

УДК 004.4

**П. П. Костенко, І. В. Левченко**

Кременчуцький національний університет імені Михайла Остроградського, Кременчук

### **ВЕБ-СЕРВІС УТОЧНЕННЯ РЕЛЕВАНТНОСТІ ВЕБ-ДОКУМЕНТІВ ПОШУКОВОЇ ВИДАЧІ GOOGLE НА ОСНОВІ ПОВЕДІНКИ КОРИСТУВАЧА**

Проаналізовано стан популярності всесвітньої мережі Інтернет та розглянуто основні проблеми пошукових систем, зважаючи на приведену статистику перевантаження інформації. Сформульовано практичну проблему встановлення релевантності в сучасних пошукових системах, розглянуто існуючі рішення для підвищення релевантності пошукової видачі та запропоновано новий підхід до уточнення релевантності веб-ресурсів. Розглянуто основні технічні засоби розробки запропонованого сервісу з аргументацією щодо вибору технічного інструментарію. Наведено опис інтерфейсу, приклад використання та представлено переваги нового методу уточнення релевантності веб-ресурсів на основі поведінки користувача.

**Ключові слова:** Інтернет, Google, пошукова система, релевантність.

**П. П. Костенко, І. В. Левченко**

Кременчугский национальный университет имени Михаила Остроградского, Кременчуг

### **ВЕБ-СЕРВИС УТОЧНЕНИЯ РЕЛЕВАНТНОСТИ ВЕБ-ДОКУМЕНТОВ ПОИСКОВОЙ ВЫДАЧИ GOOGLE НА ОСНОВНИИ ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЯ**

Проанализировано состояние популярности всемирной сети Интернет и рассмотрены основные проблемы поисковых систем основываясь на приведенной статистике перегрузки информации. Сформулирована практическая проблема определения релевантности в современных поисковых системах, рассмотрены существующие решения для увеличения релевантности поисковой выдачи и предложен новый подход к уточнению релевантности запросов. Рассмотрены основные технические средства разработки предложенного сервиса с аргументацией выбора технического инструментария. Приведено описание интерфейса, пример использования и представлены преимущества нового метода уточнения релевантности веб-ресурсов на основании поведения пользователя.

**Ключевые слова:** Интернет, Google, поисковая система, релевантность.

**АКТУАЛЬНІСТЬ РОБОТИ.** Всесвітня мережа Інтернет стає основною бібліотекою знань сучасної людини. Доступ до всесвітньої павутини є навіть на бюджетних мобільних телефонах, що вже дозволяє користувачам переглядати елементарну текстову інформацію – новини, статті, книги тощо. Аналітики із організації «PricewaterhouseCoopers» передбачають що в найближчому майбутньому електронні видання можуть відвоювати у друкованих книг почесний титул основного джерела доходів видавництв [1].

Значно зростає кількість користувачів-учасників всесвітньої мережі, що, в свою чергу, пропорціонально зростанню кількості інформації. За даними статистичного порталу Internet Live Status (ILS) в поточному році кількість інтернет-сайтів перевищило один мільярд [2].

Автори сайту ILS заявляють, що за такої кількості веб-ресурсів реально активними є лише близько 75 відсотків (рис. 1). Звіт Міжнародного союзу електрозв'язку (МСЕ) [3], згідно якого на планеті інтернетом користуються більше 3 мільярдів людей, підтверджує прогнози стрімкого розвитку всесвітньої інформаційної мережі.

Зважаючи на вище наведену статистику щодо кількості веб-ресурсів, задача пошуку інформації постає на перше місце. За даними «ФОМнибус» [4] пошук інформації в інтернеті – основна потреба користувачів. Автором інформації може виступати будь-хто, а зважаючи

на велику та постійно зростаючу кількість веб-ресурсів в мережі інтернет – проблема пошуку потрібної інформації загострюється. Вище зазначене висуває високі вимоги до пошукових сайтів – надавати максимально відповідну користувачькому запиту інформацію.

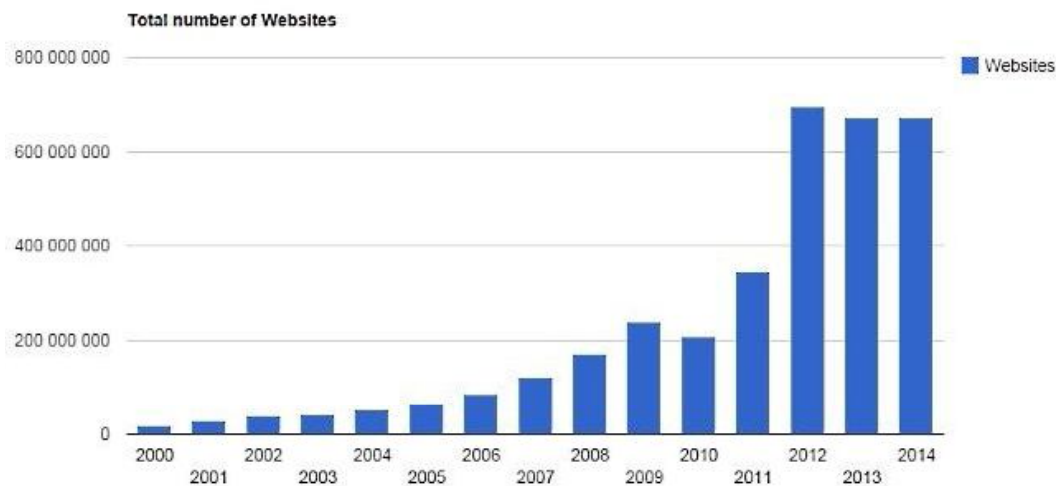


Рисунок 1 – Стовпчикова діаграма зростання кількості інтернет-сайтів за останні 14 років за даними статистичного порталу ILS

Пошукова система – певна база даних – онлайн-служба (програмно-апаратний комплекс з веб-інтерфейсом), що надає можливість пошуку інформації в Інтернеті [5]. У просторіччі під пошуковою системою розуміють веб-сайт, на якому розміщено інтерфейс (фронт-енд) системи. Програмною частиною пошукової системи є пошукова машина (пошуковий рушій) – комплекс програм, що забезпечує функціональність пошукової системи і, зазвичай, є комерційною таємницею компанії-розробника пошукової системи.

Існують 4 типи пошукових систем [6].

