Науковий вісник НЛТУ України
Scientific Bulletin of UNFU
ISSN 1994-7836 (print)
ISSN 2519-2477 (online)
https://doi.org/10.36930/40340410
Article received 24.04.2024
Article accepted 23.05.2024
UDC 004.056.5
Correspondence author: I. V. Pikh pikhirena@gmail.com

І. В. Піх[1,2], В. М. Сеньківський[1], Р. Р. Андріїв[3]

[1] Українська академія друкарства, м. Львів, Україна
[2] Національний університет "Львівська політехніка", м. Львів, Україна
[3] Львівський державний університет безпеки життєдіяльності, м. Львів, Україна

ОПТИМІЗОВАНА МОДЕЛЬ ЧИННИКІВ ДОСТОВІРНОСТІ ТЕКСТОВИХ ДАНИХ

На підставі аналізу літературних джерел описано основні характеристики чинників впливу на ступінь достовірності текстових даних, оскільки обсяги та швидкість поширення новин створюють складнощі у визначенні їх правдивості. З'ясовано, що ймовірність інформації, особливо в соціальних медіа, часто ставиться під сумнів через поширення фейкових новин, маніпуляції та дезінформацію, що може змінити загальний образ подій і вплинути на суспільство. Навіть без спеціального спотворення, інформація може бути неточною через помилки в джерелах, неправильне тлумачення чи недостатню перевірку фактів. Виокремлено із загальної множини чинників достовірності даних деяку їх підмножину, для якої виконано формалізоване відтворення взаємних зв'язків між елементами з використанням засобів семантичних мереж, що забезпечило відображення в одній графічній структурі впливів і залежностей між чинниками та лінгвістичної семантики їх суті. Застосовано для визначення рівнів пріоритетності чинників стосовно впливу на достовірність даних метод математичного моделювання ієрархій, згідно з алгоритмом реалізації якого запроектовано квадратну бінарну матрицю досяжності, що ідентифікує характерні зв'язки між чинниками семантичної мережі: прямі залежності та прямі впливи. Побудовано на підставі матриці досяжності таблиці ітераційного процесу, опрацювання яких забезпечило встановлення рівнів важливості чинників. Розроблено базову багаторівневу модель впливу чинників на ступінь достовірності текстових даних. Запроектовано за методом попарних порівнянь, шкалою відносної важливості об'єктів та моделлю чинників достовірності текстових даних обернено-симетричну матрицю попарних порівнянь, опрацювання якої за програмою розрахунку вагових пріоритетів чинників забезпечило отримання числових вагових переваг чинників досліджуваного процесу. Розроблено багаторівневу оптимізовану графічну модель чинників пріоритетного впливу чинників на достовірність текстових даних. Проведено перевірку адекватності отриманих результатів за критеріями методу попарних порівнянь, до яких віднесено: максимальне власне значення додатної обернено-симетричної матриці; показник узгодженості; відношення узгодженості.

Ключові слова: чинники ймовірності інформації; семантична мережа; метод аналізу ієрархій; матриця попарних порівнянь; критерії оптимізації; багаторівнева модель.

Вступ / Introduction

Проблема достовірності текстових даних становить серйозне завдання для багатьох галузей, таких як наука, журналістика, суспільствознавство та інші. Зростання обсягів інформації в мережі Інтернет, швидкість поширення новин створюють складнощі у визначенні правдивості даних. Згідно з даними американської компанії IDC (англ. *International Data Corporation*), яка спеціалізується на дослідженні ринків інформаційних технологій та послуг зв'язку, передбачено, що до 2025 р. загальний обсяг цифрової інформації в мережі Інтернет збільшиться до 175 зеттабайт (1 зеттабайт = 1 мільярд терабайт), що становить більше ніж удвічі від обсягів 2021 року.

Дані можуть бути надані різними джерелами, враховуючи офіційні засоби масової інформації, соціальні мережі, власні дослідження, експертні висловлювання тощо. Важливо оцінювати достовірність кожного джерела перед тим, як вважати інформацію достовірною. Інформацію можна підтвердити за допомогою незалежних джерел або аналізувати за допомогою критичного мислення та додаткових досліджень. Застосування останнім часом технологій штучного інтелекту для створення текстового контенту може призвести до появи автоматизованих систем, які створюють недостовірну інформацію. Інформація може бути неправдивою через недостатню перевірку або використання неперевірених фактів. Ще одним важливим чинником може бути намагання владних структур, корпорацій та інших організацій маніпулювати інформацією, щоб захистити свої інтереси або вплинути на громадську думку.

Дослідження ступеня достовірності текстових даних повинно бути спрямоване на розроблення методів, моделей та інструментів для визначення правдивості інформації, фільтрації фейкових новин, виявлення мані-

Інформація про авторів:

Піх Ірина Всеволодівна, д-р техн. наук, професор, кафедра інформаційних систем та технологій у видавничо-поліграфічних процесах. **Email:** pikhirena@gmail.com; https://orsid.org/0000-0002-9909-8444

Сеньківський Всеволод Миколайович, д-р техн. наук, професор, кафедра інформаційних систем та технологій у видавничо-поліграфічних процесах. **Email:** senk.vm@gmail.com; https://orsid.org/0000-0002-4510-540X

Андріїв Роман Ростиславович, асистент, кафедра управління інформаційною безпекою. **Email:** roman.andriiv@icloud.com

Цитування за ДСТУ: Піх І. В., Сеньківський В. М., Андріїв Р. Р. Оптимізована модель чинників достовірності текстових даних. Науковий вісник НЛТУ України. 2024, т. 34, № 4. С. 78–85.

Citation APA: Pikh, I. V., Senkivskyy, V. M., & Andriiv, R. R. (2024). Optimized model of textual data reliability factors. *Scientific Bulletin of UNFU*, 34(4), 78–85. https://doi.org/10.36930/40340410

пуляцій та максимізації точності даних. Воно вимагає міждисциплінарного підходу, що об'єднує зусилля з галузей комп'ютерних наук, інформаційних технологій, лінгвістики, соціології та інших наук для розвитку надійних методів аналізу та перевірки текстової інформації з метою встановлення ступеня достовірності текстових даних.

