УДК 004.912 **DOI: 10.20998/2079-0023.2021.01.14**

С. В. ПЕТРАСОВА, Н. Ф. ХАЙРОВА, А. С. КОЛЕСНИК

ТЕХНОЛОГІЯ ВИЗНАЧЕННЯ ІНФОРМАЦІЙНОГО ПОРЯДКУ ДЕННОГО В ПОТОКАХ НОВИННИХ ДАНИХ

З кожним днем обсяг потоків новинних даних зростає, що збільшує інтерес до систем, які дозволяють автоматизувати обробку великих потоків даних. Визначення смислової подібності текстової інформації на основі використання інтелектуальних засобів обробки даних дозволить виділяти спільні інформаційні простори новин. У статті проаналізовані сучасні статистичні метрики для визначення зв'язних фрагментів, зокрема, новинних текстів, що відображають порядок денний (agenda), вказані основні переваги та недоліки. Пропонується інформаційна технологія виявлення спільного інформаційного простору актуальних новин в потоці даних за певний період часу. Технологія включає логіко-лінгвістичну і дистрибутивно-статистичну модель ідентифікації колокацій. Модель дистрибутивної семантики MI застосовується на етапі вилучення потенційних колокацій. При цьому регулярні вирази, розроблені відповідно до граматики англійської мови, дозволяють виявляти граматично правильні конструкції. Перевагою розробленої логіко-лінгвістичної моделі формалізації семантико-граматичних характеристик колокацій на основі використання алгебро-предикатних операцій і предиката семантичної еквівалентності, є врахування аналізу як граматичної структури мови, так і смислу слів (колокатів). Тезаурус WordNet застосовується на етапі визначення відношення синонімії між головними і залежними компонентами колокацій. На основі досліджуваного корпусу новинних текстів служб CNN і BBC проведена оцінка ефективності розробленої технології. Аналіз показав, що коефіцієнт точності precision дорівнює 0,96. Застосування запропонованої технології дозволить поліпшити якість обробки потоків новинних повідомлень. Вирішення завдання автоматичного визначення смислової близькості може застосовуватися при виявленні текстів однієї тематики, актуальної інформації, добуванні фактів і усунення смислової неоднозначності та ін.

Ключові слова: потік даних, порядок денний, логіко-лінгвістична модель, дистрибутивно-статистична модель, колокація, смислова близькість, WordNet, корпус новинних текстів, precision.

С. В. ПЕТРАСОВА, Н. Ф. ХАЙРОВА, А. С. КОЛЕСНИК

ТЕХНОЛОГИЯ ОПРЕДЕЛЕНИЯ ИНФОРМАЦИОННОЙ ПОВЕСТКИ ДНЯ В ПОТОКАХ НОВОСТНЫХ ДАННЫХ

В настоящее время объем потоков новостных данных возрастает, что увеличивает интерес к системам, которые позволяют автоматизировать обработку больших потоков данных. Определение смысловой близости текстовой информации на основе использования интеллектуальных средств обработки данных позволит выделять общие информационные пространства новостей. В статье проанализированы современные статистические метрики для определения связных фрагментов, в частности, новостных текстов, отображающих повестку дня (agenda), указаны основные преимущества и недостатки. Предлагается информационная технология выявления общего информационного пространства актуальных новостей в потоке данных за определенный период времени. Технология включает логико-лингвистическую и дистрибутивно-статистическую модели идентификации коллокаций. Модель дистрибутивной семантики MI применяется на этапе извлечения потенциальных коллокаций. При этом регулярные выражения, разработанные в соответствии с грамматикой английского языка, позволяют выявлять грамматически правильные конструкции. Преимуществом разработанной логико-лингвистической модели формализации семантико-грамматических характеристик коллокаций на основе использовании алгебро-предикатных операций и предиката семантической эквивалентности, является возможность анализировать как грамматическую структуру языка, так и смысл слов (коллокатов). Тезаурус WordNet применяется на этапе определения отношения синонимии между главными и зависимыми компонентами коллокаций. На основе исследуемого корпуса новостных текстов служб CNN и BBC проведена оценка эффективности разработанной технологии. Анализ показал, что коэффициент точности precision равен 0,96. Применение предлагаемой технологии позволит улучшить качество обработки потоков новостных сообщений. Решение задачи автоматического определения смысловой близости может применяться при выявлении текстов одной тематики, актуальной информации, извлечении фактов и устранении смысловой неоднозначности и др.

