

ФОРМАЛЬНІ МЕТОДИ ІДЕНТИФІКАЦІЇ АВТОРА ТЕКСТУ

Гвозденко М.В., ст. викл.

*Національний університет «Юридична академія України
імені Ярослава Мудрого»*

Розвиток мережевих технологій привів до небувалого збільшення поширюваних текстів, їх несанкціонованого використання і плагіату.

Величезний штат копірайтерів займається редагуванням чужих текстів і рерайтингом, що приводить до масового порушення авторських прав.

Ідентифікація автора тексту – це набір методів встановлення автора по приватних особливостях тексту.

Основною проблемою встановлення авторства є саме визначення приватних ознак, причому ці ознаки повинні відповідати декільком умовам: ознака повинна відображати ті характеристики тексту, які автор використовує підсвідомо, зберігати постійне значення для одного автора і мати істотно різні значення для різних авторів.

Формальні методи ідентифікації автора тексту широко використовуються в розробці комп'ютерних програм визначення авторства, що приводить до значного підвищення якості і швидкості проведення авторведчеських експертиз.

У доповіді наведені формальні методи визначення автора тексту, розглянуті метод порівняння гістограм частот слів різної довжини, 'наївний байесовский' (НБА) метод, метод розподілу параметра, порівняння кількостей нових слів в суперечливому тексті, метод відносної ентропії, метод стійкості частот, індекс Флеша, FOG-індекс, підхід Колтарда, лінгво-статистичний аналіз неповнозначної лексики, розпізнавання автора тексту з використанням ланцюгів А.А. Маркова, атрибуція Минимальной Условной Сложности Сжатия (МУСС — атрибуція), апарат машини опорних векторів (Support Vector Machines, SVM). Доведена беззаперечна перевага апарата нейронних мереж порівняно зі статистичними методами.