

Державна наукова установа “Інститут інформації, безпеки і права
Національної академії правових наук України”

Національна бібліотека України ім. В.І. Вернадського
Національної академії наук України

Відкритий міжнародний університет розвитку людини “Україна”

ISSN 2616-6798

ІНФОРМАЦІЯ І ПРАВО

НАУКОВИЙ ФАХОВИЙ ЖУРНАЛ

№ 2(41)/2022

Зареєстрований Міністерством юстиції України
(Свідоцтво про державну реєстрацію друкованого засобу масової інформації:
Серія КВ № 20117-9917ПР від 05.07.13 р.)

Згідно з Наказом МОН України від 02.07.20 р. № 886 (додаток 4) журнал включено до Переліку наукових фахових видань України, категорія “Б”, галузь науки - юридичні, спеціальність - 081. У журналі можуть публікуватися матеріали стосовно дисертаційних робіт на здобуття наукових ступенів кандидата наук (доктора філософії – Ph.D.) і доктора наук у галузі юридичних наук. Друковане періодичне видання ІНФОРМАЦІЯ І ПРАВО внесене в міжнародну базу даних періодичних видань, згідно відповідного номеру ISSN, розміщується на інформаційній платформі “Наукова періодика України”, через яку здійснюється інтеграція з регіональним Реєстром DOI, Системою CrossRef, Міжнародним реєстром ORCID.

м. Київ

State Scientific Institution “Institute of Informatics, Security and Law of
National Academy of Law Sciences of Ukraine”

Vernadsky National Library of Ukraine of
National Academy of Sciences of Ukraine

Open International University of Human Development “Ukraine”

ISSN 2616-6798

INFORMATION AND LAW

SCIENTIFIC PROFESSIONAL JOURNAL

№ 2(41)/2022

Registered by Ministry of Justice of Ukraine
(Certificate of state registration of printed communication media:
KV Series № 20117-9917PR dated 05.07.13)

Pursuant to Order of the Ministry of Education and Science of Ukraine dated 02.07.20 № 886
(Annex 4), the journal is included in the List of scientific professional publications of Ukraine,
category “B”, branch of science - legal, specialty - 081.

The journal can publish materials related to thesis works aimed on the receipt of scientific degrees of
Doctor of Philosophy – Ph.D. (candidate of sciences) and Doctor of Sciences
in the area of Juridical Science.

The printed journal INFORMATION AND LAW is included in the international database of
journal, in accordance with relevant ISSN number, is placed on the information platform “Scientific
Periodicals of Ukraine”, through which integration with the regional DOI Register, CrossRef System,
ORCID International Register is carried out.

УДК 002:340+316.4+338.46

Наукова рада журналу

- Пилипчук Володимир Григорович**, доктор юридичних наук, професор,
академік НАПрН України – *голова наукової ради.*
- Бебик Валерій Михайлович**, доктор політичних наук, професор – *зас. голови наукової ради.*
- Дубровіна Любов Андріївна**, доктор історичних наук, професор, член-кореспондент
НАН України – *зас. голови наукової ради.*
- Копан Олексій Володимирович**, доктор юридичних наук, професор.
- Куйбіда Василь Степанович**, доктор наук з державного управління, професор.
- Марущак Анатолій Іванович**, доктор юридичних наук, професор.
- Нор Василь Тимофійович**, доктор юридичних наук, професор, академік НАПрН України.
- Онщенко Олексій Семенович**, доктор філософських наук, професор, академік НАН України.
- Петришин Олександр Віталійович**, доктор юридичних наук, професор, академік НАПрН України.
- Покутний Сергій Іванович**, доктор фізико-математичних наук, професор.
- Савінова Наталія Андріївна**, доктор юридичних наук, с.н.с.
- Скулиш Євген Деонізієвич**, доктор юридичних наук, професор.
- Таланчук Петро Михайлович**, доктор технічних наук, професор.
- Тихий Володимир Павлович**, доктор юридичних наук, професор, академік НАПрН України.
- Фурашев Володимир Миколайович**, кандидат технічних наук, доцент, с.н.с.
- Шемшученко Юрій Сергійович**, доктор юридичних наук, професор, академік НАН України.

Редакційна колегія

- Буханевич Олександр Миколайович**, доктор юридичних наук, професор,
член-кореспондент НАПрН України
– *голова редакційної колегії.*
- Брижко Валерій Михайлович**, доктор філософії з юридичних наук, с.н.с.
– *зас. голови редакційної колегії.*
- Довгань Олександр Дмитрович**, доктор юридичних наук, професор
– *зас. голови редакційної колегії.*
- Арістова Ірина Василівна**, доктор юридичних наук, професор.
- Баранов Олександр Андрійович**, доктор юридичних наук, с.н.с.
- Беднарук Вальдемар**, доктор габілітований (Люблінський католицький університет, Польща).
- Беляков Костянтин Іванович**, доктор юридичних наук, професор.
- Вронська Тамара Василівна**, доктор історичних наук, с.н.с.
- Дзьобань Олександр Петрович**, доктор філософських наук, професор.
- Доронін Іван Михайлович**, доктор юридичних наук, доцент.
- Золотар Ольга Олексіївна**, доктор юридичних наук, с.н.с.
- Корж Ігор Федорович**, доктор юридичних наук, с.н.с.
- Ланде Дмитро Володимирович**, доктор технічних наук, професор.
- Настюк Василь Якович**, доктор юридичних наук, професор, член-кореспондент НАПрН України.
- Ткачук Тарас Юрійович**, доктор юридичних наук, доцент.
- Чистоклетов Леонтій Григорович**, доктор юридичних наук, професор.
- Шевчук Олександр Михайлович**, доктор юридичних наук, доцент.
- Шеффлер Томаш**, доктор філософії з юридичних наук (Вроцлавський університет, Польща).

