його ресурсів. При цьому сам механізм аутентифікації є прихованим від користувача і відбувається на стороні сервера.

Авторизація відбувається за таким алгоритмом.

1. Користувач передає на сайт-сателіт свої логін і пароль.

2. Сайт-сателіт здійснює запит на сервер авторизації центрального сайту системи.

3. Сервер авторизації обробляє запит від сайту-сателіту, і надає як результат в разі невдачі – код помилки, в разі успіху - масив даних користувача, в який обов'язково входять ідентифікатор користувача в системі і унікальний для даного сайту-сателіту і користувача код, який використовується для ідентифікації сесії користувача на сайті.

4. Сайт-сателіт обробляє відповідь сервера. При необхідності, в локальній базі даних сайту створюється новий користувач, або оновлюються дані існуючого користувача.

5. Користувач здійснює вхід на сайт або отримує повідомлення про помилку входу.

Все вищеозначене виконується за допомогою протоколу HTTPS.

Таким чином, аутентифікація і створення локального профілю відбувається непомітно для користувача, без зовнішнього використання додаткових серверів аутентифікації. При кожному входженні користувача в систему здійснюється перевірка його даних і оновлення застарілої інформації в локальній базі сайту (рис. 1). Таким чином, дані користувача, такі як його ім'я, що відображується, логін, контактна інформація, завжди є актуальними.

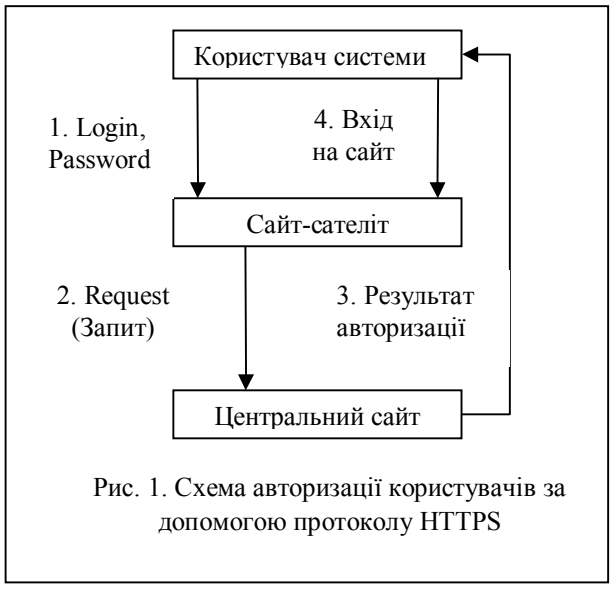

Рис. 1. Схема авторизації користувачів за допомогою протоколу HTTPS

Плюсами прямої авторизації користувачів є видима простота реалізації з точки зору користувача, використання звичного механізму логіну і паролю, відсутність додаткових громіздких бібліотек авторизації. До мінусів подібного алгоритму необхідно віднести проблеми з безпекою використання логіну і паролю на сторонніх сайтах і підтвердження достовірності сайту і приналежності його до системи до введення користувачем своїх персональних даних.

Центральний репозиторій

Репозиторій – спеціальний сервер, з якого можна завантажити програмне забезпечення. На сервері зберігається архів програмних продуктів, які доступні для завантаження. Це місце, де зберігаються і підтримуються які-небудь дані. Найчастіше дані в репозиторії зберігаються у вигляді файлів, доступних для подальшого розповсюдження по мережі [4].

Для організації множини сайтів-сателітів і підтримки їх в працездатному стані, потрібен центральний репозиторій і система оновлення програмного забезпечення, яке використовується на сайтах-сателітах.

Передбачається, що для всіх сайтів-сателітів в системі використовується єдина система управління контентом (CMS). CMS – інформаційна система або комп'ютерна програма, що використовується для забезпечення та організації спільного процесу створення, редагування і управління контентом (вмістом) сайту. На основі єдиної CMS будується уніфікована взаємодія сайтів з центральним сервером і користувачами системи. Допускається, що центральний сайт може також базуватися на цій CMS і відрізнятися від сайтів-сателітів налаштуванням методу аутентифікації користувачів.

Обов'язковою вимогою до системи є модульність її побудови і можливість збірки необхідного програмного комплексу для сайту на основі готових існуючих модулів з центрального репозиторію. Таким чином, досягається гнучкість системи і можливість оптимізації окремих сайтів шляхом відключення тих або інших модулів.

