

ПРИНЦИПИ ПОБУДОВИ СИСТЕМИ АНАЛІЗУ ТА ПРОСУВАННЯ ІНТЕРНЕТ-РЕСУРСІВ

© Басюк Т., 2012

Проаналізовано основні принципи побудови системи аналізу та просування інтернет-ресурсів, наведено дерево цілей, контекстну діаграму та сформульовано висновки щодо можливості її проектування із використанням сучасних технологій програмування.

Ключові слова: структура, аналіз, база знань, веб-сайт.

The article analyzes the basic principles of system analysis and promotion of online resources are tree purposes konteksnu chart and formulate conclusions on the possibility of designing using modern programming technologies.

Key words: structure, analysis, knowledge base, web site.

Вступ

Останніми роками спостерігається невинне зростання кількості веб-ресурсів у межах глобальної мережі Інтернет, що, з одного боку, сприяє доступності інформації, а з іншого – посилює конкуренцію між сайтами. Результатом цієї ситуації є поступова втрата відвідуваності та занепад окремих інтернет-джерел. Описана ситуація виникає доволі часто і не лише через об'єктивні чинники: відсутність оновлень, застарілу інформацію чи повільний хостинг, а як наслідок втрати «цікавості» з боку пошукових систем, що сприяє розташуванню ресурсу на «віддалених» сторінках пошукового запиту [1, 2].

Позиція інтернет-ресурсу в результаті пошукового запиту є дуже важливим фактором, що напряму впливає на кількість відвідувачів. При цьому пошукова оптимізація сприяє його відображенню у верхніх позиціях рейтингу пошукової системи. Досягнення зазначеної цілі можливе лише у випадку досконалого знання алгоритмів ранжування та вмілого застосування технологій оптимізації пошукових механізмів, що є неможливим без детального огляду проблематики цієї галузі, аналізу основних методів роботи та основних алгоритмів оцінювання.

Зв'язок висвітленої проблеми із важливими науковими та практичними завданнями

Особливістю систем аналізу та просування інтернет-ресурсів є те, що вони дають змогу проаналізувати основні характеристики веб-сайтів та звернути увагу розробників й власників на можливі проблеми оптимізації із метою підвищення позицій у пошукових запитах. Зазначена особливість передбачає використання методів та засобів побудови відповідних систем та формування на їх основі «конгломерату», що дасть змогу практично реалізовувати задачі науки й практики у галузі підтримки прийняття рішень у процесі аналізу та просування інтернет-ресурсів [3, 4].

Аналіз останніх досліджень та публікацій

Внаслідок проведеного дослідження та аналізу публікацій [1–4] виявилось, що особливими рисами сучасних пошукових систем є наявність певних складових частин та етапів роботи, які необхідно враховувати під час проектування систем аналізу та просування інтернет-ресурсів, а саме:

– збирання та опрацювання даних – містить засоби аналізу ресурсу за допомогою роботів. А саме, як тільки розробник розташував веб-ресурс у глобальній мережі інтернет (додав зовнішні посилання, зареєстрував у каталогах тощо), відповідний робот-павук пошукової системи, згідно з пріоритетами, переходить на створений ресурс й здійснює збирання даних. При цьому для веб-павуків існує обмеження на глибину проникнення (рівень вкладеності) і максимальний розмір сканованого тексту, що зазвичай дорівнює 256 Кб. Керування доступом для пошукового робота здійснюється за допомогою файла інструкцій robots.txt;

– індексація отриманих результатів – складання для кожної сторінки ресурсу зворотного (інвертованого) файлу індексу. Індекс слугує для здійснення процесу пошуку інформації і зазвичай складається зі списку слів та інформації про них (позиції в тексті, розміру тощо). Процес індексації триває короткий проміжок часу, а його результатом є можливість здійснення пошуку за ключовими словами;

– пошук інформації – аналіз запиту, введеного користувачем. У результаті цієї процедури обчислюється вага кожного слова й здійснюється пошук за інвертованими індексами, що найбільш підходять для цього запиту (1):

$$\text{similarity}(Q, D) = \text{SUM}(wqk * wdk), \quad (1)$$

де $\text{similarity}(Q, D)$ – схожість запиту Q відносно документа D ; wqk – вага k -го слова у запиті; wdk – вага k -го слова в документі.