1. Системи, які використовують пошукових ботів. Такі системи побудовані з трьох частин: краулер (бот, робот або павук), індекс та програмне забезпечення пошукової системи. Краулер необхідний для обходу мережі та створення списку веб-сторінок. Індекс являється величезним набором копій веб-сторінок. Ціль програмного забезпечення – оцінювати результати пошуку. Завдяки тому що пошуковий робот в цьому механізмі постійно досліджує мережу, інформація завжди оновлюється. Більшість сучасних пошукових систем являються системами даного типу.

2. Системи, які керуються людиною (каталоги ресурсів). Ці пошукові системи отримують списки веб-сторінок. Каталог містить адресу, заголовок та короткий опис сайту. Каталог ресурсів шукає результати тільки відповідно до опису сторінок, представлених йому веб-майстрами. Перевагою каталогів є те, що всі ресурси перевіряються вручну, виходячи з цього якість інформації буде значно кращою, ніж результати, отримані за допомогою системи першого типу. Головний недолік системи полягає в оновленні даних каталогів. Відповідно, ранжування сторінок не може бути змінено миттєво. В якості прикладів таких систем можна навести dmoz [15] та Galaxy [16].

3. Гібридні системи. Такі пошукові системи, як Yahoo [17], Google [18], MSN [19] вміщують в собі функції систем, які використовують пошукових роботів, та систем, які керуються людиною, щоб забезпечити ефективні результати пошуку.

4. Мета-системи. Такі системи були корисні коли кожна пошукова система мала унікальний індекс, та пошукові системи були менш «досвідчені». Оскільки зараз пошук

значно покращився, потреба в них зменшилась. Приклади: MetaCrawler (Zoo Search) [20] та пошук MSN.

Головною задачею пошукової системи – формування пошукової видачі (списку веб-ресурсів) максимально відповідної користувачькому запиту. В контексті пошукових інформаційних систем ступенем відповідності користувачьких запитів пошуковій видачі називають релевантністю.

Релевантність – визначена пошуковим алгоритмом міра відповідності результатів пошуку завданню, поставленому в пошуковому запиті [7].

Релевантність сторінок сайту є найважливішим компонентом пошукової оптимізації.

Алгоритм, відповідно до якого визначаються релевантні сторінки у кожній пошуковій системі різний, але загалом він має однаковий набір критеріїв. Розрізняють внутрішні та зовнішні критерії визначення релевантності веб-сторінки.

Внутрішні критерії:

- частота (так звана щільність) ключових слів у тексті;
- місцезнаходження ключових слів;
- ступінь ідентичності пошукового запиту фрагменту тексту на сторінці;
- наявність збігів в метатеггах (title, description та keywords);
- наявність словоформ і синонімів ключових слів;
- оновлюваність сторінок;
- кількість сторінок, загальний обсяг інформації;
- внутрішнє перелінкування.

Зовнішні критерії:

- показник цитування або посилальна популярність (ТИЦ у Яндекс та PageRank у Google) – основний критерій;
- кількість звернень до даної сторінки за запитом або запитами;
- трафік ресурсу;
- вік домену.

Рівень релевантності пошукової видачі зростає у всіх найбільших пошукових системах – це пов'язано з ускладненням алгоритму формування пошукової видачі. Проте, він як і раніше, далекий від ідеалу, бо навіть найпотужніші пошукові системи, що мають властивості штучного інтелекту для об'єктивно людського аналізу матеріалу сторінки та визначення її істинної релевантності, не спроможні адаптуватися до особливостей семантичного змісту користувачького запиту.

Практична проблема полягає в тому, що на заданий користувачем запит пошукова система не завжди відповідає релевантними сторінками в пошуковій видачі, або ж результат, який задовольнить користувача знаходиться далеко не на першому місці списку сторінок, що визначає пошуковий алгоритм. Звідси можна сказати, що релевантність сторінки – це досить індивідуальний показник для кожного користувача пошукової системи, що залежить, в першу чергу, від особливостей мовленнєвої групи, до якої він відноситься, здатності висловлювати свої думки та об'єму словникового запасу користувача.

Мета розробки – підвищення рівня відповідності пошукової видачі пошуковому запиту користувача за рахунок розробки веб-сервісу у вигляді віджета для браузера Google Chrome, який визначає релевантність документів пошукової видачі на основі поведінки користувача на цих сторінках.

Існуючі рішення пошукових систем значно спрощують пошук релевантних запитів шляхом підказок їх правильного формулювання. Сьогодні пошукова система Google вже не сприймає лише прямі входження пошукових фраз, існує величезна база даних, де зібрані всі семантичні зв'язки слів та варіанти їх вживання [8]. Також Google використовує і доповнює завдяки новому патенту, в своїй базі інформацію про синоніми та антоніми, слова, близькі за

значенням, гіпонімії, відношенні членства, метонімії, атрибутах часу, причинно-наслідкових зв'язках, контекстуальна подібність. Кожне значення представляє собою деяку логічну петлю, до якої, на основі вище наведених значень, пошукова система підбирає відповідності (зазвичай 1-2), найбільш часто використовуваних в тексті. Це значно підвищило якість пошуку в пошуковій системі Google.

**МАТЕРІАЛ І РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ.** За даними дослідницької компанії «NetMarkerShare» на травень 2014-го року Google щомісячно обробляє 68,69 % всіх пошукових запитів у світі (рис. 2) [9].

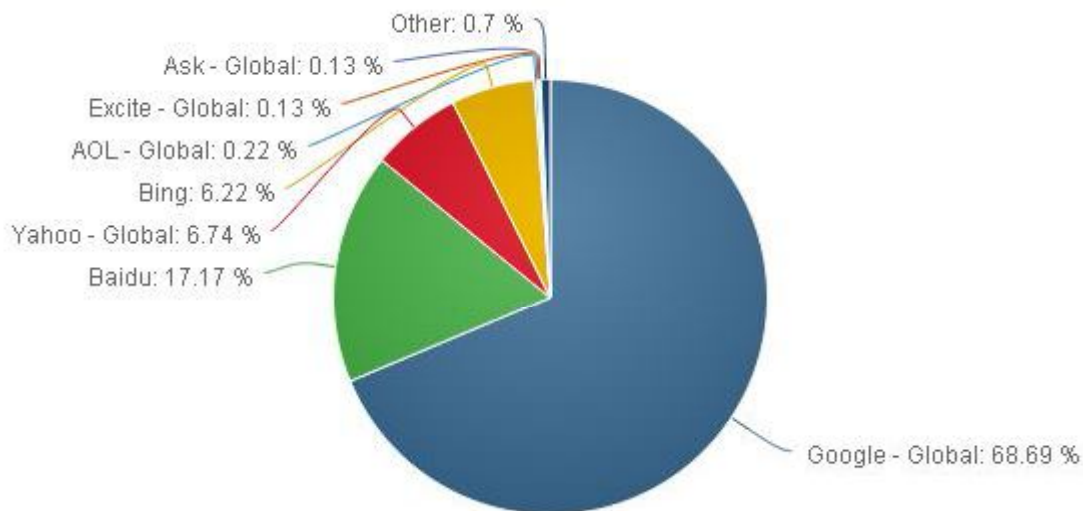


Рисунок 2 – Популярність пошукових систем за даними дослідницької компанії «NetMarkerShare»

Виходячи з вищенаведеної статистики, сервіс уточнення релевантності пошукової видачі розроблено для роботи лише з пошуковою системою Google.

