

ІНФОРМАТИКА, ОБЧИСЛЮВАЛЬНА ТЕХНІКА ТА АВТОМАТИЗАЦІЯ

УДК 681.3

А.В. Вдовіченко, аспір.

Житомирський інженерно-технологічний інститут

ХАРАКТЕРИСТИКИ ІНТЕЛЕКТУАЛІЗОВАНИХ ПОШУКОВИХ СИСТЕМ: КЛАСИФІКАЦІЯ ТА ПОРІВНЯННЯ

(Представлено *é.đ.í., äïð. A.İ. Данильченко*)

Розглянуті характеристики існуючих пошукових систем, запропонована їх класифікація відповідно до цих характеристик. Наведені порівняльні приклади для закордонних та вітчизняних пошукових машин, зроблені висновки щодо необхідності їх вдосконалення.

Âïðîï

Світова мережа Інтернет може бути представлена як гетерогенний інформаційний простір, що утворений з ресурсів різних типів та має довільну інформаційну структуру. Інформаційний простір змінюється безупинно, обсяг інформації в світі невпинно зростає, інформаційні структури стають більш складними, рівень хаосу в інформаційній системі постійно збільшується. В світі, де інформаційні ресурси необмежені, головною проблемою та найважливішою цінністю стає не сама інформація, а її пошук і ранжування знайденої інформації найвідповідніше до поставленого запиту [1–6]. Така постійно зростаюча практична потреба в пошуку інформації, що зумовлена бурхливим розвитком інформаційних технологій та глобальних мереж, стимулює зацікавленість в пошукових системах і в сервісах, що вони реалізують (рис. 1).

Пошукові машини, наприклад "crawlers", "spiders", "worms", індексують Мережу на підставі алгоритмів штучного інтелекту та за заданими запитами користувачів [4]. Але, наприклад, у чому різниця (не схожість, а саме різниця) між запитами "Мені потрібний новий автомобіль" і "Хочу нову машину" для алгоритмів штучного інтелекту? Її немає. Сьогоднішні пошукові алгоритми, як правило, не розділяють, а поєднують синоніми. "Автомобіль" логічно прирівнюється до "автомашини", хоча соціологічні дослідження зовсім чітко виділяють фінансові й інші особливості груп (читай – "груп потенційних клієнтів"), що використовують ті чи інші слова і вираження. Таким чином, алгоритмізація, впорядкування мови знищує відтінки значення. Зі збільшенням кількості проіндексованих сторінок, на жаль, часто знищується і сам зміст. *Релевантність** пошукової машини різко падає.

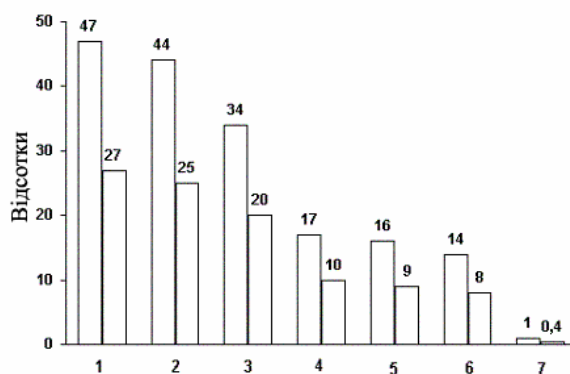


Рис. 1. Зміна кількості заіндексованих на травень 2001 року документів (правий стовпець) у відсотках від їхньої загальної кількості в Павутині до квітня 2000 року (лівий стовпець) для різних пошукових машин: 1 – AltaVista, 2 – Northern Light, 3 – HotBot, 4 – Excite, 5 – Lycos, 6 – Infoseek, 7 – WebCrawler (за матеріалами Science magazine і Forrester Research)

Каталожна організація інформації припускає, що укладач каталога добре розбирається в предметній області, що охоплюється каталогом. "Предметна область" сьогоднішньої мережі Інтернет – досвід, накопичений людством за свою історію. До цього необхідно додати всю динамічну інформацію, що відображається, про сьогоднішній день і, можливо, дещо про завтрашній. Найбільш зручними для користувача є не занадто великі, спеціалізовані тематичні каталоги з малою кількістю піддиректорій та будь-які спроби створення глобального інформаційного каталога неминує призводити до того, що безліч

* Під релевантністю ми розуміємо ступінь відповідності документа, знайденого пошуковою машиною за запитом користувача, цьому запиту.

