

## ПРОГРАМНА СИСТЕМА АВТОРСЬКОЇ АТРИБУЦІЇ ТЕКСТІВ НА ФОНОЛОГІЧНОМУ РІВНІ

Встановлення авторства текстів є актуальним у судочинстві (анонімні листи, заповіти, інші судові документи); у визначенні відсотку творчого внеску кожного із кількох авторів в наукову статтю; в документах офіційно-ділового стилю та творах художнього стилю.

При встановленні авторства тексту застосовують кількісний та статистичний аналіз. Перший передбачає підрахунок мовних одиниць з метою виявлення певних кількісних закономірностей їх функціонування. В статистичному аналізі використовують формули математичної статистики. Авторську атрибуцію здійснювали з давніх давен, однак методологічну базу розроблено в середині ХХ століття. Атрибуцію авторства виконують на різних рівнях мови. У більшості робіт використовують лексичний (Ш. Аргамон, Г. Єрмоленко, М. Коппель, Дж. Пенбейкер, Дж. Шлер) [1 - 3] та синтаксичний (Ю. Бойко, П. Вашак) рівні мови [4], [5].

У даному дослідженні для здійснення авторської атрибуції вибрано фонологічний рівень. Кількість одиниць цього рівня є незмінною, зв'язки між ними є стійкими. Тому структура є більш строгою. Це полегшує формалізацію та математизацію мовних явищ. Метою дослідження є підвищення ефективності авторської атрибуції. Новизною дослідження є розроблені: 1) метод комплексного аналізу диференціації фоностатистичних структур стилів; 2) багатофакторний метод визначення ступенів дії факторів стилю, підстилю та авторської манери викладу.

Метод комплексного аналізу ґрунтується на запропонованому поєднанні таких статистичних методів: гіпотез, ранжування та визначення відстаней між стилями. За методом гіпотез тексти різних авторів диференціюються за критерієм Стьюдента. Ступінь розбіжності між текстами визначено за кількістю груп приголосних фонем, за якими встановлено істотні відмінності. За методом ранжування тексти диференціюються за різницею рангових показників. За методом визначення відстаней між стилями визначають величину  $l$ :  $l = \frac{t - t_0}{t_0}$  (величина  $t_0$  є табульованою для рівня значущості 5%).

Багатофакторний метод передбачає визначення ступеня дії факторів: 1) стилю (стилева атрибуція); 2) підстилю (підстилева атрибуція); 3) авторської манери викладу (авторська атрибуція). Розроблені методи програмно реалізовано на мові програмування java. Структура програмної системи складається з наступних модулів: 1) введення/виведення даних; 2) перетворення тексту в транскрипційний варіант; 3) визначення кількості фонем у вибірці; 4) визначення середніх частот груп приголосних фонем; 5) перевірки тексту на нормальність (критерій Пірсона); 6) визначення критерія Стьюдента. Для тестування програмної системи вибрано два тексти з художньої прози Дж. Роулінг. Тексти одного автора, стилю, підстилю та схожій тематики повинні мати незначні відмінності. За критерієм Пірсона встановлено, що частоти груп приголосних фонем підлягають нормальному розподілу (рис. 1).

Результати, отримані за критерієм Стьюдента, підтверджують, що розбіжності між текстами є неістотними (рис. 2).

Отже, встановлено фоностатистичні параметри авторського стилю Дж. Роулінг, за якими можна відрізнити тексти даного автора від текстів інших авторів.

