

ОСОБЛИВОСТІ ПРОЕКТУВАННЯ ВИСОКОРЕЙТИНГОВИХ ІНТЕРНЕТ-РЕСУРСІВ

© Басюк Т. М., Василюк А. С., 2014

Описано особливості проектування високорейтингових ресурсів у глобальній мережі Інтернет, проаналізовано відомі засоби оцінювання рейтингу та наведено й проаналізовано чинники, які формують їх структуру. Досліджено показники пошукової оптимізації, що дало змогу аргументувати їх застосування при побудові інтернет-ресурсу. Розглянуто особливості побудови інформаційної архітектури та визначено «вузькі місця», які виникають у процесі проектування. Проаналізовано можливі схеми організації ресурсів із використанням плоскої та глибокої архітектури.

Ключові слова: вебметрика, інтернет-ресурс, рейтинг, архітектура.

This paper describes the design features of the high ranking of resources in the global Internet network, the analysis of known means of ranking estimation are presented, and the factors that shape their structure are analyzed. Indicators of search optimization are studied, which help to substantiate their use in the construction of Internet resources. Features of construction of information architecture and "bottlenecks" are identified that arise in the design process. The possible schemes of resources using a flat and deep architectures are analysed.

Key words: webometrics, internet resources, ranking, architecture.

Вступ. Загальна постановка проблеми

У сучасному інформаційному суспільстві інтернет-ресурси є однією з найважливіших частин мережових інформаційних технологій, що формують імідж як окремої особистості, так і всієї корпорації. При цьому важливо, щоб створений ресурс був не лише грамотно спроектований та зручний для сприйняття, але й залучав множину відвідувачів, а це можливо лише у випадку наявності високих рейтингових оцінок, які використовуються всіма пошуковими системами [1, 2].

Наявність рейтингових механізмів, які розробляються і застосовуються найбільшими пошуковими системами, такими як Google, Bing, Yandex [2], при видачі результатів пошуку дає можливість користувачам всього світу знаходити найбільш якісну та релевантну запиту інформацію. Результати запиту впорядковуються пошуковими машинами і виводяться користувачам у порядку зменшення рейтингів. Таким чином, наявність високих рейтингів підвищує рівень популярності інтернет-ресурсу та сприяє його пошуковій видимості [3].

Зв'язок висвітленої проблеми із важливими науковими та практичними завданнями

Створення високорейтингового ресурсу є надзвичайно важливою та водночас складною задачею. Оскільки з однієї сторони передбачає проектування веб-сайту, який буде часто оновлюватися та містити актуальну інформацію, що надасть передумови із підтримання рейтингу [3], а з іншої – відсутність єдиних норм оцінки рейтингу ресурсу [4] вимагає від розробника застосування відмінних механізмів оптимізації для різних пошукових систем.

З огляду на це окреслена науково-практична задача являє собою складний та безперервний процес, під час здійснення якого широко застосовуються як алгоритми оптимізації й пошуку, так і засоби вебметрики [5]. Розв'язання поставленої задачі дасть змогу визначити особливості проектування високорейтингових інтернет-ресурсів та надасть засоби із адаптації відомих методів та алгоритмів до задач підтримки прийняття рішень у визначеній предметній галузі.

Аналіз останніх досліджень та публікацій

Існує множина методик визначення рейтингу інтернет-ресурсу, починаючи від локальних алгоритмів, які використовуються пошуковими системами (PR, TiЦ) до глобальних, що являють собою сукупність множини показників [6,7]. Одним з найвідоміших вебметричних засобів є проект іспанської дослідницької групи Cybermetrics Lab «Webometrics Ranking» [5], присвячений вивченню вебметричних індикаторів та ранжування на їх основі інтернет-ресурсів. Іншим популярним ресурсом є портал «Ranking Web of World Research Centers», що визначає рейтинг науково-дослідних інститутів [8]. На однойменному ресурсі опубліковано рейтинг «Top 2000 R&D», до якого входять лише десять українських наукових веб-сайтів, що займають невисокі позиції. Зокрема портал Національної академії наук України займає 499-те місце, сайт Інституту теоретичної фізики ім. М. М. Боголюбова – 815-те, сайт Національного інституту стратегічних досліджень – 906 позицію, що свідчить про низьку їх популярність в глобальній мережі.

Особливістю даних досліджень є те, що користувачу надається інформація щодо позиції відповідного ресурсу, проте відсутні механізми, які надають рекомендації із початкового конструювання високорейтингового ресурсу, що робить актуальною задачу проектування системи оцінювання та надання рекомендацій, яка надасть засоби із аналізу та покращення пошукової видимості веб-сайту в глобальній мережі Інтернет.

