

Сергій НАЗАРОВЕЦЬ,

бібліотекар II категорії, НТБ НУ «Львівська політехніка»

**ВЕБОМЕТРИЧНЕ ДОСЛІДЖЕННЯ
УКРАЇНОМОВНОГО СЕГМЕНТА «ВІКІПЕДІЇ»**

Проведено порівняльний аналіз україномовного сегмента онлайн-енциклопедії «Вікіпедія» з іншими мовними сегментами. За використання вебометричного інструментарію дано кількісну та якісну оцінку україномовного сегмента «Вікіпедії».

Ключові слова: вебометрія, «Вікіпедія», Уанет.

«Вікіпедія» (Wikipedia) – онлайн-енциклопедія, побудована навколо ідеї, що енциклопедична стаття може бути додана будь-яким користувачем і відредагована іншим. Це радикальний експеримент у сфері довіри, що на практиці застосував для створення контенту афоризм Еріка Реймонда «У семи тисяч няньок – дитя в шоколаді» [9].

«Вікіпедію» започатковано 15 січня 2001 р. як англomовний проєкт для створення онлайн-енциклопедії. Організатором проєкту є американська компанія Wikimedia Foundation Inc. Метою проєкту стало створення повної, нейтральної, відкритої енциклопедії всіма мовами світу. Проєкт швидко набув популярності серед користувачів мережі, і пізніше з'явилися розділи «Вікіпедії» іншими мовами, у тому числі й українською.

Англomовний розділ є найбільшим за обсягом у «Вікіпедії» і однією з найбільших енциклопедій взагалі (рис. 1). Станом на квітень 2010 р. його обсяг становить вже понад 3,2 млн статей.

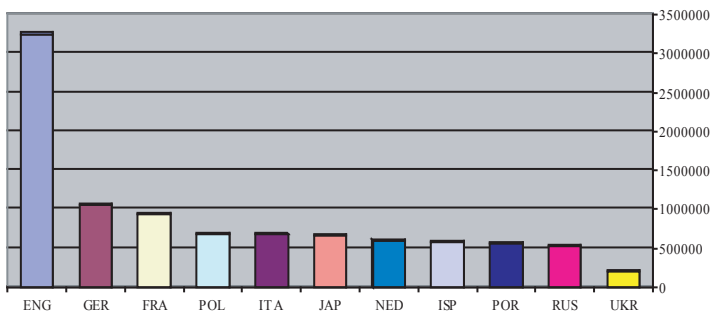


Рис.1. Кількість статей «Вікіпедії» різними мовами станом на 12.04.2010
(за даними «Вікіпедії»)

Другим за розміром є німецький розділ з більш ніж 1 млн статей. Усього ж є близько 270 розділів вікіпедій різними мовами, хоча лише 18 містять понад 150 000 статей, серед яких і український, який має 200 091 статтю [15].

Ставлення наукової спільноти до «Вікіпедії» неоднозначне. Хоча, на думку експертів, статті «Вікіпедії» в середньому містять «лише» на третину більше неточностей, ніж статті в енциклопедії Britannica [5], велика кількість професорів скаржилася на відсутність точності або повноти записів, а деякі з них рекомендували або намагалися заборонити студентам використовувати «Вікіпедію». Наприклад, викладачі історичного факультету Middlebury College забороняють студентам посилатися на «Вікіпедію» в наукових роботах, розповідають студентам про політику цього проекту і пояснюють, чому матеріали з «Вікіпедії» зручно використовувати, але, водночас, їм не можна довіряти [7].

Проте, як показують останні дослідження, проведені в шести різних коледжах США, студенти часто використовують «Вікіпедію» у своїх наукових розвідках. Більшість респондентів часто отримують довідкову інформацію з «Вікіпедії», проте рідше, ніж з інших джерел (таких, як лекції, чи пошуковець Google). «Вікіпедію» також часто використовують у комбінації з іншими інформаційними ресурсами. Матеріали з «Вікіпедії» відповідають потребам студентів, оскільки вони зручні й зрозумілі у використанні [1].

