

В. О. Рахуба, к. т. н.

ЗАДАЧІ АВТОМАТИЗОВАНОГО УПРАВЛІННЯ ПОШУКОВОЮ ОПТИМІЗАЦІЄЮ ІНТЕРНЕТ-РЕСУРСІВ

Статтю присвячено дослідженню пошукової оптимізації інтернет-ресурсів як об'єкта управління. Формалізовано задачу оптимізації та виділено її складники. Наведено результати експерименту щодо впливу основних чинників на позиції сайтів у пошуковій системі Яндекс.

Ключові слова: пошукова оптимізація web-сайтів (SEO), автоматизоване управління, математична модель.

Постановка проблеми. У сучасному середовищі ефективність роботи будь-якого підприємства значною мірою залежить від спроможності знаходити нові ринки збуту своєї продукції. В умовах високої конкуренції за клієнта, яка характерна й для економічних відносин у нашій країні та в багатьох країнах світу, одним із найважливіших чинників успішної роботи підприємства є оптимізація його інтернет-ресурса за позицією у видачі пошукових систем. Серед спеціалістів, що працюють над цією проблемою, ця задача отримала сталу назву SEO (search engine optimization).

Аналіз досягнень. Аналіз робіт, присвячених цій проблемі [1, 2, 3], яких сьогодні є чимало, дозволяє зробити висновок, що, незважаючи на спроби автоматизації SEO, переважна більшість спеціалістів з оптимізації сайтів виконує свою роботу «в ручному режимі», спираючись на емпіричні результати власних експериментів. Однак слід одразу зазначити, що процес SEO неодмінно має творчий складник, який практично не може бути автоматизований; отже, актуальним є завдання розробки алгоритмів автоматизованого управління тими етапами SEO, що можуть бути формалізовані, та створення на основі цих алгоритмів програмного забезпечення для управління процесом пошукової оптимізації сайтів.

Постановка завдання. Метою цього дослідження є: 1) визначення складників SEO, що можуть бути автоматизовані; 2) постановка задач автоматизованого управління; 3) проведення експерименту з дослідження впливу основних чинників оптимізації на позиції сайтів у пошуковій системі Яндекс.

Основний матеріал дослідження. Процес пошукової оптимізації web-ресурсу як об'єкт автоматизації можна представити схемою (див. рис. 1).

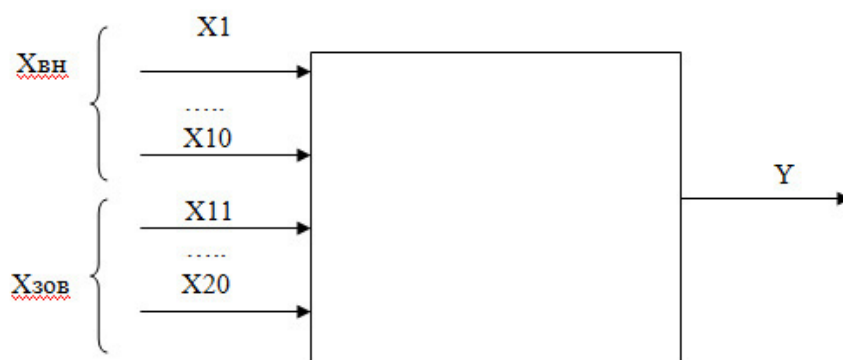


Рис. 1. SEO як об'єкт управління

Відгуком Y у цій моделі є позиція, яку займає сайт у видачі певної пошукової системи, щодо якої здійснюється оптимізація.

Усі чинники можна поділити на дві групи: внутрішні $X_{вн}$, які характеризують внутрішні

властивості web-сторінки, що має бути просунена в пошуковій видачі, та зовнішні Хзов, які характеризують конкурентність пошукового запиту, за яким просують сторінку.

До Хвн належать [1 – 5]:

X1 – вік домену; X2 – ТІЦ (або PR) сайту; X3 – кількість і якість зовнішніх посилань на сторінку; X4 – кількість посилань, що виходять зі сторінки (зокрема, так звана «внутрішня перелінковка»); X5 – точне входження пошукового запиту в теги «Title» та «Description»; X6 – наявність сайту в каталозі цієї пошукової системи; X7 – наявність файлу robots.txt; X8 – наявність файлу sitemap.xml; X9 – кількість точних входжень пошукового запиту в текст сторінки; X10 – якість представлення інформації на сторінці (змістовність, наявність нумерованих списків, підзаголовки, зображення та інше).