ефективних малих списків губляться в плутанині розділів і підрозділів. Таким чином, ускладнення каталожних систем знищує простоту користування ними. І знову ж – як результат – погіршується *релевантність* пошукової машини [5].

Основні вимоги до якісного пошуку інформації

Можна виділити деякі критерії щодо визначення ефективності пошуку інформації. Такими критеріями є [1]:

- контроль повноти охоплення ресурсів;
- контроль вірогідності інформації, отриманої з Мережі;
- висока швидкість проведення пошуку;
- ранжування отриманої інформації.

Контроль повноти охоплення ресурсів є закономірною вимогою, якщо потрібно забезпечити якісний пошук інформації [1]. Для більшості користувачів Мережі вона повністю асоціюється із Web-простором, хоча інформаційний простір Мережі значно ширший. Повний збір інформації в Інтернеті з будь-якого питання в багатьох випадках виводить за межі Web-простору, а пошук проводиться у telnet-доступних базах даних, регіональних телеконференціях, базах даних gopher та ftp-архівах, в інших структурах інформації. Використання *всіх* основних існуючих типів ресурсів Мережі обов'язкове для проведення професійного пошуку інформації.

Контроль вірогідності інформації, отриманої з Мережі в результаті пошуку, може бути виконаний різними способами: отримання інформації з альтернативних джерел; урахування популярності отриманої інформації; врахування статусу (рейтингу) вузла, на якому міститься документ; отримання даних про авторів матеріалу та інше.

Швидкість проведення пошуку в Мережі. Обраховується незалежно від технічних характеристик підключення окремого користувача. Залежить від пошукової процедури та вміння користувача працювати з вибраним ресурсом.

Ранжування отриманої інформації. Оскільки інформаційний простір Інтернет дуже великий, кількість документів, що в будь-якій мірі відповідають запиту користувача, може бути також дуже великою. Для пошукової системи доцільно виконувати ранжування отриманої інформації за певними критеріями, наприклад, релевантність знайдених документів і запит користувача. В цьому випадку увесь набір отриманих документів буде відсортовано таким чином, що найбільш релевантні документи будуть знаходитись на початку списку, найменш – у його кінці.

Будь-яку пошукову систему можна описати, виділивши певний набір характеристик. Спробуємо відокремити деякі характеристики та визначити їх.

Характеристики, що залежать від структури документа

Більшість інформаційних ресурсів сьогодишнього Інтернет представлені у вигляді статичних чи динамічних HTML, і пошукові системи повинні вміти індексувати ці документи, враховуючи їх внутрішню будову [2].

Підтримка фреймів. Пошукова система повинна вміти обробляти документи фреймової структури. Такі документи є дуже розповсюдженими і втрата можливості їх індексування призведе до неможливості індексування багатьох документів, на які посилається фреймова структура.

Підтримка ImageMap. Тут приблизно та ж проблема, що і з фреймовими структурами серверів.

Частота появи посилань. Пошукові машини можуть визначити популярність документа за посиланнями на нього з інших документів мережі – за частотою посилань, і таким чином визначають релевантність документа.

Стоп-слова. Деякі слова недоречно включати до індексу пошукової машини. Це, як правило, слова, що дуже часто використовуються чи в звичайній мові, чи в певній предметній області. Не включають такі слова для економії місця на носіях. Наприклад, Altavista ігнорує слово *web*.

