

Релевантність посилань в Інтернеті за різними критеріями на рівні задачі оптимального розміщення контекстної реклами з нелінійною складовою

К.т.н.; старший викладач кафедри ММСА, ННК «ІПСА», НТУУ
«КПІ» Діденко Д. Г.

Студентка (магістр) кафедри ММСА, ННК «ІПСА», НТУУ
«КПІ» Климковська С.Д.

ВСТУП

Пошукова оптимізація SEO – комплекс заходів для підняття позицій сайту в результатах видачі пошукових систем по певних запитах користувачів з метою просування сайту.

Робота по оптимізації включає роботу з *внутрішніми і зовнішніми чинниками*. Робота присвячена вивченню відомих методів підвищення релевантності сайту, впливаючи на зовнішні чинники, та підвищення ефективності існуючих методів за допомогою підвищення рівня автоматизації.

Постановка задачі

Звичайні біржі посилань (Sape, Blogun, MainLink, XAP) просто надають можливість користувачам-оптимізаторам закуповувати посилання централізовано, не витрачаючи час на самотійні пошуки. Користувач заковує певну кількість посилань з погляду на їх характеристики.

Метою є створення біржі посилань, яка самотійно буде заковувати необхідну кількість посилань з певними характеристиками таким чином, щоб ефективність від такої закупки була максимальною в сенсі підвищення релевантності сайту.

Поняття біржі посилань

- Біржа посилань – система для купівлі/продажу посилань на сайти.
- Веб-майстри – беруть участь в системі для заробітку грошей за допомогою свого сайту шляхом продажу місць під посилання (посилань) зі свого сайту на сайти оптимізаторів.
- Оптимізатори – закупають посилання з інших сайтів (беклінков) на свої сайти для поліпшення позицій цих сайтів в пошукових системах.

Фактори, які впливають на вартість посилання

- рівень вкладеності сторінки (головна, 2-й, 3-й і т.д.);
- тематичний індекс цитування сайту (ТІЦ – визначається сумарною вагою сайтів, що посилаються);
- Google PageRank сторінки розміщення (метод обчислення ваги сторінки шляхом підрахунку важливості посилань на неї);
- кількість вихідних посилань зі сторінки;
- якість сторінки і сайту в цілому;
- тематика сайту;

Фактори, які впливають на вартість ПОСИЛАННЯ

- вік домену;
- наявність фільтрів накладених на сайт пошуковими системами;
- якість місця наданого для розміщення посилання (чим вище до початку сторінки, оточення іншими посиланнями та інше);
- наявність сайту в каталогах DMOZ, YACSA;
- рівня домену;
- зони домену.

Дані про біржі посилань

Назва	тЩ Яндекса	Google PageRank	Кількість сайтів	Кількість сторінок	Кількість користува чів
Sape.ru	1600	5	595 325	416 961 593	457 357
Blogun.ru	1100	4	315 257	—	239 915
MainLink.ru	350	5	643 000	383 000 000	—
SetLinks.ru	300	5	105 849	20 106 598	131 397
SeoZavr.ru	230	4	20 006	—	41 880
LinkFeed.ru	230	4	397 707	205 689 795	208 055
XAP.ru	180	4	—	55 637 026	186 801

Google PageRank

PR являє собою число, яке ставиться у відповідність сторінці в Інтернеті.

Існує 2 шкали вимірювання PR:

- Перший із варіантів є цілим дійсним числом і має лінійний характер, тобто збільшення цього значення буде пропорційно збільшенню статичної ваги, яка передається на даний документ по посиланням з інших ресурсів.
- Другий варіант являється похідним від першого. Це значення називається тулбарним значенням і має діапазон від 0 до 10.

Тулбарне значення являє собою \log_5 від верхньої межі відповідного інтервалу.

Залежність між шкалами PR

Дійсне число	Тулбарне значення
0,00000001 – 5	1
5 – 25	2
26 – 125	3
126 – 625	4
625 – 3125	5
3125 – 15625	6
15626 – 78125	7
78126 – 390625	8
390626 – 1953125	9
1953126 – нескінченість	10

Тематичний індекс цитування

тІЦ – тематичний індекс цитування, алгоритм пошукової системи Яндекс, призначений для визначення авторитетності сайтів шляхом підрахунку кількості ресурсів, які посилаються на нього та є схожої тематики.

Яндекс розраховує тІЦ сайтів за спеціальним алгоритмом, який враховує як подібність сайтів, що посилаються з тематики, так і рівень їх власного тІЦ і являє собою додатне ціле число (округлене додатне дійсне – аналогічне першій шкалі PR)

AdWords API

AdWords API – це набір веб-послуг, які використовуються для створення додатків, призначених для управління обліковими записами AdWords і пов'язаними з ними даними. Для отримання необхідних даних будемо використовувати такі сервіси, які надає AdWords API:

- CampaignService (надає права на додавання, редагування та призупинення дій компаній);
- TrafficEstimatorService (надає дані по трафіку для компаній, статичні дані, а саме, PR).