До Хзов належать [1 – 5]:

X1 – загальна кількість відповідей за даним запитом; X2 – кількість головних сторінок у ТОПі пошукової видачі; X3 – середня кількість і якість посилань на сайти у видачі; X4 – середній вік сайтів у видачі; X5 – кількість точних входжень пошукового запиту в теги «Title» та «Description» сайтів у видачі; X6 – наявність блоків контекстної реклами у видачі пошукової системи; X7 – наявність сторінок Вікіпедії та Ютуба за певним запитом; X8 – наявність географічної карти в результатах видачі; X9 – кількість слів у запиті; X10 – інші внутрішні характеристики сайтів у пошуковій видачі.

Отже, задачу оптимізації пошукового просунення web-ресурсу доцільно представити у вигляді

$$Y = f(X1, \dots, X22)$$

$$Y \rightarrow \min$$

при чому обмеження на процес визначають, виходячи з бюджету, виділеного на цей проект.

Загалом у процесі пошукової оптимізації можна виділити такі складники:

1. Визначення обмежень для задачі пошукової оптимізації певного інформаційного ресурсу:

а) аналіз предметної галузі, у якій просуватимуть сайт із метою дослідження частотності та конкурентності пошукових запитів;

б) розробка варіантів просунення сайту та розрахунок варіантів бюджету для їхньої реалізації;

с) вибір варіанту пошукового просунення сайту.

2. Синтез семантичного ядра інформаційного ресурсу:

а) формування бази потенційних пошукових запитів, за якими може просуватися цей ресурс;

б) відбір оптимальних запитів за критерієм управління;

с) побудова дво- або тривірневої структури семантичного ядра за частотністю запитів.

3. Власне оптимізація сторінок сайту за вибраними запитами шляхом управління внутрішніми чинниками в наведеній вище моделі:

а) створення раціонального дизайну сайту;

б) написання статей з якісним представленням інформації та оптимальною кількістю входжень запитів;

с) раціональне заповнення тегів сторінок та головної сторінки сайту;

д) створення файлів robots.txt та sitemap.xml, реєстрація в каталогах;

е) створення раціональної внутрішньої перелінковки;

ф) закупівля рекламних посилань з інших сайтів відповідно до вибраного варіанту бюджету.

Для задач автоматизованого управління найбільший інтерес представляють пункти 1б, 2б

і 2с, адже вони потребують розробки моделей просунення web-ресурсів у пошукових системах, що враховують вплив наведених вище численних чинників на позицію у видачі, а також вимагають синтезу критеріїв відбору запитів, що забезпечать розв'язок поставленої задачі оптимізації. Також частково можна автоматизувати пункти 1а, 2а, 3е і 3f, тоді як більшість складників пункту 3 потребують творчої праці людей, що не може бути формалізована. На цьому етапі роботою системи управління оптимізацією web-ресурсів має бути надання персоналу масиву пошукових запитів, сформованого у вигляді ієрархічної структури, необхідної кількості точних входжень відповідних запитів на відповідних сторінках та схеми перелінковки сторінок із зазначенням потрібних анкорів. Пункт 1с передбачає узгодження із замовником проекту. Також є актуальною розробка підсистеми прогнозування трафіку на сайт залежно від обраного варіанту оптимізації його пошукового просунення.

Для встановлення залежностей, що мають бути покладені в основу алгоритмів управління, необхідно провести дослідження видачі пошукових систем. Очевидно, що для попереднього дослідження впливу наведених вище чинників на відгук можливо обрати тільки метод пасивного експерименту. При чому доцільно провести декілька різних експериментів, обираючи для кожного певний набір внутрішніх або зовнішніх чинників. Для того щоб результати обробки пасивного експерименту мали сенс, доцільно обирати не більш, ніж чотири чинники для однієї серії досліджень [6].

Для такого експерименту обрано освітню тематику, зокрема предметну область – профорієнтацію. Результати такого дослідження можуть бути цікаві ВНЗ та іншим організаціям, що надають освітні послуги.

Проведено дві групи по три серії досліджень ТОП видачі пошукової системи Яндекс. Перша група дозволила врахувати вплив внутрішніх чинників у кожній серії, експерименти проводили за запитамі Z1 – «*Как выбрать профессию*», Z2 – «*Какую профессию выбрать*», Z3 – «*Тест какую профессию выбрать*», що мають частотність відповідно 1542, 637 та 213 показів на місяць. У якості чинників обрано X5, X9, X3/10, X2/10. Дослідження другої групи враховували вплив зовнішніх чинників, експерименти проводили для сайтів *moeobrazovanie.ru*, *kartaznaniy.ru* та *moya-pamyat.ru*; чинниками виступали X1/млн, X9/10, X3/10, X5.

