

ПОБУДОВА СЕМАНТИЧНОЇ МЕРЕЖІ ЗНАНЬ У АВТОМАТИЧНОМУ РЕЖИМІ З НАКОПИЧЕННЯМ ЗНАНЬ

Савченко М.М., Крячок О.С.

*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»,
м. Київ*

Знання, які закладені у величезній кількості текстової інформації, які здобуті людством, дуже складно вилучати і обробляти саме через обсяги існуючої та нової інформації. Також існує необхідність розробки систем реального часу, які працюють в повністю автоматичному режимі [1, 2].

Припустимо що нам необхідно проаналізувати текстовий корпус із більше ніж 2500 окремих документів (статей) та отримати граф знань, побудований на основі цих документів. Вилучення усіх можливих концептів та зв'язків між ними утворить величезний граф, який матиме велику кількість абсолютно нерелевантної інформації. Виникає необхідність розробки алгоритмів та програмного забезпечення, які б дозволяли виконувати наступне:

1. Вилучати не всі, а лише найбільш релевантні знання із кожного тексту, поєднуючи їх у єдину загальну семантичну мережу знань.

2. Накопичувати знання: загальновідомі факти мають доповнювати задану інформацію для виявлення нових зв'язків між концептами текстів.

3. Використовувати систему у автоматичному режимі без експертів.

Розроблено програмний комплекс (PoC, Proof of Concept), який вирішує поставлені проблеми та демонструє побудову графів знань із застосуванням:

1. Для вилучення найбільш релевантних у даному контексті, використовується метрики TF-IDF для оцінки концептів (окремих термінів).

2. Розроблений алгоритм не обмежений аналізом одного тексту і легко комбінує декілька графів у один шляхом оцінки схожості концептів.

3. Алгоритм передбачає комбіноване використання інших алгоритмів (TF-IDF для оцінки концептів, PoS Tagging для ідентифікації частин мови на основі нейронних мереж та розроблений алгоритм побудови семантичної мережі за заданими правилами граматики природної мови), які не потребують втручання експертів у систему і є повністю автономними.

Приклад результату роботи алгоритму показано на рис. 1.

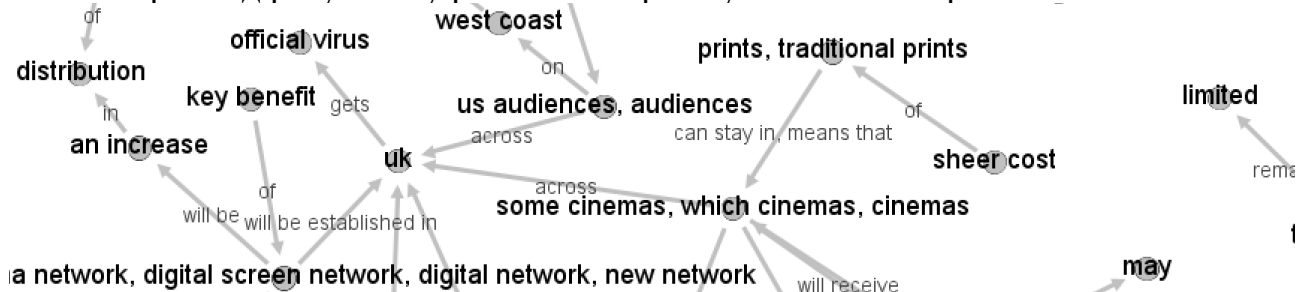


Рисунок 1 – приклад роботи алгоритму побудови семантичної мережі

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. Савченко М.М. Analysis of the relationships between entities in unstructured text based on previously extracted general knowledge [Текст] / М.М. Савченко // Наука та техніка XXI століття – Part V. – Kyiv, 2017. – С. 90-91.