

ТЕХНОЛОГІЇ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

Коноваленко О.Є.¹, Брусенцев В.О.²

¹*Національний технічний університет
«Харківський політехнічний інститут»*

²*Харківська державна академія культури, м. Харків*

Результатом розвитку інформаційних технологій та мережі Internet є кількість інформації, що накопичена людством в електронному вигляді: тексти, зображення, аудіо, відео, гіпертекстові документи, бази даних і т. ін. Сучасні системи вилучення інформації використовують засновані на методах штучного інтелекту засоби уявлення і інтерпретації для пошуку в терабайтних сховищах вельми цінну інформацію.

Більшість сучасних програм контент-аналізу обмежені обробкою тексту, проте їхні можливості набагато ширші. Прикладом технології цього покоління є технологія "видобування" даних або Text Mining. Взагалі результатом природної еволюції інформаційних технологій стали хмарні технології та методи (класифікація, кластеризація, прогнозування) і технології (Data Mining, Text Mining, Web Mining, OLAP) інтелектуального аналізу даних. Причинами їх популярності стали наступні: стрімке накопичення даних; загальна комп'ютеризація; проникнення Інтернет у всі сфери діяльності; прогрес в області інформаційних технологій (вдосконалення СУБД і сховищ даних); прогрес в області виробничих технологій (зростання продуктивності комп'ютерів, об'ємів накопичувачів, впровадження Grid-систем).

Не дивлячись на кількість методів Data Mining, пріоритет все більше зміщується в сторону логічних алгоритмів пошуку даних if-then алгоритмів, за допомогою яких вирішуються задачі прогнозування, класифікації, розпізнавання образів, сегментації БД, вилучення з даних прихованих знань, інтерпретації даних, встановлення асоціацій в БД та інше. Результати таких алгоритмів є ефективними та легко інтерпретуються. Але головною проблемою логічних методів виявлення закономірностей є проблема перебору варіантів за обмежений час. Ці методи штучно обмежують такий перебір та будують дерева рішень з принциповими обмеженнями ефективності пошуку if-then правил.

Основна особливість Data Mining – це поєднання широкого математичного інструментарію (від класичного статистичного аналізу до нових кібернетичних методів) і останніх досягнень у сфері інформаційних технологій. У технології Data Mining гармонійно об'єдналися строго формалізовані методи і методи неформального аналізу, тобто кількісний і якісний аналізи даних. Більшість аналітичних методів, які використовуються в технології Data Mining, – це відомі математичні алгоритми і методи. Новим є те, що їх можна застосовувати при рішенні тих або інших конкретних проблем. Це обумовлено новими властивостями технічних і програмних засобів.