

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).

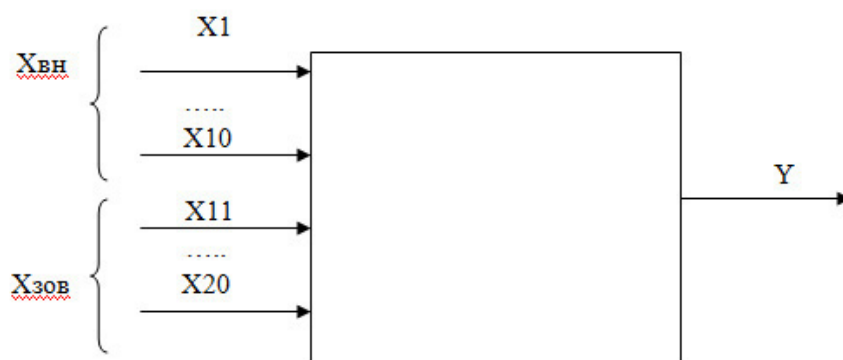


Рис. 1. SEO як об'єкт управління

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 1а, 1б.

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.

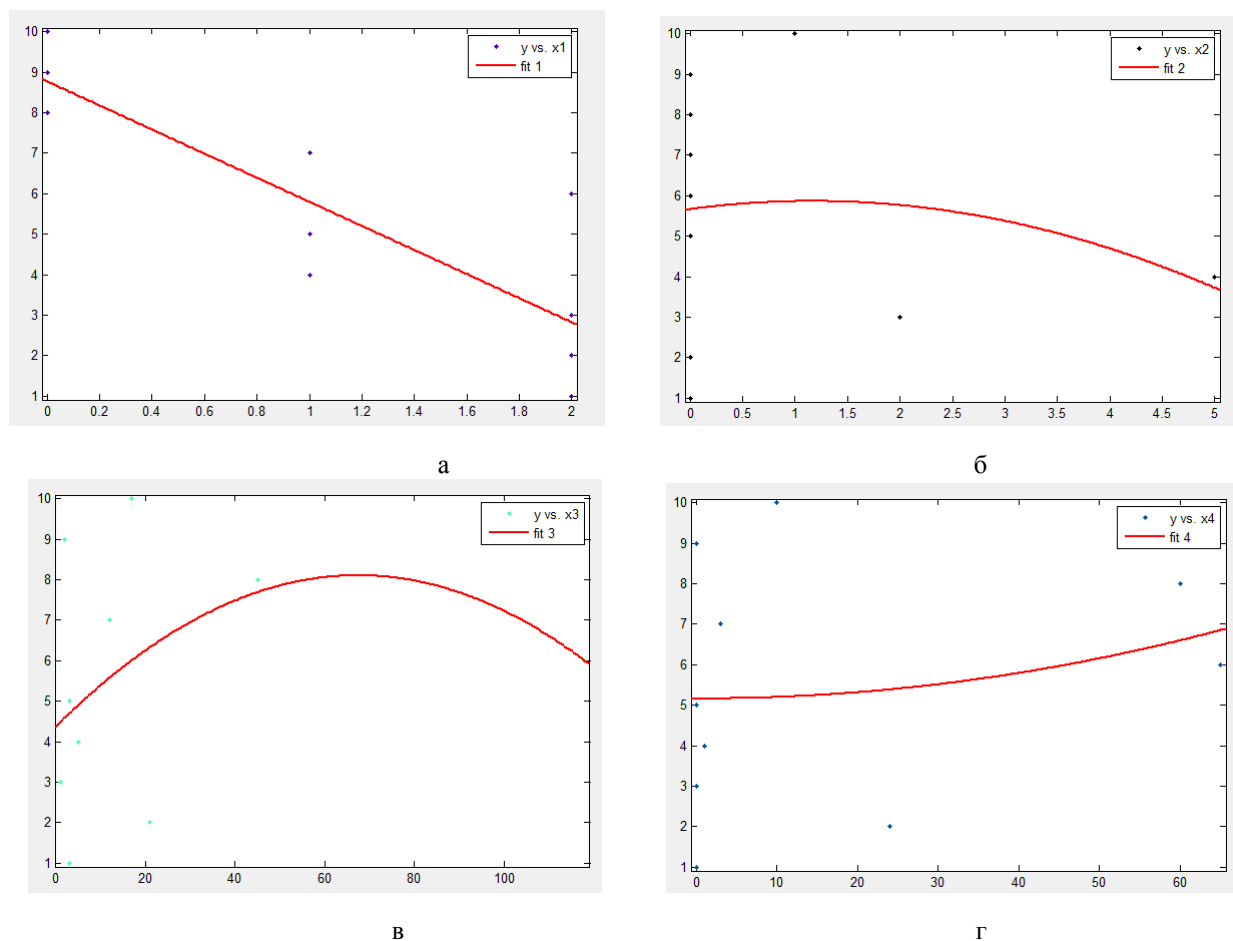


Рис. 2. Уплив внутрішніх чинників оптимізації X5, X9, X3 / 10, X2 / 10 (рис. а, б, в, г відповідно) на позицію сайту у видачі ПС Яндекс (точками показано експериментальні дані, а кривими – апроксимовані)

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.

V. O. Rakhuba, Cand. Sc. (Eng.)

PROBLEMS OF AUTOMATED CONTROL OF SEARCH OPTIMIZATION OF INTERNET - RESOURCES

The paper studies the problem of searching organization of the Internet – resources as the object of control. Optimization problem is formalized and its components are allocated. The paper contains the results of the experiment, dealing with the impact of main factors on sites locations in Yandex searching system.

Key words: searching optimization of web-sites (SEO), automated control, mathematical model.

Problem setting. Nowadays, the efficiency of operation of any enterprise greatly depends on the ability to find new markets for its products. In conditions of high competition for the client, that characterize the economic relations in our country and in other countries of the world, one of the main factors of successful operation of the enterprise is the optimization of its Internet – resource in the position of searching systems reporting

Among specialists, working in this sphere, the given problem is called SEO (search engine optimization).

