

Змінюючи величину натягу веденої гілки нитки P_0 можна за формулою (11) отримати конкретні значення натягу на ділянці «опушка-ремізі» P_1^1 . Потім отримані результати апроксимувалися поліномом. Розроблена програма апроксимації дозволяла визначати необхідні коефіцієнти полінома з необхідною точністю.

$$P_1^1 = a_0 + a_1 P_0 + a_2 P_0^2 + a_3 P_0^3 + \dots + a_n P_0^n,$$

де $a_0, a_1, a_2, a_3, \dots, a_n$ - коефіцієнти полінома.

При цьому ступінь полінома вибирався автоматично залежно від необхідної точності результатів.

Висновки

Отримано систему диференціальних рівнянь рівноваги основної нитки на поверхні напрямної. Її розв'язання дозволило отримати систему трансцендентних рівнянь, які відображають залежність $P_1^1 = f(P_0)$. Розв'язання системи трансцендентних рівнянь, з використанням чисельних методів та ЕОМ, дозволило визначити натяг основних ниток при їх взаємодії з глазками галев ремізних рамок.

ЛІТЕРАТУРА

1. Щербань В.Ю., Слизков А.Н., Озадовский А.Б., Щербань Ю.Ю. Программные и математические компоненты проектирующих подсистем технологических процессов, оборудования, свойств материалов легкой и текстильной промышленности. – К.: Конус-Ю, – 2009. – 327 с.
2. Щербань В.Ю., Хомяк О.Н., Щербань Ю.Ю. Механика нити. – К.: КНУТД, 2002. – 196 с.

Надійшла 23.03.2010

УДК 62-50

АНАЛІЗ І МОДЕЛЮВАННЯ РОБОТИ ПОШУКОВИХ СИСТЕМ

О. Ю. МАРКЕВИЧ

Національний технічний університет України «Київський політехнічний інститут»

Виконано огляд груп найзначущих факторів задіяних у формуванні пошукової видачі. Розроблено оцінку обсягу тематичного трафіку з пошукової системи, спрямованого на оптимізований сайт. Виконано експерименти над факторами та побудовано відповідні математичні моделі пошукової системи Яндекс

Тенденція до інтеграції окремих глобальних мереж та стрімкий розвиток технологій глобальних інформаційних мереж спричинили перехід від ряду автономних мереж до єдиної глобальної мережі Інтернет. З часом ресурси Інтернету перетворилися у незамінний інструмент для повсякденної роботи людей багатьох професій. Швидке зростання інформації у мережі зробили його гігантом найрізноманітніших даних, важливість яких зростає пропорційно їх обсягу. Експерти стверджують, що обсяг інформації, що передається каналами Інтернету, подвоюється кожні півроку [1].

Об'єкти та методи дослідження

Гостро стоїть проблема пошуку інформації в Інтернеті. Найбільш популярним і практичним вирішенням цієї проблеми є використання пошукових систем. Кожна пошукова система – це комплекс програм тісно пов'язаних між собою, основними з яких є:

1. Робот Spider (павук). Автономно працююча програма, яка перебирає та завантажує на диск пошукового сервера вміст досліджуваних сторінок сайтів, що стоять в черзі на індексацію.
2. Робот Crawler (мандрівний павук). Його завдання – збирати всі посилання на досліджуваній сторінці, знаходити серед них нові, невідомі пошуковій системі, і додавати їх в чергу індексації.
3. Індикатор. Обробляє завантажені сторінки, фіксує слова, що зустрічаються на сторінці. Робить спеціальні помітки слів, що використовуються в заголовках, виділені в тексті жирним шрифтом. Після чого виконується аналіз та приймається рішення заносити сторінку до індексу, чи ні. Індекс – це структурована база даних серед яких безпосередньо і проходить пошук за пошуковою фразою.
4. Система обробки запитів і видачі результатів. Приймає запит користувача, формує запит до бази даних, звідки отримує результат і передає його користувачеві.

