

## ВИЗНАЧЕННЯ СМИСЛОВОЇ БЛИЗЬКОСТІ ТЕКСТІВ ПРИ АНАЛІЗІ КОНТЕНТУ

### В РЕКОМЕНДАЦІЙНИХ СИСТЕМАХ

Крученюк А.Ю., Хайрова Н.Ф.

*Національний технічний університет  
«Харківський політехнічний інститут»,  
м. Харків*

З'явившись більш ніж 20 років тому, рекомендаційні системи залишаються затребуваними і сьогодні. Від звичайних пошукових систем їх вирізняє можливість надання персоналізованих рекомендацій з врахуванням потреб користувачів без їх явного запиту. На сьогодні існують декілька видів таких систем, що класифіковані за підходом фільтрації, на якому вони базовані, а саме Collaborative Filtering (з використанням колаборативної фільтрації), Content-based (фільтрація по контенту) та системи з гібридний підходом.

У нашому дослідженні ми зосереджуємося на задачі фільтрації по контенту, яка оцінює схожості за ознаками вмісту об'єктів. За таким підходом наступна рекомендація користувачеві залежить від якості опису метаданих об'єктів та і уподобань користувача, тоді як інші підходи спираються лише на поведінку користувача.

В якості dataset було розглянуто англomовні описи фільмів починаючи з 60-х років ХХ століття і до сьогодні у жанрі детективу та криміналу. Об'єм корпусу складає 745 текстових документів, з середньою кількістю слів –530 у кожному.

У цьому дослідженні ми визначаємо смислову близькість текстів, що містять описи фільмів, за допомогою визначення семантичної близькості дій героїв фільмів та їхнього опису (їх якісних характеристик). Для цього використовується лінгвістичний аналіз тексту. А саме: (1) виділення в текстах речень, що містять інформацію про головних героїв; (2) виділення граматичних відношень між словами, що представлені іменами героїв, та пов'язаними з ними лемами; (3) побудова вектору, який видобувається з матриці термін-контекст, де термін – слово, що називає головного героя, а контекст представлений у вигляді кортежу, який включає слово та тип відношення; (4) обчислення косинусної схожості між векторами.

#### **Література:**

1. Marie-Catherine de Marneffe and Christopher D. Manning Stanford typed dependencies manual, Revised for the Stanford Parser v. 3.7.0, USA, 2016.
2. David Bamman, Brendan O'Connor Noah, A. Smith Learning Latent Personas of Film Characters, 352–361, Pittsburgh, USA, 2013.
3. Content-Based Recommendation System – URL: <https://medium.com/towards-artificial-intelligence/content-based-recommender-system-4db1b3de03e7> (дата звернення 11.03.2020).