– ранжування ресурсів – полягає у розташуванні «найкорисніших» інтернет-ресурсів у верхніх позиціях пошукової видачі. Цю процедуру здійснюють відповідно до формули ранжування та містить такі чинники: вагу сторінки (тПЦ, PageRank); авторитетність домену; релевантність тексту запиту; релевантність текстів зовнішніх посилань.

З огляду на зазначені чинники, важливими задачами дослідження є визначення принципів побудови системи аналізу та просування інтернет-ресурсів й формування рекомендацій щодо її практичної реалізації за допомогою сучасних технологій програмування.

Основні завдання дослідження та їх значення

Відповідно до описаних задач дослідження для їх вирішення необхідно: проаналізувати технології побудови системи аналізу та просування інтернет-ресурсів; адаптувати вибраний підхід до поставленої задачі та сформулювати рекомендації й методологічний базис, необхідний для побудови прикладної системи аналізу та просування інтернет-ресурсів. Вирішення зазначених завдань дасть змогу спроектувати прикладну систему, яка надасть засоби:

- накопичення даних про інтернет-ресурси та передбачить можливість їх систематизації;
- аналізу та впровадження послуг пошукових систем;
- виведення додаткових знань із вже наявних у базі даних;
- пошуку додаткової інформації про параметри, що аналізуються;
- виведення рекомендацій щодо просування ресурсу.

Оптимізувавши веб-сайт, його власник може розраховувати на покращення позицій у пошукових системах, що, своєю чергою, призведе до збільшення кількості відвідувачів – це є вирішальним в інтернет-просторі.

Основні результати досліджень

Для проектування системи аналізу та просування інтернет-ресурсів важливою вимогою є надання рекомендацій, які ґрунтуються на принципах побудови систем підтримки прийняття рішень та методології системного аналізу. Серед них можна виділити [5, 6]:

– *принцип остаточної мети* – визначає глобальну мету системи та має абсолютний пріоритет, тобто визначається основна ціль системи – аналіз та просування інтернет-ресурсу, тобто надання рекомендацій користувачам щодо підвищення релевантності веб-сайту;

– *принцип єдності* – передбачає розгляд системи як одного цілого, тобто всі елементи системи є взаємопов'язані. У нашому випадку це є: підсистема, що забезпечує внесення нової інформації у програму (нових фактів); підсистема надання рекомендацій на основі збережених знань (фактів), машина логічного виведення; підсистема, яка дає змогу вивести рекомендації у відповідній формі (система виведення);

– *принцип зв'язності* – містить сформовані рекомендації, що ґрунтуються на внесених та збережених знаннях (фактах) у базі даних;

– *принцип модульності* – передбачає конструювання системи із сукупності модулів. Проектовану систему можна поділити на такі модулі: модуль внесення даних користувача системи, модуль внесення нових правил, головний модуль логічного виведення, модуль виведення сформованої системи рекомендацій;

– *принцип функціональності* – забезпечує загальну функціональність системи, а саме, функціональність внесення даних користувачів, внесення нових знань й формування рекомендацій, збереження звітів користувачів;

– *принцип розвитку* – полягає у внесенні нових знань протягом функціонування системи.

Наведені принципи дають змогу відобразити особливості побудови системи та врахувати важливі взаємні зв'язки між її елементами. Опис сутності системи, яка досліджується, розглянемо відповідно до мети [7, 8], яка визначає призначення системи та конкретизується за допомогою дерева цілей (рис. 1).