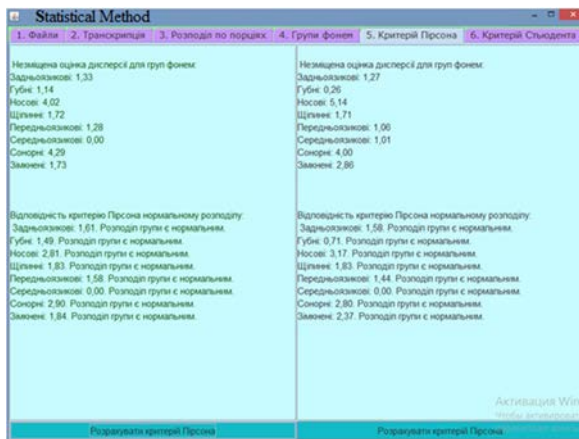


Рис. 1. Результати тестування за критерієм Пірсона

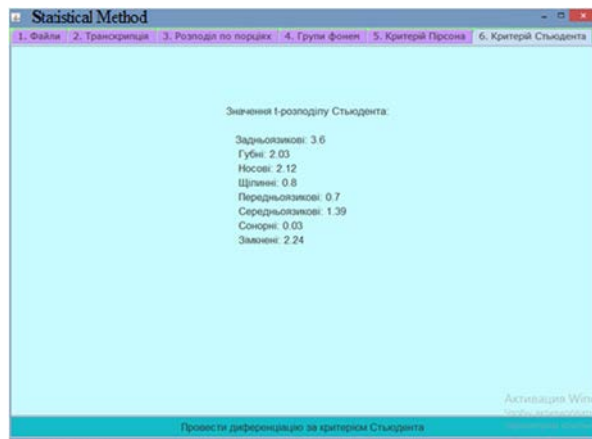


Рис. 2. Результати тестування текстів Дж. Роулінг за критерієм Стьюдента

Розроблена програмна система диференціації фоностатистичних структур стилів базується на розроблених методах: 1) комплексного аналізу диференціації фоностатистичних структур стилів та 2) багатofакторному. Система дає змогу ефективніше здійснити авторську атрибуцію текстів.

1. Ермоленко Г. В. *Анонимные произведения и их авторы.* – Минск: Изд-во «Университетское», 1988. 2. Sh. Argamon, M. Koppel, J. Pennebaker and J. Schler. *Automatically Profiling the Author of an Anonymous Text. Communications of the ACM, vol. 52, issue 2, USA, 2009, pp. 119–123.* 3. M. Koppel, J. Schler and Sh. Argamon. *Computational Methods in Authorship Attribution. Journal of the Association for Information Science and Technology, vol. 60(1), USA, 2009, pp. 9–26.* 4. Бойко Ю. Диференційні параметри речення як детермінанта авторського стилю // *Проблеми квантитативної лінгвістики.* – Чернівці: Рута, 2005 – С. 292–305. 5. Вашак П. Длина слова и длина предложения в текстах одного автора // *Вопросы статистической стилистики. Сб. науч. ст.* – К.: Наукова думка, 1974. – 331 с.

Процько І. О., Городисько В. Є.  
Національний університет «Львівська політехніка»

## АНАЛІЗ ПРОДУКТИВНОСТІ GPGPU ГРАФІЧНИХ ПРОЦЕСОРІВ НА ПРИКЛАДІ ВІДЕОКАРТИ GEFORCE GTX 1050

Високопродуктивні обчислювальні системи все більш стають затребуваними в різних застосуваннях та дослідженнях: атомній фізиці, астрономії, вивченні клімату, створенні нових ліків, оборонній промисловості, фінансових системах та моніторингу. Ще зовсім недавно матеріальний поріг входження в такі обчислення для інженерів та дослідників з обчислювально ресурсоемними завданнями був непереборно високий. Наприклад, для задач обробки і візуалізації потокового аудіовізуального контенту високої якості, не могло бути й мови. Знизити цей поріг змогли розробники відеокарт, реалізувавши можливість здійснювати GPGPU (General Programming on Graphics Processing Unit) обчислення загального призначення. Графічні процесори (GPU) з тисячами обчислювальних ядер, зробили доступнішими дану гібридну модель обчислень реалізовані за принципом SIMT (single instruction multiple threads) завдяки кратному зменшенню енергоспоживання та кратному зниженню кінцевої вартості [1].

На кафедрі ICT в ІППТ в рамках курсу "Паралельні обчислення та Grid-технології", який входить в базову частину професійного циклу освітньої програми магістра, було