

## WEBOMETRICS РЕЙТИНГ УНІВЕРСИТЕТІВ СВІТУ ТА ШЛЯХИ ПОКРАЩЕННЯ РЕЙТИНГУ

Складання різноманітних рейтингів – надзвичайно популярна справа в усьому світі. Конкуренція зростає, інформації стає все більше, а стислий список “найкращих” дозволяє заощадити час, а відповідно й гроші. До того ж, як зазначає Міністерство освіти і науки України, рейтинги сприяють відкритості та прозорості вищої освіти. “Ранжування вищих навчальних закладів потрібно абітурієнтам та їх батькам для вибору Вищого Навчального Закладу (ВНЗ), адміністрації ВНЗ для ефективного адміністрування; роботодавцям для вибору якісної робочої сили; уряду та політикам для формування стабільної нормативно-правової бази...”.

Один з доволі відомих рейтингів запропонувала іспанська лабораторія Cybermetrics, яка належить до дослідницької групи CSIC – складової іспанського Міністерства Освіти цей рейтинг. "Webometrics Ranking of World Universities" (<http://www.webometrics.info>), який підтримано UNESCO [1, 1300]. Це рейтинг не сайтів, а як у адресному полі ВНЗ відображується наукова робота.

Одним із важливих компонентів рейтингу є вибір параметрів по яких оцінюється ВНЗ. Для аналізу взято 4 параметри:

- **Розмір (Size)** - число сторінок регенерованих з чотирьох пошукових ресурсів – Google.com, Yahoo.com, Live.com (msn.com), Exalead.com. Відповідно, матеріали сайту мають бути публічними і досяжними із зовнішніх джерел, включаючи досяжність для пошукових роботів із зазначених ресурсів. Тобто, сайти з внутрішньої мережі, ресурси FTP та бази

з авторизованим доступом або неправильно оформленим файлом robots.txt не враховуються;

- **Видимість (Visibility)** - число унікальних зовнішніх зв'язків (external links) на університетський домен. Видимість визначається за допомогою 4 пошуковиків: Google, Yahoo Search, Live Search та Exalead. Для кожного пошуковика результати нормалізуються. Для кожного домену максимальний та мінімальний результат відкидаються (для досягнення максимальної об'єктивності), результати, що залишилися додаються. Наприклад, у Вікіпедії (глобальній відкритій енциклопедії) є матеріал про наш університет і посилання на наш сайт до нього – маємо зовнішнє посилання. Чим частіше на сторонніх ресурсах вказують наші адреси – тим більше у нас зовнішніх зв'язків;
- **Цінні файли (Rich files)** - файли у форматах, які зазвичай використовують автори для представлення та поширення власних робіт. Враховуються файли MS Word doc PowerPoint ppt, Adobe Acrobat pdf, PostScript ps, а також деякі інші (дані отримують від Google). Автори рейтингу вважають, що велика кількість таких документів засвідчує, що на сайті зберігаються не тільки адміністративні звіти, але й наукові матеріали;
- **Scholar** – Google Scholar дозволяє оцінити кількість наукових матеріалів і їх цитованість для кожного ВНЗ. Ці результати враховують публікації, звіти та інші академічні матеріали.

Вага параметрів: **Розмір-0.2, Видимість-0.5, Цінні файли-0.15, Scholar-0.15**

Як видно зі змісту наведених параметрів за допомогою їх робиться спроба оцінити наукову діяльність академічної установи. Можна вказати на багато недоліків такої оцінки. Це і формалізм та просування програмного забезпечення певних комерційних компаній, і неможливість оцінити реальне значення результатів наукових досліджень, і пріоритетність англомовних публікацій та інше. Але ця оцінка охоплює важливі сторони діяльності наукових організацій і вимагає присутності цих організацій в InterNet. Треба також відзначити, що провідні світові ВНЗ займають перші місця.

Є цікава фраза в описі Webometrics. “Якщо Ви вважаєте, що отримали не гідне для Вас місце, то змініть мережеву політику”. Аналізуючи стан Web-сайтів деяких ВНЗ, до цих слів варто додати вимогу змінити ставлення до електронних засобів інформації.

Ціллю статті є надати рекомендації по структурі та складу Web-сайтів ВНЗ, організації їх супроводження з метою підняття їх Webometrics рейтингу.

Виникає досить актуальна для багатьох ВНЗ України задача не тільки увійти у 13000 ВНЗ світу, але й зайняти пристойне місце. Для цього слід визначити як організується представлення інформації на сайтах підрозділів ВНЗ і яка інформація на цих сайтах впливає на його рейтинг [3, 109].