Підтримка META-тегів. За ідеєю, усі пошукові машини повинні враховувати співпадіння при індексації сторінок, однак на практиці не всі це роблять.

Title. Цей параметр показує, як пошукові машини генерують заголовки посилань для користувача у відповідь на його запит.

Description. Цей параметр показує, як пошукові машини генерують опис посилань для користувача у відповідь на його запит.

Підтримка META-тегів. Усі пошукові машини при індексації сторінок повинні враховувати співпадання, як засіб опису документа, однак на практиці не все це роблять.

Title. Показує, як пошукові машини генерують заголовки посилань для користувача у відповідь на його запит.

Description. Показує, як пошукові машини генерують опис посилань для користувача у відповідь на його запит.

Перевірка статусу URL. Визначає, чи можна перевірити наскільки глибоко проіндексований сервер і чи є він взагалі в індексі пошукової машини.

Мовні характеристики

Пошукові системи можуть підтримувати одну чи більше мов документів для індексації й, відповідно, мов запитів користувача. Чим більше мов підтримує така система, тим вона більш розвинута, так як розширюється інформаційний простір для пошуку. Для якісного пошуку також важливо, які запити підтримує пошукова система, чи дозволяє вона використовувати повнотекстові та розширені запити.

Прямий запит. Пряме співпадання фрази запиту та інформації в документі.

Запит "ЧИ"/"Г". Документ повинен містити усі слова запиту ("Г"); документ повинен містити хоча б одно слово запиту ("ЧИ").

Повнотекстовий запит. Задається на природній мові у вигляді речення, побудованого за правилами цієї мови.

Розширений запит. Використання розширених виразів у запиті: використання метасимволів та регулярних виразів.

Генерація словоформ. Параметр визначає, чи спроможна пошукова система генерувати словоформи заданих слів при побудові запиту.

Технічні характеристики індексування

Зазначені (submitted) сторінки. Сторінки серверів набагато раніш з'являються в індексах систем, якщо їх прямо вказати (Add URL), незважаючи на те, що в ідеальному інформаційному просторі пошукова машина повинна проіндексувати будь-яку сторінку будь-якого серверу.

Не зазначені (non-submitted) сторінки. Зазначивши хоча б одну сторінку сервера, в ідеалі пошукова машина обов'язково знайде і проіндексує усі інші сторінки за посиланням із зазначеної. Це, звичайно, потребує часу, і в реальному світі пошукові машини залишають індексування таких сторінок на майбутнє [6].

Контроль індексації. Визначає засоби, за допомогою яких відбувається керування пошуковою машиною. Для "розвинутих" пошукових систем такий контроль виконується розпорядженнями файла robots.txt. Деякі також підтримують контроль за допомогою META-тагов із самих документів, що індексуються.

Індексування захищених пароллями директорій та серверів. Деякі пошукові машини можуть індексувати такі сервери, якщо їм вказати Username та Password. Це може стати у пригоді користувачам, принаймні вони зможуть побачити, що є на сервері.

Глибина індексування. Визначає, скільки незазначених сторінок після зазначеної буде проіндексовано пошуковою системою. Більшість великих машин не мають обмежень за глибиною індексування, але існують причини, з яких сторінки можуть бути не проіндексовані, наприклад, занадто акуратне використання фреймових структур (без дублювання посилань у керуючому (frameset) файлі чи використання imgetagар без дублювання їх звичайними посиланнями).

Видалення старих даних. Визначає дії, що виконуються при закритті сервера чи переміщенні його на іншу адресу. Можливі дві дії: просто видалити старий зміст і переписати файл robots.txt.

- *Видалення вмісту:* проведення вилучення даних в індексі пошукової машини при спробі реіндексувати документ. Залежить від періоду відновлення даних.

- *Перепиання файла robots.txt:* всі посилання на файли цього сервера будуть вилучені з індексу при звертанні пошукової машини до файла, якщо сервер закритий від індексації.