В якості CMS є можливість використовувати безкоштовні (ліцензія GNU) системи які базуються на базах даних (WordPress, Joomla, DataLife Engine, E107) або на текстових файлах (Linkor CMS, ORKZ, WCPS, Flatnux CMS) із спеціалізованими модулями аутентифікації і підтримки зв'язку між сайтами.

Обов'язковою вимогою до системи є автоматичне оновлення CMS всіх сайтів в її складі для підтримки актуальної версії і швидкого впровадження нового функціонала. У випадку з WordPress, автоматичне оновлення і встановлення модулів (плагінів) може здійснюватися засобами самої CMS і її центрального репозиторія WordPress.org з можливістю ініціації оновлення системи і її модулів відповідальним користувачем сайту-сателіту.

Система HostCMS пропонує використання центрального сервера оновлень, що відповідає за автоматичну установку нових версій ядра і модулів системи на всіх сайтах мережі. Такий спосіб дозволяє детально контролювати процес оновлення і знизити навантаження на сервер оновлень за рахунок контролю потоку сайтів, що оновлюються. Модуль «Автоматическое обновление системы» системи HostCMS дозволяє автоматично встановлювати оновлення з розширенням функціоналу і новими модулями, підтримуючи систему управління в актуаль-

ному стані. Завантаження оновлень відбувається автоматично через центр адміністрування [5].

Головною вимогою до центрального репозиторію є наявність системи автоматичного запуску сайтів. Це означає наявність інструментів, що дозволяють здійснювати підготовку сервера, встановлення і запуск системи управління контентом, попереднє налаштування сайту-сателіта, в автоматичному режимі без участі технічних фахівців. При вказаному способі, розширення мережі сайтів може відбуватися автоматично, шляхом подачі заявки на реєстрацію сайту структурною одиницею організації або співтовариства, і подальшим схваленням заявки адміністратором.

Інформаційний обмін між сайтами

Центральний сайт системи і сайти-сателіти зв'язані в єдиний інформаційний простір. Для реалізації цього зв'язку, необхідне здійснення швидкого і зручного обміну інформацією між сайтами в автоматичному режимі. В даному випадку застосовується механізм запитів – Data Ping.

Data Ping – механізм реалізації запитів між сайтами системи, що дозволяє обмінюватися масивами даних необхідного формату. Він є спрощеною реалізацією стандарту XML-RPC. Технічно Data Ping реалізується на базі стандартного GET- або POST-запиту до HTTP-серверу сайту. Обов'язковими полями в запиті є його тип (ідентифікатор) і формат результату, що повертається. Ідентифікатор запиту визначає функцію, яка відповідає за його обробку. Як формат представлення результату, що повертається, можуть використовуватися такі варіанти.

При обміні інформацією між скриптами, написаними на мові PHP, найбільш зручним способом передачі даних є серіалізація. Вона дозволяє описати будь-які дані як представлення їх у вигляді рядка символів, з можливістю подальшого розгортання в аналогічну структуру даних на стороні сервера. Даний спосіб є найбільш зручним для внутрішнього обміну інформацією, оскільки не вимагає використання додаткових інструментів, реалізується вбудованими засобами PHP (стандартні функції serialize і unserialize) і є найбільш компактним з точки зору об'єму переданих даних.

Для взаємодії сайтів з інструментами, описаними на JavaScript, оптимальним є використання стандарту опису даних JSON. JSON (англ. JavaScript Object Notation) – текстовий формат обміну даними, заснований на JavaScript і зазвичай використовуваний саме з цією мовою. Не дивлячись на походження від JavaScript, формат вважається мовонезалежним і може використовуватися практично з будь-якою мовою програмування. Для багатьох мов існує готовий код для створення і обробки даних у форматі JSON.

Альтернативним варіантом для взаємодії з JavaScript є використання даних, представлених безпосередньо на мові JS, для подальшого їх «виконання». Даний спосіб не вимагає використання додаткових бібліотек підтримки формату JSON і дозволяє отримати дані на будь-якому зовнішньому сайті.

Взаємодія з будь-якими сторонніми сервісами, що працюють на будь-яких мовах і платформах, може бути реалізована на основі стандартизованого опису даних у вигляді XML. Засоби обробки XML-документів в даний час реалізовані для всіх поширених мов програмування і дозволяють отримувати інформацію з будь-яких джерел даних. Недоліком використання формату XML є вживання додаткових бібліотек обробки даних цього стандарту і значне збільшення об'єму даних, що передаються.