Останнім часом помітно зросло зацікавлення до україномовного сегмента «Вікіпедії» з боку українських науковців. Екс-міністр освіти та науки України І. Вакарчук офіційно закликав науковців до творення української «Вікіпедії»: «Український сегмент енциклопедії може істотно розширитися, якщо кожен з українських науковців напише хоча б по дві-три статті з його напрямку діяльності. До роботи можна залучати й студентів старших курсів, які під керівництвом своїх викладачів змогли б долучитися до створення потужної української “Вікіпедії”» [16].

Схожі переконання висловлював академік НАН України В. Грінченко, який назвав процес формування україномовного сегмента «Вікіпедії» завданням загальнонаціонального значення [19]. Групою українських дослідників було навіть зроблено припущення, що україномовна «Вікіпедія» вже увійшла в таку фазу розвитку, що подальше наповнення лише ресурсами ентузіастів видається малоперспективним. Тому саме великі наукові установи (академічні інститути та ВНЗ) повинні якісно й кількісно покращити україномовний сегмент «Вікіпедії» [18].

Ці висловлювання мають істотний недолік: «Вікіпедію» розглядають як онлайн-енциклопедію, звертаючи увагу лише на загальну кількість статей. Проте «Вікіпедія» – це передусім інтернет-проект, і видається доречним поглянути на «Вікіпедію» також з точки зору вебметричних показників.

Мета даного дослідження – за допомогою вебметричного інструментарію провести порівняльний аналіз україномовного сегмента з іншими мовними сегментами «Вікіпедії», дати кількісну та якісну оцінку україномовного сегмента «Вікіпедії», спрогнозувати шляхи розвитку цього проекту в майбутньому.

Зі стрімким зростанням World Wide Web (WWW), найбільшим сховищем інформації в електронному вигляді, перед ученими постало питання – чи зможе людство дослідити те, що воно створило? Саме вебметрія намагається виміряти World Wide Web, щоб отримати знання про кількість і типи гіперпосилань, структуру World Wide Web і його використання.

Згідно з Л. Бйорнеборном і П. Інгерсеном (Björneborn & Ingwersen), завданням вебметрії є вивчення кількості створення та використання інформаційних ресурсів, конструкцій та технологій за допомогою бібліометричного та інформометричного підходів [3]. Сам термін «вебметрія» був уперше введений Т. Алміндом (Almind) та П. Інгерсеном (Ingwersen) у 1997 р. [2].

Друге визначення вебметрії було введено М. Телволом (Thelwall), як вивчення веб-вмісту за допомогою кількісних методів соціальних досліджень, які не належать до однієї галузі знань [11]. Метою цього альтернативного визначення було ввести додаткові методи дослідження WWW, а не для заміни первісної дефініції.

WWW в основному складається з великої кількості веб-сторінок з посиланнями одна на одну [6]. Тому, окрім загальної кількості сторінок мовних сегментів «Вікіпедії», важливим показником для даного дослідження є також загальна кількість зовнішніх посилань на ці сторінки з інших електронних ресурсів.

Неформальний лідер веб-стандартів, World Wide Web Consortium, не накладає ніяких правил на створення гіперпосилань [13]. У певному сенсі, веб-посилання між сайтами або сторінками аналогічні до посилань в опублікованих роботах. Натхненний цією аналогією, Маккірнан (McKiernan) творчо запропонував уживати термін «sitiation» [8].

Однак, порівняно з цитатами, посилання між веб-сайтами радикально відрізняються. Водночас як, приміром, журнальний статті може бути відмовлено в публікації через невідповідності до певних вимог, веб-посилання потребує для створення набагато менш формальних моментів, і часто веб-цитування залежить від власних уподобань автора. Цю думку підтверджує дослідження Телвола, яке доводить, що велика кількість зовнішніх посилань на електронний ресурс, як правило, не пов'язана з якістю цього ресурсу [12].

Посилання можна розглядати як природний побічний продукт наукової діяльності, але, імовірно, що в деяких випадках і як абсолютно необхідний. Група науковців, яка досліджувала британські веб-сайти

університетів, встановила, що низький рівень посилань може вказувати на недоліки в побудові електронного ресурсу, а не на рівень наукової діяльності в навчальному закладі [14]. Причиною цього може бути поганий дизайн веб-сайту або слабка підтримка ресурсу.

