УДК 004.912

Д.Є. Костенко, Є.Д. Замотаєв, В.О. Широченко
Університет митної справи та фінансів, Україна.

ПРОБЛЕМИ ПОШУКУ РЕЛЕВАНТНОЇ ІНФОРМАЦІЇ

D.E.Kostenko, Y.D.Zamotaev, V.A.Shirochenko

THE PROBLEMS OF INFORMATION SEARCHING

Інформаційний пошук – процес отримання документальної інформації, що задовольняє інформаційним потребам.

Еволюція поняття і процесу інформаційного пошуку стимулювалася різноманітними проблемами, пов'язаними із забезпеченням знаходження та доступу до інформаційних джерел. Спочатку інформаційний пошук стосувався наукових публікацій і бібліотечних каталогів, однак незабаром він поширився і на інші сфери.

Основне завдання інформаційного пошуку – допомогти користувачу знайти інформацію, в якій він зацікавлений.

З наявної сукупності інформації потрібно відібрати підмножину, відповідну інформаційній потребі користувача. Визначальною компонентою зазвичай вважається набір ключових слів. Але ця компонента не працює, якщо немає запиту. Запит являє собою осмислену фразу або набір слів, що описують інформаційну потребу.

Результат пошуку – список документів і web-сторінок які відібрані системою і містять корисну для користувача інформацію. Цей список, як правило, впорядкований по мірі зменшення метрики, яку називають “вагою”, “ступенем релевантності запиту” або оцінкою ймовірності того, що документ задовольняє запиту.

Весь прогрес цивілізації в цілому та конкретно сучасний рівень інформаційно-аналітичної роботи, показують тенденцію зменшення ролі природнього інтелекту в процесах інтелектуальної діяльності, перекладання її на автомати, а також підвищення інтелекту в системах, які повинні допомагати, а потім і направляти дослідження даних залежно від їхнього контенту.

Зростання об’ємів інформації, яка є необхідною для прийняття рішень, призводить до різкого збільшення кількості документів. Отже, традиційні методи роботи з документами стають менш ефективними.

На сьогоднішній день інформаційний пошук швидко стає основною формою доступу населення до інформації у суспільстві. Користувачу потрібні досконалі засоби отримання релевантної інформації. У даному випадку річ може йти про швидкий доступ до потрібних даних або послуг.

Не завжди вдається знайти необхідну інформацію з першого разу. Запити повинні складатися так, щоб область пошуку була максимально конкретизована і звужена, тобто перевагу слід віддавати використанню декількох “вузьких” запитів в порівнянні з одним розширеним. Але не треба вважати, що “вузький” запит – повноцінне вирішення всіх проблем. Спробуємо змоделювати проблеми, шо заважають створенню релевантної пошуковій видачі:

1. Накопичення в Інтернеті “порожньої” і застарілої інформації. Наприклад, на форумах, на яких люди обговорюють проблему, але ніяких конкретних рішень не наводять, існують посилання на сайти з вирішенням проблеми, які вже застаріли (інформація була видалена з сайту, сайт був закритий або заблокований).

2. Ресурси і сервіси з малою текстовою інформацією (наприклад, сайти для фотографів, ілюстраторів, дизайнерів або спецсервіси, що несуть корисну інформацію,

але не мають на своїх сторінках відповідних "міток", щоб їх знайшли пошукові алгоритми).

3. Чорний копірайтинг або сайти для "роботів". Рівень SEO-оптимізаторів зростає з кожним роком, як і їх кількість, і разом з ним зростає чисельність сайтів для заробітку на контекстній рекламі. Здебільшого такі статті – це вода в чистому вигляді, що не представляє практичної цінності для читача.

Які ж можна запропонувати шляхи вирішення таких проблем?

По-перше, в Інтернеті (і не тільки, адже мова може йти і про локальні БД великих масштабів) потрібно створювати більше спеціалізованих і тематичних ресурсів з якісним контентом. З IT-тематики є, наприклад, Stackoverflow. Чим більше таких порталів, тим швидше і якісніше буде проходити пошук інформації.

По-друге, необхідно розвивати сервіси порівняння. Це сервіси, що надають зведену таблицю по певній проблемі, грунтуючись на результатах, отриманих з багатьох сайтів і інших сервісів. Яскравий тому приклад: сервіс порівняння цін eKatalog. Написавши назву товару, сервіс видасть порівняльну таблицю цін і відгуків з усіх можливих магазинів. IT-сфера значною мірою потребує подібного сервісу. При введенні проблеми, користувач отримував би звіт готових рішень з усіх IT-сайтів без необхідності заходити на кожен з сайтів і перечитувати масу інформації.

По-третє, по можливості використовувати сервіси геолокації і засновувати пошукові видачі на основі місця розташування користувача. Наскільки відомо, така технологія застосовується дуже рідко. Або майже не застосовується. Але тут, якщо чесно казати, є один невеличкий нюанс – перш ніж використовувати такий підхід, необхідно створити умови для "регіонального" (якщо так можна сказати) зберігання інформації.

По-четверте, необхідно розвивати пошукові системи. Зараз необхідно робити упор на якість інформації і розділяти її в залежності від завдань. Це складний процес, але немає нічого неможливого. Якщо користувач налаштований на роботу і йому необхідно максимально швидко знайти потрібний контент, то варто прибирати з пошукової видачі розважальні ресурси. А коли користувач хоче відпочити – то навпаки.

Задача-максимум полягає в тому, щоб зробити пошук динамічним і зручним для користувача. Для будь-якого типу запиту, що виникає в практичній діяльності, повинні бути знайдені адекватні знання в інформаційному просторі. При цьому мова для формулювання пошукової вимоги не повинна бути занадто складною.

Система пошуку також повинна реалізовувати так званий процес "самонавчання" системи. Цей процес, або подібний до нього дозволить ліквідувати ситуації з термінами або назвами, які записуються некоректно або дублюються у базі даних.

Література

1. Baeza-Yates R. Modern Information Retrieval / Ricardo Baeza-Yates, Berthier Ribeiro-Neto – New-York: Addison-Wesley, ACM Press, 1999. – 501 p.

2. Маннинг, К.Д. Введение в информационный поиск / К. Д. Маннинг, П. Рагхаван, Х. Шютце. – М.: ООО "И.Д. Вильямс", 2011. – 128 с.