Існує ряд пошукових систем, які постійно розвиваються, удосконалюються і вже зараз алгоритми провідних пошукових систем показують відмінні результати за багатьма критеріями [2, 3].

Власнику тієї чи іншої інформації надзвичайно важливо, зробити її легкодоступною через пошукові системи, особливо якщо вона несе комерційний характер. Конкуренція за перші позиції пошукової видачі дуже велика. Також загострює ситуацію більш стрімке зростання джерел інформації, доступних через Інтернет, порівняно із зростанням кількості користувачів, яка фактично обмежена кількістю населення планети.

Особливої актуальності набуває отримання конкретних, практичних рекомендацій подання інформації в Інтернеті для максимальної її доступності. Важливим завданням є аналіз існуючих інформаційно-пошукових систем, розкриття їх особливостей, пошук і визначення значущості різних факторів, що впливають на пошукову видачу. Більшість факторів які враховують пошукові системи є загальновідомими, проте їх рівні значущості та метрики залишаються невідомими. Внести ясність в цю ситуацію допоможуть проведення експериментів, щодо цих факторів.

Постановка завдання

Метою цієї роботи є визначення найбільш значущих факторів, що впливають на формування видачі пошукових систем, а також формування рекомендацій стосовно вибору значень цих параметрів на початкових стадіях запуску та просування сайту. Побудувати модель регресійного типу для визначення залежності відносного обсягу тематичного трафіку для конкретного сайту однієї пошукової системи.

Результати та їх обговорення

Тематичний трафік з пошукових систем. Визначимо параметр, який необхідно оптимізувати дотримуючись вимог, що накладаються на параметр оптимізації згідно теорії планування експерименту. Основною метою якісного позиціонування сайту у пошукових системах є максимізація обсягу тематичного трафіку у цих пошукових систем направлено на оптимізований сайт. Основним завданням є аналіз існуючих інформаційно-пошукових систем, розкриття їх особливостей, пошук і визначення значущості різних факторів, що впливають на пошукову видачу. Необхідним параметром є оцінка обсягу тематичного трафіку, направлено на сайт з пошукової системи, виражена у відсотках по

відношенню до максимально можливого тематичного пошукового трафіку цієї пошукової системи. Проте необхідно врахувати той факт, що для сайтів різних тематик аналогічні методи просування можуть дати різні результати. Це викликано різними рівнями конкуренції та різними рівнями розвитку лідируючих сайтів різних тематик. Також багато суперечок викликає думка про те, що алгоритми пошукових систем однаково працюють в різних тематиках. Саме тому кожний окремий експеримент слід ставити в межах однієї тематики.

Побудуємо метрику, необхідну для оцінювання відносного трафіку. У основу метрики закладемо твердження, що будь-яку тематику можна описати скінченною множиною пошукових фраз. Кожному елементу цієї множини ставимо відповідно кількісну характеристику опису конкретної тематики цим елементом, число в межах від 0 до 1. Назвемо це число тематичним коефіцієнтом. Причому, зменшимо множину таким чином, щоб до неї входили тільки ті фрази, тематичні коефіцієнти, які більші деякого обмежувального коефіцієнта, скажімо 0,01. Назвемо цю множину семантичною множиною ключових фраз.

Опис тематики сайту є однією з ключових задач, що розв'язуються власниками та адміністрацією сайту під час його розробки та функціонування. Тому побудова множини семантичних фраз тематики сайту не буде проблемою. Але в силу присутності значного впливу людського фактора множини семантичних фраз, побудовані різними людьми для однієї тематики можуть значно відрізнятися і бути неповними, що негативно відрізняється на якості експерименту. Тому необхідне автоматизоване і формалізоване розширення цієї множини. Детальніший аналіз можливих пошукових фраз вимагає детальнішої класифікації та частоти запитів користувачів. Пошукові системи надають таку інформацію, як інструмент для зручного та швидкого створення контекстних рекламних компаній [4]. Пошукова система Google в межах використання даного інструменту дозволяє отримати інформацію про оцінку вартості кліку за заданою пошуковою фразою та рівень конкуренції, пов'язані між собою пошукові фрази та кількість середньомісячних запитів за пошуковою фразою в залежності від типу. Всього виділено три типи запитів: широкий, фразовий, точний. Широкий запит вказує на зв'язок пошукової фрази зі схожими фразами та іншими релевантними варіантами. Фразовий – з словосполученнями, що містять у собі пошукову фразу. Точний – становить точну відповідність пошуковій фразі.

