

УДК 004.738.52

## ПРОБЛЕМИ ПОШУКУ ІНФОРМАЦІЇ В ГЛОБАЛЬНІЙ МЕРЕЖІ ІНТЕРНЕТ

Поташова А. В.

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Україна, Київ

*Розкриваються проблеми пошуку спеціалізованої інформації в інтернеті, пов'язані із її великою кількістю та слабкою структурованістю. Розглядається предметна онтологія, як сучасний засіб підвищення пертинентності результатів пошуку, шляхом структурування інформації та створення віртуальних каталогів документів.*

*Ключові слова: інформація, інформаційний пошук, проблеми пошуку інформації, спеціалізована інформація, розширений пошук, технологія пошуку, онтологія.*

*A. Potashova The problems of information search in the global Internet/ National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Ukraine, Kyiv*

*The article reveals the problems of searching for specialized information on the Internet, related to its quantity and weak structuredness. The subject ontology is considered as a modern means of increasing the pertinence of search results, by structuring information and creating virtual catalogs of documents.*

*Keywords: information, information search, information search problems, specialized information, advanced search, search technology, ontology.*

**Введення.** В Окінавській хартії глобального інформаційного

суспільства, яка була прийнята лідерами «вісімки» 22 липня 2000 року, інформаційно-комунікаційні технології (ІТ) виступають сьогодні як фактор, що впливає на формування суспільства ХХІ століття. Наступає новий етап у розвитку процесів обміну інформацією. Стрімкий розвиток і розповсюдження нових інформаційних і телекомунікаційних технологій набуває характер глобальної інформаційної революції, яка здійснює зростаючий вплив на політику, економіку, управління, фінанси, науку, культуру та інші сфери життєдіяльності суспільства у рамках національних кордонів і світу в цілому [1].

Питання автоматизації процесу збору й обробки інформації розкриваються у наукових працях вчених К. Н. Нарібаєва, В. І. Ісакова, М. Я. Розенберга, О. М. Островського. Проблемам використання інформаційних технологій у навчальній діяльності присвячені роботи Г. А. Бордовського, Р. С. Гуревича, А. М. Довгялло, І. Б. Горбунової, В. О. Ізвозчикова, М. Ю. Кадемії, В. І. Клочка, Л. Л. Коношевського, С. В. Панюкової, І. В. Роберт, А. В. Смірнова, В. І. Сумського та ін.

Згідно результатів досліджень, проведених Каліфорнійським університетом, понад 90% світової інформації зберігається в цифровій формі в комп'ютерах, таким чином займаючи менше місця, ніж на папері, у фільмах і відеозаписах. У цілому в мережі Інтернет зібрано близько 550 мільярдів документів, і щодня створюється ще 7,3 мільйони нових сторінок. За словами вчених університету, кожний житель в будь-якій точці земної кулі може мати доступ до інформації в мережі Інтернет, і абсолютно є очевидним ризик «потонути в цьому океані інформації» [2].

**Мета статті.** Показати необхідність підвищення ефективності пошукових засобів в мережі Інтернет, та запропонувати використання

спеціалізованих предметних онтологій, як одного із методів підвищення ефективності інформаційного пошуку.

**Виклад основного матеріалу.** Інформаційні потоки представляють собою сукупність повідомлень, які об'єктивно відображають розвиток процесів у господарській діяльності і передаються по каналам зв'язку для здійснення оперативного управління [3]. У сучасному світі характерним є постійне збільшення потоків інформації, що несе у собі суперечливість. Тобто, чим більший обсяг інформації, тим більше можливість використання корисної її частини для прийняття управлінського рішення. Сучасні технічні засоби: комп'ютери, мережа Інтернет тощо значно спрощують доступ до величезної кількості матеріалів, що знаходяться у вільному доступі. Причому ця величезна кількість постійно зростає, подвоюючись у середньому раз на 2 роки. Саме в такому зростанні і приховано «підводні камені»: знаходження потрібної інформації часто перетворюється у досить не просту проблему. При роботі із інформаційними потоками діє закон Парето – згідно статистичним дослідженням, якщо інформація збільшується вдвічі, її корисність становить не більш 20%, а 80%, які залишилися, не носять корисного характеру [4]. Виокремити ж навіть із корисної інформації необхідну стає дедалі складніше.