Архітектура веб-сервісу включає в себе три головні частини – сервер бази даних, сервер роботи з базою та формування пошукової видачі, та тонкий веб-клієнт – віджет для браузера Google Chrome (рис. 3).



Рисунок 3 – Загальна архітектура веб-сервісу

В якості СКБД обрано NoSQL рішення MongoDB. MongoDB — документо-орієнтована система керування базами даних (СУБД) з відкритим вихідним кодом, яка не потребує опису схеми таблиць. MongoDB займає нішу між швидкими і масштабованими системами, що оперують даними у форматі ключ/значення, і реляційними СУБД, функціональними і зручними у формуванні запитів [10].

В контексті представленої розробки дана СУБД оптимально підходить під реалізацію веб-сервісу, це обумовлено наступними причинами: по-перше, зберіганням даних у форматі JSON [11], що збільшує зручність роботи з БД, зберігання даних в наглядно зручному вигляді; по-друге – значно більша швидкодія в порівнянні з SQL базами даних. Так, в порівнянні з PostgreSQL швидкодія вставки інформації відносно вибірки майже в два рази

вища [12], адже саме ця операція виконується сервером роботи з БД найчастіше – постійне додавання інформації про дії на певній веб-сторінці. Також MongoDB за результатами тестів значно швидша у пошуку інформації не по ключу, а по тексту (у реалізації клієнтської частини сервісу є підказки у полі вводу запиту в реальному часі).

В якості серверу сервісу обрано програмну платформу Node.js. Node або Node.js – програмна платформа, яка основана на двигуні Google V8 (транслюючим JavaScript в машинний код), перетворює JavaScript із вузькоспеціальної мови на мову загального призначення та загалом використовується для реалізації серверної частини веб-додатків.

Обираючи платформу реалізації серверної частини, висувалися вимоги до високої швидкодії формування відповідей на запити. За даними порівняльного аналізу Node.js із PHP [13], швидкодія відчутна при високому навантаженні великою кількістю запитів за короткий час. Також можна виділити наступні переваги Node.JS: асинхронність, неблокуючий ввід/вивід, система модулів CommonJS, двигун JavaScript Google V8.

Асинхронність надає можливість формувати відповідь на запит клієнтської частини незалежно від поточного стану виконання інших методів, а система модулів CommonJS надає можливість підключати додаткові модулі (наприклад, express.js), де абстрагування у використанні стандартних методів значно спрощує зовнішній вигляд коду, не впливаючи на швидкодію, та спрощує розробку.

Так як метод уточнення релевантності передбачає собою відстеження поведінки користувача на певних веб-ресурсах, тому вибір платформи для здійснення такої операції обумовлений популярністю програми перегляду веб-сторінок та наявністю якісного API – набором інструментів, класів, методів для роботи з функціоналом браузера (відстеження вкладок, дій користувача, скачування файлів, робота з закладками тощо).

За даними статистичного порталу «StatCounter Global Stats» [14] станом на 2014-ий поточний рік найпопулярнішим браузером у всьому світі є Google Chrome (рис. 4).

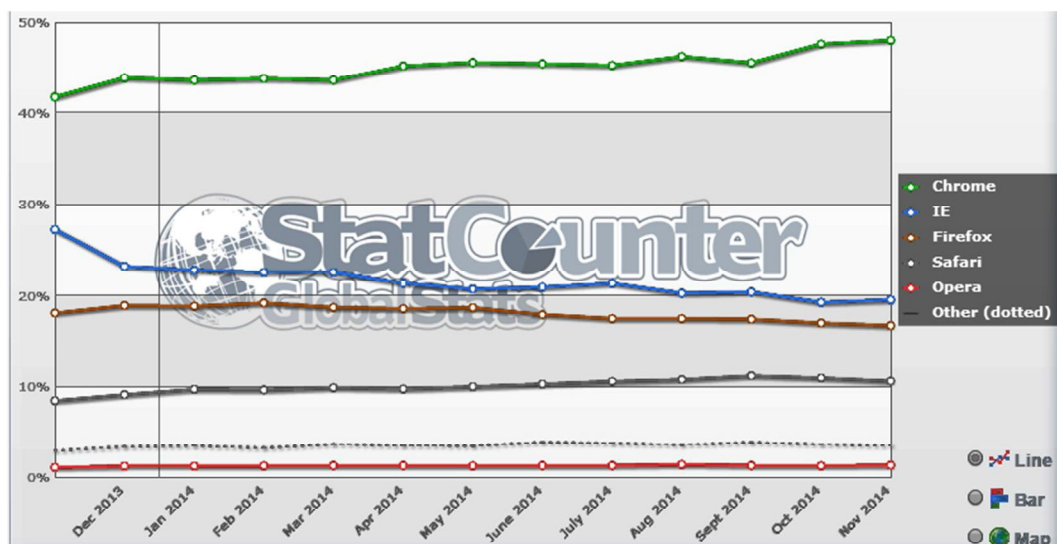


Рисунок 4 – Популярність сучасних веб-браузерів станом на 2014 рік за даними порталу «StatCounter Global Stats»

Тому, зважаючи на вище перераховані вимоги, браузер Google Chrome обрано для реалізації тонкого веб-клієнту сервісу. Клієнт-серверна архітектура надає перевагу в зберіганні даних незалежно від місця користування.

Для реалізації даного проекту використано тонкий клієнт клієнт-серверної архітектури за наступних його переваг:

- перенос обчислювальної потужності на сервер, що зменшує навантаження на клієнта;
- простота в налаштуванні;
- так як основна логіка роботи додатка переноситься на сервер, зміна чи поліпшення алгоритмів роботи не призводять до повного перевстановлення чи переналаштування клієнта.

