

В.В. ВОЙТКО, Д.В. МЕЛЬНИК, С.А. ЯРЕМКО

Вінницький національний технічний університет
Вінницький торговельно-економічний інститут Київського національного торговельно-економічного університету

ПОРІВНЯЛЬНИЙ АНАЛІЗ ПОШУКОВИХ СИСТЕМ ЯК СЕРВІСІВ СУЧАСНОГО ДОКУМЕНТООБІГУ НАУКОВИХ ТА НАВЧАЛЬНИХ ОРГАНІЗАЦІЙ

Розглянуто популярні пошукові системи як сервіси сучасного документообігу наукових та навчальних установ, розроблено систему критеріїв оцінювання ефективності використання професійно орієнтованих пошукових систем, проведено порівняльний аналіз пошукових систем, визначено переваги та недоліки їх використання.

Рассмотрены популярные поисковые системы как сервисы современного документооборота научных и учебных учреждений, разработана система критериев оценивания эффективности использования профессионально ориентированных поисковых систем, проведен сравнительный анализ поисковых систем, определены преимущества и недостатки их использования.

Ключові слова: спеціалізований пошук, інформаційно-пошукова система, пошуковий запит.

V.V. VOJTKO, D.V.MEL'NYK, S.A. YAREMKO

Vinnitsya national technical university
Vinnitsya trading-economy institute Kyiv national trading-economy university

COMPARING ANALYSIS OF SEARCH SYSTEMS AS A MODERN DOCUMENT SERVICES OF SCIENTIFIC AND EDUCATIONAL ORGANIZATIONS

The task of finding data in fast evolving conditions of information technology is one of the most important in modern information theory and makes the fundamental value in data computing processes. Traditional methods of finding information in global network are becoming less and less effective because of fast evolving number of information resources and because of that the creation of new methods and instruments of search are needed that makes resolving finding information tasks faster and more effective.

Results of the comparing analysis of the modern information searching systems is saying about the high importance of development and usage of the effective specialized search systems, oriented on the identified user requests in the limited professional information environment.

Analyzed popular search systems as a modern document services of scientific and educational organizations, developed the criteria grade system for the usage effectiveness of professional oriented search systems, was done a comparing analysis of search systems, showed the advantages and disadvantages of their usage.

Keywords: specialized search, informational search system, search request.

Вступ

У наш час, в умовах стрімкого розвитку інформаційних технологій, постійного збільшення обсягу інформаційних ресурсів, постають задачі швидкого та коректного пошуку потрібної інформації [1]. Порівняльний аналіз сучасних пошукових систем дозволяє визначити їх можливості, оцінити базові характеристики та спрогнозувати перспективність їх використання в автоматизованих системах ведення документообігу організацій освітянського простору.

Метою роботи є підвищення ефективності пошукових процесів у глобальній мережі.

Об'єктом дослідження постають пошукові системи.

Під предметом дослідження розуміємо методи та засоби реалізації мережевих пошукових процесів.

Головними задачами вбачаємо порівняльний аналіз пошукових систем як сервісів сучасного документообігу наукових і навчальних організацій.

Аналіз сучасних реалізацій пошукових систем

У сучасному інформаційному просторі розроблено і використовується велика кількість пошукових систем та сервісів. Значна кількість пошуковиків сьогодні вже стала «ветеранами» глобальної мережі Інтернет, інші системи лише зовсім недавно розпочали свою роботу. Суттєвою в них є різниця якості пошукових процесів, розміри бази Web-ресурсів тощо. Тому для розробки спеціалізованої пошукової системи, яка базується на пошуку інформації в базі наукової чи навчальної літератури, потрібно провести огляд існуючих пошукових систем, здійснити їх порівняльний аналіз, визначити переваги та недоліки їх спеціалізованого використання.

Одним із найпопулярніших сучасних пошуковиків є Google. У його індексі зберігається більше 3 мільярдів документів [1]. Новаторська технологія PageRank дозволила проведення якісного пошуку із забезпеченням високої релевантності результатів. Крім того, важливою особливістю Google, на відміну від інших пошукових систем, є пошуковий павук Google (або Googlebot), якому можна вказати на особливості індексування ресурсу (наприклад, вимогу пропустити індексацію деяких сторінок чи замовити їх обов'язкове індексування). Використання такої можливості дозволяє користувачу переглянути знайдену сторінку безпосередньо в кеші Google. З іншого боку, коли часто оновлюється інформаційний ресурс, індексування сторінок краще обмежити, щоб відвідувачі завжди отримували тільки свіжу інформацію з сайту [1].