Побудуємо математичну модель семантичної множини фраз:

$P^{(0)} = \{ phrase_i, t_coef_i \}, i = 1..N_0$ – початкова множина семантичних фраз, побудована адміністратором сайту. $P = Accord(P^{(0)})$ – множина семантичних фраз після її розширення.

Процедура розширення множини семантичних фраз складається з двох етапів. На першому етапі знаходимо нові фрази і підраховуємо для них коефіцієнти тематичності. Розрахунок коефіцієнта тематичності j -ї, що отримана шляхом семантичного розширення i -ї фрази:

$$t_coef_j = t_coef_i * norm(Google_accord_coef_{i,j}),$$

де $norm(Google_accord_coef_{i,j})$ – нормалізований коефіцієнт семантичного зв'язку між i -ю та j -ю фразами, отриманими з інструменту контекстної реклами. На другому етапі перерахуємо всі коефіцієнти тематичності:

$$t_coef_{i(new)} = \frac{t_coef_i * \sqrt{G_price_i * G_competition_i} * (G_exact_i)^{\frac{7}{3}}}{(G_broad_i * G_phrase_i)^{\frac{2}{3}}}, \quad (1)$$

та виконаємо процедуру нормування:

$$t_coef_{i(new)}' = \frac{t_coef_{i(new)}}{\max(t_coef_{j(new)})}, j = 1..N, \quad (2)$$

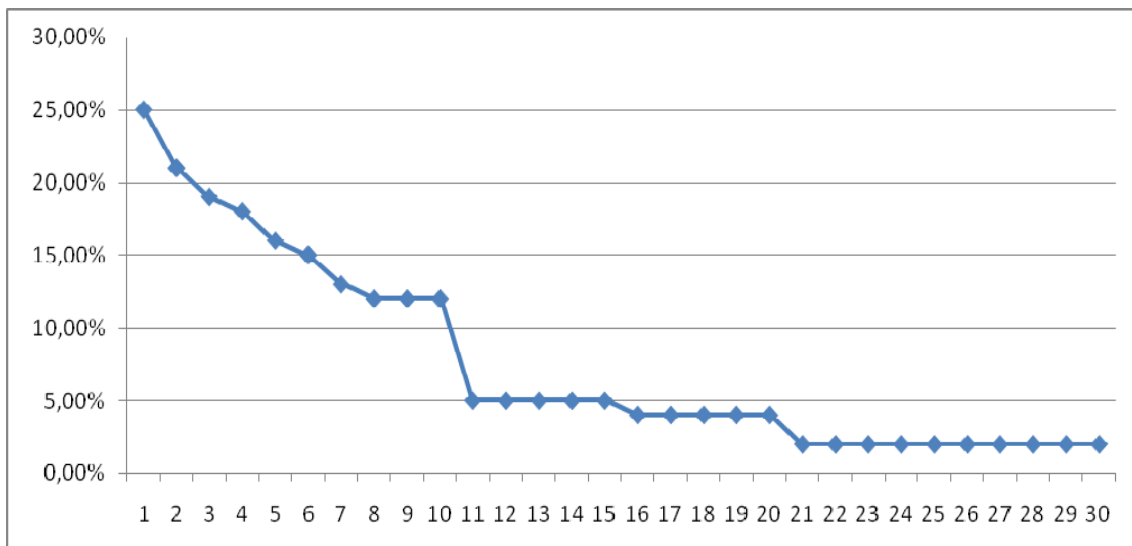
де G_price_i – оцінку вартості кліку i -ої фрази; $G_competition$ – рівень конкуренції за i -ю фразою; G_broad – середньомісячна кількість запитів широкого типу за i -ю фразу; G_phrase – середньомісячна кількість запитів фразового типу за i -ю фразу; G_exact – середньомісячна кількість запитів точного типу за i -ю фразу.

