

УДК 681.5.004.77

Ю.Е. Парфьонов

Харківський національний економічний університет, Харків

ВИБІР ПОКАЗНИКІВ ЕФЕКТИВНОСТІ ІНФОРМАЦІЙНОГО ПОШУКУ В МЕРЕЖІ ІНТЕРНЕТ

Пошук інформації в мережі Інтернет стає все більш складним і трудомістким. Це пов'язано з ускладненням структури інформаційних ресурсів Інтернет та постійним збільшенням обсягу наявної інформації.

Відомо, що поняття інформаційного пошуку базується на необхідності задоволення інформаційної потреби користувачів. Для досягнення даної мети важливим є застосування методів інформаційного пошуку та відповідних інформаційно-пошукових систем, які забезпечують високу ступень відповідності результатів інформаційного пошуку запиту користувача.

При цьому виникає задача оцінювання ефективності інформаційного пошуку. Для вирішення даної задачі використовується ряд показників: повнота, точність, втрата інформації, інформаційний шум, F-міра тощо[1].

Найбільш важливими є такі показники ефективності інформаційного пошуку як повнота та точність. Ці показники базуються на оцінюванні семантичної відповідності пошукового запиту та пошукового образу документа, тобто релевантності.

Для розрахунку цих показників потрібно скористатися матрицею класифікації [2] (табл. 1).

Таблиця 1

Основні категорії відповідей інформаційно-пошукової системи

	Релевантні документи	Не релевантні документи
Знайдено системою	A	B
Не знайдено системою	C	D

Точність визначається як відношення кількості знайдених релевантних документів до загальної кількості знайдених документів (1):

$$\text{precision} = \frac{A}{A+B} \quad (1)$$

Вона характеризує здатність інформаційно-пошукової системи щодо пошуку релевантних документів.

Повнота – це відношення кількості знайдених релевантних документів, до загального числа релевантних документів у базі інформаційно-пошукової системи (2):

$$\text{recall} = \frac{A}{A+C} \quad (2)$$

Повнота характеризує здатність системи знаходити документи, потрібні користувачеві, але не враховує кількість не релевантних документів, які йому видаються.

Іноді точність і повноту інформаційного пошуку об'єднують. Доброю мірою для спільної оцінки точності і повноти є F-міра, яка визначається як їх середнє гармонійне(3):

$$F = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}} \quad (3)$$

Таким чином, існує багато показників ефективності інформаційного пошуку, найбільш важливими з яких є семантичні показники точності та повноти.

Список літератури

1. Ландэ Д.В. *Интернетика. Навигация в сложных сетях. Модели и алгоритмы* / Д.В. Ландэ, А.А. Снарский, И.В. Безсуднов. – М.: Либроком, 2009. – 264 с.
2. *Российский семинар по оценке методов информационного поиска [Электронный ресурс]. – Режим доступа до ресурсу: <http://romip.ru/>*