За останні два десятиліття в наше віртуальне життя впевнено увійшли пошукові засоби інформації, головними з яких стали так звані пошукові системи. Пошукові системи являють собою певний інструмент для пошуку інформації в мережі та її ранжування за запитом користувача. Найбільш популярними з яких на сьогодні є Google, Rambler, Meta та ін.

Основне призначення цього інструментарію збирати інформацію про документи (об'єкти) мережі, зберігати отримані дані та формувати

відповіді на запит користувача на основі аналізу збережених даних. Тобто створювати. Кожна з таких машин має свої переваги та недоліки, проте, спільним їх недоліком можна назвати те, що вони всі ... добре виконують свою роботу і знаходять багато потрібної інформації.

Означимо основні дефініції, пов'язані із пошуком інформації.

Поняття **інформаційного пошуку** вперше запровадив в інформатиці американський математик Келвін Муерс в 1947 році. Інформаційним пошуком (ІП) називається деяка послідовність операцій, яка виконується з метою відшукування документів, які містять певну інформацію (з подальшою видачею цих документів або їх копій), або з метою видачі фактичних даних, які надають відповіді на задані питання [5].

Основним засобом передачі інформації в часі і просторі є **документ**. Документ визначається як засіб закріплення будь-яким чином на спеціальному матеріалі будь-якої (деякої) інформації про факти, події, явища об'єктивної дійсності і розумової діяльності людини [6].

**Запит** зазвичай являє собою осмислену фразу або набір слів, що описують інформаційну потребу

**Результат пошуку** - список документів, які відібрані системою як потенційно містять корисну для користувача інформацію. Цей список, як правило, впорядкований по мірі зменшення якоїсь метрики, яку називають «вагою», «ступенем релевантності запиту» або оцінкою ймовірності того, що документ відповідає запиту.

Для прикладу приведемо запит на пошукову машину Google «Екологія підручник». Результатом такої дії стало знаходження 3 360 000 документів, що на думку пошукового робота Google мають відношення до створеного нами запиту. Користувачу пропонується

самостійно визначитись, які ж саме файли його цікавлять.

Те ж саме відбудеться і при пошуку за допомогою інших засобів. Певною і досить дієвою допомогою в такій ситуації є спеціалізовані пошукові машини. Порівняємо результат того ж запиту на пошуковій машині Google Академія (<http://scholar.google.com>) – безкоштовна спеціалізована пошукова система, що дозволяє здійснювати пошук академічної літератури. В цьому випадку вибір обмежується 18 700 web-сторінок.

Таким чином, можна зробити висновок, що пошук необхідної інформації із конкретних питань як мінімум займає у користувача певний час, а то й потребує спеціальних знань.

Необхідність вдосконалення процесу пошуку обумовлена наступними причинами:

- як правило, користувачі як інструмент пошуку використовують або пошукові машини (AltaVista, HotBot), або каталоги ресурсів (Yahoo). Незважаючи на ряд переваг, вищезгадані пошукові системи не забезпечують науковий контроль якості (тобто процес включення інформації в БД цих рейтингів не знаходиться під контролем фахівців-предметників), що може призвести до отримання в результаті пошуку недостовірної інформації;

- відсутність знань і навичок зі складання запиту до пошукової системи призводить до отримання великої кількості інформаційного шуму;

- користувач часто не усвідомлює той факт, що, незважаючи на те, що інформація існує в інтернеті, вона знаходиться поза полем зору пошукових машин. Відсутність елементарних знань з проведення пошуку призводить до зниження повноти отримуваної інформації.

Тому перший крок пошуку інформації - у формулюванні запиту. Відповідність між інформаційною потребою користувача і

формалізованим запитом визначає успіх інформаційного пошуку. Назвемо відповідність між інформаційною потребою користувача і формалізованим запитом **адекватністю** запиту.

Ми визначаємо **релевантність** як бінарне відношення між формалізованим запитом і відповіддю на цей запит. Числове значення релевантності залежить від трьох параметрів - **точності**, **повноти** і **ранжування**.

*Ранжування* - правильність порядку, в якому представлений список результатів інформаційного пошуку.

*Точність* - частка релевантних ресурсів серед всіх ресурсів, присутніх у видачі.

*Повнота* - це частка релевантних ресурсів присутніх у видачі серед всіх релевантних ресурсів, наявних в мережі Інтернет.