Представлено початковий етап реалізації інформаційного підходу для оцінювання міри впливу певних чинників на ступінь достовірності текстових даних. В основі пропонованого підходу використано засоби системного аналізу, теорії ієрархічних багаторівневих систем і теорії моделювання [1, 14, 15, 16, 21]. Його застосування передбачає вирізнення із загальної множини чинників об'єктивності інформації певного набору чинників, процедури над якими забезпечать розроблення багаторівневої моделі їх впливу на загальний ступінь достовірності текстових даних.

Об'єкт дослідження – оцінювання ступеня достовірності текстових даних.

Предмет дослідження – методи та засоби проектування моделі чинників впливу на ступінь достовірності текстових даних, що дасть змогу оцінити міру їх важливості у процесі отримання інформації.

Мета роботи – розробити оптимізовану багаторівневу модель пріоритетного впливу чинників на ступінь достовірності текстових даних, що дасть змогу виконати надалі проектування альтернативних і здійснити розрахунок оптимального варіанта процесу визначення ступеня достовірності даних і обчислити його прогнозоване значення.

Для досягнення зазначеної мети визначено такі основні *завдання дослідження*:

1. Виконати формування та здійснити опис загальної множини чинників впливу на достовірність текстових даних, що дасть змогу використати графічний спосіб формалізованого подання зв'язків між чинниками;
2. Виокремити підмножину найбільш істотних чинників дії на об'єктивність інформації та побудувати семантичну мережу зв'язків між ними, що дасть змогу виконати ранжування чинників;
3. Визначити рівні важливості чинників та розробити базову модель пріоритетного їх впливу на ступінь достовірності текстових даних, що уможливить створення передумов для розрахунку вагових пріоритетів чинників;
4. Розрахувати на підставі матриці попарних порівнянь числові вагові переваги чинників та оптимізувати розроблену базову модель, що забезпечить отримання вихідної бази для продовження досліджень в напрямі прогностичного оцінювання величини ступеня достовірності текстових даних засобами теорії нечітких множин.

Аналіз останніх досліджень та публікацій. Ознайомлення з особливостями предметної області передбачає аналіз наукових праць, що відтворюють результати досліджень процесів забезпечення достовірності текстових даних. Оскільки актуальність зазначеної проблеми зумовила наявність великої кількості публікацій, наведемо деякі з них, що стосуються тематики статті.

Серед іноземних публікацій однією з програмних є праця [3], в якій описано методологію автоматизованого виявлення фейкових новин за допомогою методу ймовірнісного аналізу головних компонентів PPCA (англ. *Probabilistic Principal Component Analysis*) та моделі довготривалої короткочасної пам'яті LSTM (англ. *Long Short-Term Memory*). Завдяки такому підходу вхідні дані опрацьовують для визначення характеристик, які можуть вказувати на можливість фальсифікації або недостовірності інформації. У роботі [19] виявлення фейкових новин та класифікацію недостовірної інформації здійснено за допомогою розширеного представлення тексту з розпізнаванням мульти-EDU-структур (англ. *Elementary Discourse Unit*). Робота [10] рекомендує реалізувати процес виявлення фейкових новинних кампаній за допомогою згорткових мереж у вигляді графів. Вказаний граф базується на профілях користувачів соціальних мереж, що підвищує надійність методу та розширює його застосування у різних контекстах. Такий підхід перевершує методи класифікації текстової інформації і надає можливість використовувати його в інших середовищах. Публікація [12] орієнтована на інтегрований підхід стосовно виявлення дезінформації в соціальних мережах з використанням фреймворку, що вміє розпізнавати інформацію та класифікувати її за типами, а саме: правдиві текстові дані; фейкові дані; такі, що подаються у вигляді сатири. Наукова праця [13] описує роботу фреймворку, який базується на графовій нейронній мережі для ідентифікації вузлів, якими можуть бути потенційні користувачі – розповсюджувачі недостовірної інформації. Такий підхід вміє надавати власні припущення щодо користувачів, які можуть розповсюджувати фейкову інформацію. Цей спосіб ефективніший від перевірки наявності фейків в опублікованих текстах.

Вітчизняні видання, що стосуються зазначеної тематики, виокремлюють, аналізують і досліджують питання щодо сучасних викликів, загроз та механізмів протидії негативним інформаційно-психологічним діям, що безпосередньо впливають на інформаційна безпеку України. Звертають увагу на контроль взаємодії суб'єктів соціальної мережі для забезпечення національної інформаційної безпеки. Розроблення синергетичної системи для контролю віртуальних спільнот, що самоорганізують, забезпечить перехід від хаосу до контролю, у такий спосіб досягаючи прогнозованого результату роботи [7]. Відзначено, що перевіреним методом уникнути неправдивих відомостей є отримувати перевірені новини з надійних джерел. Для виявлення фейкових новин доцільно використовувати такі методи, як: аналіз джерела, змісту й заголовка новин; перевірка інформації про автора та джерел, на які посилаються в повідомленні; перевірка "свіжості" новини; використання фактчекінгових інструментів; консультування з експертом; аналіз власної емоційної реакції на новину [17]. Проаналізовано нові прогнозовані загрози кібербезпеці організаціям, особливу увагу приділено захисту кінцевих точок. Встановлено загрози у сфері розвитку штучного інтелекту, голосове шахрайство для соціальної інженерії. Для виявлення майбутніх загроз акцентовано на потребі стратегічного планування впровадження нових технологій і платформ [9]. Введено типи залежностей між чинниками інтенсивності процесу вакцинації у семантичній мережі. Засобами методології моделювання ієрархій та методу ранжування визначено рівні переваг чинників та синтезовано модель пріоритетного їх впливу на процес [11].

Виконаний огляд невеликої частини публікацій свідчить про надзвичайно важливу потребу розроблення засобів протидії спотворенню текстових, усних чи електронних повідомлень, а в разі виявлення подібних фак-

тів забезпечити належний ступінь достовірності даних в інформаційному просторі.

Результати дослідження та їх обговорення / Research results and their discussion

Формулювання вимог та вихідних даних. Аналіз літературних джерел підтвердив наявність різнопланових чинників, що мають істотний вплив на достовірність інформації, переважну частку якої становлять текстові дані. У цьому контексті наведемо одне з важливих положень, суть якого має пряме відношення до обґрунтування прийнятої для дослідження концепції [4].