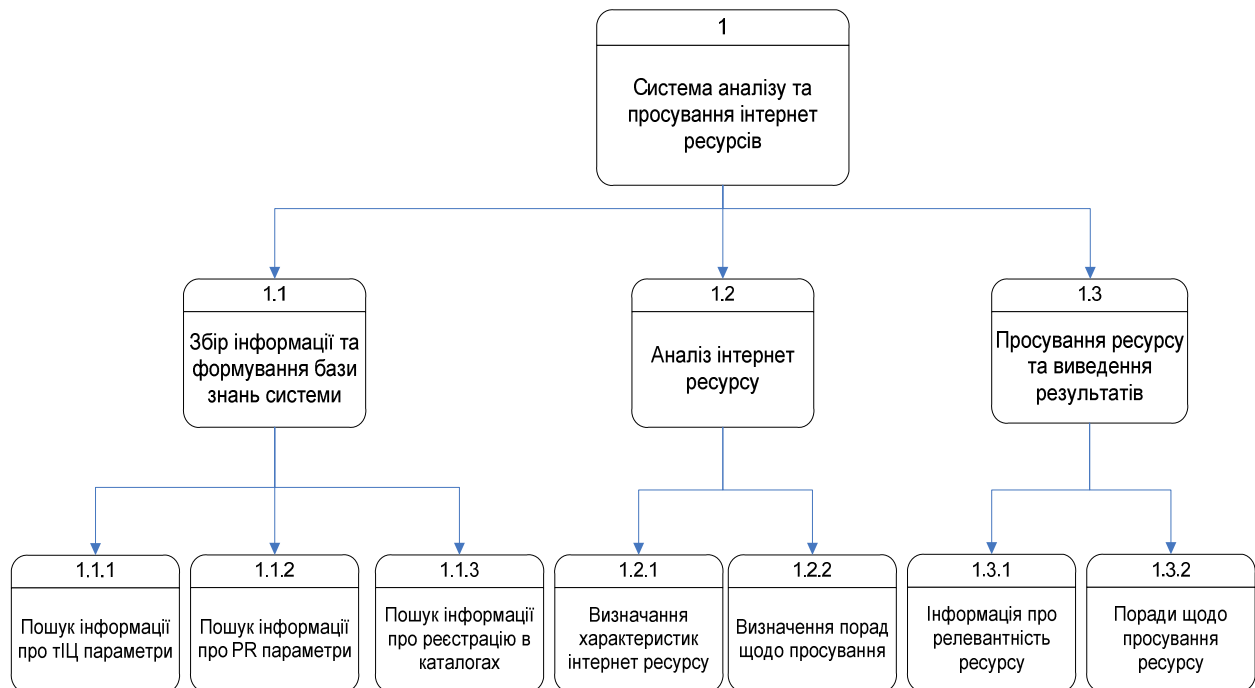


Рис. 1. Дерево цілей системи

Згідно з методологією системного аналізу головною ціллю є створення системи аналізу та просування інтернет-ресурсів. Це завдання можливо виконати тільки після виконання всіх підцілей. Від вершини дерева цілей відходять три гілки, які поділяють головну ціль на три підцілі.

Першою підціллю є “Збирання інформації та формування бази знань”. На цьому етапі необхідно виділити необхідні характеристики та визначити структуру знань, які необхідні для роботи системи. Зазначена ціль поділяється на три підцілі, метою яких є створення логічної моделі бази знань, вибір методів подання знань, проектування засобів автоматизованого опитування експертів, що дасть змогу наповнювати та розширювати базу знань автоматизовано.

Другою підціллю є “Аналіз інтернет-ресурсу”. Метою цієї гілки є збирання та відображення інформації, що стосується вибраного інтернет-ресурсу, а також методів та засобів аналізу цих даних. Ця підціль поділяється на дві підцілі: визначення характеристик інтернет-ресурсу (здійснює визначення поточної релевантності ресурсу) та визначення порад щодо просування (відобразатиме за допомогою продукційної моделі можливі варіанти підвищення позиції в пошукових видачах).

Третя підціль є “Просування ресурсу та виведення результатів”. Під час виконання цієї підцілі будуть реалізовані спеціалізовані засоби відображення інформації щодо релевантності сайту та поради (рекомендації) щодо його просування.