Невміння більшості користувачів правильно формувати запити і отримувати прийнятні обсяги відгуку породило в кінці 20 століття думку про Інтернет, як про величезне інформаційне звалище. Саме для максимального задоволення інформаційних потреб користувачів отримали широке практичне застосування теорії і методи семантичних мереж, контент-аналізу і глибинного аналізу текстів (Text Mining). При цьому виник новий клас систем, який все ж дозволяє справлятися з «проблемою розмірності» мережі. Сьогодні можна розглядати в якості одного з дивовижних феноменів той факт, що змістовні, семантично наповнені результати формуються без безпосереднього залучення методів штучного інтелекту, об'ємних баз знань і навіть експертів як таких, а шляхом використання частотно-лінгвістичних і евристичних методів. І сьогодні ефективно працюють в основному системи, що базуються саме на таких методах.

Виходячи із наведеного, найбільш важливою величиною, від якої залежить задоволення інформаційних потреб користувача є

**пертинентність** – співвідношення обсягу корисної для нього інформації до загального обсягу отриманої інформації [7]. При цьому слід враховувати, що формальний запит до системи є предметом творчого осмислення інформаційної потреби і не завжди точно відображає останню.

Таким чином, під підвищенням ефективності інформаційного пошуку ми будемо розуміти зростання пертинентності інформації, отриманої користувачем.

До засобів підвищення пертинентності можна віднести структурування знайденої інформації, і тому спрощення її сприйняття користувачем. Одним із найбільш вдалих мені вважається рішення, пропоноване російськими дослідниками Д.Е. Пальчуновим та Е.С. Сидоровою, яке полягає у синтезі інформаційно-пошукових систем і Інтернет-каталогів – віртуальний каталог [8]. Віртуальний каталог будується на основі певних предметних онтологій та забезпечує простоту і ясність каталогів документів, а також повноту і актуальність знайденої інформації, що забезпечується інформаційно-пошуковими системами.

Знайдена за допомогою пошукових машин інформація структурується у відповідності до розроблених предметних онтологій, використовуючи методи контент-аналізу, семантичного аналізу, штучного інтелекту тощо.

Інтерфейс віртуального каталогу зовні має нагадувати інтерфейс звичайного Інтернет-каталогу. В його основі система рубрик, відповідна ієрархії підобластей даної предметної області. У якості назв рубрик каталогу беруться розділи та підрозділи даної предметної області. Кожній рубриці порівнювати пояснення її змісту на природній мові. Наявність опису рубрик усуває один з недоліків каталогів - відсутність довідкової інформації. Користувач вибирає певну рубрику і

отримує список Інтернет-ресурсів, які їй відповідають [9].

Іншим прикладом використання онтологічної системи є ініціатива (KA) 2 [7] (Knowledge Annotation Initiative of the Knowledge Acquisition Community). Це міжнародний проект, метою якого є організація інтелектуального пошуку в Інтернет і автоматичне накопичення нових знань. У цій ініціативі виділяють наступні напрямки:

- анотація web сторінок інтелектуальної інформацією.
- онтологічний інжиніринг.
- організація інтерфейсу запитів і виведення з розподіленою онтології.

Нові технології інформаційного пошуку можуть допомогти зробити пошук більш значущим, суб'єктивним і прив'язаним до задач (task-based), що стоять перед користувачем [6].

**Висновки.** Величезна швидкість зростання кількості інформації в інтернеті приводить до все більших ускладнень щодо її пошуку. Сучасні пошукові засоби у намаганні забезпечити повноту результатів пошуку знаходять сотні тисяч документів, що ускладнює користувачу задоволення його інформаційної потреби. Такий стан речей зумовлює дослідження нових шляхів підвищення ефективності пошуку інформації.

Перспективним напрямком сучасних досліджень є створення онтологій предметних областей та розробка механізмів їх взаємодії із існуючими пошуковими засобами.

### ***Література:***

1. Пожуреєв В. І. Глобальне інформаційне суспільство як новий соціальний та економічний феномен XXI століття / В. І. Пожуреєв // Гуманіт. вісн. Запорізьк. держ. інженерної академії: зб. наук. пр. – Запоріжжя: ЗДІА, 2013. – Вип. 52. – С. 5 – 14.