Analysis of achievements. Analysis of the materials, studying the given problem [1,2,3] allows to make a conclusion, that in spite of the attempts of SEO automation, the majority of the specialists of sites optimization perform their work in “manual mode”, proceeding from empiric results of their own experiments. However, it should be noted that SEO process has creative component, that practically can not be automated; thus, the problem of development of the algorithms of automatic control of those SEO stages, that could be formalized and creation on the basis of these algorithms the software for the control of the process of search engine optimization is very urgent.

Problem setting. The aim of the given research is 1). Determination of SEO components, that can be automated, 2). Problems of automated control setting, 3). Determination of the problems of SEO automation and searching of the ways of their solution.

Main material and research. The process of search optimization of web – resource as the object of automation may be represented by the scheme (see Fig. 1).

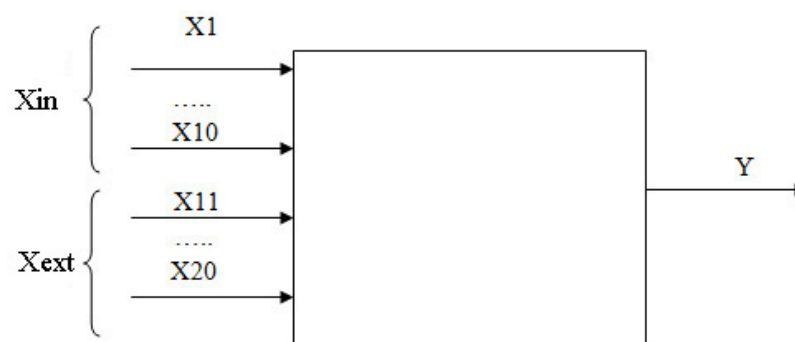


Fig. 1. SEO the object of control

The response Y in the given model is the position, the site occupies in the reporting of a certain searching system, regarding which optimization is realized

All the factors may be divided into 2 groups: internal X_{in} , characterizing internal properties of web – page, suggested in search reporting, and external X_{ext} , characterizing competitiveness of the searching request.