Під час проектування прикладної програми важливими завданнями, які необхідно вирішити, є визначення всіх функцій розроблюваного продукту, інтерфейсу зв'язку між складовими модулями та набір проектних рішень для забезпечення безвідмовної роботи програми. Архітектура програми забезпечує поєднання розроблених компонентів, які утворюють модулі та задає контекст, у межах якого приймаються наступні проектні рішення. Розроблення архітектури вимагає прийняття таких проектних рішень: створення програми з модулів та керування базами даних.

Створення програми з модулів визначає виділення набору окремих компонентів, що взаємозалежні і мають добре визначені інтерфейсні зв'язки з іншими компонентами програми. Вони визначають форму всіх взаємодій між ними, проте не містять з'єднань на їх внутрішню структуру. З огляду на те, проєктована програма повинна складатись зі своєрідних блоків, які зображуються за допомогою контекстної діаграми.

Керування базами даних здійснюватиметься у двох аспектах: керування внутрішніми та зовнішніми базами. Під внутрішніми базами даних у розроблюваній системі розумітимемо зберігання внутрішньої структури програми в основній пам'яті комп'ютера. При цьому використання файлів дасть змогу звертатись до них зі сторони багатьох користувачів та легко переносити їх на різні платформи. Своєю чергою, зовнішні бази даних (зберігаються в зовнішній пам'яті комп'ютера) використовуватимуться для збереження даних користувачів, оскільки зазначені типи даних погано піддаються структуризації та мають низьку інформаційну щільність.

Формулювання методологічного підґрунтя проєктування системи аналізу та просування інтернет-ресурсів неможливе без побудови контекстної діаграми, яка відображає взаємодію системи із зовнішнім середовищем. Представлена контекстна діаграма (рис. 2) описує вхідні та вихідні потоки системи та дає змогу систематизувати та автоматизувати етапи розроблення програмного забезпечення.



Рис. 2. Контекстна діаграма потоків даних

Застосування контекстної діаграми є обмеженим, оскільки вона відображає лише основні сутності розроблюваної системи, не роблячи наголосу на їх внутрішню побудову. З огляду на те відображається діаграма декомпозиції потоків даних першого рівня, яка показує основні процеси системи аналізу та просування інтернет-ресурсів (рис. 3). Основними процесами є:

- аналіз параметрів інтернет-ресурсу полягає у комплексному аналізі параметрів веб-сайту, його Google PR та Яндекс тІЦ показників, наявність в каталогах (DMOZ) та кількість проіндексованих сторінок;
- вибір рекомендацій щодо просування полягає у збиранні проаналізованої інформації та виборі рекомендацій щодо оптимізації веб-сайту;
- виведення результатів полягає у поєднанні отриманих результатів та виведенні інформації про сайт із рекомендаціями щодо просування.