Перенаправлення (redirect). Деякі сайти перенаправляють відвідувачів з одного сервера на інший. Цей параметр показує, який URL буде пов'язаний з вашими документами. Це важливо, оскільки, якщо пошукова машина не відпрацьовує перенаправлення, то можуть виникнути проблеми з неіснуючими файлами.

Період відновлення. Динаміка зміни документів у Інтернет дуже велика і пошукові машини повинні індексувати усі без обліку дати. Однак у кожен момент часу посилання, видані у відповідь на запити користувачів, можуть бути одноденної давнини, а можуть бути і місячної, а те і більше давнини. Деякі пошукові машини відразу індексують одну сторінку, а лише потім продовжують індексувати ще непроіндексовані сторінки, інші можуть частіше сканувати найбільш популярні сторінки Мережі.

Дата індексування документа. Деякі пошукові машини показують дату, коли був проіндексований той чи інший документ. Це допомагає користувачу зрозуміти, якої "свіжості" посилання видає пошукова система.

Спат-штрафи. Підвищити рейтинг сайту можна, вказавши свій URL багато разів через Add URL чи завдяки багаторазовому згадуванню того самого ключового слова тощо. Звичайно, що пошукові системи намагаються уникати подібних дій та, навпаки, понижують рейтинг такого сайту.

Інтелектуальні характеристики

Crc32. Можна легко побачити, як при використанні пошукової системи вона на запит видає декілька посилань на один й теж самий документ, що не тільки знижує зручність використання системою, а й значно впливає на її загальну релевантність.

Звертання до інших систем. Деякі пошукові системи не мають власної бази індексів, а звертаються за інформацією до інших пошукових систем, передаючи їм запит і отримуючи результат пошуку, а потім виконують ранжування отриманих даних.

Врахування популярності документа. Розвинуті пошукові машини при обрахуванні релевантності документа враховують його популярність – як часто на нього посилаються інші документи.

"Здатність до навчання".

- Якщо сервер обновляється часто, то пошукова машина частіше буде його реіндексувати, якщо рідко – рідше.

- Деякі пошукові системи здатні запам'ятовувати успішно виконані запити, щоб надалі виконувати їх більш ефективно.

- Настроювання на роботу з окремим користувачем пошукової системи, створення персонального облікового запису з метою адаптації системи на стиль окремого користувача.

Вплив на алгоритм визначення релевантності. Пошукові машини обов'язково використовують розташування і частоту повторення ключових слів у документі. При цьому методи підвищення релевантності різні для кожної окремої машини.

Загальні характеристики

Тип пошукової машини. Повнотекстові пошукові машини індексують кожне слово на веб-сторінці, виключаючи *стоп-слова*. Абстрактні пошукові машини створюють деякий екстракт кожної сторінки. Для *вебмайстрів* повнотекстові машини більш корисніші, оскільки будь-яке слово, що зустрічається на веб-сторінці, піддається аналізу при визначенні його релевантності до запитів користувачів. Однак може трапитися й таке, що для абстрактних пошукових машин сторінки проіндексовані краще, ніж для повнотекстових. Це може виходити з алгоритму екстрагування, наприклад, за частотою вживання на сторінці тих самих слів.

Ім'я пошукового робота. Ім'я пошукового робота – слово чи словосполучення, яким пошукова система відповідає на HTTP-запит.

Розмір. Розмір пошукової машини – кількість проіндексованих нею сторінок. Пошукові машини з великим розміром можуть проіндексувати майже усі сторінки сайту, машини середнього розміру проіндексують його лише частково, а при малому розмірі сторінки можуть взагалі не потрапити в каталоги пошукової машини.

Визначивши характеристики, згрупуємо їх, отримуючи таким чином класифікацію пошукових машин за даними характеристиками.

Дана класифікація відображена у табл. 1.

