

## РЕФЕРАТ

*Магістерська дисертація: 104 с., 39 рис, 26 таб., 3 додатки, 21 використаних джерел.*

**Актуальність теми:** задача адаптації методів NLP до українських текстів досі не вирішена. Саме тому створення якісного і уніфікованого методу аналізу змішаної прямої мови (української та російської, а також суржика та сленгових виразів) є нагальною.

**Мета дослідження:** адаптувати методи аналізу прямої мульти-лінгвістичної мови для застосування до україномовних чатів, підвищити якість аналізу прямої мови, що містить русизми та сленгові вирази

**Об'єкт дослідження:** підходи та процеси створення програмного забезпечення для класифікації та категоризації повідомлень соціальних мереж та месенджерів.

**Предмет дослідження:** методи, моделі, програмна архітектура засоби створення програмного забезпечення з аналізу, класифікації та категоризації мульти-лінгвістичної та насиченої сленговими виразами прямої мови.

Для реалізації поставленої мети сформульовані наступні **завдання:**

- проаналізувати наявні рішення для аналізу прямої мультилінгвістичної мови;
- провести аналіз та адаптувати існуючі методології по опрацюванню мультилінгвістичного тексту для застосування до мов слов'янської групи. А саме української та російської;
- удосконалити та адаптувати методи аналізу мультилінгвістичних даних для застосування до мов слов'янської групи, а саме української та російської, з урахуванням опрацювання сленгових виразів сучасної української мови;
- розробити архітектуру та програмне забезпечення з реалізованими методами аналізу мультилінгвістичних даних;
- розробити план по масштабуванню та розгортанню програмного забезпечення у хмарі;

- провести дослідження ефективності розроблених методів та програмного забезпечення.

**Наукова новизна:** у рамках роботи було удосконалено наявні методи опрацювання української прямої мови, також було формалізовано та адаптовано методи класифікації прямої мови що містить русизми, суржик та сленгові вирази.

**Практичне значення отриманих результатів** визначається тим, що розроблена програмна бібліотека дозволить автоматизувати класифікацію та категоризацію мультилінгвістичних повідомлень з соціальних мереж та месенджерів, а також гарантує високу ефективність при застосуванні до слов'янських мов (а саме української та російської, а також суржику і сленговим виразам)

**Зв'язок роботи з науковими програмами, планами, темами:** Зв'язок з науковими програмами, планами, темами. Робота виконувалась на кафедрі інформатики та програмної інженерії Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського" в рамках теми «Методи та технології високопродуктивних обчислень та обробки надвеликих масивів даних». Державний реєстраційний номер 0117U000924.

**Апробація:** Наукові положення дисертації пройшли апробацію на V всеукраїнській науково-практичній конференції молодих вчених та студентів «Інженерія програмного забезпечення і передові інформаційні технології SoftTech-2023» – м. Київ.

**Публікації:** Наукові положення дисертації опубліковані в тезах конференції «Інженерія програмного забезпечення і передові інформаційні технології SoftTech-2023»

**Ключові слова:** АНАЛІЗ ПРЯМОЇ МУЛЬТИЛІНГВІСТИЧНОЇ МОВИ, НЕЙРОННІ МЕРЕЖІ, МОДЕЛЬ МІШКА СЛІВ, КЛАСИФІКАЦІЯ, КАТЕГОРИЗАЦІЯ