Ключевые слова: поток данных, повестка дня, логико-лингвистическая модель, дистрибутивно-статистическая модель, коллокация, смысловая близость, WordNet, корпус новостных текстов, precision.

S. V. PETRASOVA, N. F. KHAIROVA, A. S. KOLESNYK

TECHNOLOGY FOR IDENTIFICATION OF INFORMATION AGENDA IN NEWS DATA STREAMS

Currently, the volume of news data streams is growing that contributes to increasing interest in systems that allow automating the big data streams processing. Based on intelligent data processing tools, the semantic similarity identification of text information will make it possible to select common information spaces of news. The article analyzes up-to-date statistical metrics for identifying coherent fragments, in particular, from news texts displaying the agenda, identifies the main advantages and disadvantages as well. The information technology is proposed for identifying the common information space of relevant news in the data stream for a certain period of time. The technology includes the logical-linguistic and distributive-statistical models for identifying collocations. The MI distributional semantic model is applied at the stage of potential collocation extraction. At the same time, regular expressions developed in accordance with the grammar of the English language make it possible to identify grammatically correct constructions. The advantage of the developed logical-linguistic model formalizing the semantic-grammatical characteristics of collocations, based on the use of algebraic-predicate operations and a semantic equivalence predicate, is that both the grammatical structure of the language and the meaning of words (collocates) are analyzed. The WordNet thesaurus is used to determine the synonymy relationship between the main and dependent collocation components. Based on the investigated corpus of news texts from the CNN and BBC services, the effectiveness of the developed technology is assessed. The analysis shows that the precision coefficient is 0.96. The use of the proposed technology could improve the quality of news streams processing. The solution to the problem of automatic identification of semantic similarity can be used to identify texts of the same domain, relevant information, extract facts and eliminate semantic ambiguity, etc.

Keywords: data stream, agenda, logical-linguistic model, distribution-statistical model, collocation, semantic similarity, WordNet, news text corpus, precision.

Вступ. Дослідження інформаційного порядку денного (information agenda) є актуальним і досить популярним завданням, в тому числі, через поширеність та широкий доступ до контенту ЗМІ.

Сучасне суспільство використовує ЗМІ, щоб формувати і описувати поточні проблеми та події, що відбуваються в усьому світі. У зв'язку з цим автоматична обробка та аналіз контенту ЗМІ (зокрема, потоків новинних повідомлень) є однією з можливостей для вивчення взаємодії громадського, політичного та медійного порядків денних, дозволяючи виявити, які теми популярні в суспільстві та ЗМІ. Таким чином, під порядком денним маємо на увазі «сукупність актуальних проблем і сюжетів, що мають ряд самостійних характеристик» [1, 2].

Існуючі сьогодні підходи до класифікації новин і розроблені додатки групують новини за досить великими тематичними групами. Найчастіше це пов'язано із застосуванням методів машинного навчання, які добре працюють на великих обсягах текстів з простором ознак класифікації, що добре поділяється. Однак, дуже часто необхідно виділити відносно вузьку сукупність новин, які мають ряд самостійних та пов'язаних характеристик.

Ми пропонуємо інформаційну технологію визначення смислової подібності новинного контенту медійних платформ BBC і CNN для виявлення актуального порядку денного за певний період часу. Технологія включає в себе розроблену модель семантичної подібності для ідентифікації синонімічних колокацій у новинному потоці даних, регулярні вирази з граматичними правилами та модель дистрибутивної семантики MI для виявлення колокацій як найбільш змістовних фрагментів новинних текстів.

Аналіз останніх досліджень і публікацій. Колокація, як невипадкове синтаксичне та семантичне сполучення двох або більше лексем, що складається з головного та залежного компонентів, містить у собі більш конкретну семантичну інформацію, ніж окремі слова. Тому автоматичне визначення близьких за смислом колокацій в новинному потоці даних дозволить виявляти спільний новинний контент, що відображає актуальний порядок денний.

