

РОЗРОБКА ТА ДОСЛІДЖЕННЯ ГІБРИДНОГО АЛГОРИТМУ ІНТЕГРАЦІЇ СКЛАДНОСТРУКТУРОВАНИХ ДЖЕРЕЛ

Ушаков Д.В.

Науковий керівник – проф. каф. «Інформаційних систем»,

докт. техн. наук.

Антощук С.Г.

При розвитку інформаційної системи підприємства виникає необхідність у вирішенні проблеми інтеграції складноструктурованих неоднорідних джерел даних (СНД-даних). В основі вирішення цієї проблеми лежить узгодження СНД-даних. До останнього часу для вирішення цієї проблеми використовувався синтаксичний аналіз [1], але такий підхід дозволяє порівнювати лише імена об'єктів та використовує чітку логіку для вирішення слабкоформалізованої задачі.

Пропонується реалізація уніфікованого методу інтеграції на основі гібридного методу, в основі якого лежить нечіткий семантичний аналіз концептів (сутностей імен об'єктів). Метод складається з наступних етапів:

- 1) здобування даних та їх перетворення до уніфікованого вигляду в формі спрямованого графу на основі впорядкованих XML-структур;
знаходження концептів міток вузлів отриманого графу;
знаходження концептів вузлів отриманого графу;
розрахунок зв'язків між концептами міток;
розрахунок зв'язків між концептами вузлів.

Таким чином, запропонований метод використовує нечіткі семантичні відносини для вирішення проблеми інтеграції СНД-даних та не залежить від кількості та типу цих джерел.

1. Madhavan J., Bernstein P., Rahm E. Generic schema matching using Cupid // In Proceedings of the International Conference on Very Large Data Bases – 2001 – P. 48–58.