Довільні процеси – виробничі, соціальні, технологічні – характеризують набором параметрів, чинників, або більш загально множинами чинників, що визначають якість їх реалізації, чи ступінь достовірності інформації. Формалізований запис вказаного твердження може мати таке вираження.

Нехай $R = \{r_i, i = \overline{1,m}\}$ – деяка множина процесів певного виду; $M = \{x_{j_m}, j = \overline{1,n_m}\}$ – множина чинників впливу на якість виконання m-го процесу, де n_m – кількість чинників m-го процесу. Вважатимемо також, що процес характеризується певною функцією якості $A(x)$. Тоді можна прийняти, що

$$A(x) \equiv \bigcup_{j=1}^{n_m} \omega(x_{j_k}),\ k = \overline{1,m}\ , \qquad (1)$$

де: $A(x)$ – числовий показник функції якості m-го процесу; $\omega(x_{jk})$ – числова міра якості, привнесеної j-тим чинником у k-ий процес.

Тоді попередній виклад можна сформулювати так: серед багатьох процесів знайдеться принаймні один, для всіх чинників якого виконується умова (1), що може бути записано таким виразом:

$$(\exists r)\ (\forall x)\ A(x);\ r \in R;\ x \in M\ . \qquad (2)$$

Вивчення чинників, що стосуються процесу визначення рівня достовірності текстових даних, уможливило їх виокремлення за типами та мірою впливу на правдивість повідомлень.

Джерело інформації. Відоме та надійне джерело збільшує достовірність інформації. Авторитетні джерела, такі як визнані видання, наукові журнали, офіційні документи чи експертні думки, зазвичай надають більше гарантій щодо точності даних і повідомлень.

Професіоналізм автора. Якщо автор тексту є експертом у відповідній галузі, це підвищує ймовірність правдивості інформації. Автори з відповідною освітою, досвідом та, що особливо важливо, відповідним рівнем доброчесності зазвичай можуть надати більше обґрунтованих і достовірних даних.

Перевірка фактів. Якщо інформація в тексті підтримана відомими фактами, це підвищує ступінь достовірності. Наявність відповідних доказів, документації та посилань на джерела може вказувати на надійність інформації.

Контекст і структура тексту. Чіткість, логічність та послідовність викладу можуть слугувати ознаками достовірності та може свідчити про її об'єктивність відносно описаної події чи певної ситуації.

Актуальність. Інформація має бути актуальною, назрілою, злободенною, оскільки застарілі дані можуть втратити свою автентичність.

Мова та стиль написання. Професійна та грамотна мова, досконалий стиль написання можуть свідчити про серйозність та фаховість автора щодо обраної теми та відповідну надійність інформації.

Перевірка на багаторазове публікування. Повторення текстових даних у кількох незалежних і солідних джерелах збільшує їх достовірність.

Спростування та критика. Інформація, яка витримує критику та спростування, має більшу ймовірність бути достовірною. Якщо інші експерти або джерела інформації можуть підтвердити дані, це додає вагомості їхній правдивості.

Об'єктивність. Нейтральність та об'єктивність автора можуть впливати на достовірність інформації. Текст, що має явно виражене суб'єктивне ставлення, може підвищити ризик прихованої упередженості, що схиляє людей до певних думок чи ідей щодо сприйняття інформації та впливати на їх рішення.

Коректність використання джерел. Правильне цитування та використання джерел є важливою особливістю достовірності текстових даних. Автор повинен чітко вказувати, звідки отримана інформація, та, якщо це можливо, надавати посилання на первинні джерела.

Технічні особливості. У випадку Інтернет-джерел важливо враховувати безпекові особливості, такі як захищеність сайту, наявність сертифікатів безпеки SSL (англ. *Secure Sockets Layer*), а також перевірка на технічні помилки чи редагування. У сучасній практиці SSL замінений протоколом TLS (англ. *Transport Layer Security*).

Соціальна довіра. Інформація, що користується великою довірою у громадськості або серед фахівців, може свідчити про її достовірність.

Методологія дослідження. Текстові дані, що містять результати досліджень, важливо оцінити з огляду на методологію проведення дослідження, оскільки це може впливати на їх надійність.

Виокремимо із загального набору чинників деяку підмножину R, що стане у майбутньому основою формування ступеня достовірності текстових даних. Виразимо залежність зазначеного показника через рівні другого порядку:

$$R = F_R(A, Z, S)\ , \qquad (3)$$

де: A – рівень сукупності чинників, що свідчать про кваліфікацію автора повідомлення; Z – рівень організаційних заходів; S – рівень суспільної лояльності.

Розкриємо зазначені рівні, прикріпивши до них певні чинники.

$$A = F_A(a_1, a_2, a_3)\ , \qquad (4)$$

де: a_1 – професіоналізм автора; a_2 – об'єктивність автора; a_3 – контекст і структура текстового повідомлення.

$$Z = F_Z(z_1, z_2, z_3)\ , \qquad (5)$$

де: z_1 – джерело інформації; z_2 – перевірка фактів; z_3 – багаторазове публікування (перевірка через пошук багаторазового публікування).

$$S = F_S(s_1, s_2)\ , \qquad (6)$$

де: s_1– спростування та критика; s_2 – соціальна довіра.

Подальші дії виконаємо над об'єднаною множиною чинників, елементи якої для спрощення матимуть однотипне позначення:

$$X = \{x_j, j = \overline{1,8}\}\ . \qquad (7)$$

Запроектуємо семантичну мережу зв'язків між чинниками у вигляді орієнтованого графа, у вершинах якого розмістимо зазначені чинники, а дуги визначатимуть зв'язки між ними (рис. 1).

