

УДК 004.75

К.т.н., доцент Петрашенко А.В., студент Шолох О. О.

Національний технічний університет України
«Київський політехнічний інститут»

МЕТОД ОРГАНІЗАЦІЇ РОЗПОДІЛЕНОГО МОНІТОРИНГУ WEB-РЕСУРСІВ З ВИКОРИСТАННЯМ ОБЧИСЛЮВАЛЬНИХ ПОТУЖНОСТЕЙ КЛІЄНТІВ ЗА ДОПОМОГОЮ WEB-БРАУЗЕРА

Abstract

*Andriy Petrashenko, Assoc. Prof., PhD; Oleksandr Sholokh, student
The method of distributed Web-monitoring with using users
computing power by Web-browser*

This paper concerns the task of creating new method of organization distributed computing Web-monitoring system. The main idea for this purpose is using users computing power by Web-browsers for Web-monitoring.

Most of the time, when you're reading a web page, or surfing the internet, the processing power required on the client is little to none. The idea of using power of billions internet users can provide large computing power for science purposes.

Researches were provided for the tasks of Web-resources monitoring, however it can be used in other areas with large computing power demand.

Вступ

Проблема недостатньої кількості обчислювальних потужностей необхідних для забезпечення якісного моніторингу Web-ресурсів є надзвичайно актуальною в час, коли кількість інформації в світовій павутині збільшується надзвичайно швидко.

В останні роки застосування Web-додатків значно розширились. Перегляд електронної пошти, фото, відео, покупки товарів, спілкування з друзями та колегами переходить з великого різноманіття окремих спеціалізованих додатків до одного – Web-браузера. Користувачі проводять години часу користуючись цим додатком. При цьому обчислювальна потужність їх комп'ютерів в більшості випадків використовується ледве на половину. Ідея використання обчислювальних ресурсів, що не використовуються користувачами персональних комп'ютерів, не нова. З кожним роком вона набирає все більшої популярності. Підходи до реалізації обчислювальних систем, що могли б використовувати дані ресурси можуть бути різні. Одні пропонують волонтерам (власникам персонального комп'ютера, який дає згоду на

використання його апаратного забезпечення) встановлювати спеціалізоване програмне забезпечення, інші дають можливість встановити собі додаток, який може виконувати інші корисні функції (наприклад відео програвач).

Сучасні підходи для асинхронної передачі даних (Ajax) дають можливість за допомогою спеціальних скриптів мовою програмування JavaScript, яка підтримується більшістю браузерів, легко використовувати обчислювальні потужності, що не використовуються користувачем.

Постановка задачі

Метою даного дослідження є розробка методу організації розподіленої системи моніторингу Web-ресурсів з використання обчислювальних потужностей Web-браузера.

Даний метод повинен забезпечити ефективне використання обчислювальної потужності користувачів. При цьому обмін даними між клієнтською та серверною частинами повинен проходити через текстовий протокол обміну даними (HTTP).

Також необхідно оцінити показники швидкості обробки інформації такої системи та розглянути можливість використання даного підходу в інших задачах, що потребують значних обчислювальних ресурсів.

Опис методу

Основною ідеєю запропонованого методу є використання обчислювальних потужностей користувачів за допомогою їх браузерів. Реалізація даної ідеї включає в себе розробку серверної та клієнтської частини.

Серверна частина повинна здійснювати контроль за роботою системи розподіляючи навантаження по розбору сторінок між вузлами клієнтської частини.

Клієнтська частина являє собою набір вузлів. Окремий вузол – це Web-браузер з інтерпретатором мови програмування JavaScript. Користувач даного Web-браузера повинен відкрити певну сторінку, на якій підключений спеціалізований код на мові програмування JavaScript для розбору HTML коду сторінки.

Специфікації безпеки сучасних браузерів не дають можливості відправляти асинхронні запити на доменні, імена відмінні від того, на якому знаходиться користувач, переглядаючи Web-сторінку. Данні специфікації не дають можливості завантажувати HTML код

безпосередньо з вузлів, тому дана задача повинна виконуватись на серверній частині.