Реалізовано програмний модуль, завданням якого є збір необхідної інформації та реалізація процедури розширення множини семантичних фраз. Розглянемо функцію густини розподілу ймовірності переходу користувача за посиланням в залежності від позиції цього посилання у пошуковій видачі, зображеній на рисунку.

Ця функція є досить складною та безумовно залежить від тематики, до якої належить пошуковий запит, а також якості роботи самої пошукової системи. Здійснимо серію запитів до пошукової системи. Як пошукові фрази будемо по черзі брати кожен елемент семантичної множини фраз і фіксуватимемо позицію оптимізованого сайту у пошуковій видачі за кожною фразою. Таким чином, параметр оптимізації будемо розраховувати за наступною формулою:

$$y = \frac{\sum_{i=1}^{i=N} t_coef_i * F(pos_i)}{\sum_{i=1}^{i=N} t_coef_i * F(1)}, \quad (3)$$

де $F(pos_i)$ – значення функції розподілу, аргументом якої є позиція оптимізованого сайту в пошуковій видачі за i -ю фразою.



Функція густини розподілу ймовірності переходу користувача за посиланням залежно від його позиції у пошуковій видачі

Слід зауважити, що семантична множина фраз у будь-який момент часу між експериментами може бути розширеною, у зв'язку з чим точність параметра оптимізації буде підвищеною.

Фактори, що впливають на формування списку посилань. Згідно теорії планування експерименту, слід перерахувати всі істотні фактори, що впливають на формування списку посилань пошукової видачі. Кількість таких факторів в популярних пошукових системах може сягати декілька сотень [5]. Тому доцільним буде об'єднання цих факторів у групи. Перерахуємо отримані групи факторів:

1. Доменне ім'я сайту
2. Сервер на якому розміщено сайт
3. Архітектура сайту
4. Параметри сторінок сайту
5. Профілі відвідувачі сайту
6. Зовнішні посилання з сайту
7. Внутрішня «перелінковка» сторінок
8. Песимізація і фільтрація
9. Контент та ключові слова
10. Зовнішні посилання на сайт

В інформаційних технологіях утворився істотний сегмент послуг, орієнтованих на поліпшення позиціонування сайту у пошукових системах. Оптимізація сайту передбачає складний, часто довготривалий комплекс робіт з урахуванням усіх побажань та вимог щодо груп факторів. Сьогодні перші вісім груп факторів є достатньо глибоко дослідженими, існує значна кількість якісних практичних та навчальних матеріалів [2, 6–9], у тому числі і рекомендацій безпосередньо від представників пошукових систем та каталогів. По відношенню до цих груп факторів параметр оптимізації – це монотонно зростаюча функція. Чим точніше та якісніше виконані інструкції та поради пошукової оптимізації, тим більшого значення набудуватиме параметр оптимізації. Проведення експериментів з урахуванням груп цих факторів потребують суттєвих матеріальних та часових затрат, тому є проблемними. Слід зауважити, що за умовами сьогоденної конкуренції в Інтернет середовищі, нехтування цими інструкціями унеможливує високий рівень позиціонування сайту у пошукових системах. Зазначимо, що дві останні групи факторів мало формалізовані і потребують додаткових досліджень.

Експеримент над факторами групи зовнішніх посилань. Розглянемо керовані фактори, що належать до групи зовнішніх посилань на сайт. Маємо такі фактори:

1. Якісний показник ресурсів, на яких розміщуємо зворотне посилання.
2. Тематичність тексту посилання.
3. Кількість посилань з соціальних мереж та соціальних закладок.
4. Тренд збільшення кількості нових посилань.