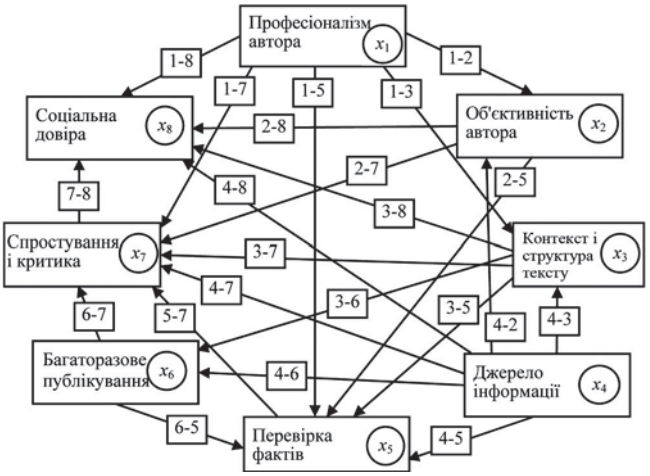

Рис. 1. Семантична мережа чинників достовірності текстових даних / Semantic network of reliability factors of textual data

Вершини графа, що відтворює семантичну мережу, позначені чинниками множини $X = \{x_j, j = \overline{1,8}\}$, зв'язки між парами вершин (x_i, x_j) відтворені дугами, на яких для кращого орієнтування додатково зазначено номер чинника – джерела впливу та номер чинника, на який спрямована дія.

Формування рівнів пріоритетності чинників впливу на достовірність текстових даних. Для визначення рівнів пріоритетності чинників використаємо метод математичного моделювання ієрархій, що розробив Томас Сааті у 1970-х роках, і відомий також як метод аналізу ієрархій [11]. Метод є потужним інструментом прийняття рішень, оскільки забезпечує систематичне структурування та аналіз складних проблем з урахуванням впливу різних чинників.

Згідно з алгоритмом реалізації зазначеного підходу будуємо квадратну бінарну матрицю досяжності B (табл. 1), що відтворює залежності між чинниками семантичної мережі. Елементи матриці визначають за таким правилом:

$$b_{ij}=\begin{cases}1, \text{якщо з } i\text{-ої вершини можна потрапити у } j\text{-ту;}\\ 0 – \text{в іншому випадку.}\end{cases} \quad (8)$$

З виразу (8) зрозуміло, що діагональні елементи такої матриці дорівнюють одиниці. Можна також стверджувати, що вершина x_j буде вважатися досяжною відносно вершини x_i, якщо з останньої "прямим шляхом" можна потрапити в x_j. Справедливим буде і зворотне твердження: вершину x_i вважатимемо попередницею вершини x_j, якщо вона досягається з неї.

Матриця досяжності у вигляді таблиці матиме такий вигляд (табл. 1).

Табл. 1. Матриця досяжності чинників достовірності текстових даних / Matrix of reachability of textual data reliability factors

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	1	1	1	0	1	0	1	1
X_2	0	1	0	0	1	0	1	1
X_3	0	0	1	0	1	1	1	1
X_4	0	1	1	1	1	1	1	1
X_5	0	0	0	0	1	0	1	0
X_6	0	0	0	0	1	1	1	0
X_7	0	0	0	0	0	0	1	1
X_8	0	0	0	0	0	0	0	1

На підставі матриці формується масив досяжних вершин $D(w_i)$, який кожному i-му рядку матриці ставить у відповідність j-ий стовпець, в якому розміщені одиниці. Сукупність вершин попередниць утворить масив $P(w_i)$, котрий, подібно до описаного вище, кожному j-му стовпцю матриці ставить у відповідність i-ий рядок, в яких розміщені одиниці.

Унаслідок цього перетин елементів масивів, що містять вершини досяжних і вершин попередниць, тобто масив визначає певний рівень пріоритетності дії чинників, віднесених до цих вершин.

$$Z(w_i) = D(w_i) \cap P(w_i), \quad (7)$$

Додатковою умовою при цьому є забезпечення рівності

$$P(w_i) = Z(w_i). \quad (8)$$

На підставі правил формування масивів даних $D(w_i)$ і $P(w_i)$ та умови (7) формуємо початкову базову таблицю ітераційного процесу визначення рівнів пріоритетності чинників. При цьому залежність (8) означатиме виконання умови рівності номерів чинників, заданих у другому і третьому стовпцях таблиці, внаслідок чого утворюється певний рівень переваг чинників (табл. 2).

Табл. 2. Дані першого рівня пріоритетності чинників / Data of the first level of priority of factors

i	$D(w_i)$	$P(w_i)$	$D(w_i)\cap P(w_i)$
1	1,2,3,5,7,8	1	1
2	2,5,7,8	1,2,3	2
3	3,5,6,7,8	1,3,4	3
4	2,3,4,5,6,7,8	4	4
5	5,7	1,2,3,4,5,6	5
6	5,6,7	3,4,6	6
7	7,8	1,2,3,4,5,6,7	7
8	8	1,2,3,4,7,8	8

Відповідно до умови (8) збіг номерів зафіксовано у третьому і четвертому стовпцях (див. табл. 2) для чинників 1 (професіоналізм автора) і чотири (джерело інформації), що зумовлює вважати їх згідно з використаним методом моделювання ієрархій чинниками найвищого рівня. Згідно з методом, вилучаємо з табл. 2 перший і четвертий рядки, а в третьому стовпці – цифри 1 і 4. Отримаємо табл. 3.

Табл. 3. Дані другого рівня пріоритетності чинників / Data of the second level of priority of factors

i	$D(w_i)$	$P(w_i)$	$D(w_i)\cap P(w_i)$
2	2,5,7,8	2,3	2
3	3,5,6,7,8	3	3
5	5,7	2,3,5,6	5
6	5,6,7	3,6	6
7	7,8	2,3,5,6,7	7
8	8	2,3,7,8	8

Табл. 3 забезпечує формування другого рівня для чинника 3 (контекст і структура тексту). Дії, аналогічні попереднім, приводять до табл. 4.

Табл. 4. Дані третього рівня пріоритетності чинників / Data of the third level of priority of factors

i	$D(w_i)$	$P(w_i)$	$D(w_i)\cap P(w_i)$
2	2,5,7,8	2	2
5	5,7	2,5,6	5
6	5,6,7	6	6
7	7,8	2,5,6,7	7
8	8	2,7,8	8

Третій рівень утворюють чинники 2 (об'єктивність автора) і 6 (багаторазове публікування). Далі матимемо табл. 5 четвертого рівня переваг чинників.

Четвертий рівень – чинник 5 (перевірка фактів). Без таблиць п'ятий рівень – чинник 7 (спростування та критика) і шостий – чинник 8 (соціальна довіра.)