Для аналізу зібраних даних було використано методику обробки результатів пасивних експериментів [6]. За результатами експериментів, отримані такі регресійні залежності:

для першої групи

$$Y = 5,5 - 2,69x_1 + 2,038x_2 + 0,003x_3 - 0,081x_4$$

$$Y = 5,5 - 4,519x_1 + 0,94x_2 + 0,009x_3 - 0,004x_4$$

$$Y = 5,5 - 3,544x_1 - 0,226x_2 + 0,075x_3 - 0,072x_4,$$

для другої групи

$$Y = 2,5 - 0,23x_1 - 0,0007x_2 - 0,0004x_3 - 0,548x_4$$

$$Y = 3,78 + 2,47x_1 + 0,087x_2 - 0,119x_3 - 1,781x_4$$

$$Y = 11,1 - 1,2x_1 - 0,117x_2 + 0,077x_3 - 4,482x_4,$$

де x_1, x_2, x_3, x_4 – центровані значення обраних чинників відповідно.

Перевірити адекватність отриманих моделей класичним способом (за критерієм Фішера) неможливо через відсутність паралельних досліджень, тому для перевірки адекватності було використано метод, згідно з яким дисперсія відтворюваності замінюється дисперсією відгуку

[6]. У результаті застосування цього методу зроблено висновок, що дві моделі, із отриманих у першій серії досліджень, можна вважати адекватними. Моделі, що характеризують вплив зовнішніх чинників, неадекватні.

Аналіз результатів експерименту дозволяє зробити такі висновки:

1. Використаний у цій роботі метод отримання регресійних моделей впливу внутрішніх чинників оптимізації ресурсу можна застосовувати для попередніх розрахунків на етапах 1a, 1b.

2. Найбільшу вагу під час ранжування ресурсу у видачі має чинник точного входження пошукового запиту в теги «Title» і «Description». На рис. 2 графічно показано залежність позиції сайту у видачі від кожного з обраних чинників.