Для визначення сили смислового зв'язку між членами заданої лексичної підсистеми, тобто колокації, пропонується застосувати дистрибутивно-статистичну методику, суть якої полягає в статистичному аналізі спільної зустрічальності лексем у великих потоках даних. Алгоритми аналізу [3, 4] використовують засоби математичної статистики і алгебри, лінгвістична інформація присутня тільки в морфологічній розмітці колокацій.

Статистичні метрики або міри асоціації ґрунтуються на частотах колокацій і окремих колокатів (слів), що входять до них, для обчислення стійкості, притаманної лексичним одиницям. Всього існує більше 80 мір, що дозволяють оцінити силу зв'язності [5–7]. Частіше за інших використовуються такі міри, як MI, PMI, t-score, log-likelihood, коефіцієнт ймовірності, критерій узгодженості Пірсона та інші.

Однак, статистичні методи витягають додаткові шумні дані та ігнорують синтаксичні зв'язки між словами на довгих відстанях. Водночас визначення колокацій вимагає на додаток до моделей статистичного аналізу використання морфологічних та синтаксичних засобів. У цьому випадку ідентифікація семантично близьких колокацій дозволяє не тільки враховувати імовірність спільної появи компонентів колокацій, а й формалізувати граматичні залежності між головним та залежним компонентами [8].

Отже, аналіз статистичних методів і моделей дистрибутивної семантики при вирішенні завдання виявлення смислових зв'язків між словами показав, що найбільш продуктивною є комбінація статистичної міри (зокрема, коефіцієнта взаємної інформації MI, який порівнює залежні контекстно-зв'язані частоти з незалежними) та розробленої моделі семантичної подібності, що базується на алгебрі скінченних предикатів [9], які описують відношення між компонентами колокації.

Метою дослідження є визначення спільного інформаційного простору (порядку денного), що представляє актуальні питання в новинних потоках даних, шляхом моделювання функцій інтелекту з розуміння смислу.

Матеріали і результати дослідження. Для визначення близького за змістом новинного контенту розроблено корпус новинних текстів, що постійно поповнюється. Корпус включає статті новинних сайтів BBC [10] і CNN [11], автоматично витягнуті за допомогою інструменту Scrapy.

На першому етапі розробленої інформаційної технології (рис. 1) проводиться пошук потенційних кандидатів – головних та залежних компонентів колокацій – на основі визначення статистичної значимості колокацій за допомогою моделі дистрибутивної семантики MI, що дозволяє виділяти рідкісні колокації і підходить для виділення термінології, власних імен та інших конструкцій, в яких показники частоти слів колокації зовсім малі.

Для встановлення лінгвістичної правильності виявлених кандидатів, а саме їх лексичної залежності, використовується морфологічна розмітка (POS-tagging) та розроблені регулярні вирази, що описують 3 типи колокацій: дієслівні, субстантивні та атрибутивні. Відповідно до корпусно-орієнтованого підходу ці колокації являють собою значимі фрагменти тексту, що найбільш часто зустрічаються в корпусах [12].

Таким чином, використовуючи Python NLTK, визначаємо дієслова (x^i): 'MD', 'VB', 'VBD', 'VBG', 'VBP', 'VBZ', 'VBN', іменники (y^i): 'NN', 'NNS', 'NNP', 'NNPS' і прикметники (z^i): 'JJ', 'JJR', 'JJS' та здобуваємо послідовності слів, які утворюють синтаксичні конструкції (біграми): xy, yy та $zy|yx$.

Другий етап інформаційної технології полягає у виділенні близьких за смислом колокацій. Для цього пропонується розроблена логічна модель семантичної подібності, заснована на використанні алгебро-предикатних операцій і предиката семантичної еквівалентності γ_i.