X_{in} includes [1 – 5]:

X1- age of the domain; X2 – PR of the site; X3 – number and quality of external references to the page; X4 – number of references, coming from the page (particularly, so-called “internal interlinking”); X5 – exact entry of the request in tags «Title» and «Description»; X6 – availability of the site in the catalogue of the given searching system; X7 – availability of the site robots.txt; X8 – availability of the file sitemap.xml; X9 – number of exact entries of the request in the text of the page; X10 – quality of information presentation on the page (content, availability of the numbered lists, sub – titles, images, etc.).

Xtext includes [1 – 5]:

X1- total amount of answers to the given request; X2 – number of main pages in theTOP of search results ; X3 – average number and quality of weblinks; X4 – average age of sites links; X5 – number of exact entries of the request in the tags «Title» and «Description» sites of search results; X6 – availability of context advertising blocks in search results system ; X7 – availability of Wikipedia and You Tube pages for the given request; X8 – availability of geographic map in search results; X9 – average value of PR sites in the result; X10 – other internal characteristics of the site in search result.

Thus, the problem of optimization of search promotion of the web-resource is expedient to present in the form:

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

$$Y \rightarrow \min$$

The constraints on the process are determined, proceeding from the budget, allocated for the given project.

On the whole, in the process of search optimization the following components may be distinguished:

1. Determination of constraints for the problem of search optimization of the given information resource.

a. Analysis of subject area, where the site will be promoted in order to investigate frequency and compatibility of search requests;

b. Development of the variants of site promotion and calculation of budget variants for their realization;

c. Selection of the variant of search promotion of the site;

2. Analysis of the semantic core of the information resource.

a. Formation of the base of potential search requests, the given resource may be promoted;

b. Selection of optimal requests in accordance with control criterion;

c. Construction of two- or three – level structure of semantic core, proceeding from the frequency of the requests;

3. Optimization of site pages, according to selected requests by means of control of internal factors in the model, given above.

a. Creation of rational design of the site.

b. Writing of the texts with qualitative presentation of information and optimal amount of requests entries.

c. Rational filling of page tags and main page of the site.

d. Creation of robots.txt and sitemap.xml files, registration in catalogues;

e. Creating of rational interlinking;

f. Purchase of advertising links from other sites in accordance with the selected variant of the budget.

For the problems of automatic control, points 1b, 2b and 2c are of great interest, because they require the development of the models of web-resources promotion in search systems, taking into account the impact of the above-mentioned factors on the position in search result and also require

the synthesis of requests selection criteria, that will provide the solution of optimization problem, put forward. Points 1a, 2a and 3f can be partially automated. However, greater part of the components of point 3, require creative human labour and can not be formalized. At this stage the operation of web-resource optimization system must include the presentation to the staff the array of searching requests, organized in the form of hierarchical structure, necessary amount of accurate entries of corresponding requests at corresponding pages and schemes of pages relinking with the indication of the needed anchor. Point 1c provides the agreement of the project with customer. Development of the subsystem of traffic forecast on the site depending on the chosen variant of optimization of its search promotion is actual.

To establish the dependences, making the basis of control algorithms, it is necessary to perform the research of search systems issue. It is obvious, that for preliminary investigation of the above-mentioned factors impact on the response only the method of passive experiment can be selected. It is expedient to carry out several experiments, choosing for each experiment certain set of internal or external factors. The results of passive experiment processing will have the sense, if not more than four factors are chosen for one series of research [6].