Проте у випадку з різними мовними сегментами «Вікіпедії», такі поняття як дизайн та підтримка відіграють не таку істотну роль, як у випадку досліджень веб-сторінок різних академічних установ. Тому кількість посилань на той чи інший мовний сегмент може нам дати інформацію про його популярність у мережі.

Для визначення загальної кількості сторінок та зовнішніх посилань було використано три пошукові машини: Yahoo!, Exalead та Alta Vista, оскільки вони дають можливість отримати необхідні показники для даного дослідження і при цьому індексують достатньо велику кількість сторінок.

Сервіс Site Explorer (siteexplorer.search.yahoo.com) дає змогу визначити кількість веб-сторінок, проіндексованих пошуковою системою Yahoo!, і знайти веб-сторінки, які посилаються на заданий сайт або сторінку (див. табл. 1).

Таблиця 1

Кількість сторінок та посилань мовних сегментів «Вікіпедії»
(дані зібрані за допомогою Yahoo!)

URL	Кількість сторінок	Кількість посилань
en.wikipedia.org	30,263,975	147,005,857
fr.wikipedia.org	11,185,030	26,927,460
de.wikipedia.org	7,601,901	34,927,246
ja.wikipedia.org	5,622,951	63,997,597
it.wikipedia.org	5,707,238	20,501,797
es.wikipedia.org	4,827,462	22,857,967
pt.wikipedia.org	3,855,210	16,302,680
ru.wikipedia.org	3,660,483	16,824,968
nl.wikipedia.org	3,364,895	17,403,892
pl.wikipedia.org	3,227,038	16,720,482
uk.wikipedia.org	1,535,424	6,919,372

Французька пошукова система Exalead володіє не лише всіма якостями сучасного пошуковика, а й дозволяє подальшу обробку інформації на сторінці результатів за різними параметрами (мова, географічне положення, тип файлів, категорії), уточнення даних щодо контенту ресурсів, отриманих у результаті пошуку (аудіо, відео, RSS) (див. табл. 2).

AltaVista (altavista.com) – один із провідних постачальників пошукових служб і технологій. AltaVista продовжує покращувати пошук в

Інтернеті, вводячи нові функції і технології, роблячи процес пошуку більш зручним для споживачів. Цей пошуковик має багату історію інновацій, втілених у 61 патенті, пов'язаному з пошуком (див. табл. 3).

Таблиця 2

Кількість сторінок та посилань мовних сегментів «Вікіпедії»
(дані зібрані за допомогою Exalead)

URL	Кількість сторінок	Кількість посилань
en.wikipedia.org	6,768,665	8,482,271
fr.wikipedia.org	1,614,149	3,599,240
de.wikipedia.org	812,800	1,759,063
es.wikipedia.org	532,140	1,138,344
it.wikipedia.org	430,189	1,115,050
pt.wikipedia.org	91,944	839,990
nl.wikipedia.org	78,888	941,712
pl.wikipedia.org	69,185	866,181
ja.wikipedia.org	68,005	779,683
ru.wikipedia.org	60,108	770,098
uk.wikipedia.org	27,475	342,931

Таблиця 3

Кількість сторінок та посилань мовних сегментів «Вікіпедії»
(дані зібрані за допомогою Alta Vista)

URL	Кількість сторінок	Кількість посилань
en.wikipedia.org	29,800,000	90,500,000
fr.wikipedia.org	11,000,000	31,400,000
de.wikipedia.org	7,450,000	31,900,000
it.wikipedia.org	5,600,000	26,700,000
ja.wikipedia.org	5,490,000	33,400,000
es.wikipedia.org	4,750,000	29,500,000
pt.wikipedia.org	3,810,000	23,300,000
ru.wikipedia.org	3,590,000	22,800,000
nl.wikipedia.org	3,320,000	23,900,000
pl.wikipedia.org	3,190,000	23,600,000
uk.wikipedia.org	1,540,000	11,400,000

Пошукові системи, як правило, неповно охоплюють вміст ресурсів. Проте, як бачимо, показники кількості сторінок та посилань, отримані за допомогою пошукових машин, здебільшого підтверджують вищенаведене ранжування популярних мовних сегментів «Вікіпедії» на основі кількості статей.

