

## РЕФЕРАТ

Магістерська дисертація: 76 с., 15 рис., 26 табл., 1 додаток, 15 джерел.

**Актуальність.** Ідентифікація та перевірка авторства є унікальною і, водночас, дуже широко розповсюдженою задачею, з огляду на можливість застосування у різних сферах діяльності людини: для боротьби з плагіатом, для встановлення авторства анонімних текстів, для експертизи та встановлення особистості в криміналістиці та у багатьох інших задачах та напрямках.

Проблеми дослідження стилю автора та застосування статистичного аналізу в дослідженні авторства розглядалася в дослідженнях О.О. Архипової та В.М. Журавльова, Л.А.Борисова, Ю.Н.Орлова та К.П. Осмініна.

Задача є також дуже складною через фундаментальну проблему формування набору ознак, за якими можна оцінити ймовірність належності тексту певному автору.

Задачу ускладнює також той факт, що до останнього часу для розроблених систем визначення авторства текстів необхідною умовою їх стійкої та якісної роботи була наявність великих об'ємів авторських текстів у навчальній вибірці. Ще однією вадою розроблених моделей є їх якісне обмеження на кількість авторів.

**Мета дослідження.** Метою дослідження є розробка алгоритму та реалізація програмного забезпечення аналізу автора художнього тексту за допомогою методів статистичного аналізу, також встановлення впливу імен героїв та слів з великої літери в цілому на частотні характеристики тексту.

**Завдання дослідження.** Для досягнення мети необхідно виконати наступні завдання:

- дослідити праці авторів в цій галузі;

- проаналізувати існуючі системи пошуку автора;
- розробити алгоритм аналізу авторства;
- розробити програмний додаток заснований на розробленому алгоритмі;
- проаналізувати точність роботи алгоритму;
- зробити висновки щодо доцільності використання алгоритму.

**Об'єкт дослідження** – художні тексти.

**Предмет дослідження** – аналіз художніх текстів з метою визначення авторства.

**Методи дослідження.** Використовувалися методи статистичного аналізу для побудови алгоритму, методи об'єктно-орієнтованого програмування для розробки програмного додатку.

**Наукова новизна.** Найбільш суттєвими науковими результатами магістерської дисертації є:

- розроблений алгоритм дає високу точність при визначенні авторства художніх текстів;
- було досліджено вплив імен героїв та слів з великої літери в цілому на частотні характеристики тексту.

**Практичне значення.** Отримане програмне забезпечення дозволяє проводити аналіз художніх текстів та з високою точністю визначати автора тексту.

**Зв'язок роботи з науковими програмами, планами і темами.** Робота виконувалась на кафедрі автоматизованих систем обробки інформації та управління Національного технічного університету України «Київський політехнічний інститут ім. Ігоря Сікорського» в рамках теми «Математичні

моделі та технології в СППР». Державний реєстраційний номер 0117U000914.

**Апробація.** Основні положення роботи доповідались і обговорювались на XVIII міжнародної науково-практичної конференції «Математичне та програмне забезпечення інтелектуальних систем МПЗІС-2019» та 3 всеукраїнської науково-практичної конференції молодих вчених та студентів «Інформаційні системи та технології управління» (ІСТУ-2019).

**Публікації.** Наукові положення дисертації опубліковані в тезах XVIII міжнародної науково-практичної конференції «Математичне та програмне забезпечення інтелектуальних систем МПЗІС-2019» та в тезах 3 всеукраїнської науково-практичної конференції молодих вчених та студентів «Інформаційні системи та технології управління» (ІСТУ-2019).

**Ключові слова:** Статистичний аналіз текстів, Визначення авторства.