2. Кудрявцева С. П., Колос В. В., Навчальний посібник — К.: Видавничий Дім «Слово», 2005. - 400с.
3. Васенев К. А. Анализ информационных потоков промышленного предприятия в контроллинге // Управление компанией. – 2003 - №2 – с. 46-56.
4. Лук'янчикова Ю. В., Попова Ю. М. Інформаційні потоки в системі управління організацією. 2010.- С. 35- 36. Доступ: [http://confer-dsum.ucoz.ua/\\_fr/0/7120405.pdf](http://confer-dsum.ucoz.ua/_fr/0/7120405.pdf)
5. Долинський Є. В., Скиба К. М. Інформаційні технології у професійній діяльності перекладача: навчальний посібник з інформаційних технологій для студентів 3 першого (бакалаврського) рівня, а також другого (магістерського) рівня вищої освіти, галузі знань «Гуманітарні науки», спеціальності «Філологія», «Переклад» / Є. В. Долинський, К. М. Скиба – Хмельницький: ХНУ, 2016. – 75 с.
6. Захаров В. П. Информационно-поисковые системы/ Учебно-методическое пособие, Санкт-Петербург, 2005
7. Товпига А. В., Маєвський О. В. Показники ефективності інформаційного пошуку/ Збірник тез доповідей IV Міжнародної науково-технічної конференції молодих учених та студентів «Актуальні задачі сучасних технологій» /с.64-65.
8. Пальчунов Д. Е., Сидорова Е. С. Виртуальный каталог. Труды Всероссийской конференции «Знания-Онтологии-Теории». Новосибирск. 2007. стр. 166-175.
9. Пальчунов Д. Е. Решение задачи поиска информации на основе онтологий / журнал Бизнес-информатика 2008. № 1. С. 3–13. Режим доступа: <https://bijournal.hse.ru/2008--1/26601231.html>

**References:**

1. PozhuEv V. I. *Globalne Informatslyne susplstvo yak noviy sotsialniy ta ekonomlchniy fenomen XXI stolittya* / V. I. Pozhuev // *Gumanlt. vlsn. Zaporizk. derzh. Inzhenernoyi akademiyi: zb. nauk. pr.* – Zaporihzhya: ZDIA, 2013. – Vip. 52. – S. 5 – 14.
2. Kudryavtseva S. P., Kolos V. V., *Navchalniy posibnik* — K.: Vidavnichiy Dim «Slovo», 2005. - 400s.
3. Vasenev K. A. *Analiz informatsionnyih potokov promyshlennogo predpriyatiya v kontrollinge* // *Upravlenie kompaniey.* – 2003 - № 2 – s. 46-56.
4. Luk'yanchikova Yu. V., Popova Yu. M. *Informatsiyi potoki v sistemi upravlinnya organizatsieyu*.2010.- S. 35- 36. Dostup: [http://confer-dsum.ucoz.ua/\\_fr/0/7120405.pdf](http://confer-dsum.ucoz.ua/_fr/0/7120405.pdf)
5. Dolinskiy E. V., Skiba K. M. *Informatsiyi tehnologiyi u profeslynly diyalnosti perekladacha: navchalniy posibnik z informatsiyi tehnologiy dlya studentiv z pershogo (bakalavrskogo) rivnya, a takozh drugogo (magisterskogo) rivnya vischoyi osviti, galuzi znan «Gumanitarni nauki», spetsialnosti «Filologiya», «Pereklad»* / E. V. Dolinskiy, K. M. Skiba – Hmelnitskiy: HNU, 2016. – 75 s.
6. Zaharov V. P. *Informatsionno-poiskovye sistemy/ Uchebno-metodicheskoe posobie*, Sankt-Peterburg, 2005
7. Tovpiga A. V., MaEvskiy O. V. *Pokazniki effektivnosti informatsiyi poshuku/ Zbirnik tez dopovldey IV Mizhnarodnoyi naukovu-tehnolchnoyi konferentsiyi molodih uchenih ta studentiv «Aktualni zadachi suchasni tehnologiy»* /s.64-65.
8. Palchunov D. E., Sidorova E. S. *Virtualnyi katalog. Trudyi Vserossiyskoy konferentsii «Znaniya-Ontologii-Teorii»*. Novosibirsk. 2007. str. 166-175.

9. *Palchunov D. E. Reshenie zadachi poiska informatsii na osnove ontologiy / zhurnal Biznes-informatika 2008. № 1. S. 3–13. Rezhim dostupu: <https://bijournal.hse.ru/2008--1/26601231.html>*