* * * * *

UDC 002:340+316.4+338.46

THE SCIENTIFIC COUNCIL OF THE JOURNAL

- Pylypchuk Volodymyr**, Doctor of Juridical Science, Professor,
Academician NALS of Ukraine – *Chairman of Editorial Board*.
- Bebyk Valerii**, Doctor of Political Sciences, Professor – *Vice-chairman of Editorial Board*.
- Dubrovina Lyubov**, Doctor of Historical Sciences, Professor, Corresponding Member National
Academy of Sciences of Ukraine – *Vice-chairman of Editorial Board*.
- Furashev Volodymyr**, Candidate of Engineering Sciences, Associate Professor,
Senior researcher fellow.
- Kopan Oleksii**, Doctor of Juridical Science, Professor.
- Kuibida Vasyl**, Doctor of Administration Science, Professor.
- Marushchak Anatolii**, Doctor of Juridical Science, Professor
- Nor Vasyl**, Doctor of Juridical Science, Professor, Academician NALS of Ukraine.
- Onishchenko Oleksii**, Doctor of Philosophical Science, Professor, Academician NAN of Ukraine.
- Petryshin Oleksandr**, Doctor of Juridical Science, Professor, Academician NALS of Ukraine.
- Pokutnyi Serhii**, Doctor of Physics and Mathematics Sciences, Professor.
- Savinova Nataliia**, Doctor of Juridical Science, Senior researcher fellow.
- Shemshuchenko Yurii**, Doctor of Juridical Science, Professor, Academician NAN of Ukraine.
- Skulysh Ievhen**, Doctor of Juridical Science, Professor.
- Talanchuk Petro**, Doctor of Engineering Sciences, Professor.
- Tykhyi Volodymyr**, Doctor of Juridical Science, Professor, Academician NALS of Ukraine.

EDITORIAL BOARD

- Bukhanevych Oleksandr**, Doctor of Juridical Science, Professor, Corresponding Member National
Academy of Sciences of Ukraine – *Editor in Chief*.
- Bryzhko Valerii**, Doctor of Philosophy of Juridical Science, Senior researcher fellow
– *Vice-Editor*.
- Dovgan Oleksandr**, Doctor of Juridical Science, Professor – *Vice-Editor*.
- Aristova Iryna**, Doctor of Juridical Science, Professor.
- Baranov Oleksandr**, Doctor of Juridical Science, Senior researcher fellow.
- Bednaruk Waldemar**, Doctor habilitowany (Catholic University of Lublin, Poland).
- Bieliakov Konstantyn**, Doctor of Juridical Science, Professor.
- Chistokletov Leontiy**, Doctor of Juridical Science, Professor.
- Dz'oban Oleksandr**, Doctor of Philosophical Science, Professor.
- Doronin Ivan**, Doctor of Juridical Science, Associate Professor.
- Zolotar Olga**, Doctor of Juridical Science, Senior researcher fellow.
- Korzh Ihor**, Doctor of Juridical Science, Senior researcher fellow.
- Lande Dmytro**, Doctor of Engineering Sciences, Professor.
- Nastiuk Vasyl**, Doctor of Juridical Science, Professor, Corresponding Member NALS of Ukraine.
- Tkachuk Taras**, Doctor of Juridical Science, Associate Professor.
- Shevchuk Oleksandr**, Doctor of Juridical Science, Associate Professor.
- Schaffler Tomasz**, Doctor of Philosophy of Juridical Science (University of Wroclaw, Poland).
- Vronska Tamara**, Doctor of Historical Science, Senior researcher fellow.