Розподілення навантаження на вузли відбувається шляхом посилення завдань щодо розбору HTML коду певної сторінки. Нова задача не буде відправлятися, доки вузол не виконав поставлену йому задачу або не пройде виділений ліміт часу. Цей ліміт часу для кожного вузла необхідно визначати індивідуально, оскільки потужність персональних комп'ютерів користувачів може бути різною.

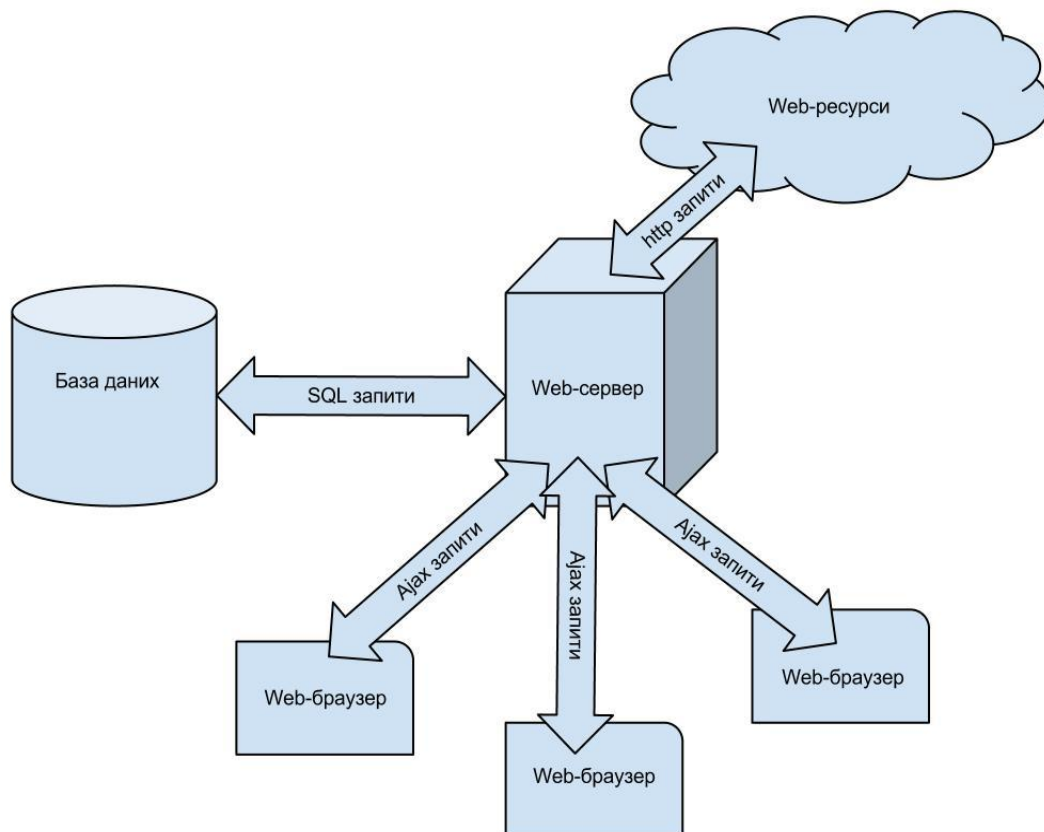


Рис. 1. Загальна схема роботи основних компонент.

Згідно запропонованого методу серверна частина являє собою Web-сервер, який виконує наступні функції:

- забезпечувати хостинг JavaScript коду клієнтської частини;
- завантажувати HTML код Web-сторінок, над якими здійснюється моніторинг, та передавати його на клієнтську частину за допомогою Ajax запитів;
- отримувати результати обробленого HTML коду та заносити їх до бази даних.

Клієнтська частина – це спеціалізований JavaScript код, який виконує наступні завдання:

- отримання HTML коду з серверної частини;
- проведення аналізу розбір даного документу;
- передача результатів на серверну частину.

Практично будь-який сучасний браузер містить інтерпретатор мови програмування JavaScript.