Предикат γ_1 показує семантичну близькість дієслівних колокацій, γ_2 – субстантивних колокацій, γ_3 – атрибутивних колокацій:

$$\gamma_1(coll_1, coll_2) = x_1^{[1-6]} y_1^{[1-4]} x_2^{[1-6]} y_2^{[1-4]} \qquad (1)$$

$$\gamma_2(coll_1, coll_2) = y_1^{[1-4]} y_1^{[1-4]} y_2^{[1-4]} y_2^{[1-4]} \quad (2)$$

$$\gamma_3(coll_1, coll_2) = z_1^{[1-3]} y_1^{[1-4]} z_2^{[1-3]} y_2^{[1-4]} \vee \vee z_1^{[1-3]} y_1^{[1-4]} y_2^{[1-4]} x_2^7 \vee \vee y_1^{[1-4]} x_1^7 y_2^{[1-4]} x_2^7 \quad (3)$$

Отже, предикат γ_i описує семантико-синтаксичні зв'язки між колокаціями, тоді як тезаурус WordNet 3.1.0 надає інформацію про синонімічність колокатів.

Наприклад, наступні фрагменти новинних повідомлень BBC і CNN були визначені як семантично близькі, оскільки містять близькі за смислом колокації, що формують спільний порядок денний:

…A flooding emergency($y_1^1 y_1^1$) in the Washington DC area left commuters in hazardous conditions. Torrential downpours led to road closures and left drivers stranded ($y_1^2 x_1^7$) as well as dangerous flooding on the underground rail-lines.

…A flash flood emergency($y_2^1 y_2^1$) in Washington left roads submerged and cars stranded ($y_2^2 x_2^7$) as heavy rains poured over the region.

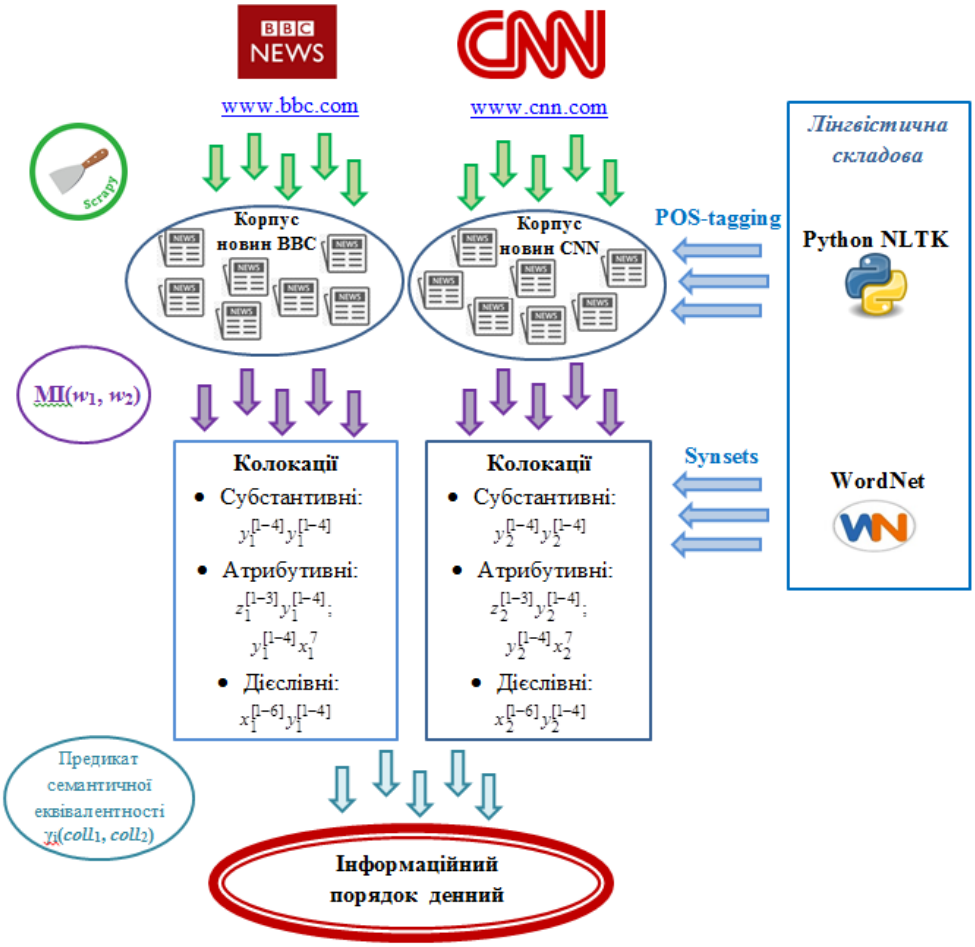

Рис. 1. Етапи інформаційної технології визначення інформаційного порядку денного в потоках новинних даних

Таким чином, текстові фрагменти вважаються семантично подібними, якщо граматичні характеристики синонімічних колокацій задовольняють предикату семантичної еквівалентності.