* * * * *

З М І С Т

Інформаційне право

ЯЩЕНКО В.А. Інформаційна свідомість як фактор війни Росії проти України.....	9
КОРЖ І.Ф., КОРЖ Т.І. Правова інформація як база для ухвалення окремих юридично значимих рішень.....	18
БРАЙЧЕВСЬКИЙ С.М. Проблема надання нормативно-правової інформації в особливих умовах.....	28
ЛІСОВСЬКА Ю.П. Правове забезпечення інтелектуально-ціннісного ресурсу в системі інформаційного капіталу кіберсучасності.....	37

Правова інформатика

ЛАНДЕ Д.В., ДМИТРЕНКО О.О. Побудова семантичних мереж та визначення ступеня розбіжності текстів.....	44
ДУБНЯК М.В., СОЛОВЙОВА В.С. Цифрова трансформація юридичних послуг.....	52
КОСТЕНКО О.В. Аналіз національних стратегій розвитку штучного інтелекту.....	58
КРИЛОВА Ю.І. Електронна демократія в Україні: теоретичний і практичний аспекти.....	70

Інформаційна і національна безпека

КУДІНОВ С.С. Державний тероризм – визначення та характеристика.....	78
МЕЛЬНИК Д.С., ЛЕОНОВ Б.Д. Поняття та зміст кримінологічної характеристики фінансування тероризму.....	85
КОВАЛЬЧУК А.Ю., ГАВЛОВСЬКИЙ В.Д. Інформаційно-психологічні впливи як засіб маніпуляції свідомістю, що застосовується організованими злочинними угрупованнями.....	94
РАДУТНИЙ О.Е. Публічні заклики та заперечення як форми інформаційної колабораційної діяльності за Кримінальним кодексом України...	99
ГУЦАЛЮК М.В., АНТОНЮК П.Є. Процесуальна спроможність використання електронної (цифрової) інформації як доказу в кримінальному провадженні.....	116
ТАРАН О.В., ГАВЛОВСЬКИЙ В.Д. Правове значення інформації, що міститься в єдиному реєстрі досудових розслідувань та окремі питання її використання у правозастосуванні.....	123
ДОРОГИХ С.О. Щодо питань інформаційної безпеки як напряду інформаційної політики України в умовах війни.....	133

Інформація за іншими предметними напрямками досліджень за спеціалізаціями в галузі знань 08 – “Право”

МАНЬГОРА В.В. Розвиток дистанційної юридичної освіти на сучасному етапі в Україні.....	138
САНДУЛ В.С., ПАЛАМАРЮК Т.Д. Окремі особливості викладання предметів в освітніх закладах в умовах воєнного стану....	149
МАНЬГОРА Т.В. Правове регулювання протидії торгівлі людьми.....	154

До відома читачів

Стратегія кібербезпеки України: “Безпечний кіберпростір – запорука успішного розвитку країни”: Указ Президента України від 26 серпня 2021 року № 447/2021.....	164
---	------------

Питання Національної ради з відновлення України від наслідків війни: Указ Президента України від 21 квітня 2022 року № 266/2022.....	180
---	------------

До відома авторів.....	185
-------------------------------	------------

Наукове редагування, створення оригінал-макета та дизайн – Брижко В.М.
Граматичне коректування – Майстренко І.А. (укр., англ.).

Формат 70 x 108/16. Спосіб друку – різнографія. Ум. друк. арк. 16.4. Тираж 100 прим.
Виготовлено з оригінал-макета в друкарні ТОВ “Видавничий дім “АртЕк”.
04050, м. Київ, вул. Мельникова, буд. 63. Свідоцтво про внесення суб’єкта видавничої справи до державного реєстру видавців, виготівників і розповсюджувачів видавничої продукції – серія № ДК № 4779 від 15.10.14 р.

Рекомендовано до друку Вченою радою ДНУ ІБП НАПрН України, протокол № 6 від 27.06.22 р.

TABLE OF CONTENTS

Informative Law

YASHCHENKO V. Information consciousness as a factor in Russia's war against Ukraine.....	9
KORZH I., KORZH T. Legal information as a basis for making certain some legally significant decisions.....	18
BRAYCHEVSKYY S. The problem of providing regulatory information under special conditions.....	28
LISOVSKA Y. Legal support of intellectual and value resource in the cyber-modern information capital system.....	37

Legal Informatics

LANDE D., DMITRENKO O. Construction of semantic networks and determining the degree of divergence of texts.....	44
DUBNYAK M., SOLOVYOVA V. Digital transformation of legal services.....	52
KOSTENKO O. Analysis of national artificial intelligence development strategies.....	58
KRYLOVA Y. Electronic democracy in Ukraine: theoretical and practical aspects.....	70

Informative and National Safety

CUDINOV S. State terrorism – definitions and characteristics.....	78
MELNYK D., LEONOV B. The concept and content of criminological characteristics of terrorism financing.....	85
KOVALCHUK A., GAVLOVSKY V. Information and psychological influences as a means of manipulating consciousness used by organized criminal groups.....	94
RADUTNIY O. Public appeals and objections as forms of information collaboration activities under the Criminal Code of Ukraine.....	99
GUTSALYUK M., ANTONIUK P. Procedural capacity of use electronic (digital) information as evidence in criminal proceedings...	116
TARAN O., GAVLOVSKY V. Legal significance of information contained in the unified register of pre-trial investigations and individual issues of its use in law enforcement.....	123
DOROGICH S. On information security issues as a direction of Ukraine's information policy in war.....	133

Information on other subject research directions by specializations in the field of knowledge 08 – “Law”

MANGORA V. Development of distance legal education at the present stage in Ukraine.....	138
SANDUL V., PALAMARYUK T. Individual features of teaching subjects in educational institutions in conditions of martial law...	149
MANGORA T. Legal regulations against human trafficking.....	154

For the consideration of readers

Ukraine’s cybersecurity strategy: “Secure cyberspace is a guarantee of successful development of the country”: Decree of the President of Ukraine, dated August 26, 2021 № 447/2021.....	164
Issues of the National Council for the Restoration of Ukraine from the Consequences of the War: Decree of the President of Ukraine, dated April 21, 2022 № 266/2022.....	180

For the consideration of authors..... 185

Recommended for publication by the IISL of the NALS of Ukraine, protocol № 6 dated 27.06.22.