For such experiment educational topic is selected, in particular – professional orientation. The results of such research will be interesting for higher educational establishments and other institutions, providing services in the sphere of education.

Two groups, each consisting of tree series of research of TOP-10 issue of Yandex searching system, are performed. The first group allowed to take into account the impact of internal factors in each series, experiments were carried out by requests Z1 – «Как выбрать профессию», Z2 – «Какую профессию выбрать», Z3 – «Тест какую профессию выбрать», having the frequency 1542, 637 and 213 demonstrations a month. As factors, X5, X9, X3/10, X2/10 were chosen. Investigations of the second group took into consideration the impact of external factors, experiments were carried out for the sites moeobrazovanie.ru, kartaznaniy.ru and moya-pamyat.ru; factors where X1/mls, X9/10, X3/10, X5.

For the analysis of the collected data the technique of the results of passive experiment processing was used, the given technique is shown in [6]. In accordance with the given technique the assumption of linear form of regression model is taken, matrix of centrated factors with additional fictitious column is composed; then regression coefficients are determined by means of solution of matrix equation, composed of transformed matrices of initial data. By the results of the experiments the following regression dependences are obtained:

for the first group:

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

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

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

for the second group:

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

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

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

where x_1, x_2, x_3, x_4 – centered values of the chosen factors, correspondingly. It is impossible to verify the adequacy of the obtained models, applying classical method (by Fisher criterion), because of the lack of parallel experiments. That is why, for verification of the adequacy, the method, according to which dispersion of reproducibility is replaced by the dispersion of the response [6].

As a result of application of the given method the conclusions is made, that two model, from the models, obtained in the first series of experiments, may be considered to be adequate. Models, characterizing the impact of external factors, are inadequate.

Thus, the analysis of the results of the research carried out, allows to make the following conclusions: 1) method of obtaining regression impact models of internal factors of resource optimization, used in the given research, can be used for preliminary calculations at the stages 1a, 1b; 2) The greatest weight in the process of resource ranking has the factor of exact entry of the search request in the tags “Title” and “Description”. The dependence of site position in the issue on each of the selected for the experiment factors is graphically shown in Fig 2.

