



## ПРОГРАМНО-ТЕХНІЧНІ КОМПЛЕКСИ

УДК 004.912

**А.С. КОЛЕСНИК**

Національний технічний університет «Харківський політехнічний інститут», Харків,  
Україна, e-mail: [kolesniknastya20@gmail.com](mailto:kolesniknastya20@gmail.com).

**Н.Ф. ХАЙРОВА**

Національний технічний університет «Харківський політехнічний інститут», Харків,  
Україна, e-mail: [nina\\_khajrova@yahoo.com](mailto:nina_khajrova@yahoo.com).

### ОБҐРУНТУВАННЯ ВИКОРИСТАННЯ СТАТИСТИКИ КАППА КОЕНА В ЕКСПЕРИМЕНТАЛЬНИХ ДОСЛІДЖЕННЯХ NLP ТА TEXT MINING

**Анотація.** Виконано порівняння сучасних метрик оцінювання узгодженості (agreement coefficients) між результатами експериментів і експертною думкою та оцінено можливість використання цих метрик під час проведення експериментальних досліджень у галузі автоматичного оброблення текстів методами машинного навчання. Обґрунтовано вибір коефіцієнта каппа Коена як міри оцінювання узгодженості думок експертів у задачах NLP та Text Mining. Наведено приклад застосування коефіцієнта каппа Коена для оцінювання рівня узгодженості між думкою експерта і результатами ML класифікації та міри узгодженості думок експертів у випадку вирівнювання речень казахсько-російського паралельного корпусу. На підставі наведеного аналізу доведено, що завдяки зручності у використанні, простоті обчислення та високій точності результатів коефіцієнт каппа Коена є одним з найкращих статистичних методів визначення рівня узгодженості в експериментальних дослідженнях текстів.

**Ключові слова:** Text Mining, NLP, статистика каппа Коена, коефіцієнт узгодженості, класифікація текстів методами машинного навчання, паралельний корпус.

#### ВСТУП

У сучасних умовах розвитку науки гостро постає питання оцінювання результатів досліджень. Для того, щоб можна було надалі спиратись і посила-тися на одержані результати експериментів, вони мають бути повністю об'єктивними та максимально точними. Ці вимоги стосуються як кількісних результатів наукової роботи, так і якісних результатів досліджень.

Оцінювання достовірності експериментів, проведених на тому чи іншому корпусі під час розв'язання задач Natural Language Processing (NLP) або Text Mining, також є одним з основних складних завдань, пов'язаних з автоматичним обробленням текстів. Зазвичай, результати застосування методів і моделей інформаційного пошуку, класифікації, кластеризації, морфологічного розмічування, синтаксичного парсера та інших задач оцінюють або із залученням експертів, або методом порівняння з так званим «золотим стандартом», тобто корпусом, розміченим деякими цільовими значеннями (target value), пов'язаними із завдан-нями Machine Learning (ML). При цьому традиційно використовують такі мет-

© А.С. Колесник, Н.Ф. Хайрова, 2022