Також можна виділити наступні недоліки тонкого клієнта:

- високі вимоги до якості зв'язку із сервером;
- неможливість роботи клієнта за відсутності з'єднання.

Логічну структуру запропонованого сервісу можна представити у вигляді наступної схеми (рис. 5).

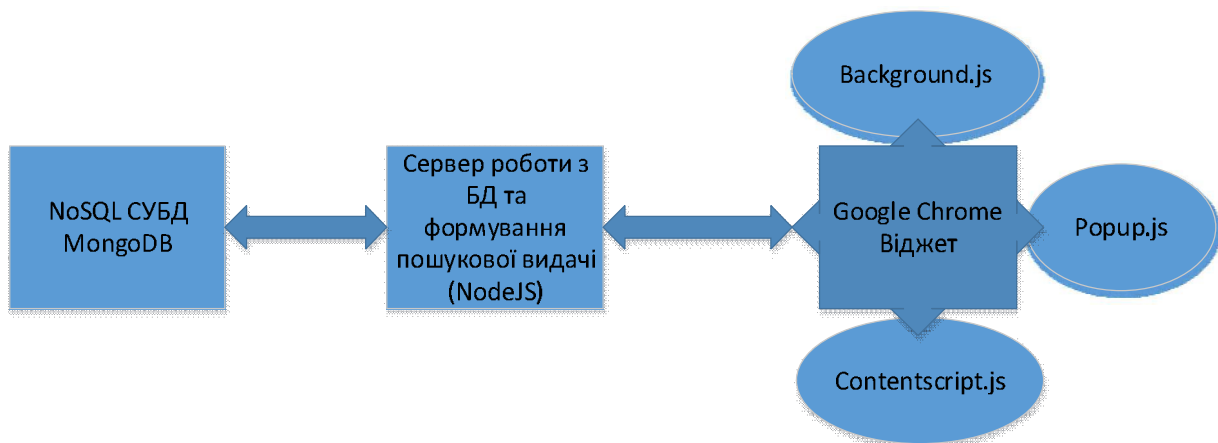


Рисунок 5 – Логічна структура сервісу уточнення релевантності пошукової видачі

Клієнт налічує в собі три основні логічні модулі роботи: background.js, contentscript.js, rorup.js.

Background.js – модуль, який працює у фоновому режимі браузера з моменту його завантаження. Його головна функція – відстежувати дії користувача, якщо ініційовано пошук в пошуковій системі Google. Нижче на рис. 6 наведено спрощену схему роботи алгоритму відстеження дій користувача у фоновому режимі роботи браузера.

Contentscript.js – модуль, що відповідає за відстеження часу, проведеного на конкретному ресурсі, та операцій з текстом (копіювання інформації) з подальшим відправленням даних до background-модуля.

Rorup.js – модуль, що виконує запит для пошуку у внутрішній накопиченій першими модулями бази запитів та відповідних їм ресурсів у вигляді форматowanego масиву пошукової видачі згідно з алгоритмом уточнення релевантності на основі поведінки користувача та зручне виведення на екран (рис. 7).

Сервер сервісу виконує форматування даних, що надходять від модуля background.js – інформацію про запити до Google з текстом запиту, про сторінки, що відкриті відносно конкретного запиту пошукової системи, та інформацію про дії користувача на відкритій з пошуку сторінці. На запит модуля rorup.js сервер форматує список сайтів згідно спеціального алгоритму для уточнення релевантності на основі отриманих даних про дії користувача на кожному сайті та відповідно сортує їх для подальшого відправлення клієнту даних у форматі JSON.



