

# Методи обробки даних у суспільних віртуальних системах

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

На сьогоднішній день системи, що автоматизують облік діяльності та документоведення, оперують величезними обсягами що зберігаються оперативних даних. Також перед усіма хто намагається використовувати суспільні інформаційно-комунікаційні системи та спільноти постає ряд проблем які потребують детального аналізу та вирішення. Це проблеми обробки неструктурованих даних, проблеми обробки великих об'ємів інформації та її аналіз та виділення корисних даних від інших та інші [4].

## МЕТОДИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

Інтелектуальний аналіз даних (Data Mining) – це напрямок в області інформаційних систем, орієнтований на рішення задач підтримки прийняття рішень на основі кількісних і якісних досліджень надвеликих масивів різномірних ретроспективних даних. Являє собою процес пошуку кореляцій, тенденцій, взаємозв'язків і закономірностей за допомогою різних математичних і статистичних алгоритмів.

При обробці даних досить часто можна стикнутися з проблемами цілісності та доступності інформації [1]. Шляхом обмеження можливостей виконання певних операцій таких як CREATE, UPDATE, DELETE буде можливість скорочення витрат отримання помилок та невідповідностей, усунення проблем порушення цілісності та відповідності інформації при роботі з великими обсягами даних [2].

### Методи статистичної обробки даних.

Попередній аналіз природи 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.

### Обробка історичної інформації та інші методи обробки даних.

Розпізнавання даних з різних сайтів є досить важливим для отримання фрагментів

документації на певні теми з різних відкритих ресурсів. Часто в таких випадках використовуються методи декомпозиції та синтезу різних наборів тематичних структур веб сторінок [3].

Також є досить важкою задачею обробка, аналіз та оперування історичною інформацією, так як досить важко отримати достовірну та об'єктивну інформацію з тих чи інших джерел різних віртуальних громад про ті чи інші історичні події [5].

В таких випадках досить часто використовують методи лінгвістичного аналізу даних отриманої інформації. Дані методи підрозумівають під собою детальне і глибоке опрацювання різних даних, наприклад постів користувачів в соціальних мережах, чи коментарів на нас сайтах різних онлайн магазинів, або ж аналіз інформації з постів, повідомлень, коментарів чи відгуків на різних туристичних сайтах, тощо, виключенням з ним шаблонів певних ключових фраз [7].

Загалом такі системи, з обробки і збору інформації мають досить складну структуру. 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 параметрах та категоріям, що в свою чергу значно спрощує подальшу роботу з вже отриманою та обробленою інформацією. Однак якщо не врахувати всі можливі компоненти та ризики це може привести до значних втрат інформації або до отримання інформації, що не буде релевантна відносно того чи іншого питання.

## ВИСНОВКИ

Як бачимо є чимало різноманітних методів для отримання різного роду інформації, що в цілому допомагатиме створенню різного роду документального забезпечення. Важливим є вміння чітко ставити відповідні задачі і підбирати для них відповідні методи.

Для вирішення питань більш загального роду виникає необхідність в компонуванні даних методів. Це допомагає в свою чергу охопити більш широкий спектр джерел

інформації однак не дає можливості уніфікувати всі під один узагальнений метод.

## 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 web-page 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.
- [5] Khymytsia, N., Lisina, S., Morushko, O., Zhezhnych, P. Peculiarities in generating historical information in virtual communities (2017) *Proceedings of the 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2017*, 1, art. no. 8098799, pp. 336-339.
- [6] Zhezhnych, P., Khymytsia, N., Lisina, S., Morushko, O. Analysis of computer-based methods for processing historical information (2017) *Proceedings of the 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2017*, 1, art. no. 8098806, pp. 365-368.
- [7] Zhezhnych, P., Markiv, O. A linguistic method of web-site content comparison with tourism documentation objects (2017) *Proceedings of the 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2017*, 1, art. no. 8098800, pp. 340-343.