Основні завдання дослідження та їх значення

Метою дослідження є визначення особливостей проектування високорейтингових ресурсів у мережі Інтернет. Проведене дослідження надасть засоби із виконання комплексного аналізу веб-сайту, що сприятиме побудові механізму надання рекомендацій з його вдосконалення з метою входження у верхні позиції рейтингу пошукових систем. Для досягнення поставленої мети необхідно вирішити такі основні завдання: проаналізувати відомі технології оцінювання рейтингу та визначити можливості їх застосування в досліджуваній задачі; проаналізувати відомі архітектури побудови ресурсів й особливості їх застосування; виявити чинники, що найбільше впливають на процес побудови високорейтингового ресурсу.

Реалізований з використанням даної методики інтернет-ресурс зможе претендувати на високі позиції в рейтингу пошукових систем, що, своєю чергою, призведе до збільшення як кількості відвідувачів, так і його популярності в інтернет-просторі.

Результати виконаної роботи у своїй сукупності вирішують актуальне завдання проектування високорейтингових ресурсів у мережі Інтернет.

Аналіз результатів досліджень

Оцінюванням рейтингу інтернет-ресурсу займається особлива галузь – вебметрика (webometrics), що визначає розділ інформатики, в межах якого досліджуються кількісні аспекти конструювання й використання інформаційних ресурсів, структур та технологій у всесвітній павутині [5]. Сучасна вебметрична методологія дослідження веб-сайтів, розроблена Кіберметричною лабораторією центру наукової інформації та документації національного дослідницького центру Іспанії, передбачає такі напрями досліджень:

- веб-індикатори (індекси цитування, показники видимості ресурсу);
- соціальні феномени у веб-спільнотах (соціальні мережі, співтовариства сайтів);
- збирання даних, що розміщується у всесвітній павутині (функціонування пошукових роботів, краулерів, особливості та критерії інформаційного пошуку);
- аналіз гіперпосилань, зокрема зв'язків між сайтами згідно з тематичними особливостями.

Рейтинг Webometrics формується на основі чотирьох показників: розміру, видимості, кількості важливих («цінних») файлів, кількості зовнішніх посилань (табл. 1.). Проаналізувавши наведені показники рейтингу Webometrics, можна зробити висновок, що при незважаючи на їх простоту та прозорість (відповідають змістовній інтерпретації рейтингу, а їх значення обчислюється поза всією сукупністю досліджуваних ресурсів), вони є доволі суб'єктивними за способом збирання даних (коло пошукових систем та аналізованих даних визначають дослідники).

При цьому, відмінною особливістю цього рейтингу є складання не кількісних значень показників, а рангових (підсумовуються місця ресурсу в процесі ранжування за кожним з чотирьох показників рейтингу), на основі якого формується кінцевий рейтинг без формування рекомендацій з його підвищення.

Таблиця 1

Характеристика показників рейтингу ресурсу

Показник	Ваговий коефіцієнт	Вміст	Спосіб підрахунку
Size (S)	25 %	Кількість знайдених посилань на сторінки ресурсу	Для визначення цього показника застосовуються чотири пошукові системи: Google, Yahoo, Live Search та Exalead. Кількість сторінок інтернет-ресурсу, знайдених за допомогою кожної пошукової системи, нормується за максимальним значенням (максимум прирівнюється до 1), потім отримані показники підсумовуються.
Visibility (V)	50 %	Кількість унікальних зовнішніх посилань на сторінки ресурсу	Пошук зовнішніх посилань на інтернет-ресурс здійснюється за допомогою описаних пошукових систем. При цьому, як і в попередньому випадку, значення показника за кожною пошуковою системою нормується за максимальним значенням та здійснюється їх підсумовування.
Rich Files (R)	12,5 %	Кількість «цінних» файлів, розміщених на ресурсі	Підраховується кількість файлів таких чотирьох форматів: Adobe Acrobat (.pdf), Adobe Postscript (.ps), Microsoft Word (.doc), Microsoft PowerPoint (.ppt). За кожним типом файлів підраховується окремий показник, який нормується та підсумовується за кожним типом файлів
Scholar (Sc)	12,5 %	Число Кількість сторінок і посилань на ресурс	Використовується пошукова система Google Scholar з допомогою якої підраховується кількість розміщених на сайті матеріалів і відповідних посилань

З огляду на наведені параметри для їх підвищення пропонується або здійснювати процедуру SEO-оптимізації (для існуючих ресурсів) або створювати ефективну структуру ресурсу (для нових проектів). У першому випадку необхідно здійснити: складання семантичного ядра, підготовку текстів сайту (копірайтинг), оптимізацію коду, розміщення посилань, моніторинг і аналіз результатів [9].