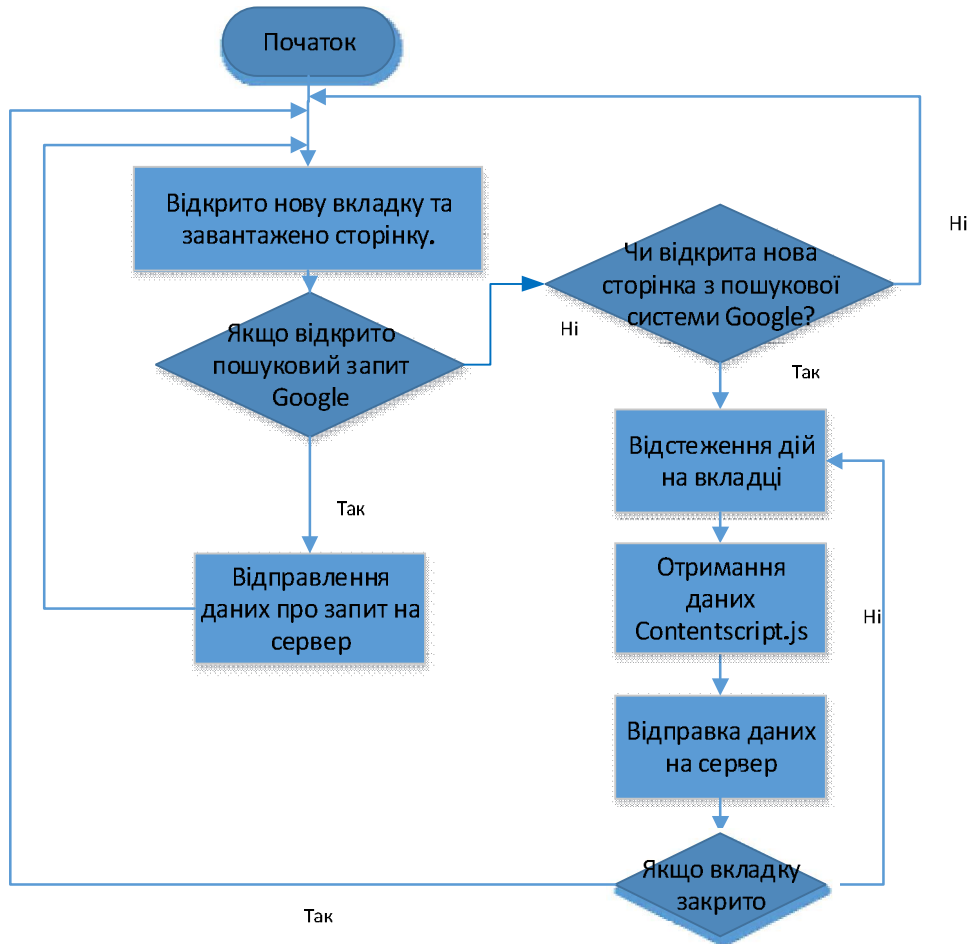


Рисунок – 6. Логічна структура роботи background-модуля клієнта

База даних MongoDB зберігає інформацію у форматі JSON про дії користувача на веб-ресурсах у наступному вигляді:

```

{
  "_id": "5480d5d0d8cd37454f37de77",
  "url":
  "https://www.google.com.ua/?gfe_rd=cr&ei=ydWAVLTmF4Ku8weJxYCYDg&gws_rd=ssl#q=%D1%81%D1%
  80%D0%B0%D0%B2%D0%BD%D0%B5%D0%BD%D0%B8%D0%B5+%D0%BF%D1%80%D0%BE%D0%
  B8%D0%B7%D0%B2%D0%BE%D0%B4%D0%B8%D1%82%D0%B5%D0%BB%D1%8C%D0%BD%D0%B
  E%D1%81%D1%82%D0%B8+mongodb",
  "s_tl": "сравнение производительности mongodb",
  "f_u": "ifree92",
  "date": 1417729488416,
  "tabs": [
    {
      "f_url": "http://habrahabr.ru/post/197590/",
      "url": "http://habrahabr.ru/post/197590/",
      "title": "Сравнение производительности MongoDB vs PostgreSQL. Часть I: No index / Хабрахабр",
      "load_t": "4186",
      "files": [
        {
          "url": "http://yoihj.ru/habr/mongo_vs_mysql.py",
          "filename": "mongo_vs_mysql.py",
          "mime": "application/octet-stream",
          "size": "2749",
          "time": "1543"
        }
      ]
    }
  ]
}
    
```

```

],
"act_t": 24158,
"bkmrk": "1417729534297"
"sel_text": [
  "MongoDB\n30% — 40% нагрузки на процесор\n704 — 706 МБ використовуваної оперативної
пам'яті\n\nВ то время как PostgreSQL\n5% — 25% нагрузки на процесор\n550 — 1000 МБ
используемой оперативной памяти"
]
}
]
}

```

де `_id` – унікальний ідентифікатор запису запита, `url` – посилання з пошукової системи Google, `s_tl` – текст запиту до пошукової системи, `f_u` – ім'я користувача, `date` – час у мілісекундах коли був виконаний запит, `tabs` – ресурси, що були відкриті з поточного запиту з подальшим відстеженням інформації про дії користувача на них: `url` – посилання відкритого ресурсу, `title` – заголовок сайту, `load_t` – час завантаження сайту в мілісекундах, `act_t` – час, проведений користувачем на веб-ресурсі в мілісекундах, `bkmrk` – час занесення до закладок, `files` – масив об'єктів з інформацією про успішно скачаний файл (посилання, ім'я файлу, тип, розмір та час скачки), `sel_text` – масив зі скопійованим текстом на веб-сторінці, `f_url` – посилання, всередині якого виконується перехід за іншим посиланням.

Зовнішній вигляд користувацького інтерфейсу клієнтської частини сервісу представлено на рис. 7.

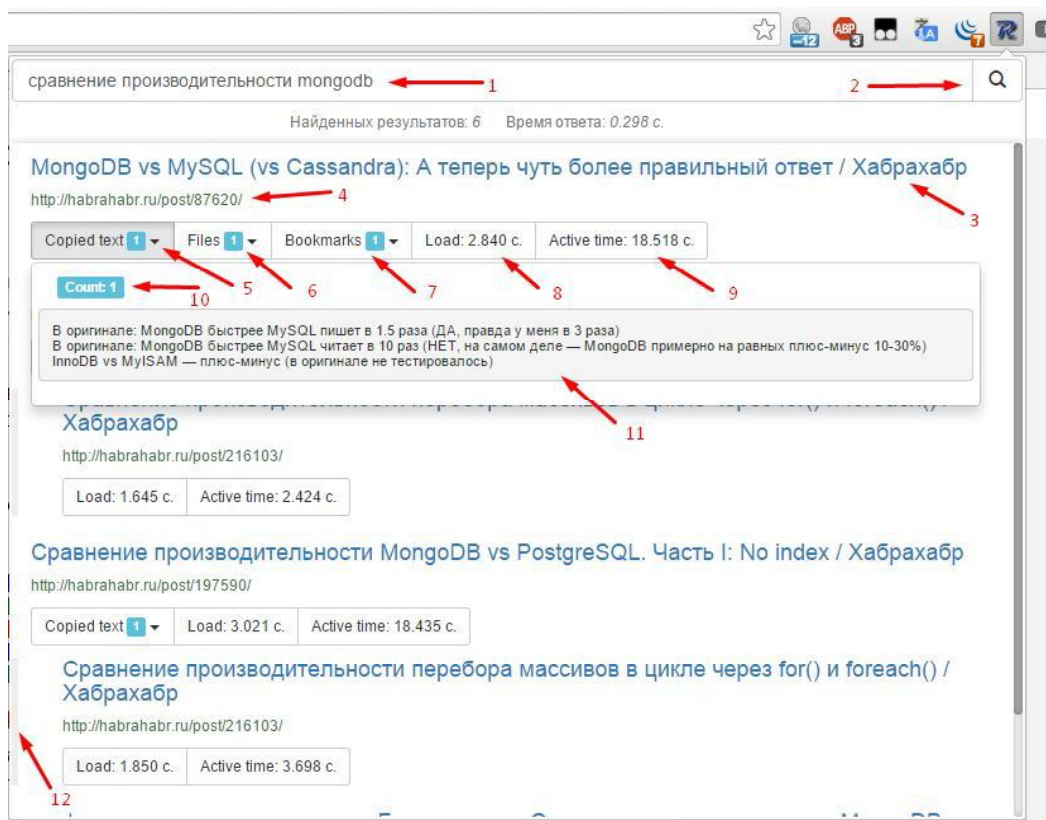


Рисунок 7 – Користувацький інтерфейс клієнта видачі пошукових запитів:

- 1 – пошуковий запит, 2 – кнопка запуску запиту до сервера для отримання уточненої інформації на основі поведінки користувача, 3 – заголовок сайту, 4 – посилання на сайт,
- 5 – скопійований текст, 6 – скачані файли, 7 - занесення сайту до закладок,
- 8 – час завантаження веб-сторінки, 9 – час, проведений на сторінці,
- 10 – кількість скопійованого однакового тексту, 11 – скопійований текст,
- 12 – перехід всередині сторінки за посиланням



Порядковий номер сайту у запропонованому сервісі визначається розрахунком на сервері коефіцієнтом релевантності, який формується з зібраних даних модулем background.js у внутрішню NoSQL MongoDB базу даних. Це дає змогу користувачеві отримувати пошукову видачу веб-сайтів максимально відповідних його власним потребам, а не тільки на перші місця пошукової видачі отримувати сайти, які мають просто високий рейтинг та велику кількість переходів, але для користувача контент цих сайтів не є релевантним.