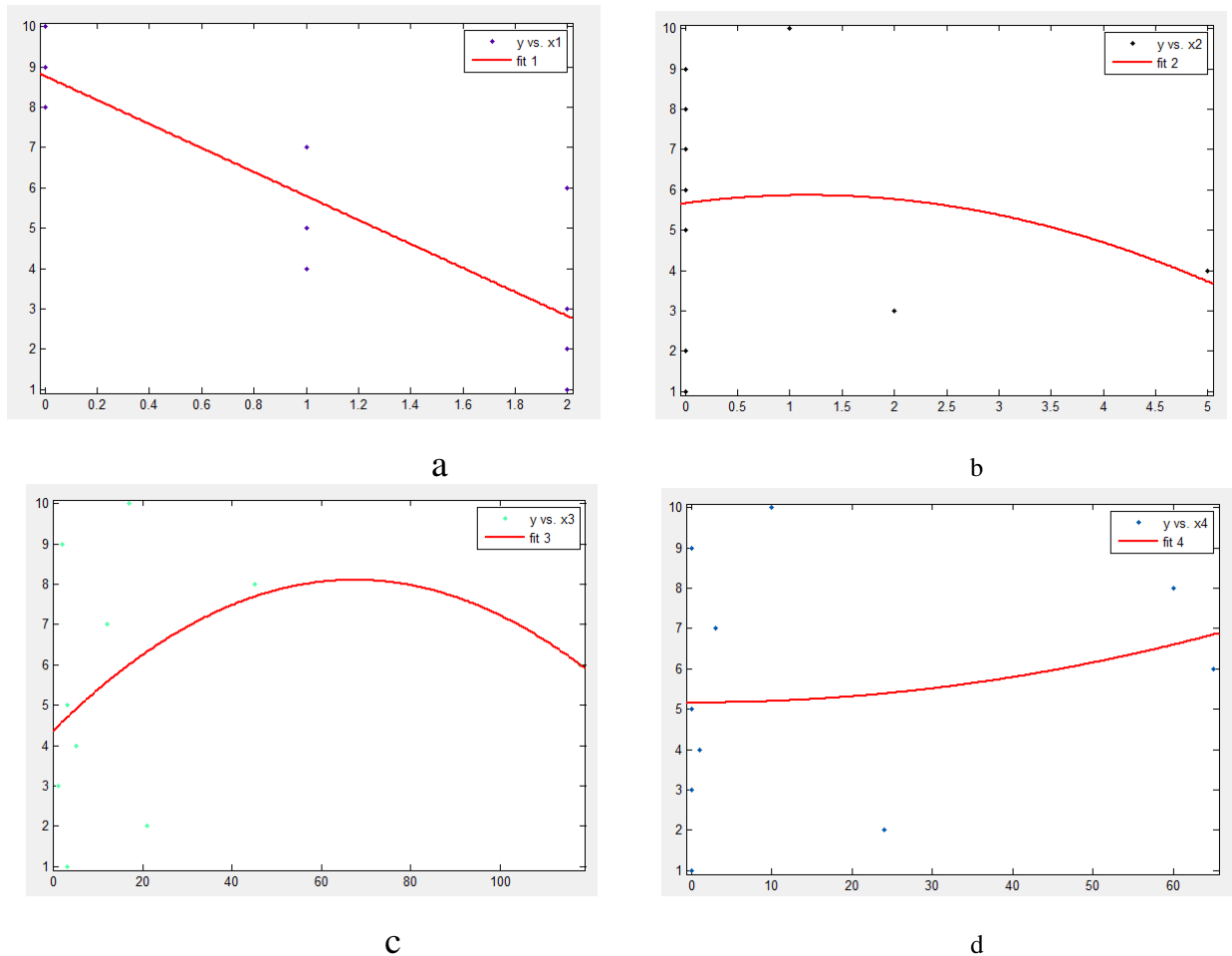


Fig. 2. Impact of internal factors of optimization of X5, X9, X3 / 10, X2 / 10 (Fig. a, b, c, d correspondingly) on the position of the site to issue in search system Yandex (dots show experimental data, curves – approximated data)

It is impossible to obtain adequate dependences for the second series of the experiments because in the formation of the value, called competitiveness of the request, more factors play role and (or) they interact in more complex connections.

Conclusions. Preliminary studies, carried out, prove the complexity of resources ranking description in search results and give grounds for the development of models with the help of modern methods of complex systems modeling. Taking into account such characteristics of ranking algorithms as high complexity, secrecy, functioning in the sphere of artificial intelligence, variability and continuous adaption, the most expedient is the usage of neural networks [7] as the method of processes modeling, on the base of which optimization of search promotion of web-

resources is realized. Thus, the results of formalization of the process of web-resources search optimization, presented in the given research, may be used for development of their promotion control systems. Problem of the development of sites ranking models in search systems remains actual, the most expedient for these systems development is the usage of neural networks.

REFERENCES

1. Актуальные факторы ранжирования [Электронный 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 [Электронный 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 [Электронный ресурс] / И. Токарев. – Режим доступа : <http://whiteseo.ru/content/view/521/>. – Назва з титул. екрану.
4. Как раскрутить сайт? [Электронный ресурс] / В. Беляев. – Режим доступа : <http://trynyty.ru/kak-raskrutit-sajt/>. – Назва з титул. екрану.
5. Алгоритмы поисковых систем [Электронный 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 [Электронный 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.

Rakhuba Viktoria – Cand. Sc. (Eng.), Assistant Professor with the Department of automated control of technological processes.

Zaporizhia State Engineering Academy.