Табл. 5. Дані четвертого рівня пріоритетності чинників / Data of the fourth level of priority of factors

i	$D(w_i)$	$P(w_i)$	$D(w_i)\cap P(w_i)$
5	5,7	5	5
7	7,8	5,7	7
8	8	7,8	8

На підставі отриманих рівнів проектуємо багаторівневу модель пріоритетного впливу чинників на ступінь достовірності текстових даних (рис. 2).

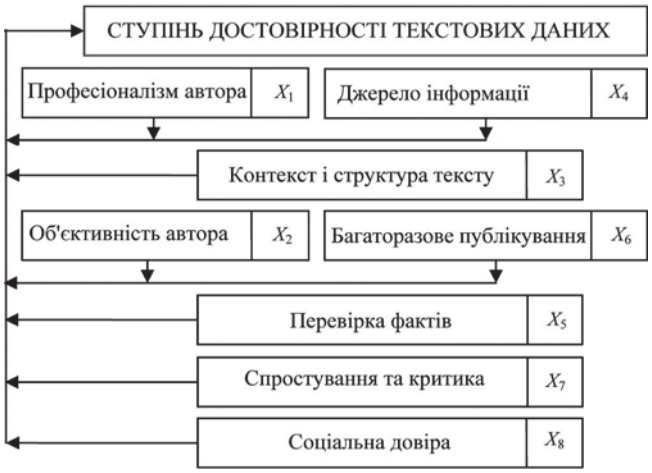

Рис. 2. Багаторівнева модель чинників достовірності текстових даних / Multilevel model of textual data reliability factors

Розрахунок вагових пріоритетів чинників та синтезу оптимізованої моделі чинників достовірності текстових даних. Оскільки у моделі (див. рис. 2) перший і третій рівні містять по два чинники, виконаємо оптимізацію, що забезпечить її переведення у багатофакторну структуру. Скористаємося при цьому методом попарних порівнянь, який, як відзначено у роботі [4], "встановлює кардинальну погодженість стосовно рівня пріоритетності чинників". Внаслідок застосування зазначеного методу для кожного з чинників будуть розраховані числові пріоритети переваг, що відповідатимуть мірі їх впливу на ступінь достовірності даних. З іншого боку, отримані величини стануть підставою синтезу оптимізованої моделі. Згідно з алгоритмом реалізації методу будуємо квадратну обернено-симетричну матрицю попарних порівнянь A, елементи якої відтворюють результати порівняння рівнів розміщення чинників у моделі (див. рис. 2) та вибираються із шкали відносної важливості об'єктів (табл. 6) [4]. Так, для чинників x_1 та x_2 оцінка важливості розміщена на перетині рядка, позначеного змінною x_1, та стовпця x_2. Діагональні елементи матриці дорівнюють одиниці, окрім цього $a_{ji} = 1/a_{ij}$.

Табл. 6. Шкала відносної важливості об'єктів / Scale of relative importance of objects

Оцінка	Критерій порівняння	Пояснення щодо критерію
1	Об'єкти рівноцінні	Відсутність переваги x_1 над x_2
3	Один об'єкт дещо переважає інший	Слабка перевага x_1 над x_2
5	Один об'єкт переважає інший	Істотна перевага x_1 над x_2
7	Об'єкт значно переважає інший	Явна перевага x_1 над x_2
9	Об'єкт абсолютно переважає інший	Абсолютна перевага x_1 над x_2
2,4,6,8	Компромісні проміжні значення	Допоміжні порівняльні оцінки

Остаточно матрицю попарних порівнянь помістимо у табл. 7.

Табл. 7. Матриця попарних порівнянь чинників достовірності даних / Matrix of pairwise comparisons of data reliability factors

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	1	4	2	1/2	5	4	8	8
X_2	1/4	1	1/2	1/4	3	2	5	7
X_3	1/2	2	1	1/3	3	4	6	7
X_4	2	4	3	1	6	5	8	9
X_5	1/5	1/3	1/3	1/6	1	1/2	2	4
X_6	1/4	1/2	1/4	1/5	2	1	4	5
X_7	1/8	1/5	1/6	1/8	1/2	1/4	1	2
X_8	1/8	1/7	1/7	1/9	1/4	1/5	1/2	1

Опрацювання матриці попарних порівнянь здійснено за програмою "Імітаційне моделювання в системному аналізі методом бінарних порівнянь" [2], інтерфейс якої наведено на рис. 3. Унаслідок цього отримуємо головний власний вектор пріоритетів, нормалізовані компоненти якого ідентифікують вагові переваги чинників впливу на ступінь достовірності текстових даних.

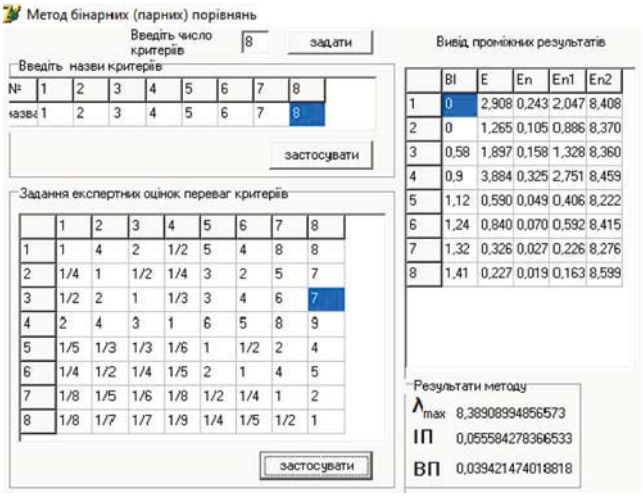

Рис. 3. Інтерфейс програми розрахунку вагових пріоритетів чинників / Interface of the program for calculating weight priorities of factors

Пояснимо коротко алгоритм отримання вагових значень чинників, користуючись даними вікна інтерфейсу "Вивід проміжних результатів".