Правова інформатика

УДК 004.912

ЛАНДЕ Д.В., доктор технічних наук, професор, керівник наукового центру правової інформації ДНУ ПБП НАПрН України.
ORCID: 0000-0003-3945-1178.

ДМИТРЕНКО О.О., аспірант Інституту проблем реєстрації інформації НАН України.
ORCID: 0000-0001-8501-5313.

ПОБУДОВА СЕМАНТИЧНИХ МЕРЕЖ ТА ВИЗНАЧЕННЯ СТУПЕНЯ РОЗБІЖНОСТІ ТЕКСТІВ

Анотація. У статті викладено методику порівняння текстових документів, що базується на побудові та порівнянні відповідних їм семантичних мереж. Ця методика може стати основою побудови систем порівняння правових документів у рамках парламентського контролю. Також розглянуто алгоритм побудови семантичних мереж як одного із видів онтологій. Цей алгоритм також може застосовуватися в системах автоматичного реферування правової інформації з метою формування лаконічних інформаційно-насичених звітів, коротких анотацій або дайджестів. Пропонована методика може бути використана в процесі обробки запитів при проведенні інформаційного пошуку, надаючи можливість визначення ступеня подібності або відмінності структури та семантики текстів.

Ключові слова: семантична мережа, аналіз природної мови, правова інформація, мережа горизонтальної видимості, порівняння текстів, комп'ютерна лінгвістика.

Summary. The article presents a method for comparing text documents, which is based on the construction and comparison of the corresponding semantic networks. This technique can become the basis for building systems for comparing legal documents in the framework of parliamentary control. An algorithm for constructing semantic networks as one of the types of ontologies is also considered. This algorithm can also be used in systems for automated summarizing legal information in order to generate concise information-rich reports, brief annotations or digests. The proposed technique can be used in processing queries during information retrieval, providing the ability to determine the degree of similarity or difference in the structure and semantics of texts.

Keywords: semantic network, natural language processing, legal information, horizontal visibility network, text comparison, computational linguistics.

Постановка проблеми. Внаслідок швидкого розвитку інформаційно-телекомунікаційних технологій відбувається стрімке накопичення даних у вигляді найрізноманітніших джерел – текстових файлів, електронних листів, веб-сторінок [1] в різноманітних форматах подання. Кількість нормативно-правових документів поданих у електронному вигляді, а отже, і кількість інформації, з якою доводиться мати справу експерту у цій сфері, теж постійно зростає. І для прийняття обґрунтованих рішень на основі існуючих нормативно-правових даних інколи необхідно ознайомлюватися з тисячами документів, відкидаючи інформаційний шум. Тож актуальною для правової галузі є задача спрощення доступу до суті тексту, виокремлення з нього головних викладок, ідей та заздалегідь заявлених змістових аспектів, без необхідності опрацьовувати величезний об'єм інформації. Також важливим є завдання виявлення подібної або дублюючої інформації та суперечностей у нормативно-правових документах.

Всі ці проблеми призводять до необхідності розвивати та удосконалювати наявні технологічні рішення та створювати нові з метою забезпечити оперативну обробку й аналіз правової інформації. Зважаючи на величезні об'єми нормативно-правових текстів, актуальною є завдання формалізації текстових даних й представлення їх у формі, яка була б зручною для автоматичної обробки [2 – 4].

Метою статті є представлення методики визначення ступеня подібності між текстовими документами, що базується на використанні направлених зважених мереж термінів, де вузлами таких мереж є ключові терміни тексту, а ребра – семантико-семантичні зв'язки між цими термінами у тексті.

Виклад основного матеріалу.

Побудова семантичних мереж.

Прикладом моделі предметної галузі (онтології), в якості якої можна представити величезний масив текстових даних, та яка буде зручною для обробки комп'ютером, є направлена зважена мережа термінів. Направлена зважена мережа термінів (Directed Weighted Network of Terms – DWNT, або просто мережа термінів) – семантична модель представлення тексту, де вузлами такої мережі є ключові терміни (слова та словосполучення), які використовуються як назви концептів певної предметної галузі, а ребра – семантико-синтаксичні зв'язки між цими термінами. Порівняння DWNT, отриманих для різних текстів, дає змогу визначити семантичну близькість відповідних текстів.