Технологія PageRank ґрунтується на кількості та якості опрацювання зовнішніх посилань на сайт. Саме від неї, в першу чергу, залежить релевантність сторінки. Крім PageRank на релевантність сторінки впливає ім'я домену, текст заголовка сторінки, текст посилань з інших серверів на сайт, вміст атрибуту ALT; наявність ключових слів у заголовках <h2>, візуальне виділення ключових слів (, розмір, колір), щільність ключових слів на сторінці, тобто частота їх появи в документі [1].

Слід окремо відмітити простий, зручний і наочний інтерфейс пошукової системи Google, який не містить зайвої інформації.

Недоліком пошукової системи, на нашу думку, є розміщення поряд з релевантними результатами запиту частини малорелевантних посилань, презентованих даними з відповідними ключовими словами, що відносяться до різноманітних тематичних категорій (рис. 1). Ще одним недоліком є розміщення поряд з результатами запиту рекламних посилань, які часто непов'язані безпосередньо з запитом користувача.

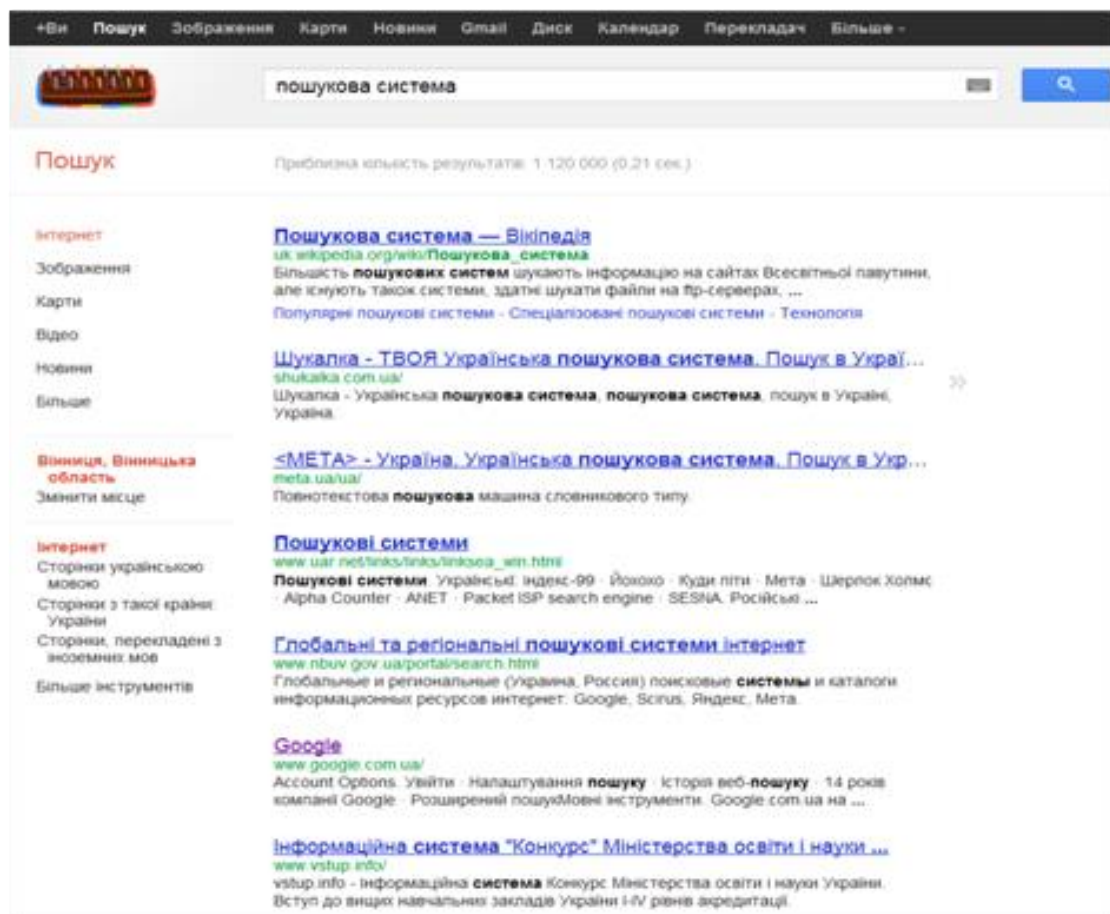


Рисунок 1 – Результати виконання пошукового процесу системою Google

Для уточнення результатів у пошуковий запит вводять додаткові ключові слова, які розкривають тему запиту. Крім того, покращення результатів можливе за рахунок розширеного пошуку чи ідентифікованого вибору певної категорії даних.

Іншою популярною пошуковою системою є Rambler – одна з перших повноцінних пошукових систем російськомовного Інтернету. Вирізняльною особливістю Rambler завжди були грамотні й чіткі пошукові алгоритми, орієнтовані на реалізацію релевантного пошуку за запитом користувачів. Базовою характеристикою пошукової системи є унікальний механізм асоціацій, який у результаті пошуку дозволяє отримати не лише класичний список посилань, розташованих у порядку спадання релевантності, а й рядок зі списком запитів,