Складання семантичного ядра полягає у відборі пошукових фраз, за якими здійснюватиметься просування інтернет-ресурсу та їх розподіл за сторінками сайту. В загальному випадку це повний набір слів і словосполучень, що описують певний предмет та його характеристики. Правильне складання семантичного ядра має важливе значення для пошукової оптимізації, тому що саме на його основі будуватиметься пошуковий механізм, без якого не можливе просування ресурсу.

Підготовка текстів сайту (копірайтинг) полягає в підготовці текстів, привабливих для відвідувачів сайту і одночасно оптимізованих під пошукові системи. Існує кілька видів копірайтингу: простий та складний, при цьому їх часто називають рерайтингом. Простий рерайтинг полягає в переробці однієї статті в іншу, в якій описується все те саме, але іншими словами. На виході формується «унікальний контент», що не порушує авторських прав джерела. Складний рерайтинг здійснюється аналогічно, але на вході існує множина джерел. У результаті цього процесу створюється оптимізований текст щодо запитів пошукових систем.

Під оптимізацією коду сайту розуміють технічний процес щодо зменшення розміру коду, який передається користувачам і пошуковим системам при під час завантаження. Оптимізація коду дозволяє зможу: зменшити розмір сторінок сайту і прискорити їх завантаження; знизити навантаження на сервер хостингу, що позитивно позначається на його стабільності (особливо це актуально для проектів з високим рівнем відвідуваності) [9].

Розміщення посилань належить до «зовнішніх чинників» оптимізації, та без врахування якого неможливе успішне просування сайту в конкурентних тематиках. Ці роботи проводяться постійно і полягають у створенні зовнішнього «посилального оточення» [8] та розміщення «необхідного тексту» навколо посилань. Моніторинг і аналіз результатів полягає в постійному контролі позицій

інтернет -ресурсу в результатах пошуку по за всіма відібраними запитами і аналіз змін, що відбуваються. Дана Ця процедура дозволяє дає змогу оперативно реагувати на зміни в алгоритмах ранжування пошукових машин, проводити оцінювання і своєчасно вносити необхідні корективи.

Другий варіант реалізується на початку проектування й передбачає створення інформаційної архітектури веб-сайту, оскільки добре спроектована архітектура надає множину переваг як для потенційних користувачів (зручність та ефективність роботи), так і для пошукових систем (швидкість опрацювання). З огляду на це, логічна і правильно спроектована архітектура веб-сайту надає механізми подолання зазначених труднощів та створює передумови як до залучення пошукового трафіку, так і зручності користування ресурсом.

Відомо [7, 8], що в основу організації інтернет -ресурсу покладено два базові принципи: зручність (легкість використання сайту) та інформаційна архітектура (логічна ієрархічна структура контенту). Що стосується першого принципу, то пошукові механізми намагаються відтворити «людський процес» сортування релевантних веб-сторінок за якістю. А саме: якби цим процесом займалась людина, то зручність користування та враження користувача були б основними для оцінювання. У випадку пошукових механізмів застосовують множину альтернативних чинників, вимірювання яких можна подати схемою (рис. 1).