Метод оптимізації

Вхідні дані:

- початковий бюджет b_1 ;
- ціна кожного посилення a_j ;
- PR для кожного посилення $p_j = 5^{PR}$ (лінійна шкала).
- тІЦ для кожного посилення t_j

Метод оптимізації

$L_1 = \sum_{j=1}^n p_j x_j \rightarrow \max (1)$ і $L_2 = \sum_{j=1}^n t_j x_j \rightarrow \max (2)$ за умови, що $\sum_{j=1}^n a_j x_j \leq b_1$ (3), при чому $x_j \in \{0; 1\}, j = 1, \dots, n, (3) a_j \geq 0, p_j \geq 0, t_j \geq 0$

За допомогою *методу зважених сум нормованих неоднорідних критеріїв* зводимо функції L_1 та L_2 до однієї функції, яку і будемо максимізувати.

Для цього нам потрібно ввести функції цінності $v_1(L_1)$ і $v_2(L_2)$:

$$v(L) = v_1(L_1) + v_2(L_2)$$

Будемо розглядати лінійні функції з коефіцієнтами k_1 і k_2 .

Нехай L_i^* і L_i^* - найбільше та найменше значення критерію $L_i(x)$ на множині варіантів X .

Функцію цінності можна подати у такому вигляді:

$$\hat{v}(L) = w_1 \hat{v}_1(L_1) + w_2 \hat{v}_2(L_2)$$

Де $\hat{v}_i(L_i) = v_i(L_i)/v_i(L_i^*)$, $w_i = v_i(L_i^*)/\mu$, $\mu = v_1(L_1^*) + v_2(L_2^*)$

Після перетворень цільова функція буде приймати вид:

$$L = \sum_{j=1}^n c_j x_j \rightarrow \max$$

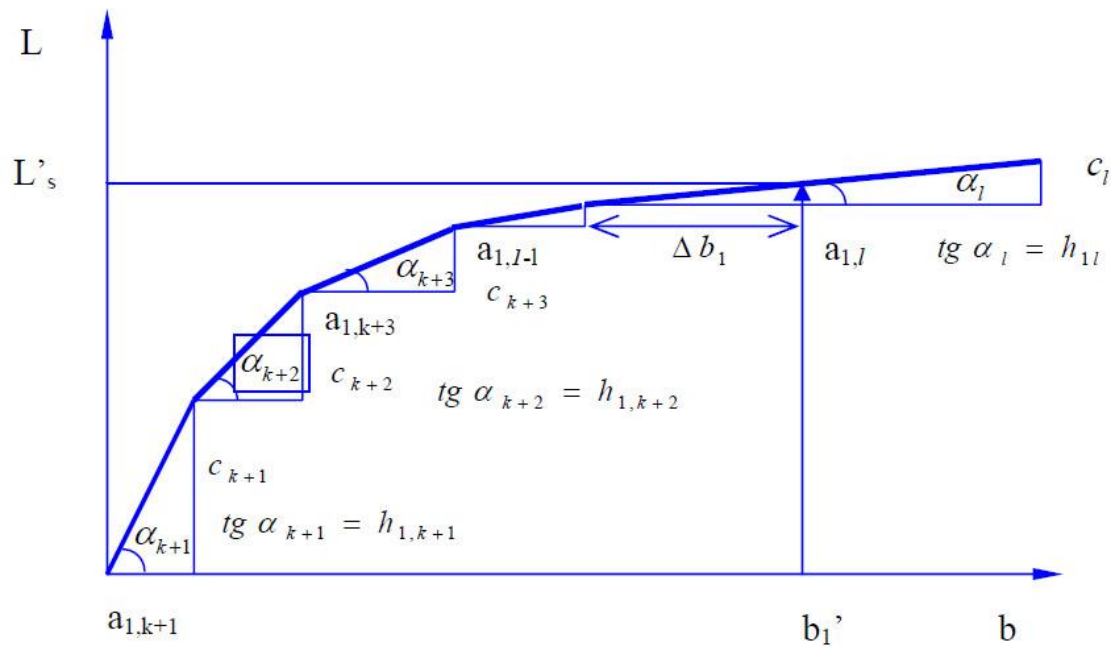
Метод оптимізації

Використовуємо алгоритм гілок і границь для розв'язку задач цілочисельного програмування з булевими змінними

