

Аналіз методів обробки даних у суспільних віртуальних системах

Ірина Форкун

Кафедра фінансів, банківської справи та страхування
Хмельницький національний університет
Хмельницький, Україна
ivforkun@gmail.com

Яворський Ярослав

Кафедра інженерії програмного забезпечення
Хмельницький національний університет
Хмельницький, Україна
torondor6@gmail.com

Abstract. *In this article was considered some of data processing methods in virtual communities. Was discussed some problems of data processing in different systems.*

Ключові слова: обробка даних, Data mining, інтелектуальний аналіз, інформаційно-комунікаційні технології.

ВСТУП

Сучасні інформаційні системи, що автоматизують облік діяльності та документоведення, оперують величезними обсягами оперативних даних що зберігаються. Керівництво і аналітики підприємств, установ та організацій стикаються з проблемами оперативного отримання аналітичної інформації, необхідної для прийняття ефективних управлінських рішень. Для реалізації даного процесу використовуються інформаційно-комунікаційні системи та спільноти, які мають широкі можливості для здійснення пошуку та фільтрації даних, однак вони не є достатньо гнучкими для забезпечення подальшої обробки та проведення аналізу великих об'ємів інформації, виділення корисних даних, що потребує подальшого ґрунтового аналізу існуючих методів для їх вирішення.

МЕТОДИ ОБРОБКИ ДАНИХ У СУСПІЛЬНИХ ВІРТУАЛЬНИХ СИСТЕМАХ

Інформаційні технології, які широко використовуються в різних сферах діяльності сучасного суспільства, дозволяють оптимізувати інформаційні процеси, в тому

числі організувати й автоматизовано обробляти електронні дані.

Аналіз даних в суспільних віртуальних системах зводиться до вирішення наступних завдань:

- цілісності та доступності інформації;
- скорочення втрат отримання помилок та невідповідностей;
- усунення проблем порушення цілісності інформації;
- усунення проблем порушення відповідності інформації при роботі з великими обсягами даних.

Розв'язання поставлених проблем вимагає підбору, компонування та узагальнення існуючих методів для створення оптимального документаційного забезпечення.

Важливу роль в даному процесі відіграють методи:

- аналізу статистичних даних. (перевірка гіпотез стаціонарності, нормальності, незалежності, однорідності, оцінка виду функції розподілу і її параметрів);
- виявлення зв'язків і закономірностей (лінійний і нелінійний регресійний аналіз, кореляційний аналіз);
- багатовимірний статистичний аналіз (лінійний і нелінійний дискримінантний аналіз, кластер-аналіз, компонентний аналіз, факторний аналіз);
- динамічні моделі і прогноз на основі часових рядів.

Найбільш відомі і популярні засоби статистичного аналізу: пакети Statistica, SPSS, Systat, Statgraphics, SAS, BMDP, TimeLab, Data-Desk, SPlus, Scenario (BI). Недоліки: вимога до спеціальної підготовки користувача, більшість методів, ґрунтуються на парадигмі, в якій

головними фігурантами виступають усереднені значення, які при дослідженні реальних складних життєвих феноменів часто виявляються фіктивними величинами.

В даному ракурсі важливими є методи розпізнавання даних з різних сайтів, що забезпечує отримання фрагментів документації на певні теми з різних відкритих ресурсів (декомпозиція та синтез різних наборів тематичних структур веб сторінок) [3].

Вирішення проблеми цілісності та доступності інформації забезпечується шляхом вмілого і правильного розуміння та оперування обмеженими наборами операцій які доступні в ERP системах. Скорочення втрат отримання помилок та невідповідностей може бути забезпечено шляхом обмеження можливостей виконання певних операцій.

До методів і алгоритмів інтелектуального аналізу даних (ІАД) обробки даних у суспільних віртуальних системах можна віднести наступні: штучні нейронні мережі, дерева рішень, символні правила, методи найближчого сусіда і к-найближчого сусіда, метод опорних векторів, байесові мережі, лінійна регресія, кореляційно-регресійний аналіз; ієрархічні методи кластерного аналізу, неієрархічні методи кластерного аналізу, зокрема і алгоритми к-середніх і к-медіани; методи пошуку асоціативних правил, зокрема алгоритм Apriori; метод обмеженого перебору, еволюційне програмування і генетичні алгоритми, різноманітні методи візуалізації даних і безліч інших методів. Варто зазначити, що більшість методів ІАД була розроблена у межах теорії штучного інтелекту. Єдиної думки щодо того, які задачі необхідно зараховувати до ІАД, немає. Більшість авторитетних джерел перераховує такі: класифікація, кластеризація, прогнозування, асоціація, візуалізація, аналіз виявлення відхилень, оцінювання, аналіз зв'язків, підведення підсумків. Загалом такі системи, з обробки і збору інформації мають досить складну структуру[4]. Більшість з них реалізують свою структуру через складні і досить об'ємні бази даних в яких можна помістити досить великі об'єми інформації розділивши їх на багато малих підрозділів по певних параметрам та категоріям, що в свою

чергу значно спрощує подальшу роботу з вже отриманою та обробленою інформацією.

ВИСНОВКИ

Як бачимо є чимало різноманітних методів для аналізу та отримання різного роду інформації, що цілому допомагатиме створенню різного роду документаційного забезпечення. Для вирішення питань більш загального роду виникає необхідність в компонуванні задач: цілісності та доступності інформації; скорочення втрат отримання помилок та невідповідностей; усунення проблем порушення цілісності інформації; усунення проблем порушення відповідності інформації при роботі з великими обсягами даних.

Це допомагає в свою чергу охопити більш широкий спектр джерел інформації, однак не дає можливості уніфікувати всі під один узагальнений метод чи засіб, що, у свою чергу, дозволяє зробити висновок про їх комплексне використання при вирішенні оставленого завдання.

ЛІТЕРАТУРА

- [1] Zhezhnych, P., Tarasov, D. On Restricted Set of DML Operations in an ERP System's Database (2019) *Advances in Intelligent Systems and Computing*, 871, pp. 256-266.
- [2] Zhezhnych, P., Tarasov, D. Methods of data processing restriction in ERP systems (2018) *2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2018 - Proceedings*, 1, art. no. 8526734, pp. 274-277.
- [3] Zhezhnych, P., Markiv, O. Recognition of tourism documentation fragments from webpage posts (2018) *14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering, TCSET 2018 - Proceedings*, 2018-April, pp. 948-951.
- [4] Berezko, O., Zhezhnych, P. Rethinking the NGO website from the knowledge management perspective (2017) *Proceedings of the 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2017*, 1, art. no. 8098812, pp. 389-392.