Компоненти головного власного вектора $A = \{a_i, i = \overline{1,n}\}$ розраховують через середні геометричні значення елементів рядків матриці за формулою

$$a_i = \sqrt[n]{\prod_{j=1}^{n} a_{ij}}, i = \overline{1,n}\,. \qquad (9)$$

Результатом буде вектор (опція E в інтерфейсі програми)

A = (2,908; 1,265; 1,897; 3,884; 0,590; 0,840; 0,326; 0,227),

нормалізація якого за формулою

$$a_{i,норм} = \sqrt[n]{\prod_{j=1}^{n} a_{ij}} \Big/ \sum_{i=1}^{n} \sqrt[n]{\prod_{j=1}^{n} a_{ij}}\,,\ i = \overline{1,n}\,. \qquad (10)$$

приводить до шуканого головного власного вектора матриці попарних порівнянь, компоненти якого відтворені в опції E_n:

$A_{норм}$ = (0,243; 0,105; 0,158; 0,325; 0,049; 0,070; 0,027; 0,019).

Як було відзначено вище, компоненти нормалізованого вектора по суті можна вважати результатом розв'язання задачі. Для відображення їх у вигляді цілих чисел помножимо компоненти вектора $A_{норм}$, наприклад, на

коефіцієнт k = 500. Отримаємо такий масштабований остаточний вектор

$$A_{норм} \cdot k = (243; 105; 158; 325; 49; 70; 27; 19). \quad (11)$$

Нормальною практикою досліджень вважається перевірка узгодженості отриманих результатів за певними еталонами. Такими критеріями для цього методу вважаються: максимальне власне значення λ_{max} додатної обернено-симетричної матриці; показник узгодженості *UI*; відношення узгодженості *WU* (опції λ_{max}, ІП, ВП вікна інтерфейсу). Для їх отримання виконаємо певні перетворення, а саме – множимо матрицю попарних порівнянь справа на вектор $A_{норм}$, що приведе до вектора $A_{норм1}$ (опція $E_{п1}$):

$$A_{норм1} = (2{,}047; 0{,}886; 1{,}328; 2{,}751; 0{,}406; 0{,}592; 0{,}226; 0{,}163).$$

Продовжуючи, ділимо компоненти вектора $A_{норм1}$ на відповідні компоненти вектора $A_{норм}$, що зумовлює вектор $A_{норм2}$ (опція $E_{п2}$):

$$A_{норм2} = (8{,}408; 8{,}370; 8{,}360; 8{,}459\ 8{,}222; 8{,}415; 8{,}276; 8{,}599).$$

Середнє арифметичне компонент вектора $A_{норм2}$ зумовлює величину максимального власного значення λ_{max} додатної обернено-симетричної матриці *A*. За програмою λ_{max} = 8,39. Вважається, що ціла частина його значення не повинна перевищувати кількості чинників досліджуваного процесу. Точніші оцінки визначають показником узгодженості, що розраховують за формулою

$$IU = \left(\lambda_{\max} - n\right) / \left(n - 1\right). \quad (12)$$

Розраховане значення *IU* = 0,056 (величина ІП у вікні "Результати методу"). Результати вважаються задовільними, якщо значення показника узгодженості не перевищує 10 % еталонного значення випадкового показника *WI*, що аналогічно виконанню нерівності $IU < 0{,}1 \cdot WI$. Для восьми об'єктів *WI* = 1,41, тобто $0{,}056 < 0{,}1 \cdot 1{,}41$. На завершення оцінюють величину відношення узгодженості, що отримується за виразом $WU = IU/WI$. Його значення оптимальне, якщо $WU \le 0{,}1$ (величина ВП у вікні "Результати методу"). Розраховане *WU* = 0,04, що остаточно підтверджує достовірність отриманих результатів.

Порівняння числових результатів опрацювання матриці попарних порівнянь підтвердило наявність таких принципових постулатів дослідження: адекватність отриманих вагових пріоритетів чинників; достатній рівень збіжності ходу розрахунків; узгодженість експертних суджень стосовно попарних порівнянь чинників. Оптимізовані вагові значення чинників достовірності текстових даних, відтворені виразом (11), вказують на сувору відповідність їх розміщення за рівнями відповідно до принципу "один рівень – один чинник".

Зазначене вище доповнимо таким твердженням. Оскільки максимальне значення власного вектора матриці попарних порівнянь та величина відношення узгодженості не виходять за допустимі межі, їх можна вважати критеріями, що зумовлюють синтезу оптимізованої багаторівневої моделі пріоритетного впливу чинників на ступінь достовірності текстових даних (рис. 4).

Результат синтезу моделі (див. рис. 4) теоретично підтвердив інтуїтивне розуміння переваги чинників "Джерело інформації" та "Професіоналізм автора" у процесі формування ступеня достовірності текстових даних. Вагові пріоритети чинників можуть бути використані для проектування альтернативних і розрахунку оптимального варіантів процесу визначення ступеня достовірності текстових даних методами лінійного згортання критеріїв чи багатокритеріальної оптимізації.

Рис. 4. Оптимізована модель пріоритетного впливу чинників на ступінь достовірності текстових даних / Optimized model of the priority influence of factors on the degree of reliability of textual data

Обговорення результатів дослідження. Особливої актуальності проблематика достовірності даних набуває в наш час для протидії новим гібридним інформаційним війнам, що постійно ведуться у боротьбі за виборців чи просто з метою просування шкідливих, але вигідних, наративів через світову мережу Інтернет. Досягнення задовільного ступеня достовірності текстових даних повинно бути спрямоване на розроблення методів, моделей та інструментів для визначення правдивості інформації, фільтрації фейкових новин, виявлення маніпуляцій та максимізації точності даних.

У роботі [14] визначено на підставі методології моделювання ієрархій рівні переваг чинників якості програмного забезпечення за побудованою бінарною матрицею досяжності, опрацювання якої виконано ітераційними таблицями. За допомогою засобів інфографіки синтезовано базову модель чинників, що, однак, не можна вважати остаточним результатом, тому завершальне ранжування ми здійснили з використанням матриці попарних порівнянь, опрацювання якої забезпечило отримання вагових пріоритетів чинників.

У публікації [15] отримана оптимізована модель стала результатом вивчення впливу певних чинників на рівень читацького попиту на книгу. На противагу цьому у рекомендованій статті вихідною базою даних слугують чинники впливу на достовірність текстових даних і їх інтерпретація стосовно процесів забезпечення ймовірності інформації.