**Роль Web-сайту ВНЗ** Треба чітко визначити роль і місце центрального (основного) сайту ВНЗ. При цьому зосередження всієї інформації на одному основному сайті і не доцільно, і для великого ВНЗ фактично не можливо. Обумовлено це тим, що супроводжувати такий сайт досить важко. Більшість потужних ВНЗ мають поле IP адрес і треба щоб воно було одне. Бо якщо Web-сайти підрозділів ВНЗ мають різні IP адреси, то вважати їх своїми може тільки ВНЗ, а для Webometrics вони такими не є. Зрозуміти це досить просто, бо аналіз виконують програмні роботи, яким задається певна адреса (адресне поле ВНЗ). Часто виникає ще одна помилка, коли робиться проста переадресація (має назву-alias) і при цьому всі гілки сайту повинні мати адресу у адресному полі ВНЗ.

Основний Web-сайт ВНЗ виконує наступні функції:

1. Виконує роль інформаційного порталу, надаючи доступ до всіх сайтів його підрозділів;
2. Відображує всі основні події ВНЗ (міжнародні конференції, семінари, зустрічі та інше);
3. Відображує все, що відноситься до рівня Ректорату;
4. Є основним елементом спілкування з абітурієнтами під час роботи приймальної комісії;

## 5. Координація роботи Web-сайтів підрозділів (допомога у розробці сайтів, розкручуванні, пошуку вразливостей та інше)

Розглянемо, яким же чином можливо покращити всі 4 показники, і, як наслідок, Web-рейтинг наукової, освітньої установи, як слід вибудовувати Інтернет-політику ВУЗу

### **Оформлення сторінок**

Після створення Web-сайтів і наповнення їх контентом звичайно йде процес їх розкручування, тоб-то комплекс заходів по збільшенню їх популярності.

Перш за все робиться правильне оформлення сторінки і для цього вводяться:

Title - заголовок сторінки;

Keywords - ключові слова сторінки;

Description - опис сторінки;

Robots - вказівки пошуковому роботу;

URL - адреса основного сайту

Всі сайти університету повинні містити якнайбільшу кількість публікацій різного характеру: рефератів, звітів, тез, анотацій, наукових статей, ін. Всі вони беруться до розрахунку при визначенні рейтингу. Кожна сторінка повинна бути доступна англійською, відсутність англійського варіанту суттєво знижує рейтинг наукової установи.

**Індексація** Всі Інтернет-сторінки Web-сайтів повинні бути проіндексовані в системах Yahoo, Open Directory Project, Live Search, Exalea, Google та Google Scholar. Завдяки цьому, всі сайти наукового закладу та всі публікації розміщені на них будуть виявлені та враховані при обчисленні параметрів Size, Rich Files та Scholar.

**Багатомовність сайтів** Сайти на теренах України не повинні бути одномовними. Без особливої втрати відвідувачів це можуть дозволити собі лише англійські країни. Обумовлено це тим, що українська мова (і російська також) для InterNet є мовами національних меншин. Крім цього

такі рекомендовані Webometrics пошуковими як Live Search, Exalead не реєструють не англomовні сайти.

**Видимість** Забезпечується не тільки індексацією та реєстрацією у основних пошуковиках, але й рекламою сайтів на конференціях, в публікаціях при відрядженнях, особливо закордонних та інше. Всі сайти підрозділів повинні мати переходи до основного сайту ВНЗ. Треба також пам'ятати, що робот не проходить через Flash та Java Script, які досить часто використовуються при побудові сайтів. Певна інформація повинна бути і при побудові Web-сайту в robots.txt Ще один суттєвий недолік роботів, що використовує Webometrics, це не глибоке проходження робота в глибину сайту. Це дуже суттєво при використанні CMS для побудови сайту, які роблять сайти 5-6 рівневі.

В зв'язку з тим, що PageRank має вагу 0.5 в оцінці Webometrics, розглянемо принаймні загальні питання його визначення.

PageRank - це метод Google для визначення «важливості» сторінки. Коли всі інші фактори, такі як тег Title і ключові слова враховані, Google використовує PageRank, щоб відкорегувати результати так, що більш «важливі» сайти піднімуться відповідно вгору на сторінці результатів пошуку користувача

Тобто, порядок ранжирування в Google працює наступним чином:

- Знайти всі сторінки, відповідні ключовим словам пошуку.
- Відранжувати відповідно «сторінковим факторам», таким, як ключові слова.
- Врахувати текст посилань на сторінки.
- Відкорегувати результати даними PageRank.

Теорія Google говорить, що якщо Сторінка А посилається на сторінку В, то Сторінка А вважає, що Сторінка В - важлива сторінка. Текст посилання не використовується в PageRank також впливає на важливість посилань на сторінку. Якщо на сторінку вказує багато важливих посилань, то її посилання на інші сторінки також стають більш важливими. Важливість кожного окремого фактора в алгоритмах пошукової системи залежить від якості

інформації, котру він забезпечує. Тому є сенс спочатку поглянути на цю якість. Проте, у PageRank є одна важлива якість. На нього важче впливати, ніж на будь-який інший фактор ранжування. Це означає, що у PageRank є можливість надати вам перевагу перед конкурентами, коли він використовується в комбінації з іншими прийомами оптимізації для пошукових машин.