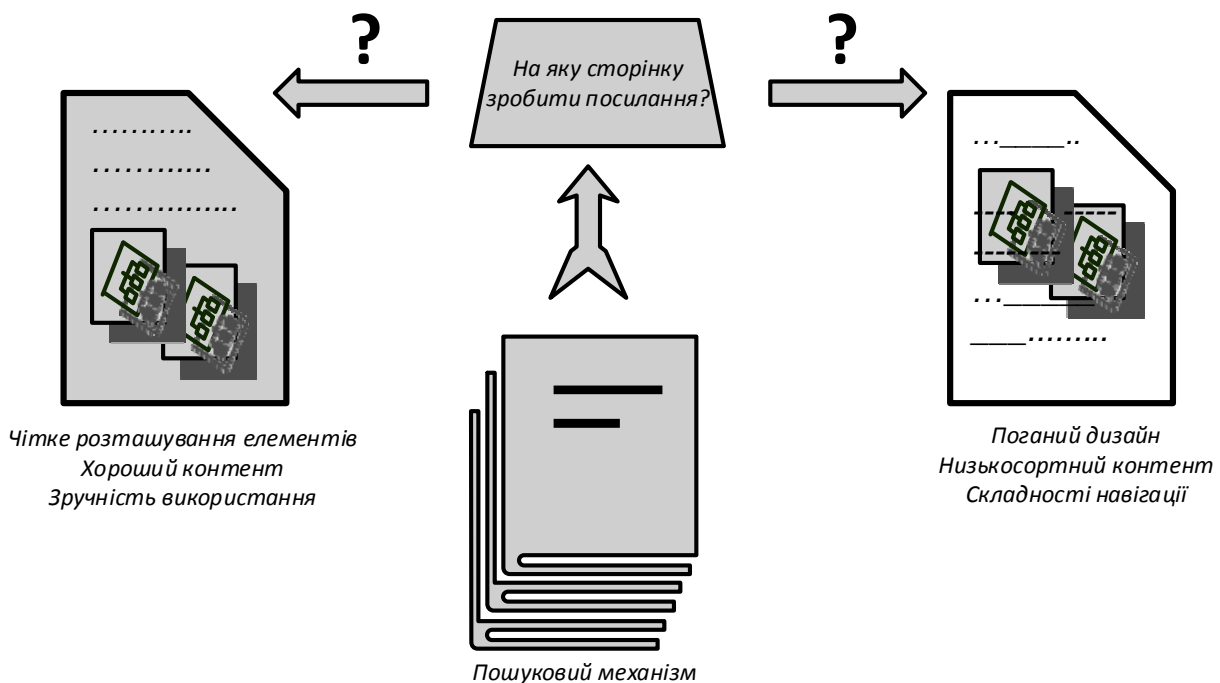


Рис. 1. Особливості проведення вибору пошуковим механізмом

З огляду на те, вдала організація веб-сайту з більшою ймовірністю сприятиме додатковим посиланням, що, своєю чергою, вплине на підвищення рейтингу інтернет -ресурсу.

Другий принцип – проектування архітектури ресурсу відіграє значну роль та вимагає від розробника знання певного рівня соціальної психології, оскільки, увійшовши на веб-сайт, користувач «не знає, куди йти та що робити». Тому найвищі шанси створити у користувача позитивне враження має така архітектура, що розуміє цю проблему, застосовує звичні стандарти зручності користування та інтуїтивно зрозумілу структуру посилань. Добре організована архітектура допомагає вирішити зазначені проблеми і дає семантичні переваги як користувачам, так і пошуковим механізмам.

На рис. 2 зображено структуру інтернет -ресурсу, що використовує структуровану архітектуру, яка, з одного боку, дає змогу виправдати очікування користувачів щодо контенту і створити позитивне враження від перегляду, а з іншого – допомагає пошуковим механізмам розглядати організацію контенту в концептуально-логічних групах та надавати «додаткові бали» веб-сайту, що логічно організований, оскільки пошукові механізми розпізнають шаблони в архітектурі та «залучають» ресурси, які підтримують інтуїтивну структуру контенту [9].



Рис. 2. Структурована архітектура інтернет -ресурсу

Наступним етапом в створенні архітектури є структурування категорій. Переглядаючи ресурси глобальної мережі, пошукові механізми накопичують значну кількість даних (структур мови, інформації за тематикою, зв'язків між контентом), при цьому формально не застосовуючи засобів штучного інтелекту, вони створили значний репозитарій, за допомогою якого реалізуються складні визначення, що побудовані на загальних шаблонах [8, 10]. Павуки пошукових механізмів можуть вивчати семантичні зв'язки у міру того, як вони опрацьовують мільйони сторінок, що охоплюють визначену предметну галузь (рис. 3).