Вилучені з новин колокації дозволяють не тільки виявити відсутні в лінгвістичних джерелах поєднання слів, а й обчислити статистичні показники їх стійкості. У такий спосіб підставами визначення певного виразу (колокації) визнаються граматична й статистична моделі, представлені певним контекстом.

Розроблена технологія реалізована у вигляді програмного застосунка (рис. 2) з виявлення колокацій з великих потоків даних. Програмна імплементація дозволяє завантажувати новинні повідомлення, вилучати семантично близькі колокації та відображати їх у порядку спадання значення коефіцієнта MI.

Результатом роботи програмного додатка є виявлені у корпусі новинних текстів близькі за смислом колокації, що дозволяють визначити спільний інформаційний простір актуальних новин.

Для оцінки ефективності розробленої технології та порівняльного аналізу з існуючими технологіями вилучення колокацій з великих потоків даних проведено розрахунок показника точності (precision). Аналіз показав, що середній коефіцієнт точності, визначений відношенням числа релевантних колокацій та неправильно прийнятих системою рішень (не релевантних колокацій) до загального числа колокацій, виданих системою, дорівнює 0,96.

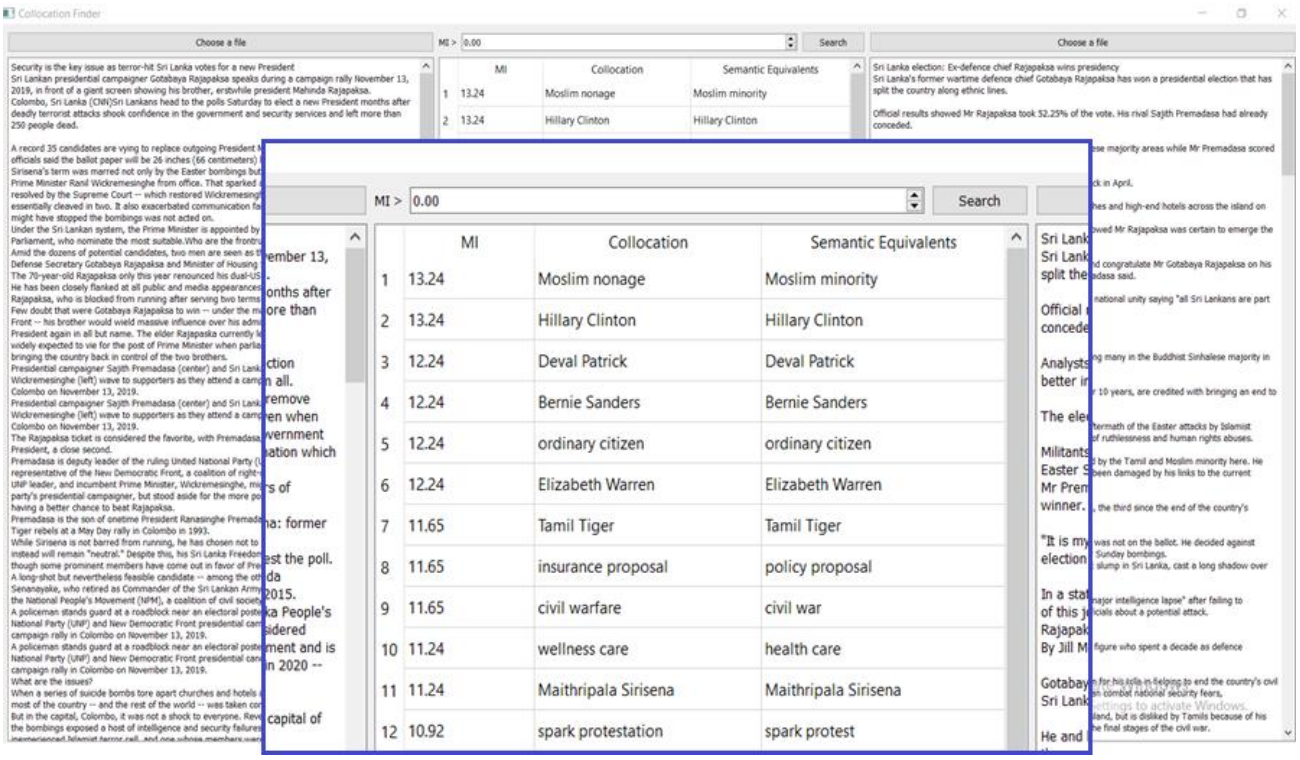

Рис. 2. Програмна реалізація виявлення колокацій з великих потоків даних

Для порівняння отриманих результатів з іншим існуючим програмним забезпеченням, що займається аналогічним завданням, було обрано web-додаток Sketch Engine [13]. Цей додаток включає в себе функцію визначення вордскетчів (word sketches), тобто типових фраз, що визначаються, з одного боку, синтаксисом, який обмежує сполучуваність слів в певній мові, та з іншого – ймовірністю, пов'язаною з частотою використання слів.

До обраного застосунка було завантажено дослідницький корпус новинних повідомлень для автоматичного вилучення колокацій. Результати порівняльного аналізу оцінки ефективності застосунків представлені в табл. 1.

Таблиця 1 – Порівняльний аналіз показників Precision

	Всього вилучено колокацій	Оціночна вибірка	Precision
Sketch Engine	1911	1000	0.57
Розроблене програмне забезпечення	3227		0.96