**Особливості PageRank.** PageRank - це число, що характеризує виключно голосуючу здібність всіх вхідних посилань на сторінку і те, як сильно вони рекомендують цю сторінку. Кожна унікальна сторінка сайту, проіндексована Google, має вагу PageRank. Люди часто помиляються, думаючи про вагу сайту, яка насправді є вагою головної сторінки цього сайту. Внутрішні посилання сайту враховуються при розрахунку ваги PageRank для інших сторінок сайту. PageRank незалежний, він не приймає до уваги текст посилань .

**Цінні файли** Джерелом публікацій у вітчизняних ВНЗ є кафедри та наукові лабораторії і тому доцільно мати розділи “публікації” на першому чи другому рівнях. Звичайно це <“Наукова робота”>=<”Публікації”>. Доцільно публікації надавати у PDF і зважаючи на те, що багато підрозділів не мають ліцензованих Adobe Acrobat, можливе використання безкоштовних Creator. Надавати повнотекстові публікації треба не порушуючи авторських прав. Досвід Західних ВНЗ показує, що розміщуються на Web-сайтах повнотекстові документи , які видано раніше (звичайно на 2-3 роки), з новими публікаціями теж можна познайомитись, але сплачуючи гроші. Дуже часто PDF документ використовується як реклама певного видання. Надається анотація, а іноді і короткий зміст, а наприкінці – де і за скільки, наприклад, цю книгу можна придбати.

**Scholar** Індексція Google Scholar, завдяки якій формується до 15% рейтингу, накладає ще жорсткіші вимоги на формат файлів. Сканер Scholar розпізнає лише публікації, викладені у вигляді одного окремого файлу в форматі Adobe Acrobat (.pdf), стиснутий PDF (pdf.gz), PostScript (ps), стиснутий PostScript

(ps.gz). При цьому розпізнаються лише ті PDF-документи, що містять доступний для пошуку текст (Це можна перевірити стандартним пошуком Acrobat Reader). Ілюстрації всередині PDF- документів мають бути в форматі OCR (Optical Character Recognition), а не просто відскановані.

Доцільно час від часу перевіряти списки сайтів, що посилаються на сторінки наукового закладу. Часто такі сайти не є проіндексованими, а отже їх «голос» не враховується рейтингом. Їх варто індексувати самостійно.

**Висновки.** Слід відзначити, що без роботи з Web-майстрами та адмінами сайтів покращити рейтинг ВНЗ важко, а може й не можливо. Тому в НТУУ"КПІ" створено сайт <http://webometr.kpi.ua> де надано досить повну інформацію про Webometrics – рейтинг, та сформульовані вимоги до сайтів підрозділів. З більшістю Web-майстрів сайтів підрозділів встановлено зв'язок та проводиться консультативна робота пов'язана з вимогами Webometrics.

### Література

1. Aguillo, I. F.; Granadino, B.; Ortega, J. L.; Prieto, J. A. (2006). Scientific research activity and communication measured with cybermetric indicators. *Journal of the American Society for the Information Science and Technology*, 57(10): 1296 - 1302.
2. Wouters, P.; Reddy, C. & Aguillo, I. F. (2006). On the visibility of information on the Web: an exploratory experimental approach. *Research Evaluation*, 15(2):107-115.
3. Цурін О.П., Іванченко Р.М. Автоматизовані засоби підняття Webometrics рейтингу Web-сайтів наукових установ.-Труди конференції SAIT-09, м.Київ, 2009, 109 с.

### Анотація

В статті розглянуто питання особливостей Webometrics рейтингу та підняття рейтингу Вищих Навчальних Закладів. Надано конкретні рекомендації по побудові Web-сайту Університету, по роботі з Web-сторінками, по

підвищенню параметрів, що входять до рейтингу (Розмір, Видимість, Цінні файли, Scholar). Надано рекомендації по роботі з Web-сайтами університету.

**Ключові слова:** Webometrics, рейтинг, університет, Web-сайт, індексація, Розмір, Видимість, Цінні файли, Scholar

#### Аннотация

В статье рассмотрены вопросы особенностей Webometrics рейтинга и поднятия рейтинга Высших Учебных Заведений. Даны конкретные рекомендации по построению Web-сайта Университета, по работе с Web-страницами, по повышению параметров, входящих в рейтинг (Размер, Видимость, Ценные файлы, Scholar). Даны рекомендации по работе с Web-сайтами университета

**Ключевые слова:** Webometrics, рейтинг, университет, Web-сайт, индексация, Размер, Видимость, Ценные файлы, Scholar

#### Summary

The article features the question and raising the Webometrics Ranking Universities. Provided specific recommendations for building Web-site of the University, to work with Web-pages to improve the parameters included in the rating (size, visibility, rich files, Scholar). The recommendations for working with Web-Site of the University

**Key words:** Webometrics, ranking, university, Web-site, indexing, size, visibility, valuable files, Scholar