

МЕТОДИ АНАЛІЗУ Й ОБРОБКИ ЕКСПЕРТНОЇ ЛІНГВІСТИЧНОЇ ІНФОРМАЦІЇ

Євсютін І. Д., Шостак І. В.

Харківський національний університет радіоелектроніки, Харків, Україна

Інтелектуальна обробка текстової інформації, яка знаходить своє застосування практично у всіх сферах інформаційно-аналітичної діяльності людини (наукові дослідження, сфера освіти, охорона здоров'я, ділова галузь та багато інші), охоплює широке коло проблем, таких як: формалізація текстових даних (наприклад, у вигляді онтологій, семантичних мереж, тезаурусів та ін.), способи витягу знань із даних (кластеризація і класифікація, отримання фактографічної інформації, асоціативні правила та ін.), методи інтеграції предметних знань, отриманих з різнорідних джерел (наприклад, відображення, злиття, вирівнювання).

Метою доповіді є опис дослідження й розробка математичних методів і алгоритмів аналізу лінгвістичної експертної інформації, отриманої з колекцій текстів заданої предметної області, з метою її формалізації для застосування в інтелектуальних системах обробки інформації.

Незважаючи на високу наукову й практичну значимість існуючих робіт у даній області, у цей час відсутній універсальний для всіх предметних областей метод формалізації витягнутих з текстової інформації знань, що дозволяє з високою точністю й повнотою витягати терміни й зв'язку між ними, що й передбачає можливість розширення формалізованих структур різних джерел шляхом їхньої інтеграції.

Одним з перспективних рішень названої проблеми є застосування онтологій [2]. Формалізоване представлення великого обсягу текстів у вигляді онтологічної моделі дозволяє виконувати автоматичну (інтелектуальну, аналітичну) обробку отриманої онтологічної інформації, що знаходить застосування в інформаційно-аналітичних і пошукових системах нового покоління й системах інтеграції даних.

В результаті має бути запропонований алгоритм формалізації лінгвістичної експертної інформації, отриманої з колекції текстів заданої предметної області, у вигляді ієрархічної структури понять, що відрізняється від відомих обчисленням схованих закономірностей у тексті, кластеризацією понять і відображенням їх ієрархічних зв'язків у вигляді дендрограми з обліком тематичної імовірнісної близькості.

Список літератури

1. Whissell John S., Clarke Charles L.A. Improving document clustering using Okapi BM25 feature weighting. Information retrieval. 2011. Т. 14, № 5, pp. 513-523.
2. Загорulyко Ю.А., Загорulyко Г.Б., Боровикова О.І. Технологія створення тематичних інтелектуальних наукових інтернет-ресурсів, що базується на онтології // Програмна інженерія. – 2016. – Т.7, №2. – С. 51–60.