

ІНФОРМАЦІЙНО-ПОШУКОВА СИСТЕМА, ОРІЄНТОВАНА НА ЗВ'ЯЗКИ ОБ'ЄКТІВ

На даний час для підтримки прийняття рішень все частіше використовуються дані з мережі Інтернет. Але ефективному застосуванню цього потужного ресурсу заважають такі фактори, як засилля інформаційного шуму, низька точність наявних інформаційно-пошукових систем, недостатня їх оперативність. Для усунення цих недоліків створюється система БДЗ (база даних зв'язків), яка орієнтована на витяг із веб-документів понять певних типів, виявлення їх взаємозв'язків, побудови відповідної бази даних, та надання її у доступ. При цьому користувачі фактично отримують базу даних об'єктів, що їх цікавлять, та відповідних інформаційних взаємозв'язків.

Передумовами створення системи БДЗ на цей час є теорія та технології глибинного аналізу текстів – Text Mining [1,2], зокрема розвинена методологія екстрагування понять [3], теорія та технології баз даних надвеликих обсягів, теорія «складних мереж» [4], яка на цей час застосовується, зокрема, для виявлення «прихованих» зв'язків в терористичних організаціях [5,6].

Технологію формування бази даних зв'язків інформаційно-пошукової системи БДЗ побудовано наступним чином (рис. 1). За допомогою програми-робота здійснюється сканування вибраних веб-ресурсів, які містять інформацію щодо об'єктів (такими ресурсами можуть бути, наприклад, сайти біографій, працевлаштування, бізнес-форуми тощо). Після цього здійснюється екстрагування необхідних користувачам понять, наприклад, найменувань брендів, компаній, електронних адрес тощо. Відібрані поняття (та відповідні відношення між ними) завантажуються у базу даних зв'язків, яка також містить посилання на документи-першоджерела.

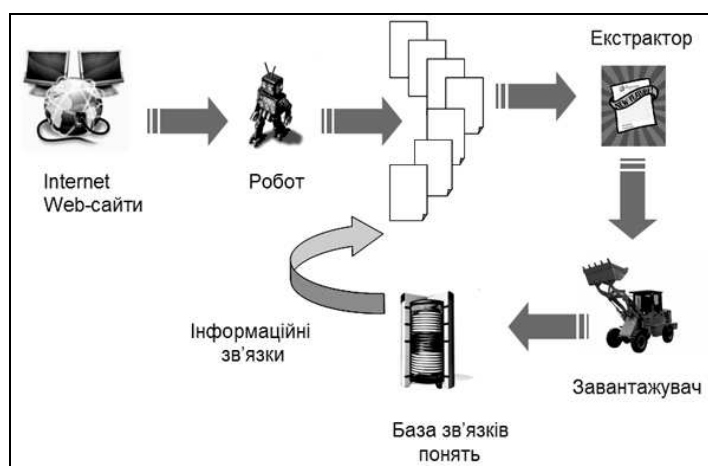


Рис. 1. Схема формування бази даних зв'язків

В корпоративній інформаційній інфраструктурі база даних зв'язків може використовуватися на різних рівнях, наприклад, окремо, щоб надавати відомості щодо можливих інформаційних взаємозв'язків об'єктів. Крім того, можливості системи БДЗ можуть бути доповнені можливостями наявних повнотекстових та/або фактографічних баз даних (рис. 2).



Рис. 2. База даних зв'язків у корпоративній інформаційній інфраструктурі

Система, що розглядається, таким чином, має властивість масштабування за трьома параметрами – по-перше, за обсягом баз даних, по-друге, за складом понять, що використовуються, і по-третє, за інфраструктурним оточенням.

Фактично пропонується інформаційно-пошукова система нового типу, в якій із самого початку вирішені питання оперативності, відсіювання інформаційного шуму, використання якої є легітимним у будь-яких умовах.

Крім того, система, принципи побудови якої розглядаються, є ланкою зв'язку між повнотекстовими та фактографічними базами даних.

Література

- [1] Ландэ Д.В. *Поиск знаний в Internet*. - М.: Диалектика-Вильямс, 2005. - 262 с.
- [2] Berry M.W. *Survey of Text Mining. Clustering, Classification, and Retrieval*. - Springer-Verlag, 2004. - 244 p.
- [3] Додонов А.Г., Ландэ Д.В. *Выявление понятий и их взаимосвязей в рамках технологии контент-мониторинга // Реєстрація, зберігання і обробка даних, 2006, - Т. 8, -N 4. – С. 45-52.*
- [4] Ландэ Д.В., Снарский А.А., Безсуднов И.В. *Интернетика: Навигация в сложных сетях: модели и алгоритмы*. - М.: Либроком, 2009. - 264 с.
- [5] Clauset A., Moore C., Newman M.E.J. *Hierarchical structure and the prediction of missing links in networks // Nature 453, 98-101 (1 May 2008).*
- [6] Ландэ Д.В., Фурашев В.М. *Дослідження терористичних зв'язків в рамках теорії складних мереж // Збірник наукових праць VIII Міжнародної науково-практичної конференції "Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях". - Київ-Харків-АР Крим, 2009. - С. 416-429.*