Таблиця 1

Характеристики пошукових машин

Загальні	Тип пошукової машини	
	Ім'я пошукового робота	
	Розмір	
Технічні характеристики індексування	Загальні	Час індексування вказаних сторінок
		Час індексування невказаних сторінок
		Контроль індексації
		Індексування захищених пароллями директорій та серверів
		Глибина індексування
		Видалення старих даних
		Перенаправлення
		Використання META-тегу для робіт
		Sram-штрафи
	Часові	Дата індексування документа
	Період відновлення	
Структура документа	Підтримка фреймів	
	Підтримка ImageMap	
	Частота появи посилань	
	Стоп-слова	
	Підтримка META-тегів	
	Опис документа	
	Заголовок документа	
	Використання robots.txt	
Перевірка статусу URL		
Інтелектуальні особливості	Crc32	
	Звертання до інших систем	
	Врахування популярності документа	

	Вплив на алгоритм визначення релевантності	
	Здатність до навчання	
Мова	Мова запити	Прямий запит
		Запит "ЧИ"/"Г"
		Повнотекстовий запит
		Розширений запит
		Генерація слівформ
	Мова документа	Одномовна
		Багатомовна

Наведемо опис існуючих інтелектуалізованих пошукових машин згідно з даною класифікацією. При цьому окремо виділимо закордонні та вітчизняні пошукові машини.

Таблиця 2

Характеристики закордонних пошукових машин

	Altavista	Excite	HotBot	InfoSeek	Lycos	OpenText	WebCrawler
Тип	Повнотекст.	Повнотекст.	Повнотекст.	Повнотекст.	Повнотекст.	Повнотекст.	Повнотекст.
Структура документа							
Підтримка фреймів	-	+	-	+	+	-	-
Підтримка ImageMap	+	-	-	+	+	-	+
Частота появи посилань	-	-	+	-	+	-	+
Стоп-слова	+	+	+	-	+	-	-
Підтримка META-тегів	+	-	+	+	+	-	-

Продовження таблиці 2

	Altavista	Excite	HotBot	InfoSeek	Lycos	OpenText	WebCrawler
Заголовок документа	Заголовок сторінки чи No Title	Заголовок сторінки чи Untitled	Заголовок сторінки чи URL	Заголовок сторінки чи перший рядок документа	Заголовок сторінки чи перший рядок документа	Перші 100 символів з документа	Заголовок сторінки чи URL
Опис документа	META-тег чи перші кілька рядків з документа	Формується з найбільш релевантних до запиту фраз документа	META-тег чи перші кілька рядків документа	META-тег чи перші 200 символів після тага <body>	META-тег чи екстракт зі змісту сторінки	Перші 100 символів документа	Створюється зі змісту; планується підтримка META-тегів у майбутньому
Використання META-тегу для роботів	+	-	+	+	+	-	Тільки тег NOINDEX
Інтелектуальні особливості							
Crc32	-	-	-	-	-	-	-
Звертання до інших систем (*)	-	-	-	-	-	-	-
Врахування популярності документа	-	-	-	-	-	-	-
Вплив на алгоритм визначення релевантності	Немає	-	Ключові слова в метаданих	Немає	Немає	Немає	Частота появи посилань
Індексація з урахуванням частоти оновлення серверу	+	-	+	+	-	-	-
Запам'ятовування запитів	+	-	-	-	-	-	+
Обліковий запис користувача	-	-	-	-	-	-	-
Мова запиту і підтримка багатомовності							
Прямий запит	+	+	+	+	+	+	+
Запит "ЦИ"/"І"	+	+	+	+	+	+	+
Повнотекстовий запит	-	-	-	-	-	-	-
Розширений запит	+	-	-	-	-	-	-
Багатомовна	+	-	-	-	+	+	+
Технічні характеристики							
Час індексування вказаних сторінок	1 день	1 тиждень	3 тижні	1 місяці	1 місяці	2-4 тижні	2-4 тижні
Час індексування невказаних сторінок	1-3 місяці	3 тижні	3 тижні	1 місяці	1 місяці	2-4 тижні	2-4 тижні
Контроль індексування	robots.txt	robots.txt (у майбутньому і метадані)	robots.txt і метадані	robots.txt	robots.txt	robots.txt	robots.txt і метадані
Індексування зхищених пароллями директорій та серверів	-	+	-	+	+	-	-