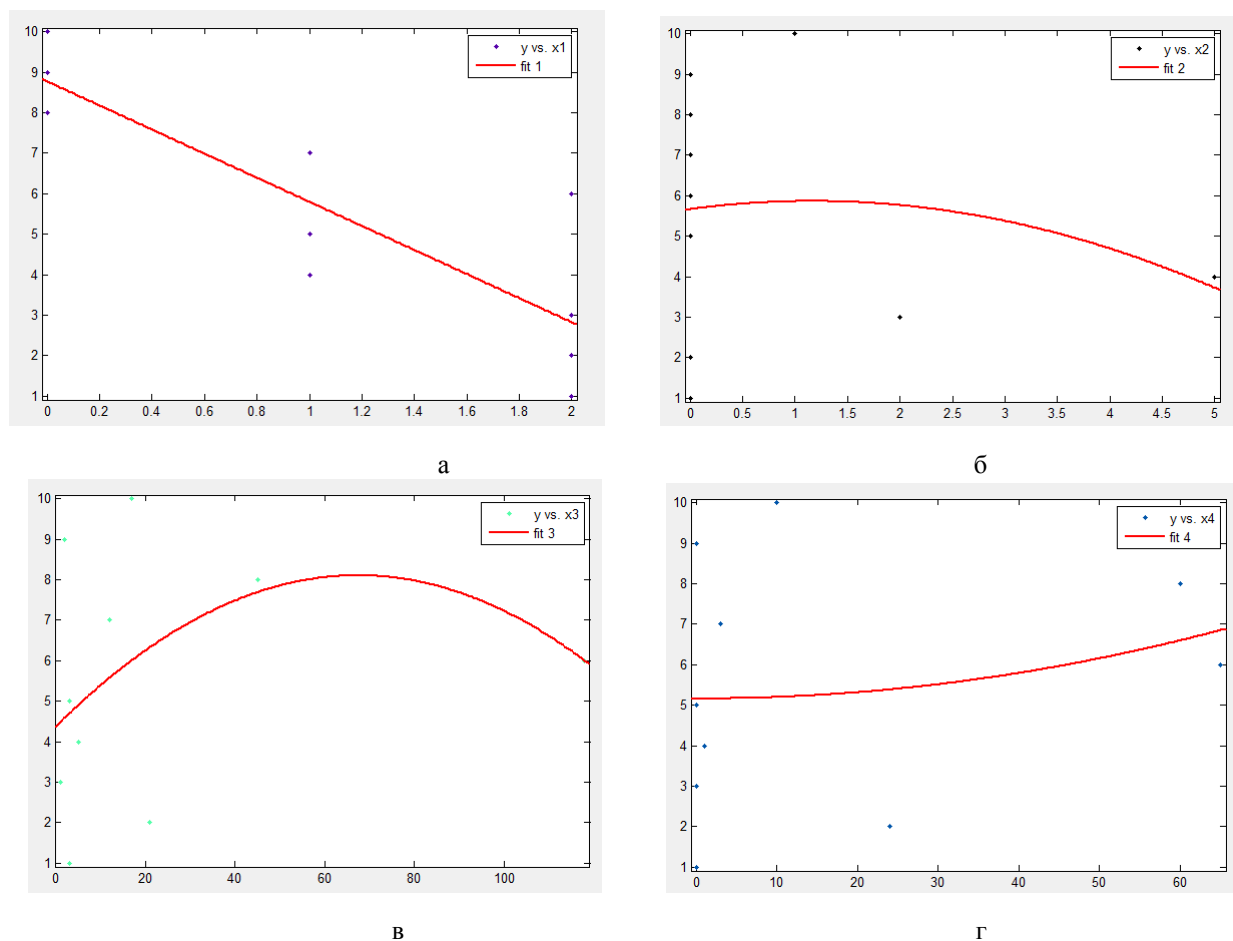


Рис. 2. Уплив внутрішніх чинників оптимізації X5, X9, X3 / 10, X2 / 10 (рис. а, б, в, г відповідно) на позицію сайту у видачі ПС Яндекс (точками показано експериментальні дані, а кривими – апроксимовані)

Неможливість отримати адекватні залежності для другої серії експериментів вказує на те, що у формуванні величини, названою конкурентністю запиту, грає роль більше число чинників і (або) вони взаємодіють у більш складному зв'язку.

Висновки. Проведені попередні дослідження підтверджують складність опису ранжування ресурсів у пошуковій видачі й дають обґрунтування розробки моделей за допомогою сучасних методів моделювання складних систем. Беручи до уваги такі характеристики алгоритмів ранжування, як висока складність, секретність, функціонування в галузі штучного інтелекту, мінливість і безперервна адаптація, найбільш доцільним вважаємо використання нейронних мереж [7] як методу моделювання процесів, на основі яких здійснюється оптимізація пошукового просування web-ресурсів.

Отже, представлені в цій роботі результати формалізації процесу пошукової оптимізації

web-ресурсів можуть бути використані в розробці систем управління їх просуванням. Актуальним залишається проблема розробки моделей ранжування сайтів у пошукових системах, для чого найбільш доцільним вважаємо використання нейронних мереж.

СПИСОК ЛІТЕРАТУРИ

1. Актуальные факторы ранжирования [Електронний ресурс] / Энциклопедия поисковых систем. – Режим доступа : http://www.searchengines.ru/articles/aktualnye_fakto.html/. – Назва з титул. екрану.
2. Три способа оценить конкурентность поискового запроса [Електронний ресурс] / Е. Камская. – Режим доступа : <http://optimizerha.ru/master-classes/konkurentnost-zaprosa>. – Назва з титул. екрану.
3. Оценка конкуренции по запросу в Яндексе. 1 часть [Електронний ресурс] / И. Токарев. – Режим доступа : <http://whiteseo.ru/content/view/521/>. – Назва з титул. екрану.
4. Как раскрутить сайт? [Електронний ресурс] / В. Беляев. – Режим доступа : <http://trynyty.ru/kak-raskrutit-sajt/>. – Назва з титул. екрану.
5. Алгоритмы поисковых систем [Електронний ресурс] / Вэбэффектор. – Режим доступа : http://www.webeffector.ru/wiki/%D0%90%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC%D1%8B_%D0%9F%D0%A1. – Назва з титул. екрану.
6. Костин В. Н. Статистические методы и модели : [учеб. пособие] / В. Н. Костин, Н. А. Тишина. – Оренбург : ГОУОГУ, 2004. – 138 с.
7. Методы и модели искусственного интеллекта [Електронний ресурс] / Jahrbuch fur EcoAnalytic und EcoPatologic. – Режим доступа до журн.: <http://www.ievbras.ru/ecostat/Kiril/Library/Book1/Content128/Content128.htm>. – Назва з титул. екрану.

Рахуба Вікторія Олегівна – к. т. н., доцент кафедри автоматизованого управління технологічними процесами.

Запорізька державна інженерна академія.