Побудова мереж термінів здійснюється в декілька етапів [3], що включають попередню обробку текстових даних, екстрагування, тобто виокремлення ключових термінів, побудова ненаправленої мережі термінів (із застосуванням алгоритму графа горизонтальної видимості), тобто встановлення ненаправлених зв'язків між термінами, а також подальше встановлення напрямків зв'язків та їх вагових значень.

Для попередньої обробки текстових даних застосовуються деякі найпоширеніші прийоми, що включають автоматичну **сегментацію на окремі речення** та подальшу **токенізацію** тексту – сегментацію вхідного тексту на елементарні одиниці (токени, лексеми) [5]. В межах кожного речення після токенізації здійснюється маркування частин мови (англ. – Part-of-Speech tagging, PoStagging) [6], що полягає у віднесенні кожного слова в тексті до певної частини мови й присвоєнні йому відповідного тега. Додатково здійснюється **лемматизація** окремих розмічених лексем з метою отримати їх канонічні, словникові форми – леми. Цей крок дозволяє додатково згрупувати різні форми одного й того слова, щоб їх можна було проаналізувати як єдиний елемент.

Для комп'ютеризованої обробки текстів, що представлені українською мовою й класифікації лексем за частинами мови й присвоєнні їм відповідних тегів використовувались функції пакету Stanza [7] мови програмування Python. Для цілей екстрагування термінів були використані слова, які відносяться до таких частин мови, як іменник (тег NOUN), зокрема загальні назви (тег PROP), прикметник (тег ADJ) та сполучник (тег CONJ).

Для побудови мережі термінів використовувались окремі слова, які належать до таких частин мови, як іменник (загальним назвам, що мають тег PROP для зручності було переприсвоєно тег NOUN). Окремі прикметники вилучались. Для побудови словосполучень використовувались наступні шаблони:

- для 2-грам: “ADJ_NOUN”;
- для 3-грам: “NOUN_CONJ_NOUN”, “ADJ_ADJ_NOUN”;
- для 4-грам: “ADJ_NOUN_CONJ_NOUN”, “ADJ_CONJ_ADJ_NOUN”.

Далі здійснюється видалення одиничних стоп-слів (окремих артиклів, прийменників, сполучників, деяких дієслів, прислівників та займенників), які не несуть ніякого інформативного навантаження. Список українських стоп-слів формувався на основі поєднання декількох стоп-словників, один з яких доступний за посиланням [8], а інший доступний у пакеті Python [9]. Також передбачається редагування стоп-словника шляхом доповнення та видалення зі списку слів, які були виявлені експертами в межах досліджуваної галузі.

На наступному етапі, щоб виокремити ключові терміни із тексту для кожного сформованого терміна послідовності будується так званий кортеж з трьох елементів: перший – термін (слово або сформоване за представленими шаблонами словосполучення); наступне – тег, який присвоюється слову в залежності від його приналежності до певної частини мови, або збірний тег для відповідного шаблона; останній елемент такого набору – числове значення *GTF* (Global Term Frequency) – глобальний показник важливості терміна [10]:

$$GTF = \frac{n_i}{\sum_k n_k},$$

де: n_i – кількість появ терміна i у тексті; $\sum \dots$ – загальна кількість сформованих термінів у всьому тексті.

Беручи до уваги розмічування частин мови, *GTF* в цьому випадку обчислюється з урахуванням двох перших елементів кортежу – терміна та тега. Кількість таких однакових кортежів у всій послідовності, що нормована на загальну кількість сформованих термінів, і визначає значення третього елемента кортежу – *GTF*. На відміну від звичайного статистичного показника *TF-IDF*, *GTF* дозволяє більш ефективно знаходити інформаційно-важливі елементи тексту під час роботи з текстовим корпусом заздалегідь визначеної теми, коли інформаційно-важливий термін зустрічається майже у кожному документі корпусу.

Для побудови ненаправленої мережі термінів, як термінологічної онтології певної предметної галузі, далі розглядається й застосовується підхід до побудови мереж на основі часового ряду – алгоритм графа горизонтальної видимості (Horizontal Visibility Graphalgorithm – HVG) [11]. Сам алгоритм графа горизонтальної видимості (Horizontal Visibility Graphalgorithm – HVG) [12], в свою чергу, є розширенням стандартного алгоритму графа видимості (Visibility Graphalgorithm – VG) [13; 14]. Графи горизонтальної видимості будуються у межах кожного окремого речення, де кожному терміну відповідає статистична оцінка *GTF*.

Ненаправлена мережа термінів з використанням алгоритму горизонтальної видимості будується у два етапи [12]. Перший етап полягає у тому, що на горизонтальній осі відмічається ряд вузлів t_i , кожен з яких відповідає термінам у тому порядку, в якому вони з'являються у тексті; а по вертикальній осі відкладаються вагові значення – числові оцінки x_i , що відповідають *GTF*. На другому етапі будується граф горизонтальної видимості. Вважається, що два вузли t_i та t_j , які відповідають елементам часового ряду x_i і x_j , знаходяться у горизонтальній видимості тоді й тільки тоді, коли $x_k \leq \min(x_i; x_j)$ для всіх t_k таких, що $t_i < t_k < t_j$, де $i < k < j$ – вершини графа.