тематично схожих з конкретним запитом користувача. Згаданий механізм також здатний виправляти синтаксичні та орфографічні помилки запитів користувачів та перекладати слова, введені з невідповідного регістру.

Сьогодні Rambler являє собою унікальну просту та ефективну пошукову систему, що обробляє близько 20% пошукових запитів російськомовного Інтернету і позиціонує себе більшою мірою як великий каталог сайтів [1].

Задавши аналогічний попередньому прикладу користувацький запит, за отриманими результатами спостерігаємо близькість пошукових операцій Rambler та Google. Деякі отримані посилання виявилися однаковими для обох пошукових систем, що свідчить про схожість алгоритмів пошуку та сортування результатів (рис. 2).

Серед недоліків Rambler можна відмітити загроможденість інтерфейсу різноформатною інформацією на початковій сторінці пошукової системи. Розширений пошук у Rambler існує, але є досить обмеженим і дозволяє виконувати лише базові уточнення. Аналогічно до системи Google, Rambler дозволяє проводити пошук за визначеними категоріями даних: новинами, картами, зображеннями, відео та ін.

На противагу пошуковим системам широкого призначення розглянемо спеціалізовану пошукову систему Scirus. Scirus – це комплексна науково-орієнтована пошукова система, що надає інформацію спеціалізованого пошуку, а також допомагає в складанні запитів і використанні визначень, дозволяє зберігати та експортувати інформацію про отримані результати. Scirus знаходить інформацію, яка може бути недосяжна у разі використання інших пошукових систем. В індексі голландського пошукового сервісу Scirus знаходиться більше 460 млн. одиниць наукової інформації [3].

