

статей з проблематики глибинного веб: «What is the Invisible Web?», «Finding the Invisible Web», «Top Places to Search the Invisible Web» та ін.

Politicalinformation.com – сервіс, що забезпечує оперативний пошук в 5000 відібраних веб-сайтів політичної спрямованості, надання новин з декількох десятків авторитетних джерел.

Infomine – сервіс забезпечує видобуток інформації з баз даних, електронних журналів (блогів), електронних доіок оголошень, електронних книг, списків розсилок, електронних каталогів тощо, переважно пізнавально-освітнього характеру. Забезпечує як загальний пошук, так і пошук по тематичним категоріям.

Особливість більшості «прихованих» ресурсів полягає у їх вузькій спеціалізації. Для пошуку в них використовуються ті ж механізми, що і для «поверхневого» веб, проте, в більшості випадків, роботи пошукових систем для глибинного веб включають унікальні для кожного такого ресурсу модулі доступу до даних.

Традиційна пошукова система найчастіше може видати адресу бази даних, але не скаже, які документи конкретно містяться в ній. Типовий приклад – інформаційно-пошукові системи з українського (zakon.rada.gov.ua) законодавства. Тисячі документів з баз даних стають доступні тільки після входу в систему, а роботи стандартних пошукових систем не в змозі заіндексувати контент баз даних.

Парадоксально, але в якості одного з ресурсів глибинного веб можна розглядати і архів матеріалів відкритого веб-простору. Такий архів – Internet Archive з 1996 року створює компанія Alexa. Сьогодні обсяг бази даних Alexa перевищує 350 млрд. веб-сторінок.

Технологія сховища Alexa включає ряд сучасних засобів управління гігантським документальним сховищем. Наприклад, за допомогою технології Alexa виконується кластеризація веб-ресурсів, тобто формування колекцій документів, близьких по тематиках. Особливий інтерес у користувачів сервісу Alexa викликає «Машина часу» (Wayback Machine), що відкриває доступ до тимчасових зрізів веб-простору. Одне з найбільш цікавих практичних застосувань цієї технології – відновлення документів, колись опублікованих в веб-просторі, але згодом видалених. При цьому зростання глибинного веб загрожує серйозними проблемами повноти в сховищі системи, пов'язаними зі зростаючою кількістю сайтів, які експлуатують різні типи технологій управління контентом, динамічною публікацією документів з баз даних тощо.

Література.

1. He B., Patel M., Zhang Z., Chang K. C.-C. Accessing the Deep Web: A Survey // Communications of the ACM (CACM), 50(5):94-101, 2007.

УДК 621.326

Рокош М.– ст. гр. СНм-51

Тернопільський національний технічний університет імені Івана Пулюя

ДОЦІЛЬНІСТЬ ВИКОРИСТАННЯ NOSQL

Науковий керівник: к.т.н., доцент Загородна Н.В.

M. Rokosh

Ternopil Ivan Pul'uj National Technical University

THE REASONABILITY OF USING NOSQL

Supervisor: Zahorodna N.

Ключові слова: NoSQL, бази даних, MongoDB, СУБД, РСУБД

Keywords: NoSQL, databases, MongoDB, DBMS, RDBMS

Поняття NoSQL (Not Only SQL або Nosql) здобуло популярність з 2009 року. Саме тоді розвиток web-технологій і соціальних сервісів дав поштовх безлічі нових підходів до зберігання і обробки даних. Розробники таких додатків зіткнулися з завданнями, для яких традиційні реляційні СУБД виявилися або занадто дорогими, або недостатньо продуктивними. Крім того, популяризаторами відмови від універсальних рішень (реляційні СУБД - РСУБД) на користь спеціалізованих стали стартапи та ті, кому доводиться працювати з Big Data. Треба розуміти, що NoSQL-рішення не обов'язково означає заміну і повну відмову від РСУБД.

Горизонтальне масштабування існуючих традиційних СУБД зазвичай є трудомісткою, дорогою і ефективною тільки до певного рівня завдачею. У той же час багато NoSQL-рішень проектувалися виходячи з необхідності масштабуватись горизонтально і робити це «на ходу». Тому ця процедура є зазвичай простішою і прозорішою в NoSQL, ніж в РСУБД.

Продуктивність БД на одному вузлі, а не в кластері, також є важливим параметром. Для багатьох завдань такі властивості традиційних СУБД, як транзакційність, ізолюваність змін, надійність в межах одного вузла або навіть сама реляційна модель, не завжди потрібні в повному обсязі. Тому відмова від цих властивостей (всіх або деяких) дозволяє NoSQL іноді досягати більшої продуктивності на одному вузлі, порівняно з традиційними рішеннями. Надійна робота в умовах, коли відмова заліза або мережева недоступність - звичайна справа, є одним з властивостей багатьох рішень NoSQL. Основний спосіб її забезпечення - це реплікація. Сама по собі реплікація аж ніяк не є унікальною особливістю NoSQL, але тут, як і при масштабуванні, важливу роль відіграють ефективність і простота внесення змін в існуючу інсталяцію. Перехід БД до роботи в режимі реплікації - це просте завдання для більшості NoSQL-рішень.

Простота розробки та адміністрування - також важливий аргумент на користь NoSQL-технологій. Цілий ряд завдань, пов'язаних з масштабуванням і реплікацією, що представляють значну складність і вимагають спеціальної експертизи на традиційних СУБД, у NoSQL займає лічені хвилини. Завдання установки і настройки, саме використання NoSQL-рішень зазвичай істотно простіше і менш трудомістке, ніж у випадку з РСУБД. Тому NoSQL-системи стали очевидним вибором для багатьох стартапів, де швидкість розробки і впровадження є ключовим фактором.

Два напрямки розвитку СУБД - SQL і NoSQL - багато в чому протиставляються, хоча і підкоряються єдиним законам і рухаються назустріч один одному. Поки цей рух майже непомітний, але він активізується в міру появи необхідності в засобах побудови високонавантажених систем, в результаті чого відбудеться злиття цих напрямків. Утворені таким чином СУБД можуть виявитися занадто громіздкими для використання на практиці, тому, можливо, відбудеться відмова від поточної архітектурної концепції СУБД в сторону появи СУБД, що представляють собою набір модулів для побудови цільової системи, яка матиме характеристики, необхідні для конкретних високонавантажених додатків.