Отримана ненаправлена мережа термінів і буде графом горизонтальної видимості (Рис. 1).

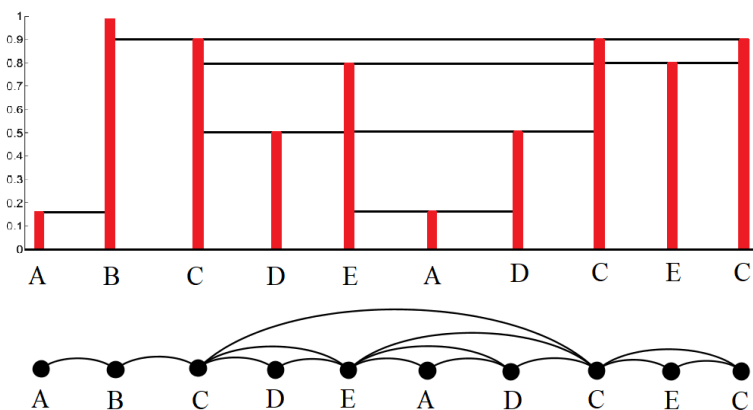


Рис. 1. Приклад побудови графа горизонтальної видимості [11].

Тож розглянутий алгоритм графа горизонтальної видимості дозволяє будувати ненаправлені мережеві структури на основі текстів у випадку, коли окремим словам або словосполученням поставлені у відповідність числові вагові значення.

Напрямки зв'язків у ненаправленій мережі із термінів встановлюються за принципом входження коротшого терміна у термін, що є його розширенням [2], якщо апіорі існує ненаправлений зв'язок між відповідними вузлами у графі горизонтальної видимості. Напрямок всіх інших ненаправлених зв'язків, що залишилися, встановлювався зліва направо (емпіричне правило).

Вагові значення зв'язків між вузлами направленої мережі визначаються за запропонованим у роботі [15; 16] принципом: вершини графа, що відповідають однаковим термінам побудованої на попередньому етапі направленої мережі, об'єднуються ("зшиваються"). Як результат, вагові значення зв'язків між парами вузлів визначаються кількістю однаково-направлених зв'язків між цими вузлами. Оскільки будь-який граф визначається матрицею суміжності, то задача визначення вагових значень зв'язків зводиться до конкатенації стовпців та відповідних рядків – зваженої компактифікації графа горизонтальної видимості [11]. Отримана матриця визначає орієнтований зважений граф сформований з вершин, що відповідають унікальним термінам у розглянутому тексті. Вагове значення ребра, що з'єднує вершину i з вершиною j , визначається кількістю появ терміна t_i перед терміном t_j у тексті (кількістю появ елемента ряду t_i перед елементом t_j).

Результуюча мережа може зберігатися у форматах graphml та json. Для візуалізації мереж, поданих у форматі graphml, застосовується пакет програмного забезпечення з відкритим кодом для мережевого аналізу та візуалізації – Gephi. Формат json може бути зручним для використань у системах побудови та візуалізації семантичних мереж. Під час візуалізації в якості міток вузлів відображаються лише текст терміна (слова чи словосполучення) без зазначення частини мови, до якої цей термін був віднесений на етапі розмічування частин мови засобами мови програмування Python.

Порівняння семантичних мереж.

При порівнянні семантичних мереж, що розглядається, застосовується загальноприйнятий підхід, який полягає у наступному. Розглядається матриця A , яка є різницею матриць, що відповідають цим семантичним мережам і оцінюється її норма, як міра розбіжності. Норма матриці відображає порядок величини матричних елементів. У даному випадку рекомендується використовувати норму Фробеніуса $\|A\|_F$, що дорівнює кореню квадратному із суми квадратів всіх елементів відповідної матриці:

$$\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2}.$$

Звісно, що розмірність двох матриць, що порівнюються має співпадати. У реальності склад термінів у різних семантичних матрицях відрізняється. Тому мережі, що порівнюються взаємно доповнюються термінами, що входять до їх загального складу.

Приклад апробації методики.

Визначення ступеня подібності текстів було здійснено на прикладі біблійських текстів, які загальновідомі і перекладені майже на всі мови (зокрема, авторами досліджувались тексти івритом, китайською, англійською, російською і українською мовами). Для побудови мереж термінів й подальших досліджень був використаний український переклад тексту священної книги Тори, П'ятикнижжя Мойсеєвого, здійснений Іваном Огієнком [17]. Загалом було опрацьовано всі п'ять книг – “Буття”, “Вихід”, “Левит”, “Числа” та “Повторення закону”.

В результаті опрацювання цих текстів було отримано онтологічні моделі у вигляді мережі із термінів. На Рис. 2 наведено фрагмент мережі термінів, що відповідає четвертій книзі “Числа”, наведеній у відомому стенфордському перекладі українською мовою. Під час опрацювання “П'ятикнижжя Мойсеєвого”, враховуючи специфіку священного письма, на етапі попередньої обробки текстів стандартний список стоп-слів корегувався: окремо формувався список слів-виключень, які не є стоп-словами та, насправді, є інформаційно-важливими; і навпаки, список стоп-слів доповнювався іншими словами, які не мають смислового навантаження в межах досліджуваного текстового документу.

