

УДК 004.912

*Ланде Д.В., д.т.н., Дмитренко О.О.*

*Інститут проблем реєстрації інформації НАН України  
Національний технічний університет «Київський політехнічний  
інститут імені Ігоря Сікорського»*

## **ВИКОРИСТАННЯ НАПРАВЛЕНИХ ЗВАЖЕНИХ МЕРЕЖ ТЕРМІНІВ ДЛЯ ВИЗНАЧЕННЯ СТУПЕНЯ ПОДІБНОСТІ ТЕКСТІВ**

Ера стрімкого розвитку інформаційно-телекомунікаційних технологій та інформаційного простору характеризується сплеском неструктурованих та дублюючих даних, зокрема текстових, що значно ускладнює пошук необхідної інформації. Тому актуальною є задача формалізації текстових даних й представлення їх у формі, яка була б зручною для комп'ютеризованої обробки з метою виявлення подібної або дублюючої інформації.

Мета роботи – представити методику визначення ступеня подібності між текстовими документами, що базується на використанні направлених зважених мереж термінів [1], де вузлами таких мереж є ключові терміни тексту, а ребра – семантико-семантичні зв'язки між цими термінами у тексті.

Побудова мереж термінів здійснюється в декілька етапів [1], що включають попередню обробку текстових даних, виокремлення ключових термінів, застосування алгоритму графа горизонтальної видимості для встановлення ненаправлених зв'язків між термінами, а також подальше встановлення напрямків зв'язків та їх вагових значень. Подальше порівняння за допомогою міри Фробеніуса матриць направлених зважених мереж, що отримані для різних текстів, дозволяє визначити семантичну близькість відповідних текстів. Представлена методика може бути використана у процесі обробки інформаційних запитів під час інформаційного пошуку, даючи змогу визначити ступінь відповідності документа до інформаційних потреб користувача та підвищити пертинентність.

### **Використані джерела**

1. *Dmytro Lande, Oleh Dmytrenko, Using Part-of-Speech Tagging for Building Networks of Terms in Legal Sphere // Proceedings of the 5th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2021). Volume I: Main Conference Lviv, Ukraine, April 22-23, 2021. CEUR Workshop Proceedings (ceur-ws.org). - Vol-2870. - pp 87-97. ISSN 1613-0073.*