В рамках експерименту відібрано запит з бази даних сервісу з наступним текстом – «среднее арифметическое javascript» та виділено дії, які користувач здійснював на запропонованих пошуковою системою Google веб-ресурсах (табл. 1).

Таблиця 1 – Поведінка користувача на першій сторінці пошукової видачі Google.

Пошукова видача Google (порядок збережено)	Час завантаження (мс)	Час проведення (мс)	Копіювання тексту	Занесення до закладок	Скачування файлів	Переходи всередині
javascript.ru	2460	23169,5	0	0	0	0
yershov.com.ua	7578	7028	1	0	0	0
java.gym5cheb.ru	1187	2680	0	0	0	0
www.cyberforum.ru	4757.6	29207.3	0	1	0	1
hashcode.ru	9389	1993	0	0	0	0
webzsl.narod.ru	958	19638	0	0	0	0
forum.dklab.ru	0	0	0	0	0	0
dumpz.org	0	0	0	0	0	0
forumpchelp.ru	0	0	0	0	0	0
otvet.mail.ru	3785	3963	0	0	0	0

На основі вихідної інформації про поведінку користувача з табл. 1 розраховано коефіцієнт релевантності для кожного веб-ресурсу (табл. 2).

Таблиця 2 – Розраховані коефіцієнти релевантності на основі поведінки користувача.

Пошукова видача Google (порядок збережено)	Коефіцієнт релевантності
javascript.ru	0,8
yershov.com.ua	1,242
java.gym5cheb.ru	0,1
www.cyberforum.ru	3,304
hashcode.ru	0,068
webzsl.narod.ru	0,681
forum.dklab.ru	0
dumpz.org	0
forumpchelp.ru	0
otvet.mail.ru	0,141

На рис. 8 зображено стовпчикову діаграму релевантності запропонованих Google веб-ресурсів згідно розрахованих коефіцієнтів зі збереженням порядку, запропонованого пошуковою системою.



Рисунок 8 – Розраховані коефіцієнти релевантності

Виходячи з розрахованих коефіцієнтів релевантності, було сформовано в порядку зменшення від максимальної релевантності, новий порядок актуальних ресурсів для користувача (рис. 9) в порівнянні з тим, що пропонується пошуковою видачею Google (рис. 10).

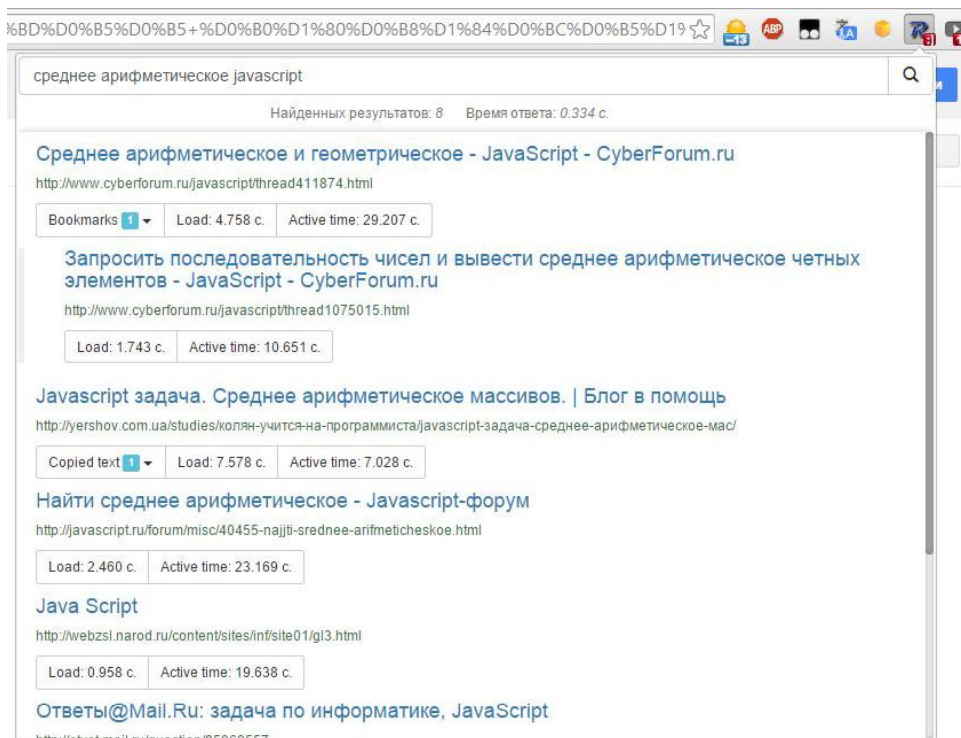


Рисунок 9 – Запропонований новий порядок видачі веб-ресурсів сервісом

Google

Поиск Видео Картинки Новости Ещё ▾ Инструменты поиска

Результатов: примерно 37 700 (0,17 сек.)

**Найти среднее арифметическое - Javascript-форум**  
javascript.ru > Форум > Javascript > Общие вопросы Javascript ▾  
6 авг. 2013 г. - Сообщений: 8  
Надо найти среднее арифметическое нескольких чисел. Вот текст программы: <b.  
Определить среднее арифметическое элементов ... Сообщений: 3 25 июн 2012  
Среднее арифметическое двумерного массива ... Сообщений: 10 24 июн 2012  
Решаем задачи в javascript - Javascript-форум ... Сообщений: 10 8 май 2012  
Вопрос по обучающим примерам ... Сообщений: 10 24 янв 2012  
Другие результаты с сайта javascript.ru

**Javascript задача. Среднее арифметическое массивов ...**  
yershov.com.ua > javascript ▾  
Javascript задача. Среднее арифметическое массивов. 01 Июл 2013. Перейти к  
комментариям Оставить отзыв ...

**Java Script 7 - Уроки java Script**  
java.gym5cheb.ru/java\_2/java\_05.htm ▾  
Сценарий в JavaScript может содержать как одну строку, так и большой листинг...  
Рассмотрим программу вычисления среднего арифметического:..