Окремо опрацьовувались найбільш частотні слова-синоніми, яким в результаті присвоювалась єдина визначена лексема. Також у зв'язку з наявністю у текстах подібного стилю архаїзмів під час PoS-tagging деяким словам могли присвоюватись невірні теги, що потребувало ручного втручання.

Глобальність під час обчислення GTF визначалася в межах всієї книги, або в межах кожного окремого розділу залежно від того, для якого тексту будувалась мережа термінів – для всієї книги чи окремого розділу. Тому одні й ті ж терміни можуть мати різні значення GTF у межах окремого розділу та всього тексту, відповідно, що впливає на побудову графу горизонтальної видимості.

Щоб досягти незначної розрідженості матриць, було також проведено видалення ребер, що мають одиничну вагу. Опісля також здійснювалось видалення вузлів, які не мають з'єднань. Такі вузли могли з'явитися, зокрема, і в результаті розрідження матриці.

Все вищесказане в результаті впливає на топологію мереж і призводить до наступних наслідків: у мережі термінів, що побудована для всієї книги, можуть бути вузли, яких не існує для кожного окремого розділу, та навпаки – мережа термінів для окремого розділу може містити вузли, яких немає у загальній мережі, що побудована для всього тексту.

Подальше порівняння за допомогою міри Фробеніуса матриць направлених зважених мереж, що отримані для різних текстів, дозволяє визначити семантичну близькість та ступінь подібності відповідних текстів.

На Рис. 4. представлено графік визначення розбіжностей семантичних мереж за правилом порівняння матриць за Фробеніусом.

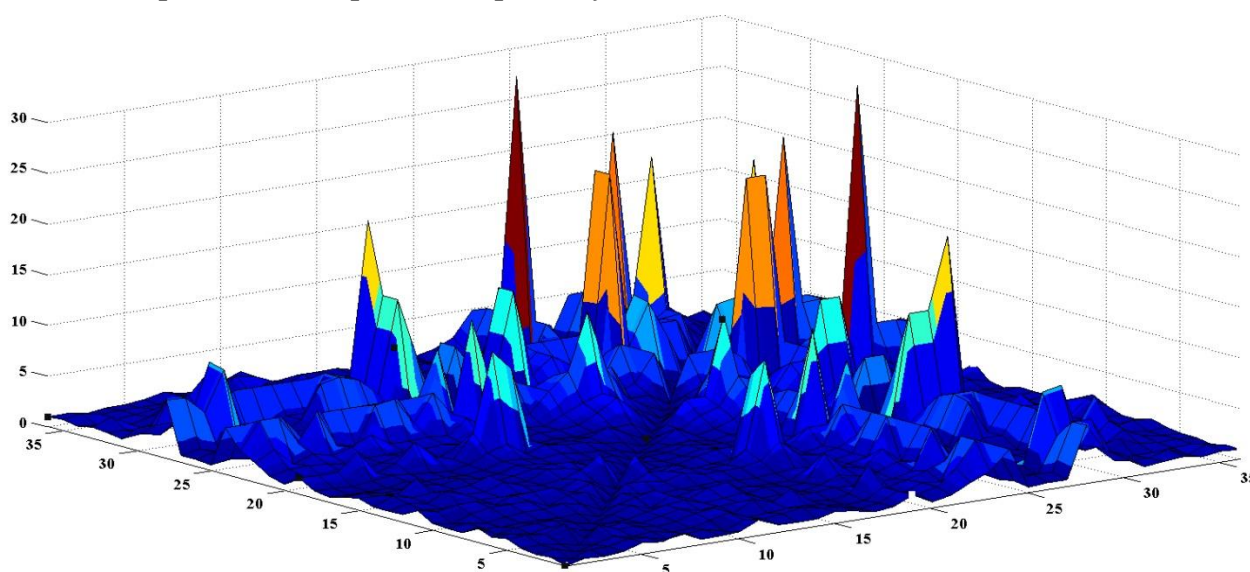


Рис. 4. Графік розбіжностей семантичних матриць, що відповідають окремим розділам книги “Числа”.

Як можна побачити на графіку, найбільші значення розбіжностей відповідають третій частині, тобто розділам 22 – 36. Суть цієї аномалії можна знайти у дослідників Святого писма. Традиційно авторство книги приписується Мойсею, як авторові П’ятикнижжя. Разом з цим, описуються події, коли наступником Мойсея вже було обрано Ісуса Навина. Суто наративні фрагменти у цій частині книги переплітаються з юридичними приписами.

Тобто зміст книги “Числа” підтверджує наведену мережеву методику дослідження текстових документів щодо виявлення структурних і термінологічних розбіжностей. Саме книга “Числа” є найбільш близькою за змістом і структурою до сучасних правових документів частиною Святого Письма, що дозволяє обґрунтовано припустити, що наведена методика може застосовуватись і до таких документів, зокрема, при здійсненні парламентського контролю.