Для кожного фактора оберемо можливі рівні значень, що будуть використовуватись в межах даного експерименту.

Якісний показник ресурсів, на яких будемо розміщувати зворотні посилання X_1 , оцінимо ранговою шкалою з двома можливими значеннями індикаторів: «погане» – відповідатиме 0, а «середнє»

– 1. Значення «добре» в експерименті задіяне не буде, адже отримання посилань з якісних, тематичних сайтів в основному вимагає значних регулярних фінансових затрат. Розміщення зворотних посилань реалізується шляхом реєстрації в базі каталогів, за допомогою спеціального програмного забезпечення (наприклад програмного продукту Allsubmitter) [10]. Якість бази визначається як комбінація середніх значень параметрів елементів бази, таких як: значення Google PageRank та аналогічних показників для інших систем (WebRank для Yahoo, ТИЦ для Yandex), кількість посилань на сайт фіксованих Yahoo, дані щодо рівня та статусу індексації сайту пошуковими системами, дані щодо відсутності штрафних санкцій на сайт.

Текст посилання на наш сайт x_2 розділимо на дві категорії, залежно від того описується тематика тексту посилання семантичною множиною ключових слів сайту, чи ні. Введемо індикатори: 1 – описується повністю, або частково, 0 – не описується. Кількість посилань з соціальних мереж та соціальних закладок x_3 являється кількісним показником, встановимо основний рівень – 15 посилань, відповідно інтервал варіювання від 0 до 30.

Тренд збільшення кількості нових посилань x_4 – опишемо цей фактор кількістю нових посилань отриманих за два тижні, приблизно стільки часу потрібно пошуковим системам для того щоб зафіксувати нові посилання. Встановимо основний рівень – 100, відповідно інтервал варіювання складатиме від 0 до 200.

Для виконання експерименту потрібно виконати мінімум п'ять дослідів. Ці досліди необхідно проводити на сайтах, у яких значення факторів, що не розглядаються в експерименті, фіксовані і не відрізняються у кожного експериментального сайту. Отже ставиться задача побудови мінімум 5 сайтів, причому вони повинні бути однотипними і обов'язково повинні бути виконані основні вимоги, щодо інструкцій пошукової оптимізації, інакше експеримент може бути безрезультатним, або сильно розтягнутим в часі. Для експерименту було обрано тип сайту «каталог тематичних ресурсів». Це сайт з фіксованим дворівневим рубрикаторм, на якому згідно кожної із рубрик зібрані посилання на тематичні сайти та короткі описи цих сайтів. Після чіткої фіксації вимог до навігації та сутності необхідних сайтів спроектована база даних. Засобами скриптової мови програмування php реалізована система керування контентом (cms). Розширення рубрикатора шляхом постановки у відповідність кожній рубриці множини із кількох тематичних тегів, реалізована подібним чином як і розширення семантичної множини тематичних слів. Розроблено скрипти, що дають можливість в автоматичному режимі зібрати необхідну інформацію і корегувати її в ручному режимі. Підбір тематичних сайтів реалізовано шляхом використання пошуку в пошукових системах за назвами підрубрик рубрикатора та тегів. Проблема наповнення баз даних контентом була вирішена. При цьому кожна наступна база формувалась з урахуванням наповнення раніше створених баз, це забезпечило відсутність дублювання інформації в межах отриманих баз. Кожна база піддавалася розробленій процедурі нормування, метою якої було забезпечення рівномірне наповнення підрубрик. Відмінність між створеними сайтами досягалася різними налаштуваннями текстових фраз навігації, описів головних сторінок, а також гамою кольорів дизайнів сайтів. Таким чином було отримано 6 сайтів. Кожному сайту було побудовано досить велику семантична множину фраз. Протягом двох тижнів всі сайти було розміщено на серверах в Інтернеті. Одразу після цього до пошукової системи Яндекс було подано відповідні заявки на індексацію сайтів.

Через два тижні (приблизно стільки часу достатньо для індексації) було запущено перший експеримент. Результати експерименту наведені в табл. 1.