Висновки. На основі розробленої логіко-лінгвістичної моделі та моделі дистрибутивної семантики запропоновано інформаційну технологію визначення смислової подібності новинного контенту з використанням алгебро-предикатних операцій для формалізації семантико-граматичних характеристик колокацій. Встановлюючи неявні смислові зв'язки між фрагментами новинного тексту, технологія дозволяє ідентифікувати спільні інформаційні простори актуальних новин – інформаційний порядок денний.

В результаті проведеного порівняльного аналізу виявлено, що показник precision розробленої технології значно перевищує аналогічний показник ефективності існуючого web-додатка Sketch Engine.

Розроблена програмна імплементація дозволить поліпшити якість обробки потоків новинних повідомлень, а також може бути запропонована як покращена модель існуючих систем семантичної обробки текстів, наприклад, використовуватись для виявлення текстів однієї тематики, актуальної інформації, добування фактів та усунення смислової неоднозначності та ін.

Список літератури

1. Каминченко Д.И. Информационная повестка дня современных сетевых СМИ: политический аспект. *Via in tempore. История. Политология*. 2019. Т. 46, № 3. С. 576–584.
2. Adams A., Harf A., Ford R. Agenda Setting Theory: A Critique of Maxwell McCombs & Donald Shaw's Theory In Em Griffin's A First Look at Communication Theory. *Meta-communicate*. 2014. Vol. 4, no. 1. URI: https://journals.chapman.edu/ojs/index.php/mc/article/view/902 (дата звернення: 23.04.2021).
3. Lenci A. Distributional Models of Word Meaning. *Annual Review of Linguistics*. 2018. Vol. 4. P. 151–171. URI: https://www.annualreviews.org/doi/pdf/10.1146/annurev-linguistics-030514-125254 (дата звернення: 23.04.2021).
4. Dinu A., Dinu L., Sorodoc I. Aggregation methods for efficient collocation detection. *Proceedings of the Ninth International Conference on Language Resources and Evaluation*. 2014. P. 4041–4045. URI: http://www.lrec-conf.org/proceedings/lrec2014/pdf/1184_Paper.pdf (дата звернення: 23.04.2021).
5. Хохлова М. В. Сопоставительный анализ статистических мер на примере частеречных преференций сочетаемости существительных. *Компьютерная лингвистика и вычислительные онтологии*. 2017. Вып. 1. С. 166–171.