Винятком є співвідношення німецькомовного й французькомовного розділів, оскільки всі пошуковики показують кращі показники для останнього, незважаючи на те що, за даними «Вікіпедії», німецькомовний розділ нараховує більшу кількість статей.

Слід також зауважити одноосібне лідерство японського мовного сегмента серед усіх азійських країн.

Треба зауважити, що при оцінюванні посилань слід розглядати не лише кількість посилань, а й досліджувати причини цих зв'язків [4]. Так, наприклад, у даному випадку необхідно зважати на загальну кількість носіїв тієї чи іншої мови у світі. Англomовний сегмент «Вікіпедії» розвивається найдовше, і це, безумовно, позитивно вплинуло на його кількість сторінок. Однак за кількістю посилань, а отже, і популярністю, він також залишається беззаперечним світовим лідером.

Україномовний сегмент «Вікіпедії» і за кількістю сторінок, і за кількістю посилань значно менший порівняно з показниками сегментів-лідерів. Однак, якщо підрахувати середню кількість посилань на сторінку україномовного сегмента, користуючись даними, отриманими за допомогою Yahoo!, то отримуємо 4,5 посилань. Для порівняння, стосовно англomовного сегмента ми отримуємо 4,9. Це, звісно, надто умовний показник для того, щоб робити далекосяжні висновки, проте він вказує на інтерес користувачів мережі україномовною «Вікіпедією».

Тому причини непотрапляння україномовного сегмента в десятку лідерів слід шукати не стільки в недостатній зацікавленості ймовірних авторів проекту «Вікіпедія», скільки в білінгвізмі українського населення і у проблемах Уанету загалом [17].

У вже згадуваному дослідженні «Як сьогоднішні студенти коледжів використовують “Вікіпедію” у своїх дослідженнях» говориться, що студенти часто розпочинають досліджувати предмет саме з «Вікіпедії», звертаючи увагу в основному на бібліографію до статті [1]. При цьому студенти дедалі менше звертають увагу на авторитет автора.

Тому написання науковцем статті для «Вікіпедії» з додаванням посилань на свої праці в списку літератури, цілком імовірно, може підняти рівень цитованості автора. Наукометричні показники серед українських авторів, які працюють в Україні, поки що не є надто популярними. Однак уже в близькому майбутньому стаття в україномовному сегменті «Вікіпедії» цілком може стати для українських науковців одним із способів промоції своїх наукових праць.

Дана стаття може слугувати матеріалом для майбутніх вебометричних досліджень. Також було б цікаво провести ревізію мовних розділів «Вікіпедії», використовуючи більшу кількість пошукових машин, а також прослідкувати зміну наповнення та динаміку зростання цих розділів із плином часу.

Вебометричні показники є лише альтернативою чи доповненням до іншого наукового інструментарію. Крім того, кількість сторінок та посилань – це не єдині показники, якими можна вимірювати якість *www-ресурсів*. Для прикладу, таким показником може бути технологія PageRank від Google [10]. Наведені ранжування є такими ж умовними, як і будь-які інші ранжування, і тому не можуть слугувати остаточним критерієм якості.

Список використаних джерел

1. *Alison J. Head*. Eisenberg How today's college students use Wikipedia for course-related research [Electronic resource] / Head J. Alison, B. Michael. – Mode of access: URL : <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2830>. – Title from the screen.

2. *Almind T*. Informetric analyses on the World Wide Web: Methodological approaches to “webometrics” / T. Almind, P. Ingwersen // *Journal of Documentation*. – 1997. – № 53 (4). – P. 404–426.

3. *Björneborn L*. Toward a basic framework for webometrics [Electronic resource] / L. Björneborn, P. Ingwersen // *Journal of the American Society for Information Science and Technology*. – 2004. – № 55 (14). – P. 1217. – Mode of access: URL : <http://www.db.dk/binaries/PerspectivesWebometrics-Jasist.pdf>. – Title from the screen.