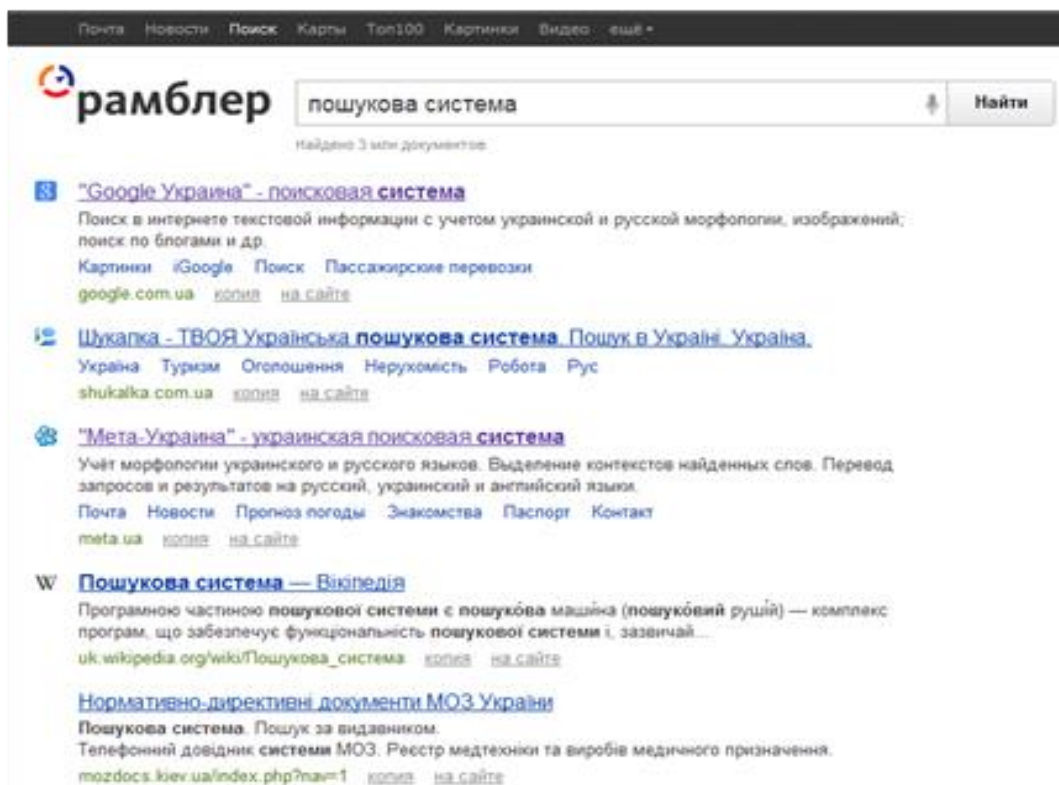


Рисунок 2 – Результати виконання запиту пошуковою системою Rambler

Scirus має зручний, наочний і зрозумілий інтерфейс, який дозволяє встановлювати індивідуальні опції. Крім того, Scirus дозволяє виконувати розширений пошук, який потребує уточнення типів шуканих даних за розміщенням (стаття, книжка, патент, сторінка вченого тощо) чи форматом файлу (PDF, DOC і т.д.) (рис. 3).

Хоча Scirus забезпечує один з кращих науково-орієнтованих пошукових сервісів і значно перевершує продуктивність пошукових систем широкого напрямку, але все ж загальноприйнятий підхід з частковим пошуком даних інколи призводить до отримання некоректних пошукових результатів особливо за однокритеріальним запитом (наприклад, пошук за заголовком чи ключовим словом) без врахування змісту основного тексту матеріалу. Так за пошуковим запитом «Пошукова система» Scirus було знайдено більшість сторінок нерелевантного характеру, оскільки при скриптовій генерації одного з веб-ресурсів у заголовок сторінок була записана відповідна фраза (рис. 4). Іншим недоліком Scirus можна вказати недостатню кількість матеріалів українською чи російською мовами, що робить використання даного пошукового ресурсу малоефективним для користувачів, що проводять пошук вказаними мовами.

Advanced search

All of the words in The complete document

AND

All of the words in The complete document

Search tips
author:smith Find results that have "smith" in the author field
DNA -sequencing Find results that have "DNA" but not "sequencing" in the text
car* Finds "car" as well as "carbon", etc. [View all search tips](#)