XML-RPC. Розширенням і уточненням використання формату XML є реалізація стандарту взаємодії на основі інтерфейсу XML-RPC. Плюсом даного стандарту є повна його підтримка будь-якими платформами. До мінусів варто віднести значне збільшення об'єму даних, що передаються в порівнянні з XML.

При реалізації механізму Data Ping мається на увазі можливість використання будь-якого з описаних вище стандартів представлення даних залежно від запиту і вільної конвертації внутрішнього представлення даних в системі в будь-якому з вказаних форматів без втрати інформації.

Використання даного механізму дозволяє отримувати з сайтів різні дані, такі як інформація про останні новини, стрічка оновлень; ініціювати різні процеси на самому сайті, такі як автоматичне оновлення ядра і модулів, оновлення даних користувача; здійснювати аутентифікацію користувачів згідно з механізмом прямої авторизації.

Додатковою можливістю інформаційного обміну є використання найбільш поширених стандартів агрегації інформації. Тому є доцільним реалізація різних протоколів альтернативного відображення даних.

RSS/Atom-протокол дозволяє відображати інформацію сайтів-сателітів як стрічку оновлень, для використання в будь-яких сервісах автоматичного збору оновлень з сайтів. Вживання протоколів RSS/Atom для експорту інформації дозволяє організувати оперативне відображення необхідної актуальної інформації сайтів-сателітів у відповідних розділах центрального сайту.

YML-опис даних і аналогічні формати дозволяють здійснювати експорт даних товарів і послуг, пропонованих центральним сайтом і сайтами-сателітами. YML (Yandex Market Language) – це стандарт, розроблений Яндексом для прийняття і розміщення інформації в базі даних Яндекс-Маркету. YML заснований на стандарті XML (XML DTD). Єдиний формат представлення даних дозволяє швидко і якісно обробляти інформацію, що надається магазинами про товари і послуги.

Sitemap представлення сайту необхідне для більшості пошукових сервісів і систем, з метою зручної побудови карти сайту і якісної індексації його пошуковими системами. Можливе використання

даних вказаного формату для побудови ієрархічної карти сайту-сателіта в рамках центрального сайту.

Також існує можливість використання додаткових форматів для передачі специфічних даних пошуковим системам і агрегаторам контента. У цей перелік можуть входити дані словників і енциклопедій, інформація про вклади і кредити, вакансії, різні оголошення і інша інформація.

Висновки

Зазначені вище принципи покладені в основу побудови розподілених мереж сайтів, що дозволить найбільш оптимальним чином організувати інформаційний простір середніх і крупних організацій і співтовариств. На основі даних принципів можлива мінімізація витрат на організацію програмної і апаратної платформи для мережі, яка побудована на основі WEB-технологій.

Список літератури

1. Что такое сателлит? [Електронний ресурс]. – Режим доступу URL: http://ip-whois.net/wtfsattelit.php.

2. OAuth. Вікіпедія [Електронний ресурс]. – Режим доступу URL: http://uk.wikipedia.org/wiki/OAuth.

3. OpenID. Вікіпедія [Електронний ресурс]. – Режим доступу URL: http://uk.wikipedia.org/wiki/ OpenID .

4. Репозиторії мережі KPI-IX. Технології Blogger. – [Електронний ресурс]. Режим доступу URL: http:// ipeacocks.blogspot.com/2009/03/kpi-ix.html.

5. Система управления сайтом HostCMS v. 5 [Електронний ресурс]. – Режим доступу URL: http://www. hostcms.ru/hostcms/.

Надійшла до редколегії 11.10.2011

Рецензент: д-р техн. наук, проф. С.Г. Удовенко, Харківський національний університет радіоелектроніки, Харків.

ПРИНЦИПЫ ПОСТРОЕНИЯ РАСПРЕДЕЛЕННЫХ WEB-СИСТЕМ

И.В. Кобзев, К.Е. Петров

В статье рассматриваются принципы построения сетей сайтов с единым информационным пространство на основе сайтов-сателлитов. Основное внимание уделено выделению оптимальных методов единой аутентификации пользователей и информационного обмена между различными компонентами сети.

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

PRINCIPLES FOR BUILDING DISTRIBUTED WEB-SYSTEMS

I.V. Kobzev, K.E. Petrov

The article deals with the principles of networking sites with a single information space based on the sites of the satellite. The focus is on best practices in the allocation of a single user authentication and data exchange between different components of the network.

Keywords: *single informative space, network of sites, methods of authentification, site, sattelite.*