6. Liu X., Huang D., Yin Zh., Ren F. Recognition of Collocation Frames from Sentences. *IEICE Trans. Inf. Syst*. 2019. 102-D. P. 620–627. URI: https://doi.org/10.1587/TRANSINF.2018EDP7255 (дата звернення: 23.04.2021)
7. Хохлова М. В. К вопросу о сходстве мер ассоциации применительно к задаче автоматического извлечения глагольных коллокаций. *Компьютерная лингвистика и вычислительные онтологии*. 2019. Вып. 3. С. 9–18.
8. Petrasova S., Khairova N., Lewoniewski W., Mamyrbayev O., Mukhsina K. Similar Text Fragments Extraction for Identifying Common Wikipedia Communities. *Data. Stream Mining and Processing*. 2018, vol. 3, issue 4, article 66. URI: https://doi.org/10.3390/data3040066 (дата звернення: 23.04.2021).
9. Бондаренко М.Ф., Шабанов-Кушнаренко Ю.П. *Теория интеллекта*. Харьков: СМИТ, 2007. 576 с.
10. *BBC*. URI: https://www.bbc.com/news (дата звернення: 23.04.2021).
11. *CNN*. URI: https://edition.cnn.com (дата звернення: 23.04.2021)
12. Бобкова Т. В. Основні підходи до ідентифікації й вилучення колокацій із текстів. *Наукові праці. Філологія. Мовознавство*. 2015. № 241 (253). С. 10–16. URI: http://linguistics.chdu.edu.ua/article/viewFile/87653/83242/ (дата звернення: 23.04.2021).
13. *Sketch Engine*. URI: https://www.sketchengine.eu (дата звернення: 23.04.2021).

References (transliterated)

1. Kaminchenko D.I. Informacionnaja povestka dnja sovremennyh setevyh SMI: politicheskij aspekt [Information Agenda of Modern Online Media: Political Aspect]. *Via in tempore. Istorija. Politologija* [Via in tempore. History. Political science]. 2019, vol. 46, no. 3, pp. 576–584
2. Adams A., Harf A., Ford R. Agenda Setting Theory: A Critique of Maxwell McCombs & Donald Shaw's Theory In Em Griffin's A First Look at Communication Theory. *Meta-communicate*. 2014, vol. 4, no. 1. URI: https://journals.chapman.edu/ojs/index.php/mc/article/view/902 (accessed 23.04.2021).
3. Lenci A. Distributional Models of Word Meaning. *Annual Review of Linguistics*. 2018, vol. 4, pp. 151–171. URI: https://www.annualreviews.org/doi/pdf/10.1146/annurev-linguistics-030514-125254
4. Dinu A., Dinu L., Sorodoc I. Aggregation methods for efficient collocation detection. *Proceedings of the Ninth International Conference on Language Resources and Evaluation*. 2014, pp. 4041–4045. URI: http://www.lrec-conf.org/proceedings/ lrec2014/pdf/ 1184_Paper.pdf (accessed 23.04.2021).
5. Hohlova M.V. Sopostavitel'nyj analiz statisticheskih mer na primere chasterechnyh preferencij sochetaemosti sushhestvitel'nyh [Comparative Analysis of Statistical Measures on the Example of Part-of-Speech Preferences for Combining Nouns]. *Komp'juternaja lingvistika i vychislitel'nye ontologii* [Computational linguistics and computational ontologies]. 2017, issue 1, pp. 166–171.
6. Liu X., Huang D., Yin Zh., Ren F. Recognition of Collocation Frames from Sentences. *IEICE Trans. Inf. Syst*. 2019, 102-D, pp. 620-627. URI: https://doi.org/10.1587/TRANSINF.2018EDP7255
7. Hohlova M.V. K voprosu o shodstve mer associacii primenitel'no k zadache avtomaticheskogo izvlechenija glagol'nyh kollokacij [To the Question of The Similarity of Association Measures Applied to the Problem of Automatic Extraction of Verb Collocations]. *Komp'juternaja lingvistika i vychislitel'nye ontologii* [Computational linguistics and computational ontologies]. 2019, issue 3, pp. 9–18.
8. Petrasova S., Khairova N., Lewoniewski W., Mamyrbayev O., Mukhsina K. Similar Text Fragments Extraction for Identifying Common Wikipedia Communities. *Data. Stream Mining and Processing*. 2018, vol. 3, issue 4, article 66. URI: https://doi.org/10.3390/data3040066 (accessed 23.04.2021).
9. Bondarenko M., Shabanov-Kushnarenko Yu. *Teorija intellekta* [The Theoryof Intelligence]. Kharkiv, SMIT Publ., 2007. 576 p.
10. *BBC*. URI: https://www.bbc.com/news (accessed 23.04.2021).
11. *CNN*. URI: https://edition.cnn.com (accessed 23.04.2021).
12. Bobkova T. V. Osnovni pidhody do identifikatsii i vyluchennia kolokatsiy iz tekstiv [Basic approaches to identification and extraction of collocations from texts]. *Naukovi pratsi. Filologia. Movoznavstvo* [Scientific works. Philology. Linguistics]. 2015, no. 241 (253), pp. 10–16. URI: http://linguistics.chdu.edu.ua/article/viewFile/87653/83242/ (accessed 23.04.2021).
13. *Sketch Engine*. URI: https://www.sketchengine.eu (accessed 23.04.2021).