Таблиця 1. Експеримент над факторами грипи зовнішніх посилань

x1	x2	x3	x4	y
0	0	5	180	0,533
1	0	5	100	1,64
0	0	25	180	1,423
0	0	15	20	0,583
1	1	25	50	3,844
0	1	5	30	1,31
c4	c3	c2	c1	c0
0,002522	0,049722	1,261049	1,401434	-0,24064
0,000781	0,004794	0,111486	0,092774	0,118151
0,911327				

Побудована модель регресії має вид:

$$y = 1,401434x_1 + 1,261049x_2 + 0,049722x_3 + 0,002522x_4 - 0,24064. \quad (4)$$

Коефіцієнт детермінації становить $0,911327$; модель є адекватною за критерієм Фішера на рівні значущості $\lambda = 0,05$.

Експеримент над факторами контенту та ключовими словами. Розглянемо керовані фактори, що належать до контенту та ключових слів:

1. Семантичність контенту.
2. Поповнення контенту.

Для проведення експерименту задіємо комбінацію двох найбільш значущих факторів групи зовнішніх посилань на сайт. Для кожного фактора оберемо можливі рівні значень, що будуть використовуватись в межах даного експерименту.

Семантичність контенту x_1 оцінимо ранговою шкалою з двома можливими значеннями: «погане» – відповідатиме 0, та «добре» – 1. Описаний вище алгоритм класифікації тексту дає можливість встановлювати значення фактора семантичності контенту в «1». Зрозуміло, що встановлення цього фактора в «0» не складає труднощів, достатньо розмістити на сайті не тематичний контент або розмістити в розділі відведеному для іншої тематики. Поповнення контенту x_2 є кількісним показником і відповідає кількості створення нових сторінок на сайті за останню добу; встановимо основний рівень – 50 сторінок, відповідно інтервал варіювання від 0 до 100. Керування якісним показником ресурсів на яких будемо розміщувати зворотні посилання x_3 буде проводитись аналогічно як і у попередньому експерименті. Тематику тексту посилання будемо формувати таким чином, щоб вона описувалася семантичною множиною ключових слів. Рекомендації щодо проведення експерименту співпадають із рекомендаціями описаними в попередньому експерименті. Для експерименту обрано тип сайту «дошка оголошень».

Це сайт з фіксованим дворівневим рубрикатором, на якому згідно кожної із рубрик зібрано приватні оголошення користувачів. Спроектвана база даних та реалізована система керування контентом. Таким чином отримано 6 сайтів. Кожному сайту була побудована досить об'ємна семантична множина фраз. Протягом двох тижнів всі сайти було розміщено на серверах в Інтернеті. Одразу після цього до пошукової системи Яндекс було подано відповідні заявки на індексацію сайтів. Через два тижні експеримент було розпочато.

Отримано лінійну модель, проте вона не пройшла перевірку адекватності. Виходячи з результатів, була спроба побувати неповну квадратичну модель, , додатково ввівши добуток факторів x_2x_3 . Результати експерименту наведені у табл. 2.

Таблиця 2. Експеримент над факторами контенту та ключовими словами

x1	x2	x3	x2*x3	y
0	20	0,2	4	0,381
1	30	0,2	6	0,549
0	70	0,8	56	5,241
0	50	0,8	40	3,432
1	80	0,4	32	3,311
1	40	0,2	8	0,769
c4	c3	c2	c1	c0
0,13764	-2,95248	-0,00932	0,097752	0,513925
0,011841	0,065272	0,00227	0,005174	0,021253
0,871947				

Побудовано модель регресії у виді:

$$y = 0,097752x_1 - 0,00932x_2 - 2,95248x_3 + 0,13764x_2x_3 + 0,513925 \quad (5)$$

Модель є адекватною за критерієм Фішера на рівні значущості $\lambda = 0,05$; коефіцієнт детермінації склав 0,871947. Найбільш значущою є одночасна оптимізація факторів x_2 та x_3 .