Закінчення таблиці 2

	Altavista	Excite	HotBot	InfoSeek	Lycos	OpenText	WebCrawler
Глибина індексування	Необмежена	Необмежена	Необмежена	Необмежена	Необмежена	–	Обмежена популярністю серверу
Видалення старих даних	Видалити зміст і вказати нову адресу	Видалити зміст, переписати robots.txt	Переписати robots.txt	Видалити зміст і вказати нову адресу чи переписати robots.txt	–	–	–
Перевірка статусу URL	+	–	–	–	+	–	–
Перенаправлення	+	+	–	–	–	–	+
Дата індексування	Явне використання	–	Неявне використання	–	–	–	–
Період поновлення	Від 1 дня до 3 місяців	1–3 тижні	Не пізніше 3 тижнів	Від хвилини до місяця	Щомісячне відновлення	1–4 тижня	Щотижневе відновлення
Спам-штрафи	+	+	+	+	+	+	+
Використання robots.txt	+	–	+	+	+	–	+
Ім'я пошукового робота	Scooter	Architext Spider	Slurp the Web Hound	Side winder	T-rex	–	Spidey
Розмір (стор.)	30 млн	55 млн	54 млн	20–50 млн	20–25 млн	5 млн	2 млн

Таблиця 3

Характеристики вітчизняних пошукових машин

	Russian Express	TELA поиск	Rambler	Яндекс	Апорт Поиск
Тип	Повнотекст.	Повнотекст.	Повнотекст.	Повнотекст.	Повнотекст.
Структура документа					
Підтримка фреймів	+	+	+	+	+
Підтримка ImageMap	+	+	+	+	+
Частота появи посилань	–	–	+	–	+
Стоп-слова	–	–	+	+	+
Підтримка META-тегів	+	+	+	+	+
Заголовок документа	Поки URL	Title	Title чи URL і відносна міра релевантності	Title і URL	Title
Опис документа	META-tag Description і частина тексту документа	Перші рядки документа	Перші 512 байт документа, крім meta, javascript, images...	Видаються перші 1024 байт тексту, міра релевантності, дата створення й обсяг документа	Пропозиції, що містять слова запиту (1, 3 чи до 10)
Використання META-тегу для роботів	+	–	+	+	+
Інтелектуальні особливості					
Crc32	–	–	–	–	–
Звертання до інших систем (*)	–	–	–	–	–
Врахування популярності документа	–	–	–	+	–

Закінчення таблиці 3

	<u>Russian Express</u>	<u>TELA поиск</u>	<u>Rambler</u>	<u>Yandex</u>	<u>Апорт Поиск</u>
Вплив на алгоритм визначення релевантності	-	-	-	-	-
Індексація з урахуванням частоти оновлення серверу	+	-	+	+	-
Запам'ятовування запитів	-	-	+	+	-
Обліковий запис користувача	-	-	-	-	-
Мова запиту і підтримка багатомовності					
Прямий запит	+	+	+	+	+
Запит "ЧИ"/"Г"	+	+	+	+	+
Повнотекстовий запит	-	-	-	-	-
Розширений запит	-	-	+	+	-
Багатомовна	-	-	+	+	+
Технічні характеристики					
Час індексування вказаних сторінок	20 днів	-	7-14 днів	1-2 дні	1-15 днів
Час індексування невказаних сторінок	20 днів	-	До 3 місяців	В залежності від популярності документів	Лімітується швидкістю відновлення індексу
Контроль індексування	-	Явно – ні, побічно – вказавши як критерій URL	+	+	+
Індексування зхищених пароллями директорій та серверів	-	+	-	+	+
Глибина індексування	5.000 документів на глибину 150	20 документів	Необмежена	Необмежена	Необмежена
Перевірка статусу URL	+	-	-	-	+
Перенаправлення	+	+	-	-	-
Дата індексування	Ні, у проєкті – так	+	Так, при розширеній видачі результатів	+	+
Период поновлення	20 днів	3-4 тижня	1 разів у тиждень	Перманентно	Раз на добу (від 10 до 40 тисяч документів)
Спам-штрафи	+	+	+	+	+
Використання robots.txt	+	+	+	+	+
Ім'я пошукового робота	www.search.ru	-	StackRambler/1.2	YandexWeb	Aport
Розмір (стор.)	500.000	500.000	2.500.000	2.000.000	2.600.000