- S – множина фіксованих змінних, які увійшли в допустимий розв'язок
- G_S – множина вільних змінних, з яких проводиться вибір в S чергової змінної.
- Позначимо $h_j = \frac{c_j}{a_j}$ і припустимо, що $x_j \in S, j = (1, \dots, k < n)$ і виконуються умови
- $h_{k+1} \geq h_{k+2} \geq \dots \geq h_l, l < n$ (4)
- $\sum_{j=k+1}^l a_j > b_1 - \sum_{x_j \in S} a_j x_j$ (5)
- $\sum_{j=k+1}^{l-1} a_j \leq b_1 - \sum_{x_j \in S} a_j x_j$ (6)
- Умови (5), (6) означають, що в множину S без порушення нерівності (2) можна додатково внести елементи $x_{k+1}, x_{k+2}, \dots, x_{l-1}$.

Метод оптимізації

$$H_S = \sum_{x_j \in S} c_j x_j + L'_S - \text{верхня границя (7)}$$



$$L'_S = \sum_{j=k+1}^{l-1} c_j + h_l \Delta b_1, \quad (8)$$

$$\Delta b_1 = b_1 - \sum_{x_j \in S} a_j x_j - \sum_{j=k+1}^{l-1} a_j \quad (9)$$

Метод оптимізації

З умов (4) – (6) випливає, що L'_S не менше максимального значення величини $\sum_{x_j \in G_S} c_j x_j$ при обмеженнях

$$\sum_{x_j \in G_S} a_j x_j \leq b_1 - \sum_{x_j \in S} a_j x_j = b_1', x_j \in \{0; 1\}, x_j \in G_S$$

Вибір чергової змінної для включення в множину S проводиться за допомогою умови

$$h_r(x_r) = \max_{x_j \in G_S} h_j(x_j) \quad (10)$$

Для вибраної змінної x_r визначаються величини $H_S(x_r)$ і $H_S(\bar{x}_r)$

тобто в S включається $x_r = 1$ або $x_r = 0$.

Метод оптимізації

Якщо в процесі розв'язку трапиться, що в множині G_S немає елементів, які можуть бути введені в множину S без порушення обмеження (2), то отриманий розв'язок $L_S = \sum_{x_j \in G_S} c_j x_j$ приймається в якості першого наближеного розв'язку L_0 .

Всі вершини дерева можливих варіантів, для яких виконується умова $H_S \leq L_0$, з подальшого розглядання виключаються.

Метод оптимізації

Із гілок, які залишилися, вибирається гілка з максимальним значенням H_S , і процес пошуку оптимального варіанту продовжується. Якщо в процесі розв'язку буде знайдено $L_S = \sum_{x_j \in G_S} e_j x_j > L_0$, то отриманий розв'язок приймається в якості нового наближеного результату. Обчислювальна процедура завершується, якщо для всіх гілок, які залишилися, виконується умова $H_S \leq L_0$.

Інтерфейс користувача

- Додаток являє собою автоматизовану біржу посилань, в якій користувачі діляться на веб-майстерів і оптимізаторів, кожен користувач може бути і першим, і другим одночасно.
- У розділі «для веб-майстера» користувач може реєструвати в системі свої посилання, задавати ціну, період дії посилання і кількість місць для розміщеної реклами на сторінці. За куплені посилання його бюджет поповнюється щодня.
- У розділі «для оптимізатора» користувач може реєструвати свої проекти, які він планує просувати в пошукових системах, закуповувати посилання.

Порівняння з Sape

Sape

- Користувач обирає посилання самостійно, переглядаючи список доступних на підставі їх характеристик релевантності та ціни, для зручності є фільтри по кожній з характеристик.
- В список включені посилання, в яких закінчились площадки.
- При обиранні наступного посилання не надано інформації про суму витрат на покупку.

Реалізована біржа

- Користувач задає суму, яку він планує витратити на закупівлю та натискає кнопку «Вуу», система автоматично заповує необхідну кількість посилань найоптимальнішим чином.