Dates	Only show results published between before 1900 and 2013		
Information types	Only show results that are <input checked="" type="checkbox"/> Any information type <input type="checkbox"/> Abstracts <input type="checkbox"/> Articles <input type="checkbox"/> Articles in Press <input type="checkbox"/> Books <input type="checkbox"/> Conferences <input type="checkbox"/> Patents <input type="checkbox"/> Preprints <input type="checkbox"/> Reviews <input type="checkbox"/> Scientist homepages <input type="checkbox"/> Theses and Dissertations		
File formats	Only show results that are <input checked="" type="checkbox"/> Any format <input type="checkbox"/> PDF <input type="button" value="List more file types"/> <input type="checkbox"/> HTML <input type="checkbox"/> Word		
Content sources	Only show results from <table border="1"> <tr> <td> Journal sources <input checked="" type="checkbox"/> All <input type="checkbox"/> American Physical Society <input type="checkbox"/> BioMed Central <input type="checkbox"/> Crystallography Journals Online <input type="checkbox"/> Hindawi Publishing Corporation <input type="checkbox"/> IOP Publishing <input type="button" value="List more sources"/> </td> <td> Preferred Web sources <input checked="" type="checkbox"/> All <input type="checkbox"/> E-Print ArXiv <input type="checkbox"/> Caltech <input type="checkbox"/> CogPrints <input type="checkbox"/> Curator <input type="checkbox"/> Digital Archives <input type="button" value="List more sources"/> </td> </tr> </table> <input checked="" type="checkbox"/> The rest of the scientific web	Journal sources <input checked="" type="checkbox"/> All <input type="checkbox"/> American Physical Society <input type="checkbox"/> BioMed Central <input type="checkbox"/> Crystallography Journals Online <input type="checkbox"/> Hindawi Publishing Corporation <input type="checkbox"/> IOP Publishing <input type="button" value="List more sources"/>	Preferred Web sources <input checked="" type="checkbox"/> All <input type="checkbox"/> E-Print ArXiv <input type="checkbox"/> Caltech <input type="checkbox"/> CogPrints <input type="checkbox"/> Curator <input type="checkbox"/> Digital Archives <input type="button" value="List more sources"/>
Journal sources <input checked="" type="checkbox"/> All <input type="checkbox"/> American Physical Society <input type="checkbox"/> BioMed Central <input type="checkbox"/> Crystallography Journals Online <input type="checkbox"/> Hindawi Publishing Corporation <input type="checkbox"/> IOP Publishing <input type="button" value="List more sources"/>	Preferred Web sources <input checked="" type="checkbox"/> All <input type="checkbox"/> E-Print ArXiv <input type="checkbox"/> Caltech <input type="checkbox"/> CogPrints <input type="checkbox"/> Curator <input type="checkbox"/> Digital Archives <input type="button" value="List more sources"/>		
Subject areas	Only show results in <input checked="" type="checkbox"/> All subject areas <input type="checkbox"/> Agricultural and Biological Sciences <input type="checkbox"/> Astronomy <input type="checkbox"/> Chemistry and Chemical Engineering <input type="checkbox"/> Computer Science <input type="checkbox"/> Earth and Planetary Sciences <input type="checkbox"/> Economics, Business and Management <input type="checkbox"/> Engineering, Energy and Technology <input type="checkbox"/> Environmental Sciences <input type="checkbox"/> Languages and Linguistics <input type="button" value="List more subject areas"/>		

Рисунок 3 – Сторінка розширених налаштувань системи Scirus

- 16. [Портал НГТУ | Поисковая система | Телефон](#) [37К]
 May 2012
 НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ Портал НГТУ Главная (Поиск) **Поисковая система** -- Телефон Быстрый доступ Все факультеты Все кафедры Все сотрудники Главная (Поиск) Поиск по ключевым словам Поиск: Указать категорию...
[\[http://www.nstu.ru/phone/object?requestObject=14612\]](http://www.nstu.ru/phone/object?requestObject=14612)
- 17. [Портал НГТУ | Поисковая система | Кафедра](#) [37К]
 Jan 2012
 НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ Портал НГТУ Главная (Поиск) **Поисковая система** -- Кафедра Быстрый доступ Все факультеты Все кафедры Все сотрудники Главная (Поиск) Поиск по ключевым словам Поиск: Указать категорию...
[\[http://www.nstu.ru/phone/chr?requestChair=8488\]](http://www.nstu.ru/phone/chr?requestChair=8488)
- 18. [Портал НГТУ | Поисковая система | Телефон](#) [31К]
 May 2012
 НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ Портал НГТУ Главная (Поиск) **Поисковая система** -- Телефон Быстрый доступ Все факультеты Все кафедры Все сотрудники Главная (Поиск) Поиск по ключевым словам Поиск: Указать категорию...
[\[http://www.nstu.ru/phone/object?requestObject=9356\]](http://www.nstu.ru/phone/object?requestObject=9356)

Рисунок 4 – Результати виконання запиту системою Scirus

BASE (Bielefeld Academic Search Engine) – науковий інтернет-пошуковик. Сервіс Base належить німецькому університету і являє собою пошукову систему міжнародного проекту Open Archives Initiative та співпрацює з європейським науково-інформаційним проектом DRIVER. Перший із названих проектів дозволяє власникам електронних бібліотек обмінюватися даними та забезпечує відкритий доступ до наукової інформації. Проект DRIVER, у свою чергу, створив зведений ресурс, об'єднуючи європейські електронні бібліотеки, котрі спеціалізуються на зберіганні наукової та навчальної інформації. Проект BASE став зручним засобом доступу до вільно поширюваної наукової інформації. Крім названих джерел, BASE звертається і до власної індексної бази тематичних веб-сайтів [4].