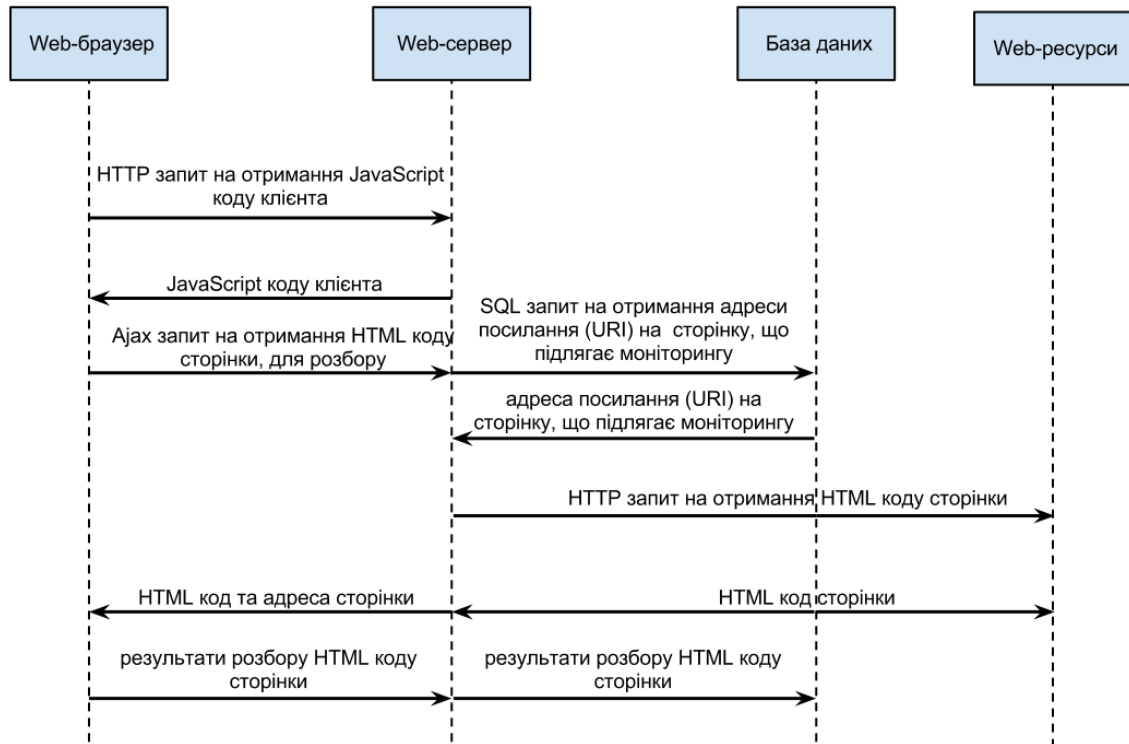


Рис. 2. Діаграма роботи основних компонент.

За допомогою Ajax можна відправляти та отримувати запити з сервера не оновлюючи сторінку, що не створює жодних перешкод для роботи користувача з основним Web-додатком, в який інтегрована дана система.

Даний підхід є особливо ефективним для роботи з HTML кодом, оскільки JavaScript включає в себе ефективні методи для роботи з доменною структурою Web-сторінки.

Для дослідження ефективності даного підходу застосовуються задача розбору HTML коду та пошуку в ньому посилань на інші Web-сторінки.

Висновки

Даний метод вводить новий підхід до реалізації розподіленого моніторингу Web-ресурсів. Він надає можливість створювати ефективні системи, що використовують спеціалізоване обладнання для проведення

текстового розбору HTML коду, оскільки дана задача делегується на Web-браузери користувачів.

Розроблений підхід може бути використаний для інших задач, що потребують великих обчислювальних потужностей.

Література

1. *D. P. Anderson, J. Cobb, E. Korpela, M. Lebofsky, and D. Werthimer.* SETI@home: An experiment in public-resource computing. *Communications of the ACM*, Nov. 2002, Vol. 45 No. 11, pp. 56-61.

2. *Згуровский М.З., Петренко А.И.* Е-наука на шляху до семантичного Грід. Частина 1: Об'єднання Web- і Грід- технологій.- //Системні дослідження і інформаційні технології.-Київ, №1, 2010.-с.26-38.

3. *Киселев Г.Д.* Мониторинг мультисервисных компьютерных сетей средствами системы Nagios / Киселев Г.Д., Шпакаускас М.С. // *Электроника и связь : тематический выпуск «Электроника и нанотехнологии».* – 2010. – № 3. – С. 178–181.

4. *Маннинг К., Рагхаван П., Шютце Х.* Введение в информационный поиск — Вильямс, 2011. – С. 121–152.