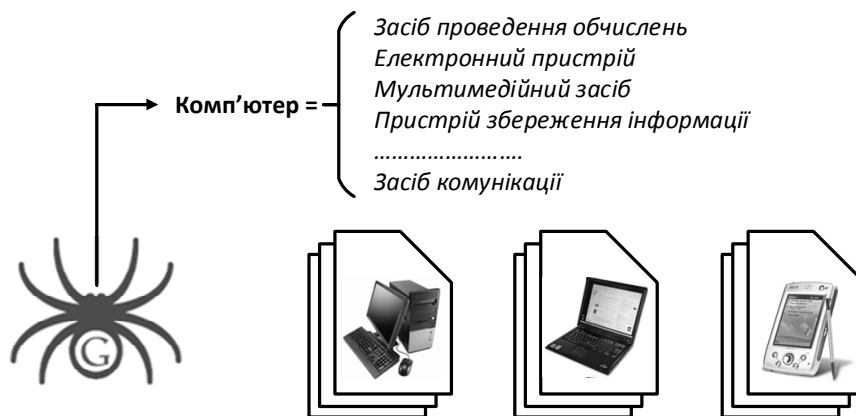


Рис. 3. «Пошукові павуки» вивчають семантичні зв'язки

Незважаючи на те, що контент не завжди потрібно структурувати за найбільш передбачуваним шаблоном (у випадку, коли інтерес для відвідувача представляє інший метод структурування), логічна організація предметів допомагає як користувачам, яким буде простіше користуватися відповідним ресурсом, так і пошуковим механізмам, котрі завдяки підвищеній тематичній релевантності нададуть веб-сайту вищий рейтинг.

Розробляючи архітектуру інтернет-ресурсу, необхідно також врахувати таксономію та онтологію майбутнього веб-сайту. Таксономія – це двовимірна ієрархічна модель архітектури, а онтологія – представлення, за допомогою якого людський мозок уявляє тематичну область. Онтологія є значно складнішою ніж таксономія, оскільки в ній наявна значна кількість типів зв'язків, а їх освоєння при проектуванні інтернет-ресурсів можна реалізувати за допомогою методики «сортування карт». За таким сортуванням користувачі записують основні концепції на

карти, з якими далі здійснюються певні маніпулювання (читання, модифікація, організація). Потім відповідальна особа (експерт), збирає картки в порядку, який утворює найбільш логічний шлях в процесі пошуку необхідної інформації. Цю методику можна застосовувати не лише для виявлення логічних маршрутів у проєктованому ресурсі, але й для пошуку місця з неоднозначною або незрозумілою термінологією, яку необхідно перефразувати.

Наступною задачею, яку необхідно розв'язати, є вибір типу архітектури: плоскої чи глибокої [8–10]. При цьому одним з найбільш строгих правил щодо інтерактивної взаємодії є побудова плоскої архітектури сайту, оскільки, щоби досягти певної сторінки, необхідно здійснити мінімальну кількість переходів (рис.4).

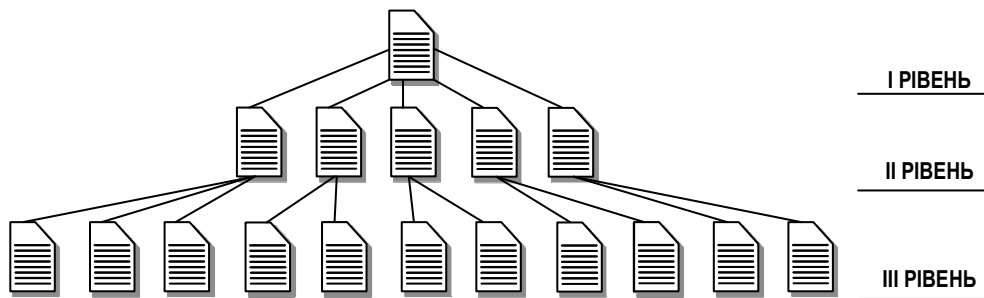


Рис. 4. Типовий приклад плоскої архітектури інтернет -ресурсу

Глибока архітектура передбачає створення довгих маршрутів-переходів, які необхідні для доступу до контенту. Вкладеність сторінок ресурсу при цьому визначається п'ятьма і більше рівнями ієрархії, тим самим створюючи значні труднощі в навігації як для користувачів, так і для пошукових роботів (рис. 5).