BASE пропонує режими простого і розширеного пошуку. У режимі простого пошуку для уточнення запиту доступна лише опція перемикавання пошукової сфери між ресурсами BASE і базою Google Scholar через інтерфейс BASE. У режимі розширеного пошуку знаходяться джерела за конкретними полями опису ресурсів (автор, заголовок, тема), а також визначається географічний регіон і рік видання публікацій. Крім того, забезпечується можливість визначення типу шуканого контенту, підтримується пошук статей, книг, текстів лекцій, тез доповідей тощо. У систему інтегрований тезаурус BASE Eurovoc, тому при обробці запиту цілком реально отримати релевантні результати пошуку відразу кількома мовами. Крім того, система характеризується наявністю зручного інтерфейсу (рис. 5).

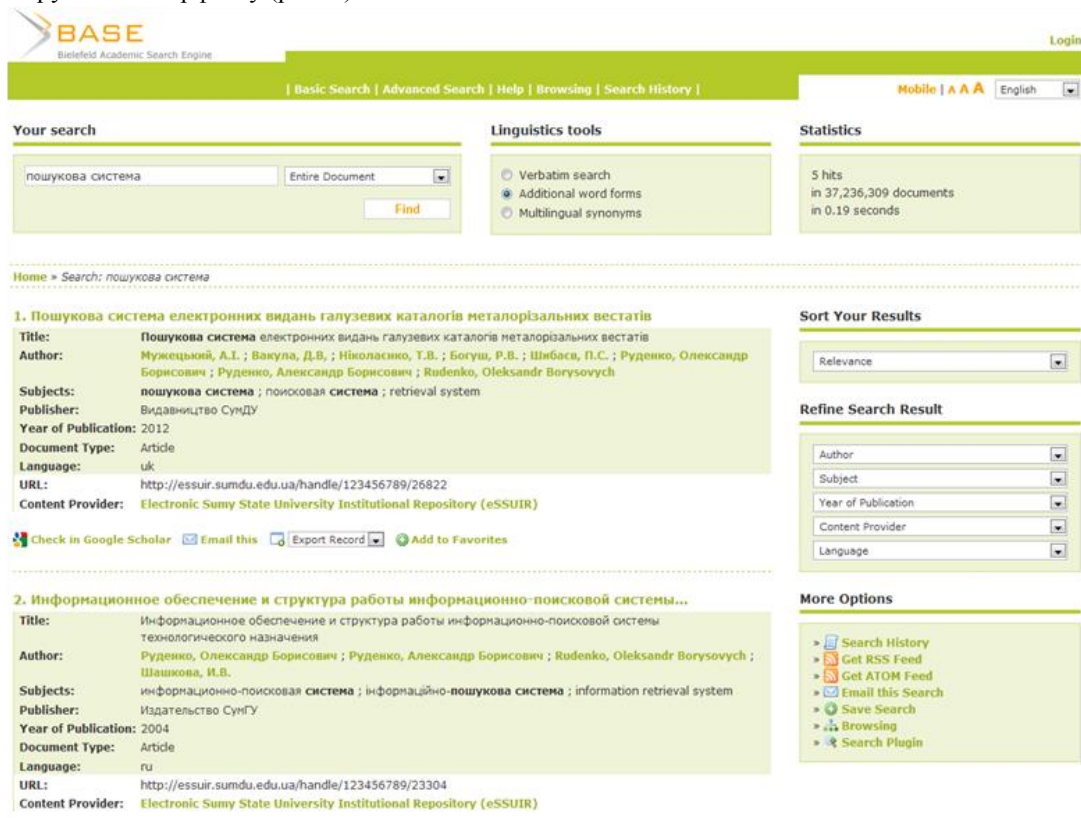


Рисунок 5 – Результати виконання запиту системою BASE

Порівняльна характеристика сучасних пошукових систем

Проаналізуємо та порівняємо характеристики розглянутих пошукових систем з метою виявлення їхніх переваг та недоліків.

