

## АНАЛІЗ НЕСТРУКТУРОВАНОГО ТЕКСТУ З ВИКОРИСТАННЯМ МОДЕЛЕЙ ЗНАНЬ НА ОСНОВІ РАНІШЕ НАБУТИХ ЗНАНЬ

Савченко М.М., Крячок О.С.

*Національний Технічний Університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ*

Із розвитком комп'ютерних технологій та мережі інтернет кількість інформації, яка була згенерована людством примножується майже кожного року. Все більш актуальнішим постає питання аналізу накопиченої інформації, адже дуже велика її частина – це неструктурована текстова інформація, яка представлена у вигляді документів, книг, звітів, статей, відгуків, тощо.

Для аналізу машинописного тексту в сучасному світі використовуються підходи та моделі, що здебільшого ефективні для обмеженого кола задач.

Однією з найбільш відомих та простих моделей для витягування інформації з тексту є модель «bag-of-words», або «мішок слів» [1], яка дозволяє провести статистичний аналіз слів в певному тексті і на основі цього виявити його найбільш виражені характеристики. Даний підхід гарно показує себе на практиці [2], але він залишатиметься вузькоспеціалізованим та застосовуваним лише до певних задач через відсутність будь-яких попередніх знань про текст.

Більш складні моделі включають в себе інші підходи до витягування знань з текстів, наприклад, розробку правил витягування інформації кодуванням шаблонів фраз, або, більш досконалий метод, що «навчається» таким правилам з тексту, самостійно визначаючи тип сутностей [3].

Але основною задачею для машини залишається не просто сприйняти текст, основувшись на певних вбудованих правилах або на тих, яким вдалося навчити машину, а і «зрозуміти» цей текст. Для досягнення цього пропонується побудувати асоціативну ймовірнісну модель знань на основі заданого тексту і обов'язково – раніше оброблених. Асоціативна пам'ять, покладена в основу нашої людської природи, дозволяє пригадувати усі пов'язані з сутностями речі, події та характеристики. Така модель дозволяє витягнути знання з тексту без потреби у будь-яких наперед визначених правил. Але тут важливо зауважити, що машині, так як і людині необхідні базові (природні) знання, або набуті асоціації, наприклад такі, яких набуває людина в процесі свого розвитку. Для прикладу, розглянемо речення «На вулиці дощить». Його аналіз машиною має створити не тільки нові асоціації «дощ-вулиця», а ще і пов'язати дані сутності з тими, які могли зустрічатися у текстах раніше: «дощ-вода», «дощ-парасолька». Серед асоціацій можна виділити стійкі асоціації – такі, які мають приблизно однакову силу відношення у різних текстах та нестійкі. Стійкими асоціаціями, наприклад, можуть бути певні правила граматики: «вчора-був», «завтра-буде».

### Література:

1. Aggarwal C.C. Mining Text Data / C.C. Aggarwal, C. Zhai. – New York, USA: Springer Science and Business Media, 2012. – 524 с. – (Library of Congress).
2. Borgen P. Boosting Sales With Machine Learning [Електронний ресурс] / Per Harald Borgen // Medium.com. – 2016. – Режим доступу до ресурсу: <https://medium.com/xeneta/boosting-sales-with-machine-learning-fbcf2e618be3>.
3. J. R. M. Mining Knowledge from Text Using Information Extraction / R. M. J., R. Bunescu. // Department of Computer Sciences. – 2005. – №1. – С. 2–5.