**Среднее арифметическое и геометрическое - JavaScript - ...**  
www.cyberforum.ru > Форум web-программистов > Форум JavaScript ▾  
20 дек. 2011 г. - Сообщений: 2 - 1 автор  
Написать сценарий на языке JavaScript, который по значениям, введенным в  
первые три поля, рассчитывает среднее арифметическое ...  
Среднее арифметическое значение массивов 1 сообщение 3 дек 2014  
Запросить последовательность чисел и вывести ... Сообщений: 4 18 янв 2014  
Среднее арифметическое 8-ми чисел и мин/макс Сообщений: 11 17 ноя 2012  
Вычислить среднее арифметическое чисел из ... 1 сообщение 16 дек 2011  
Другие результаты с сайта www.cyberforum.ru

**циклы / Найти среднее арифметическое - Javascript ...**  
hashcode.ru/questions/.../циклы-найти-среднее-арифметическое-javasc... ▾  
24 февр. 2014 г. - Добрый день помогите составить программу, которая  
определяла среднее арифметическое вводимых чисел, только количество ...

**Циклы - Java Script**  
webzsl.narod.ru/content/sites/inf/site01/gj3.html ▾  
{alert("Среднее арифметическое введенных вами чисел равно "+s/3);} </script> ...  
а после 170! JavaScript вообще перестанет считать и выдаст "Infinity".

**Задача. Нахождение среднего арифметического. :: Разно...**  
forum.dklab.ru/viewtopic.php?p=155997 ▾  
18 мая 2008 г. - Сообщений: 4 - Авторы: 4  
Сгенерировать случайным образом числа(оценки) 2,3,4,5. Вывести на экран около  
15 оценок и вывести их среднее арифметическое.

Рисунок 10 – Пошукова видача Google

**ВИСНОВКИ.** Запропоновано новий метод уточнення релевантності документів пошукової видачі що базується на поведінці користувача браузера Google Chrome. Обґрунтовано вибір технічних засобів реалізації даного методу, детально розглянуто архітектуру запропонованого віджета. В основу уточнення релевантності веб-ресурсів покладено характеристики поведінки користувача: часу, проведеного на певній сторінці; скачування файлів; занесення сторінки до закладок; копіювання тексту; перехід всередині

ресурсу за посиланнями; швидкість завантаження сторінки (найменш значущий параметр). Наведені в роботі експериментальні дослідження використання запропонованого підходу вказують на покращення рівня відповідності пошукової видачі запиту користувача та дозволяють більш ефективно використовувати пошукові засоби за рахунок індивідуальної адаптації до користувача веб-браузера.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Электронные книги могут стать популярнее печатных уже в 2018 году - 4PDA [Электронный ресурс]. – Режим доступа : <http://4pda.ru/2014/11/24/187258/>
2. Количество веб-сайтов в мире перевалило за миллиард | Техно [Электронный ресурс]. – Режим доступа : <http://podrobnosti.ua/tech/2014/09/23/994736.html>
3. Всего 3 миллиарда жителей Земли имеют доступ в Интернет - новости интернета, сетевых и высоких технологий на Hi-News.ru [Электронный ресурс]. – Режим доступа : <http://hi-news.ru/internet/vsego-3-milliarda-zhitelej-zemli-imeyut-dostup-v-internet.html>
4. Для чего люди используют интернет? / ФОМ [Электронный ресурс]. – Режим доступа : <http://fom.ru/smi-i-internet/11088>
5. Пошукова система – Вікіпедія [Електронний ресурс]. – Режим доступу : [http://uk.wikipedia.org/wiki/Пошукова\\_система](http://uk.wikipedia.org/wiki/Пошукова_система)
6. K. Tarakeswar & D. Kavitha, «Search Engines:A Study» // Journal of Computer Applications (JCA), ISSN: 0974-1925, Volume IV, Issue 1, 2011 [Электронный ресурс]. – Режим доступа : [http://www.jcaksrce.org/upload/39115\\_jcaksrcev4i1p7.pdf](http://www.jcaksrce.org/upload/39115_jcaksrcev4i1p7.pdf)
7. Релевантність – SEO Словник – iGroup Україна [Електронний ресурс]. – Режим доступу : <http://igroup.com.ua/seo-articles/relevantnist/>
8. Патент Google на уточнение результатов поиска – блог Indigo Marketing [Электронный ресурс]. – Режим доступа : <http://blog.indigo-marketing.com.ua/novye-standarty-optimizatsii-poiska-google>
9. Рейтинг лучших поисковых систем мира на 2014 год | BEST SEO BLOG [Электронный ресурс]. – Режим доступа : <http://www.bestseoblog.ru/rejting-poiskovyx-sistem/>
10. MongoDB – Вікіпедія [Електронний ресурс]. – Режим доступу : <https://uk.wikipedia.org/wiki/MongoDB>
11. JSON – Вікіпедія [Електронний ресурс]. – Режим доступу : <https://uk.wikipedia.org/wiki/JSON>
12. Сравнение производительности MongoDB vs PostgreSQL. Часть II: Index / Хабрахабр [Электронный ресурс]. – Режим доступа : <http://habrahabr.ru/post/197630/>
13. Nodejs vs PHP — скорость выполнения, сравнительный анализ. | Thoughts of the person [Электронный ресурс]. – Режим доступа : <http://b.brainscode.com/2011/04/nodejs-vs-php.html>
14. Top 5 Desktop, Tablet & Console Browsers from Nov 2013 to Nov 2014 | StatCounter Global Stats [Электронный ресурс]. – Режим доступа : <http://gs.statcounter.com/>
15. DMOZ – the Open Directory Project [Электронный ресурс]. – Режим доступа : <http://www.dmoz.org/>
16. Galaxy Directory [Электронный ресурс]. – Режим доступа : <http://www.galaxy.com/>
17. Yahoo [Электронный ресурс]. – Режим доступа : <https://www.yahoo.com/>
18. Google [Электронный ресурс]. – Режим доступа : <https://www.google.com>
19. msn [Электронный ресурс]. – Режим доступа : <http://www.msn.com/>
20. Zoo Search – a new, one-stop search engine [Электронный ресурс]. – Режим доступа : <http://www.zoo.com/Zoo-Site/>

**P. Kostenko, I. Levchenko**

Kremenchuk Mykhailo Ostrohradskyi National University, Kremenchuk

## WEB-SERVICE FOR CLARIFICATION RELEVANT WEB-DOCUMENTS OF SEARCH RESULTS OF GOOGLE BASED ON USER BEHAVIOR

Current status of World Wide Web popularity was analyzed. Basic problems of search systems meaning the overload information statistic are presented. Practical problem of relevance web-resources definition in a modern search engines is defined. Exiting solutions to increase relevance of search results are presented. A new approach for update requests relevance of web-resources is offered. Modern programming tools for development presented method reviewed and argued. A widget-interface and it functional parts with detail description are presented.