4. *Chu Heting*. Taxonomy of inlinked Web entities: What does it imply for webometric research? / Heting Chu // *Library & Information Science Research*. – 2005. – № 27. – P. 8–27.

5. *Nicholson P. J*. The changing role of intellectual authority [Electronic resource] / P. J. Nicholson. – Mode of access: URL : <http://www.arl.org/arldocs/resources/pubs/mmproceedings/148/nicholson.pdf>. – Title from the screen.

6. *Ingwersen P*. The calculation of web impact factors [Electronic resource] / P. Ingwersen // *Journal of Documentation*. – 1998. – № 54 (2). – P. 236–243.

7. *Jaschik S*. A Stand Against Wikipedia [Electronic resource] / S. Jaschik. – Mode of access: URL : <http://www.insidehighered.com/news/2007/01/26/wiki>. – Title from the screen.

8. *Mckiernan G*. CitedSites(sm): Citation Indexing of Web resources [Electronic resource] / G. Mckiernan. – Mode of access: URL : <http://www.public.iastate.edu/~CYBERSTACKS/Cited.htm>. – Title from the screen.

9. *O'Reilly T*. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software [Electronic resource] / T. O'Reilly. – Mode

of access: URL : <http://oreilly.com/web2/archive/what-is-web-20.html>. – Title from the screen.

10. *Page L.* The pagerank citation ranking: Bringing order to the web [Electronic resource] / L. Page, S. Brin, R. Motwani, T. Winograd. – Mode of access: URL : <http://en.scientificcommons.org/42893894>. – Title from the screen.

11. *Thelwall M.* Introduction to Webometrics: Quantitative Web Research for the Social Sciences [Electronic resource] / M. Thelwall. – Mode of access: URL : <http://www.morganclaypool.com/doi/abs/10.2200/S00176ED1V01Y200903ICR004>. – Title from the screen.

12. *Thelwall M.* Methodologies for crawler based Web surveys. Internet Research – Electronic Networking Applications and Policy [Electronic resource] / M. Thelwall. – Mode of access: URL : http://wlv.openrepository.com/wlv/bitstream/2436/4012/1/2002_%20Methodologies%20for%20Crawler%20Based%20Web%20Surveys%20IRENA.pdf. – Title from the screen.

13. W3C (2003). About the World Wide Web Consortium (W3C). Cambridge, MA: World Wide Web Consortium [Electronic resource]. – Mode of access: URL : <http://www.w3.org/Consortium/>. – Title from the screen.

14. *Wilkinson D.* Motivations for academic web site interlinking: evidence for the Web as a novel source of information on informal scholarly communication [Electronic resource] / D. Wilkinson, G. Harries, M. Thelwall, L. Price. – Mode of access: URL : <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.83.2571&rep=rep1&type=pdf>. – Title from the screen.

15. Вікіпедія. Матеріал з Вікіпедії – вільної енциклопедії [Електронний ресурс]. – Режим доступу: URL : <http://uk.wikipedia.org/wiki/Вікіпедія>. – Назва з екрана.

16. Іван Вакарчук закликав до творення української Вікіпедії. – Пресслужба МОН [Електронний ресурс]. – Режим доступу: URL : http://www.mon.gov.ua/main.php?query=newstmp/2010_1/09_03/. – Назва з екрана.

17. *Пелешишин А. М.* Веб 2.0 – другий шанс для Уанету [Електронний ресурс] / А. М. Пелешишин. – Режим доступу: URL : www.ntsh.org/ua/web2. – Назва з екрана.

18. *Пелешишин А. М.* Формування суспільного авторитету ВНЗ шляхом комплексного подання в системі Вікіпедія / А. М. Пелешишин, Ю. Й. Пероганич, О. В. Марковець, Н. О. Думанський // Інформаційні системи та мережі № 653. – Л. : Вид-во Нац. ун-ту «Львів. політехніка», 2009. – 187 с.

19. *Переяславець М.* Учені все ще сподіваються, що їх почують / М. Переяславець // Дзеркало тижня. – 2010. – № 6.