За даною класифікацією та за приведеними характеристиками реальних пошукових систем, що функціонують сьогодні, можна зробити висновки щодо невеликої ефективності їх функціонування. Більшість існуючих сьогодні пошукових систем орієнтовані, перш за все, на так званий "масовий" пошук: увага приділяється кількості проіндексованих документів та швидкості проведення пошуку. Пошукові системи не охоплюють усі інформаційні ресурси і майже не виконують контроль вірогідності отриманої інформації. Можна стверджувати, що *якість* пошуку дуже низька – характеристики, що визначають релевантність системи (інтелектуальні характеристики, мовні характеристики), показують,

що існуючі пошукові системи потребують вдосконалення. Пошукові системи потребують інтелектуалізації для досягнення більшої релевантності: розширення інтелектуальних та мовних характеристик, інтелектуалізації обробки текстової інформації та покращення методів ранжування знайдених документів.

ЛІТЕРАТУРА:

1. *Талантов М.* Профессиональный поиск в Интернет. Полнота. Достоверность. Скорость // КомпьютерПресс. – 1999. – № 7.
2. *Солтон Дж.* Динамические библиотечно-информационные системы. – М.: Мир, 1979.
3. *Pinkerton Brain Finding What People Want: Experiences with the WebCrawler.* – <http://info.webcrawler.com/bp/WWW94.html>.
4. *Тихонов В.* Поисковые системы в сети Интернет. – atomzone.hypermart.net
5. *Francisco Pinto, Claudio Baptistay, Nick Ryanz.* Using Semantic Searching for Web Portal Interoperability.
6. *Аликберов А.А.* Подборка статей и технической информации по работе и характеристикам поисковых машин. – <http://www.citforum.ru/internet/search/ips.shtml>.
7. *Alta Vista.* <http://www.altavista.digital.com/> / Digital Equipment Corporation, 1996.

ВДОВІЧЕНКО Андрій Володимирович – аїї³даіò Æèòìèðñüêîîâî ðæâîîäîî-òâîîîêîî³-îîîî ðîèòòòò.
 Íâòêîî³ ðîäâîîè:

- комп'ютерні технології;
- інтелектуалізація Інтернет-дâîîîêîî³й;
- текстові процесори (обробка природних мов).

Подано 12.02.2002

Вдовіченко А.В. Характеристики інтелектуалізованих пошукових систем: класифікація та порівняння

Вдовиченко А.В. Характеристики интеллектуализированных поисковых систем: классификация и сравнение

Vdovachenko A.V. The features of the intellectual searching systems: the classification and comparison

УДК 681.3

Характеристики интеллектуализированных поисковых систем: классификация и сравнение. / Вдовиченко А.В.

Рассмотрены характеристики существующих поисковых систем и предложена их классификация согласно этим характеристикам. Приведены сравнительные примеры для зарубежных и отечественных поисковых систем и сделаны выводы о необходимости их усовершенствования.

УДК 681.3

The features of the intellectual searching systems: the classification and comparison

The features of existing searching systems and their classification are considered. The comparative examples of foreign and home searching systems are presented. The conclusions of their needs of improvement are presented.