Надійшла (received) 05.05.2021

Відомості про авторів / Сведения об авторах / About the Authors

Петрасова Світлана Валентинівна – кандидат технічних наук, доцент, Національний технічний університет «Харківський політехнічний інститут», доцент кафедри інтелектуальних комп'ютерних систем; м. Харків, Україна; ORCID: https://orcid.org/ 0000-0001-6011-135X; e-mail: svetapetrasova@gmail.com.

Хайрова Ніна Феліксівна – доктор технічних наук, професор, Національний технічний університет «Харківський політехнічний інститут», професор кафедри інтелектуальних комп'ютерних систем; м. Харків, Україна; ORCID: https://orcid.org/0000-0002-9826-0286; e-mail: nina_khajrova@yahoo.com.

Колесник Анастасія Сергіївна – Національний технічний університет «Харківський політехнічний інститут», аспірантка кафедри інтелектуальних комп'ютерних систем; м. Харків, Україна; ORCID: https://orcid.org/0000-0001-5817-0844; e-mail: kolesniknastya20@gmail.com.

Петрасова Светлана Валентиновна – кандидат технических наук, доцент, Национальный технический университет «Харьковский политехнический институт», доцент кафедры интеллектуальных компьютерных систем; г. Харьков, Украина; ORCID: https://orcid.org/ 0000-0001-6011-135X; e-mail: svetapetrasova@gmail.com.

Хайрова Нина Феликсовна – доктор технических наук, профессор, Национальный технический университет «Харьковский политехнический институт», профессор кафедры интеллектуальных компьютерных систем; г. Харьков, Украина; ORCID: https://orcid.org/0000-0002-9826-0286; e-mail: nina_khajrova@yahoo.com.

Колесник Анастасия Сергеевна – Национальный технический университет «Харьковский политехнический институт», аспирант кафедры интеллектуальных компьютерных систем; г. Харьков, Украина; ORCID: https://orcid.org/0000-0001-5817-0844;e-mail: kolesniknastya20@gmail.com.

Petrasova Svitlana Valentynivna – candidate of technical sciences, docent, National Technical University «Kharkiv Polytechnic Institute», associate professor of the Department of Intelligent Computer Systems; Kharkiv, Ukraine; ORCID: https://orcid.org/ 0000-0001-6011-135X; e-mail: svetapetrasova@gmail.com.

Khairova Nina Feliksivna – doctor of technical sciences, professor, National Technical University «Kharkiv Polytechnic Institute», professor of the Department of Intelligent Computer Systems; Kharkiv, Ukraine; ORCID: https://orcid.org/0000-0002-9826-0286; e-mail: nina_khajrova@yahoo.com.

Kolesnyk Anastasiia Sergiivna – National Technical University «Kharkiv Polytechnic Institute», PhD student of the Department of Intelligent Computer Systems; Kharkiv, Ukraine; ORCID: https://orcid.org/0000-0001-5817-0844; e-mail: kolesniknastya20@gmail.com.