Список доступных посылань в Sape

OK	13.05.10	подписка на блог. Прочитать seo продвижение . Заводской раб seo блогер	gonzi.ru/news/the_mystery_of_the_druids_tajna_druidov_irus/20100513/ URL: http://zavrab.com/ ID: 512152 / 361375721 / Развлечения	10	3 3	3 0	0 0	0,99	<input checked="" type="checkbox"/>
			WL YAP HOST YAL GS GL Я: [7903] G: [649]		±				
OK	12.05.10	seo блог. Хороший seo блог заводского раба. Читать seo статьи	pitomec.ru/qaz URL: http://zavrab.com/ ID: 207920 / 61874078 / Дом и семья	350	3 4	3 0	0 0	8,25	<input checked="" type="checkbox"/>
			WL YAP HOST YAL GS GL YACA Я: [254000] G: [17600]		±				
OK	11.05.10	seo блог. Читаем про seo продвижение . Поисковая оптимизация	medikcenter.com/tabletkad/19-danazol-danazol.html URL: http://zavrab.com/ ID: 405965 / 130926218 / Медицина	0	4 3	3 0	0 0	0,66	<input checked="" type="checkbox"/>
			WL YAP HOST YAL GS GL Я: [359] G: [327]		±				
OK	11.05.10	статьи про seo. Тут про seo обучение . Заводской раб seo блогер	oe-group.ru/?page=269 URL: http://zavrab.com/ ID: 508610 / 377514539 / Интернет	30	6 9	2 2	0 0	1,65	<input checked="" type="checkbox"/>
			WL YAP HOST YAL GS GL Я: [3021] G: [2330]		±				
OK	11.05.10	Читать. Хороший SEO блог о SEO тематике. Блог читать	my-daina.ru/Slipy-azalia-p-837.html URL: http://zavrab.com/ ID: 227319 / 321942366 / Товары и услуги	0	2 2	3 0	0 0	1,38	<input checked="" type="checkbox"/>
			WL YAP HOST YAL GS GL Я: [3037] G: [5050]		±				
OK	09.05.10	статьи про seo. Тут про seo продвижение . Заводской раб seo блогер	varez-tyt.ru/filmi_on-line/treileri_k_film/31895-treiler-bitva-titanov/ URL: http://zavrab.com/ ID: 461755 / 346803882 / Новости и СМИ	10	1 1	3 0	0 0	0,75	<input checked="" type="checkbox"/>
			WL YAP HOST YAL GS GL Я: [28000] G: [29200]		±				
OK	08.05.10	статьи про seo. Тут про поисковая оптимизацию . Заводской раб seo блогер	vdox.ru/avtomobili-mototocikly/avtosignalizacii/15347.html URL: http://zavrab.com/ ID: 8338 / 1547817 / Товары и услуги	10	4 4	3 0	0 0	1,98	<input type="checkbox"/>
			WL YAP HOST YAL GS GL Я: [11000] G: [267]		±				
OK	08.05.10	seo блог. Читаем про seo обучение . Только не удалять отзывы	reclama-uq.ru/Board/page/431 URL: http://zavrab.com/	30	5 5	3 0	0 0	2,19	<input checked="" type="checkbox"/>

Сторінка стану бюджету користувача

HOME

MY BUDGET

LOG OUT



Budget table

	Day	Week
Expense	0.0	0.0
Profit	100.0	700.0

On Account: **5000.0**

Fill account



Сторінка оптимізатора

HOME

MY BUDGET

LOG OUT



Available Links

Link Address	PR	тИЦ	Price	Valid From	Valid To	Amount Of Places
www.google.com.ua	7	19000	220.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	1
http://www.wikipedia.org/	9	150000	100.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	2
www.facebook.com	9	330000	80.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	2
http://vk.com	9	360000	50.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	1
http://ipsa.kpi.ua	5	130	1.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	3
http://studsovet.ipsa.kpi.ua/studsovet	3	20	2.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	1
http://www.gismeteo.ua	6	5700	30.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	2
http://www.gismeteo.ua/weather-kyiv-4944	0	5700	9.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	1
http://vk.com/id47393806	0	360000	1.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	1
developers.google.com	7	2400	130.0	2014-06-01 00:00:00.0	2015-06-01 00:00:00.0	1

On Account

5000.0

400

Buy

Куплені посилання

HOME

MY BUDGET

LOG OUT



Bought Links

Link Address	PR	ТІЦ	Price
http://www.gismeteo.ua	6	5700	30.0
http://ipsa.kpi.ua	5	130	1.0
http://vk.com	9	360000	50.0
www.facebook.com	9	330000	80.0
http://studsovet.ipsa.kpi.ua/studsovet	3	20	2.0
www.google.com.ua	7	19000	220.0

total: **383**

On Account

4617.0

Наукова новизна

Новизною реалізованої біржі у порівнянні з існуючими є автоматизація закупки по обмеженому бюджету, тобто користувач повинен лише задати кількість коштів, які він згоден витратити на закупку посилянь і натиснути кнопку, далі система вирішить питання оптимальної закупки самостійно.

Перспективи та рекомендації для подальших досліджень

З даної тематики можна продовжувати дослідження таким чином:

- До аналізу посилань перед закупкою додати такі критерії як вік та зона домену, якість місця наданого для розміщення посилання, наявність сайту в каталогах DMOZ, YASA.

Дякую за увагу!
Ваші запитання?