У роботі [3] описано методику виявлення недостовірних новин за допомого методу ймовірнісного аналізу та моделі короткочасної пам'яті, що, однак, не завжди доступно для широкого кола користувачів, оскільки передбачає наявність та вміння користуватися програмними додатками.

Найбільш близькою до тематики статті є робота [10], що рекомендує реалізувати не просто апостеріорний процес виявлення недостовірних новин, але фейкових новинних кампаній за допомогою згорткових мереж у вигляді графів. При цьому відстежується джерело походження та достовірність його інформації. Зазначений граф базується на профілях користувачів соціальних мереж, що підвищує надійність методу. На відміну від описаного методу ми рекомендуємо фільтрувати вже надіслані повідомлення, оскільки фейкові компанії для свого виживання можуть передавати і деякі достовірні новини. Пропонований авторами інформаційний підхід ґрунтується на мережах семантичних і є початковим етапом повнішого дослідження, що передбачає розроблення автоматизованої системи прогностичного оцінювання ступеня достовірності текстових даних з використанням методів і засобів теорії нечітких множин.

У роботі [20] відзначено, що сьогоднішній цифровий світ приніс широкий спектр ризиків, таких як атаки та шкідливі програми, що спотворюють отримані дані. Освітній сектор вищої освіти близький до цих негативних впливів, які перешкоджають навчальному процесу. У доповнення до рекомендованого підходу досліджено важливість кібербезпеки в цьому секторі та запропоновано стратегії, які студенти, викладачі та співробітники можуть використовувати для отримання достовірних даних у вищих навчальних закладах.

Публікація [18] звертає увагу на потенціал використання штучного інтелекту та веб-сервісів Amazon для покращення виявлення загроз, створення високоякісних і достовірних навчальних даних й оптимізації використання ресурсів. Ці дані потрібні для навчання систем виявлення загроз і посилення стійкості організації проти кібератак. Рекомендована стаття обмежується чинникним аналізом проблеми достовірності інформації в загальних ознаках, тому важливі питання ймовірності даних стосовно навчального процесу можна дослідити у майбутньому.

Отже, за результатами виконаної роботи можна сформулювати такі наукову новизну та практичну значущість результатів дослідження.

Наукова новизна отриманих результатів дослідження – вперше у проблематиці, орієнтованій на процес формування ймовірності інформації, здійснено формалізоване подання та ранжування чинників за пріоритетністю впливу на достовірність текстових даних у вигляді багаторівневої моделі, що уможливить подальше їх опрацювання засобами нечіткої логіки.

Практична значущість результатів дослідження – систематизовано вихідну інформаційну базу та розроблено модель, що забезпечить для користувачів можливості здійснювати обґрунтованіше фільтрування текстових даних, вилучаючи фейкові повідомлення.

Висновки / Conclusions

Розроблено оптимізовану багаторівневу модель пріоритетного впливу чинників на ступінь достовірність текстових даних, що дало змогу виконати у подальшому проектування альтернативних і здійснити розрахунок оптимального варіанта процесу визначення ступеня достовірності даних і обчислити його прогнозоване значення. За результатами проведеного дослідження можна зробити такі основні висновки.

1. Розроблено оптимізовану багаторівневу модель пріоритетного впливу чинників на достовірність текстових даних – основи для формування та прогностичного оцінювання ступеня вірогідності інформації. Наведено та описано на підставі аналізу публікацій, що стосуються тематики статті, чинники впливу на ступінь достовірності текстових даних.
2. Виокремлено достатню для дослідження підмножину чинників та виконано формалізоване відтворення зв'язків між ними за допомогою семантичної мережі. Визначено за методом математичного моделювання ієрархій рівні пріоритетності чинників, що забезпечило синтез базової моделі впливу чинників на ступінь достовірності текстових даних.
3. Встановлено на підставі методу попарних порівнянь та шкали відносної важливості об'єктів вагові переваги чинників, що зумовило створення оптимізованої моделі пріоритетного впливу чинників на достовірність текстових даних. Виконано перевірку адекватності отриманих результатів за критеріями методу попарних порівнянь.
4. Використання результатів дослідження забезпечить користувачів новими ефективними засобами для визначення достовірності текстових даних, що в час стрімкого зростання обсягів та швидкості поширення інформації набуває першочергового значення.

References

1. Bartish, M. Ya., & Dudzianyi, I. M. (2009). Operations research. Part 3. Decision making and game theory. Lviv: Publishing center of Ivan Franko National University, 278 p. [In Ukrainian]. URL: https://ami.lnu.edu.ua
2. Certificate of copyright registration for work No. 41832. Ukraine. Simulation modeling in system analysis using binary comparisons method (computer program). The copyrights are owned by I. V. Hileta, V. M. Senkivskyy, O. V. Melnykov. Registered 17.01.2012.
3. Dixit, D. K., Bhagat, A., & Dangi, D. (2022). Automating fake news detection using PPCA and levy flight-based LSTM. *Soft Computing A Fusion of Foundations, Methodologies and Applications,* 26, 12545–12557. https://doi.org/10.1007/s00500-022-07215-4
4. Durniak, B. V., Pikh, I. V., & Senkivskyy, V. M. (2022). Theoretical foundations of the information concept for the formation and evaluation of the quality of publishing and printing processes. Monograph. Lviv: Ukrainian Academy of Printing, 356 p. [In Ukrainian]. URL: https://biblio.uad.lviv.ua
5. Dyak, T. P., Hrytsiuk, Yu. I., & Horvat, P. P. (2022). The problem of fake news detection on Internet websites. *Scientific Bulletin of UNFU*, 32(6), 78-94. https://doi.org/10.36930/40320612
6. Eight effective ways to recognize fake information. URL: https://www.prostir.ua/?news=8-dijevyh-sposobiv-rozpiznaty-fejkovu-informatsiyu
7. Hryshchuk, R., & Molodetska, K. (2017), Synergetic Control of Social Networking Services Actors Interactions. *Recent Advances in Systems, Control and Information Technology Advances in Intelligent Systems and Computing*, 543, 34–42. https://doi.org/10.1007/978-3-319-48923-0_5
8. Kovbasyuk, O. M., & Hrytsiuk, Yu. I. (2022). Detection of fake news using machine learning methods. *New directions of science development during martial law. Collection of scientific materials of the CXIV International Scientific and Practical Internet Conference*, (pp. 207–2018), December 12, 2022, Odessa (el-conf.com.ua), 404 p. URL: https://el-conf.com.ua/wp-content/uploads/2023/01/Odesa_12222022.pdf#page=207
9. Legominova, S., & Gaidur, G. (2023). Analysis of modern threats to the information security of organizations and the formation of an information platform to counter them. *Electronic professional scientific publication. Cybersecurity: Education, Science, Technology*, *2*(22), 54–67. https://doi.org/10.28925/2663-4023.2023.22.5467
10. Michail, D., Kanakaris, N., & Varlamis, I. (2022). Detection of fake news campaigns using graph convolutional networks. *Inter-*