Висновки

Виконано огляд існуючих пошукових систем та груп найбільш значущих факторів, задіяних у формуванні пошукової видачі. Розроблено оцінку обсягу тематичного трафіку з пошукової системи, спрямованого на оптимізований сайт. Для виконання експериментів було спроектовано, розроблено та запущено в дію понад 10 сайтів. Виконано експеримент над факторами групи зовнішніх посилань на сайт. За результатами експерименту побудовано математичну модель роботи пошукової системи Яндекс у межах цієї групи. Також було виконано експеримент над факторами групи контенту та ключових слів і найбільш значущими факторами групи зовнішніх посилань, за результатами якого була побудована відповідна адекватна математична модель пошукової системи Яндекс у виді регресійного рівняння.

У подальших дослідженнях планується врахування додаткових факторів та створення ефективної методики автоматизованого аналізу функціонування пошукових систем.

ЛІТЕРАТУРА

1. «SEO moz» Online Journal – URL: <http://www.seomoz.org/articles/size-of-Internet.php>
2. Ашманов И.С., Иванов А.А. Продвижение сайта в поисковых системах.– М.: – Вильямс, – 2009.– 318 с.
3. Гусев В. С. Google. Эффективный поиск.– М.: – Вильямс, – 2006.– 240с.
4. Google AdWords: Инструмент подсказки ключевых слов
// <https://adwords.google.com/select/KeywordToolExternal?forceLegacy=true>.
5. Search Engine Journal – URL: <http://www.searchenginejournal.com/>
6. Буренина Т. А. Маркетинг на базе Интернет-технологий. – М.: Благовест-В, – 2005.– 152 с.
7. Евдокимов Н.В. Основы контентной оптимизации. Эффективная Интернет-коммерция и продвижение сайтов в Интернет.– М.: – Вильямс, – 2007.– 160 с.
8. Зиссер Ю. А. Маркетинг on-line. Как превратить сайт компании в эффективный инструмент продаж. – М.: – Издательство Гревцова, – 2007.– 304 с.
9. Зуев М. Б., Маурис П. А., Прокофьев А. Г. Продвижение сайтов в поисковых системах. Спасательный круг для малого бизнеса.– М.: Бином. Лаборатория знаний, – 2006.– 304 с.
- 10 Сайт розробників програмного продукту Allsubmitter // <http://www.allsubmitter.ru/>

Надійшла 12.04.2010

УДК 621.9.048.6

**РОЗРОБКА СИСТЕМИ ВИЗНАЧЕННЯ КІЛЬКОСТІ БІЛИЗНИ В
АВТОМАТИЧНІЙ ПРАЛЬНІЙ МАШИНИ НА ОСНОВІ ВИКОРИСТАННЯ
УЛЬТРАЗВУКОВИХ ДАТЧИКІВ**

Д.С. БІГУН, В.В. СТАЦЕНКО

Київський національний університет технологій та дизайну

У статті розглянуто принципи побудови системи автоматичного визначення кількості білизни в пральній машині на базі ультразвукового датчика відстані. Проведено дослідження будови та принципу роботи ультразвукових датчиків відстані

У наш час автоматичні пральні машини знайшли широке застосування як в побуті, так і у сфері послуг. У зв'язку з економічними вимогами до процесу прання одним з напрямків їх розвитку є вдосконалення електронних систем керування параметрами прання та мінімізація витрат електроенергії і миючої рідини. В нашій роботі розглянуто можливість застосування безконтактного способу визначення кількості білизни за допомогою ультразвукового датчика відстані.

Існують різні методи та системи автоматичного визначення кількості завантаженого одягу в пральну машину і знаходження оптимальної кількості миючої рідини. Одним з таких методів є інерційний метод, який полягає у визначенні моменту інерції завантаженого барабану. Перед вимірюванням двигун розганяють до заданої швидкості і відключають. Після чого, визначають швидкості обертання через задані проміжки часу. На підставі цих даних розраховують момент інерції