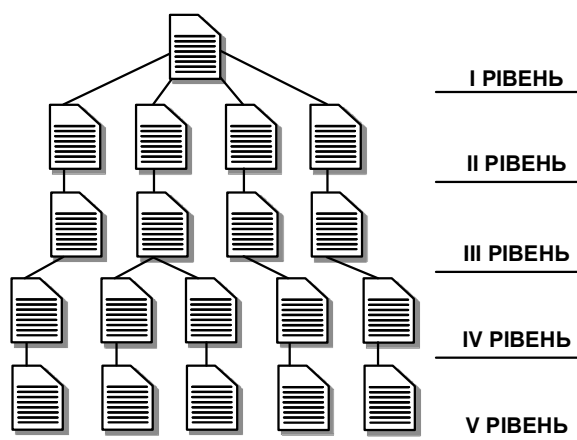


Рис. 5. Типовий приклад глибокої архітектури інтернет -ресурсу

Тому, проєктуючи інформаційні ресурси, доцільно застосовувати плоску архітектуру ресурсу, оскільки вона зменшує відсоток «перерваних шляхів» та сприяє повторним візитам. Іншою стороною плоскої архітектури є те, що пошукові механізми обмежують кількість посилань, що проглядаються з однієї сторінки (зазвичай це 100), тому виконання цієї вимоги означає, що кількість посилань на сторінках категорій необхідно обмежувати відповідною кількістю. Проблема «сотні» посилань на сторінку безпосередньо стосується й іншого правила – необхідності уникати значного розбиття на сторінки. А саме існує практика створення списку елементів на сторінках, поділених лише номерами, яка є достатньо проблематичною, оскільки таке розбиття не забезпечує тематичної релевантності, сприяє частковому дублюванню контенту та створює «пастки» для пошукових механізмів у вигляді сотень або навіть тисяч зайвих сторінок низької якості, які значно зменшують ефективність пошуку (рис.6).

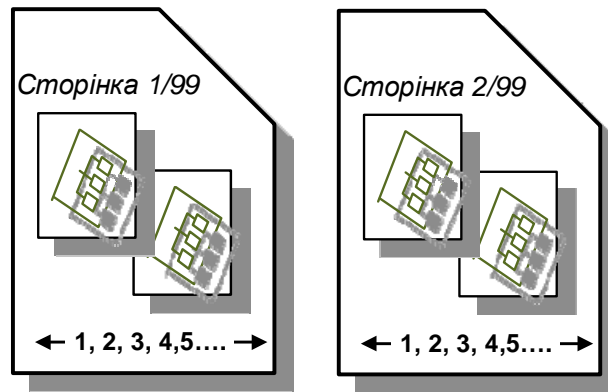


Рис. 6. Структури розбиття сторінок інтернет -ресурсу

Висновок та перспективи наукових розвідок

У результаті проведеного дослідження виявилось, що для здійснення успішного процесу побудови високорейтингових ресурсів необхідними чинниками є: зручність користування (досягається за рахунок створення відповідної архітектури ресурсу), структурування категорій (реалізується пошуковими механізмами) та врахування онтології майбутнього ресурсу.

Використання описаних заходів сприятиме створенню високорейтингових ресурсів для подальшого їх входження у верхні позиції рейтингу пошукових систем. Подальші дослідження будуть скеровані на побудову відповідних алгоритмів ранжування та верифікацію їх роботи.

1. Байков В. Д. *Интернет: поиск информации и продвижение сайтов* / В. Д. Байков. – СПб.: ВHV-СПб, 2000. – 288 с.
2. Grappone J. *Search Engine Optimization (SEO): An Hour a Day*. / J. Grappone – United States: Wiley Publishing, Inc., 2013. – 435 p.
3. Basyuk T. M. *Optimization of internet resource under plurality of user requests* / Т. М. Basyuk // Тези доповідей міжнародної науково-практичної конференції «Інформаційні управляючі системи та технології». – Одеса, 2013. – С. 305–307.
4. Зуєв М. *Просування сайтів у пошукових системах* / М. Б. Зуєв, П. А. Маурис. – К. : Біном. Лабораторія знань, 2007. – 304 с.
5. Almind T. C. *Informetric analyses on the World Wide Web: Methodological approaches to «webometrics»* / Т. С. Almind, P. Ingwersen // *Journal of Documentation* – 1997. – Vol. 27, No. 2. – P. 404–426.
6. Пелецишин А. М. *Позиціонування сайтів у глобальному інформаційному середовищі* / А. М. Пелецишин. – Львів : Видавництво Львівської політехніки, 2007. – 260 с.
7. Яновский М. В. *Системы SEO-оптимизации* / М. В. Яновский // *Computer World Україна*. – К. : Препрес, 2011. – № 8(2011). – С. 72–75.
8. Яковлев А. А. *Раскрутка и продвижение сайтов: основы, секреты, трюки* / А. А. Яковлев. – СПб.: БХВ-Петербург, 2007. – 336 с.
9. Басюк Т. М. *Принципи побудови системи аналізу та просування інтернет ресурсів* / Т. М. Басюк // *Комп'ютерні науки та інформаційні технології*. – Львів : Нац. ун-т «Львівська політехніка». – 2012. – № 784. – С. 43–48.
10. Энж Э. *Искусство раскрутки сайтов* / Э. Энж. – СПб.: ВHV-СПб, 2011. – 592 с.