national Journal of Information Management Data Insights, 2(2). https://doi.org/10.1016/j.jjimei.2022.100104
11. Pih, I. V., Senkivskyy, V. M., Teslyuk, V. M., & Tsmots, I. G. (2022). Models of factors of the intensity of vaccination against COVID-19 taking into account the predicates of semantic networks. *Proceedings*, *1*(64), 63–75. https://doi.org/10.32403/1998-6912-2022-1-64-63-75
12. Rastogi, S., & Bansal, D. (2022). Disinformation detection on social media: An integrated approach. *Multimedia Tools and Applications*, 81, 40675–40707. https://doi.org/10.1007/s11042-022-13129-y
13. Rath, B., Salecha, A., & Srivastava, J. (2022). Fake news spreader detection using trust-based strategies in social networks with bot filtration. *Social Network Analysis and Mining, 12,* 66–68. https://doi.org/10.1007/s13278-022-00890-z
14. Senkivskyy, V. M., Pikh, I. V., Kalynii, I. V., Senkivskyy, N. Y., & Drahomirov, M. A. (2022). Methodological Foundations of Software Quality Formation (Part 1: Basic Model of Quality Factors). *Printing and Publishing*, 2(84), 9–21. https://doi.org/10.32403/0554-4866-2022-2-84-9-21
15. Senkivskyy, V. M., Pikh, I. V., Kudriashova, A. V., Senkivska, N. E., Kalynii, I. V. (2021). Optimization of the model of reader demand factors for the book. *Printing and Publishing*, *2*(84), 11–20. https://doi.org/10.32403/0554-4866-2021-1-81-11-20
16. Senkivskyy, V., Pikh, I., Kudriashova, A., Senkivska, N., & Tupychak, L. (2022). Models of Factors of the Design Process of Reference and Encyclopedic Book Editions. *Lecture Notes on Data Engineering and Communications Technologies*, 77, 217–229. https://doi.org/10.1007/978-3-030-82014-5
17. Tyshchenko, V., & Muzhanova, T. (2022). Disinformation and fake news: signs and methods of detection on the Internet. *Electronic professional scientific publication. Cybersecurity: Education, Science, Technology*, *2*(18), 175–186. https://doi.org/10.28925/2663-4023.2022.18.175186
18. Villegas-Ch, W., Govea, J., & Ortiz-Garces, I. (2024). Developing a Cybersecurity Training Environment through the Integration of OpenAI and AWS. *Computing and Artificial Intelligence a section of Applied Sciences, 14*(2), 67 p. https://doi.org/10.3390/app14020679
19. Wang, Y., Wang, L., Yang, Y., & Zhang, Y. (2022). Detecting fake news by enhanced text representation with multi-EDU-structure awareness. *An International Journal Expert Systems with Applications*, 206 p. https://doi.org/10.1016/j.eswa.2022.117781
20. Yousif Yaseen, K. A. (2022). Importance of Cybersecurity in The Higher Education Sector 2022. *Asian Journal of Computer Science and Technology*, *11*(2), 20–24. https://doi.org/10.51983/ajcst-2022.11.2.3448
21. Zgurovskyi, M. Z., & Pankratova, N. D. (2007). Fundamentals of system analysis. Kyiv: VNU Publishing Group, 544 p. [In Ukrainian]. URL: https://iszzi.kpi.ua

I. V. Pikh[1,2], V. M. Senkivskyy[2], R. R. Andriiv[3]

[1] Ukrainian Academy of Printing, Lviv, Ukraine
[2] Lviv Polytechnic National University, Lviv, Ukraine
[3] Lviv State University of Life Safety, Lviv, Ukraine

OPTIMIZED MODEL OF TEXTUAL DATA RELIABILITY FACTORS

Based on the analysis of literature sources, description of the main characteristics of the factors influencing the degree of reliability of textual data was performed, since the volume and speed of news dissemination create difficulties in determining their reliability. In the course of research, the reliability of information, especially in social media, is found to be often questioned due to the spread of fake news, manipulation and misinformation, which can change the overall image of events and, therefore, affect society. Even without deliberate distortion, information can be inaccurate due to errors in sources, misinterpretation, or insufficient fact-checking. A subset of data reliability factors has been identified from the general set of factors, for which a formalized reproduction of the relationships between elements has been performed using semantic networks, which ensured the reflection of the influences and dependencies between factors and the linguistic semantics of their essence in one graphical structure. The method of mathematical modeling of hierarchies is used to determine the levels of priority of factors in terms of their impact on data reliability, according to the algorithm of which a quadratic binary reachability matrix is designed, which identifies the characteristic relationships between the factors of the semantic network, in particular, direct dependencies and direct influences. The tables of the iterative process are constructed on the basis of the reachability matrix, processing of which ensured the establishment of the levels of importance of the factors. Furthermore, a basic multilevel model of the influence of factors on the degree of reliability of textual data is developed. An inversely symmetric matrix of pairwise comparisons was designed using the method of pairwise comparisons, the scale of relative importance of objects and the model of textual data reliability factors, processing of which, according to the program for calculating the weight priorities of factors, provided for obtaining numerical weight advantages of the factors of the studied process. A multi-level optimized graphical model of the factors of priority influence on the reliability of textual data has been developed. The adequacy of the obtained results is checked according to the criteria of the method of pairwise comparisons, which include as follows: the maximum eigenvalue of a positive inverse symmetric matrix; consistency index; consistency ratio.

Keywords: probability factors; semantic network; analytic hierarchy process method; pairwise comparison matrix; optimization criteria; multilevel model.