**Key words:** Internet, Google, search engine, relevance.

### REFERENCES

1. *Elektronnye knigi mogut stat' populyarnee pechatnykh uzhe v 2018 godu* [eBooks Could Finally Inch Past Print In 2018], available at: <http://4pda.ru/2014/11/24/187258/> (accessed December 1, 2014) [in Russian]
2. *Kolichestvo veb-saytov v mire perevalilo za milliard* [The number of websites in the world has exceeded one billion], available at: <http://podrobnosti.ua/tech/2014/09/23/994736.html> (accessed December 1, 2014) [in Russian]
3. *Vsego 3 miliarda zhiteley Zemli imeyut dostup v Internet – novosti interneta, setevykh i vysokikh tekhnologiy na Hi-News.ru* [Total 3 billion people on Earth have access to the Internet - Internet News, networking and high-tech on Hi-News.ru], available at: <http://hi-news.ru/internet/vsego-3-miliarda-zhitelej-zemli-imeyut-dostup-v-internet.html> (accessed December 1, 2014) [in Russian]
4. *Dlya chego lyudi ispol'zuyut internet?* [What do people use the Internet?], available at: <http://fom.ru/smi-i-internet/11088> (accessed December 1, 2014) [in Russian]
5. *Poshukova systema* [A search engine], available at: [http://uk.wikipedia.org/wiki/Пошукова\\_система](http://uk.wikipedia.org/wiki/Пошукова_система) (accessed December 1, 2014) [in Ukrainian]
6. Tarakeswar, K. and Kavitha, D. (2011) "Search Engines: A Study", Journal of Computer Applications (JCA), Vol. IV, Iss. 1, available at: [http://www.jcaksrce.org/upload/39115\\_jcaksrcev4i1p7.pdf](http://www.jcaksrce.org/upload/39115_jcaksrcev4i1p7.pdf) (accessed December 1, 2014)
7. *Relevantnist' – SEO Slovnyk – iGroup Ukrayina* [Relevance – SEO Glossary – iGroup Ukraine], available at: <http://igroup.com.ua/seo-articles/relevantnist/> (accessed December 1, 2014) [in Russian]
8. Google's patent on the refinement of search results – blog Indigo Marketing, available at: <http://blog.indigo-marketing.com.ua/novye-standarty-optimizatsii-poiska-google> (accessed December 1, 2014) [in Russian]
9. *Reyting luchshikh poiskovykh sistem mira na 2014 god* [Ranking of the best search engines of the world in 2014], available at: <http://www.bestseoblog.ru/rejting-poiskovykh-sistem/> (accessed December 1, 2014) [in Russian]
10. MongoDB – Wikipedia, the free encyclopedia, available at: <https://uk.wikipedia.org/wiki/MongoDB> (accessed December 1, 2014) [in Ukrainian]
11. JSON – Wikipedia, the free encyclopedia, available at: <https://uk.wikipedia.org/wiki/JSON> (accessed December 1, 2014) [in Ukrainian]
12. *Sravnienie proizvoditel'nosti MongoDB vs PostgreSQL. Chast' II: Index* [Performance Comparison MongoDB vs PostgreSQL. Part II: Index], available at: <http://habrahabr.ru/post/197630/> (accessed December 1, 2014) [in Russian]

13. *Nodejs vs PHP – skorost' vypolneniya, sravnitel'nyy analiz*. [Nodejs vs PHP – execute script time, comparative analysis], available at: <http://b.brainscode.com/2011/04/nodejs-vs-php.html> (accessed December 1, 2014) [in Russian]

14. Top 5 Desktop, Tablet & Console Browsers from Nov 2013 to Nov 2014, available at: <http://gs.statcounter.com/> (accessed December 1, 2014)

15. DMOZ – the Open Directory Project, available at: <http://www.dmoz.org/> (accessed December 1, 2014)

16. Galaxy Directory, available at: <http://www.galaxy.com/> (accessed December 1, 2014)

17. Yahoo, available at: <https://www.yahoo.com/> (accessed December 1, 2014)

18. Google, available at: <https://www.google.com> (accessed December 1, 2014)

19. msn, available at: <http://www.msn.com/> (accessed December 1, 2014)

20. Zoo Search – a new, one-stop search engine, available at: <http://www.zoo.com/Zoo-Site/> (accessed December 1, 2014)

**Костенко Павло Павлович,**

канд. техн. наук,  
старший викладач кафедри комп'ютерних та інформаційних систем,  
Кременчуцький національний університет імені Михайла Остроградського,  
вул. Першотравнева, 20, м. Кременчук  
Полтавської обл., Україна, 39600.  
Тел. +38(050) 958-15-39.  
E-mail: [ppkostenko@gmail.com](mailto:ppkostenko@gmail.com)



**Kostenko Pavlo Pavlovych,**

Cand.Sc. (Eng.),  
Senior Lecturer of Computer and Information  
Systems Department,  
Kremenchuk Mykhailo Ostrohradskyy  
National University,  
vul. Pershotravneva, 20, Kremenchuk,  
Poltava Region, Ukraine, 39600.  
Tel. +38(050) 958-15-39.  
E-mail: [ppkostenko@gmail.com](mailto:ppkostenko@gmail.com)

**Левченко Ігор Володимирович,**

магістрант кафедри комп'ютерних та інформаційних систем,  
Кременчуцький національний університет імені Михайла Остроградського,  
вул. Першотравнева, 20, м. Кременчук  
Полтавської обл., Україна, 39600.  
Тел. +38(096) 06-284-06.  
E-mail: [ifree92@gmail.com](mailto:ifree92@gmail.com)



**Levchenko Ihor Volodymyrovych,**

Student of Master Degree of Computer and  
Information Systems Department,  
Kremenchuk Mykhailo Ostrohradskyy  
National University,  
vul. Pershotravneva, 20, Kremenchuk,  
Poltava Region, Ukraine, 39600.  
Tel. +38(096) 06-284-06.  
E-mail: [ifree92@gmail.com](mailto:ifree92@gmail.com)

Стаття надійшла 05.12.2014