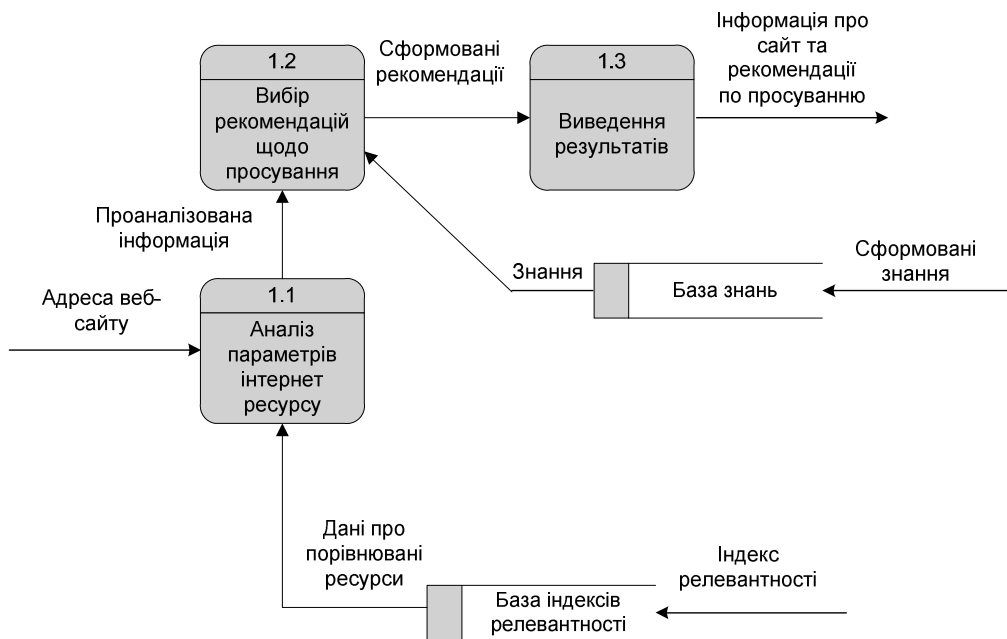


Рис. 3. Декомпозиція діаграми потоків даних

Деталізований процес аналізу параметрів інтернет-ресурсу відображено на рис.4.

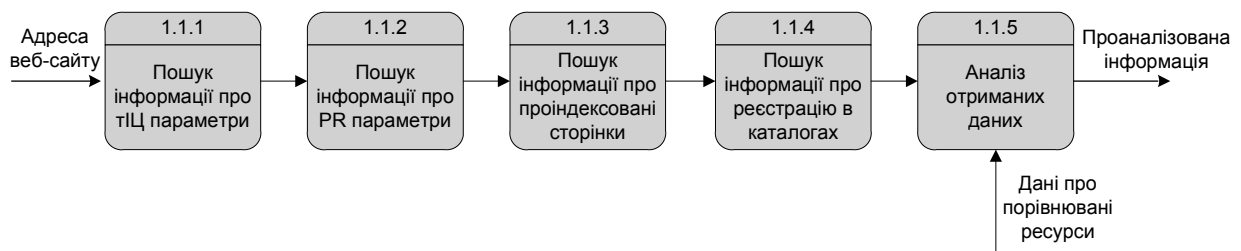


Рис. 4. Деталізація процесу «Аналіз параметрів ресурсу»

Деталізація процесу демонструє процеси пошуку інформації щодо параметрів веб-сайту та є складовою частиною системи аналізу та просування інтернет-ресурсу.

Висновки

Запропоновано принципи побудови системи аналізу та просування інтернет-ресурсів. Спроектвані контексні діаграми та приведена їх декомпозиція надає необхідний апарат із побудови зазначеної системи, що дасть змогу здійснювати підтримку прийняття рішень у галузі оптимізації інтернет-ресурсів. Зазначений підхід дасть змогу: застосовувати нові адаптивні алгоритми визначення релевантності ресурсу; використовувати в створеній системі інтерактивні елементи; спростити адміністрування інтернет-ресурсів; підвищити інформативність та конкурентоздатність ресурсу. Подальші дослідження будуть скеровані на вивчення методів організації та проектування кожної з компонент системи аналізу та просування інтернет-ресурсів.

1. Ашманов І.С. Просування сайту в пошукових системах / І.С. Ашманов, А.А. Іванов. – К.: Собор, 2010. – 340 с.
2. Тєроу Ш. Видимість в Інтернетє. Поисковая оптимизация сайтов. Издание / Ш. Тєроу – М.: Символ, 2009. – 288с.
3. Зуєв М. Просування сайтів у пошукових системах / Б. М. Б. Зуєв, П. А. Маурус – К.: Біном. Лабораторія знань, 2007. – 304 с.
4. Райцин М. Ефективне просування сайтів в Яндексї / М. Райцин , О. Кураков – К.:Діалектика, 2009. – 456 с.
5. Катренко А.В. Системний аналіз об'єктів та процесів комп'ютеризації / А.В. Катренко. – Львів: Новий світ – 2000, 2005. – 424 с.
6. Налімов В.В. Вероятностная модель представления знаний / В.В. Налімов. – М.: Наука, 2003. – 403 с.
7. Энж Э. Искусство раскрутки сайтов / Э. Энж. – СПб.: ВHV-СПб, 2011. – 592 с.
8. Кристиан Д. Поисковая оптимизация на ASP.NET / Д. Кристиан. – М.: Вильямс, 2008. – 400 с.