Наведемо систему критеріїв для порівняння пошукових систем:

- розширений пошук – дозволяє встановлювати налаштування логічних характеристик пошуку даних, використовувати додаткові критерії та параметри фільтрування результатів;
- спеціалізований пошук – визначає можливість виконання спеціалізованого пошуку наукової літератури;
- повнота пошуку – враховує обсяг даних, включених в індекс пошукової системи, за якими безпосередньо здійснюється пошук;
- релевантність пошуку наукової літератури – визначає, наскільки відповідають результати пошуку запиту наукової тематики;
- індивідуальні налаштування – дозволяють здійснювати індивідуальні налаштування режимів відображення результатів пошуку за ідентифікованими потребами користувача (кількістю презентованих пошукових результатів на сторінці, способу відображення та перегляду, збереження результатів в особистий профіль і т.п.);
- зручність інтерфейсу і простота використання – показує, наскільки інтерфейс пошукової системи є зручним та наочним, а також вказує на простоту використання програми недосвідченими користувачами.

Результати порівняльного аналізу характеристик розглянутих пошукових систем зведемо в табл.1, де:

- + – критерій виконується;
- – критерій не виконується;
- +/- – критерій виконується частково.

Таблиця 1 – Порівняння характеристик сучасних пошукових систем

Назва інформаційно-пошукової системи	Розширений пошук	Спеціалізований пошук	Повнога пошуку	Релевантність пошуку наукової літератури	Індивідуальні налаштування	Зручність інтерфейсу і простота використання
Google	+/-	+/-	+/-	+/-	+/-	+
Rambler	+/-	-	+/-	+/-	-	+/-
Scirus	+	+	+/-	+	+	+/-
BASE	+/-	+	+/-	+	+	+

З таблиці 1 видно, що з урахуванням усіх критеріїв, спеціалізовані пошукові системи показують кращий результат у порівнянні з традиційними системами широкоформатного пошуку.

Висновки

Задача пошуку даних в умовах швидкого розвитку сучасних інформаційних технологій займає одне з центральних місць у сучасній теорії інформації і має фундаментальне значення в процесах обробки цифрових даних. Оскільки традиційні підходи до пошуку інформації в глобальній мережі стають все менш ефективними через вибухове зростання кількості та об'єму інформаційних ресурсів, то постає необхідність у створенні методів та засобів пошуку, які дозволяють ефективніше та швидше розв'язувати поставлені задачі пошуку даних, враховуючи сучасні тенденції зміни інформаційного середовища. Результати проведеного порівняльного аналізу сучасних інформаційно-пошукових систем підтверджують важливість розробки та використання ефективних систем спеціалізованого пошуку даних, орієнтованих на ідентифіковані запити користувачів у професійно обмеженому інформаційному середовищі.

Література

1. Полянко, А. Особенности поисковой системы Google. / А. Полянко // [Електронний ресурс]: Site-seo.ru – поисковая оптимизация и интернет-маркетинг – Режим доступа: <http://site-seo.ru/info/seostati/GoogleMSNYa/40/>.
2. История поисковой системы Rambler или особенности Rambler [Електронний ресурс]: Seo-alabis.ru – Search engine optimization. Режим доступа: http://seo-alabis.ru/seo/rambler_istoriya.php.
3. Scirus – for scientific information only. [Електронний ресурс]: Scirus – for scientific information only. – Режим доступа: <http://www.scirus.com>.
4. Кутюренко, А. Научный интернет-поиск / А. Кутюренко // Мир ПК – 2010. – № 5. – С.64-67. – ISSN 0235-3520.

References

1. Polianko, A. Osobennosti poiskovoi sistemy Google. / A. Polianko // [Elektronnyi resurs]: Site-seo.ru – poiskovaya optimizatsiya i internet-marketing – Rezhym dostupu: <http://site-seo.ru/info/seostati/GoogleMSNYa/40/>.
2. Istoriya poiskovoi sistemy Rambler ili osobennosti Rambler [Elektronnyi resurs]: Seo-alabis.ru – Search engine optimization. Rezhym dostupu: http://seo-alabis.ru/seo/rambler_istoriya.php.
3. Scirus – for scientific information only. [Elektronnyi resurs]: Scirus – for scientific information only. – Rezhym dostupu: <http://www.scirus.com>.
4. Kutovenko, A. Nauchnyi internet-poisk / A. Kutovenko // Mir PK, 2010, № 5, pp.64-67. – ISSN 0235-3520.