Висновки.

В статті описана методика порівняння текстових документів, що базується на формуванні і подальшому порівнянні відповідних їм семантичних мереж (онтологій). Ця методика може стати основою для побудови системи порівняння правових документів в рамках здійснення парламентського контролю.

Також розглянуто алгоритм формування семантичних мереж як одного із видів онтологій. Цей алгоритм може використовуватись також у системах автоматичного реферування правової інформації з метою формування лаконічних інформаційно-насичених звітів, коротких анотацій або дайджестів. Методика, що запропонована, може бути використана у процесі обробки інформаційних запитів під час інформаційного пошуку, даючи змогу визначити ступінь подібності або розбіжності складу та семантики текстів для подальшого визначення відповідності документа до інформаційних потреб користувача. Як наслідок це дозволить підвищити пертинентність таких систем.

Отже, використання методики побудови семантичних мереж та визначення ступеня подібності текстів у сучасних інформаційно-пошукових системах та системах

автоматичного реферування інформації (зокрема, нормативно-правових документів) сприятиме формуванню й удосконаленню понятійного і термінологічного апарату у правовій галузі та гармонізації національного і міжнародного права.

Використана література

1. Mayer-Schönberger V., Cukier K. Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt, 2013.
2. Ланде Д.В., Дмитренко О.О., Радзівська О.Г. Побудова онтологій в галузі права за даними сервісу Google Scholar. *Інформація і право*. № 1(28)/2019. С. 74-85.
3. Lande D.V., Dmytrenko O.O., Radziievaska O.H. Subject Domain Models of Jurisprudence According to Google Scholar Scientometrics Data. Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2020). Volume I: Main Conference. Lviv, Ukraine, April 23-24, 2020. CEUR Workshop Proceedings (ceur-ws.org). Vol-2604. Pp 32-43. ISSN 1613-0073.
4. Lande D.V., Dmytrenko O.O. 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.
5. Manning C.D., Raghavan P., &Schütze H. An Introduction to Information Retrieval. Cambridge University Press, 2009. P. 22-36.
6. B. Santorini, Part-of-speech tagging guidelines for the Penn Treebank Project, Department of Computer and Information Science School of Engineering and Applied Science University of Pennsylvania Philadelphia, PA 19104, 1990.
7. Stanza – A Python NLP Package for Many Human Languages. URL: <https://stanfordnlp.github.io/stanza>
8. Ukrainian-Stopwords. URL: <https://github.com/skupriienko/Ukrainian-Stopwords>
9. Stop-words 2018.7.23. URL: <https://pypi.org/project/stop-words>
10. Ланде Д.В., Дмитренко О.О., Радзівська О.Г. Визначення напрямків зв'язків у мережі термінів: матеріали XIX Міжнародної науково-практичної конференції *Інформаційні технології та безпека*, ІТБ-2019. Київ: ООО “Инжиниринг”, 2019. С. 103-112.
11. Lande, D.V., Snarskii, A.A., Yagunova, E.V., & Pronoza, E. V.: The use of horizontal visibility graphs to identify the words that define the informational structure of a text. In: 2013 12th Mexican International Conference on Artificial Intelligence. Pp. 209-215 (2013).
12. Luque, B., Lacasa, L., Ballesteros, F., & Luque, J.: Horizontal visibility graphs: Exact results for random time series. *Physical Review E*, 80(4), (2009). doi: 10.1103/PhysRevE. 80.046103.
13. Gutin, G., Mansour, T., &Severini, S.: A characterization of horizontal visibility graphs and combinatorics on words. *Physica A: Statistical Mechanics and its Applications*, 390(12), 2421-2428 (2011). doi: 10.1016/j.physa.2011.02.031.
14. Lacasa, L., Luque, B., Ballesteros, F., Luque, J., & Nuno, J.C.: From time series to complex networks: The visibility graph. *Proceedings of the National Academy of Sciences*, 105(13), 4972-4975 (2008). doi: 10.1073/pnas.0709247105
15. Дмитренко О.О. Побудова направлених зважених мереж термінів із застосуванням Part-of-speech tagging. *Реєстрація, зберігання і обробка даних*, 2020. Т. 22, № 4. С. 47-55. DOI: 10.35681/1560-9189.2020.22.4.225914.
16. Dmytro Lande, Oleh Dmytrenko: Methodology for Extracting of Key Words and Phrases and Building Directed Weighted Networks of Terms with Using Part-of-speech Tagging. Selected Papers of the XX International Scientific and Practical Conference *Information Technologies and Security* (ITS 2020). CEUR Workshop Proceedings (ceur-ws.org). Vol-2859. Pp. 168-177. ISSN 1613-0073. URL: <http://ceur-ws.org/Vol-2859/paper14.pdf>
17. Біблія_(Огієнко). URL: [https://uk.wikisource.org/wiki/Біблія_\(Огієнко\)](https://uk.wikisource.org/wiki/Біблія_(Огієнко))