

## References (transliterated)

1. Shutova M. O. Neologizmy v suchasni angliyskii movi [Neologisms in modern English]. *Naukovyi visnyk kafedry UNESCO Kyivskogo natsionalnogo lingvistychnogo universitetu. Filologiya, pedagogika, psykholohiia* [Scientific Herald of the Department of the UNESCO of the Kyiv National Linguistic University. Philology, pedagogy, psychology.]. 2010, issue 21, pp. 79–85.
2. Rey A. *Semantics*. Cambridge: Cambridge University Press, 1995. 184 p.
3. Koshmanova T. V. *Neologizmy yak osnovnyi zasib popovnennya slobnikovogo skladu ekonomichnoi sfery movy* [Neologisms as the main means of replenishing the vocabulary of the economic sphere of language]. *Visnyk studentskogo naukovo tovarystva Gorlivskogo derzhavnogo pedagogichnogo instytutu inozemnyh mov* [Bulletin of Student Scientific Society of Gorlovka State Pedagogical Institute of Foreign Languages]. Gorlivka, 2008, issue 5. 148 p.
4. Andrusyak I. V. *Angliyski neologizms kintsya XX stilitya yak skladova movnoi kartyny svitu : avtoref. dys. ... kand. filol. nauk: 10.02.04 "Germanski movy"* [English neologisms of the late twentieth century as a component of the linguistic picture of the world. Abstract of a thesis candidate filol. sci. diss. (Ph. D.). – Kyiv, 2003. 20 p.
5. Arnold I. V. *Leksikologiya sovremenno angliiskogo yazyka* [Lexicology of modern English]. Moscow, Vysshaya shkola, 1986. 295 p.
6. Voloh O. T., Chemyrsov M. H. *Suchasna ukrainska literaturna mova* [Contemporary Ukrainian Literary Language]. Kyiv, Vyshcha shkola, 1976. 453 p.
7. Chertoryzka T. K. *Novi znachennya staryh sliv* [New meanings of old words]. *Ridne slovo* [Native word] 1974, no. 9, pp. 20–26.
8. Yenikieva S. M. *Systema slovtvoru suchasnoi angliyskoi movy: synergetychnyi aspekt (na materialii novoutvoren kintsya XX – npochatky XXI stolit): avtoref. dys. ... d-ra filiol. nauk : 10.02.04* [System of word formation of contemporary English: synergetic aspect (on the material of neoplasm of the end of XX - beginning of the XXI centuries). Abstract of a thesis doctor filol. sci. diss.]. Kyiv, 2011. 32 p.
9. Zatsnyi Yu. A. *Rozvytok slovnkovogo skladu suchasnoi angliyskoi movy* [Development of the vocabulary of contemporary English]. Zaporizhzhya, Vyd-vo Zaporizkogo derzh. un-tu, 1998. 430 p.
10. Zatsnyi Yu. A., Yankov A. V. *Innovatsii u slovnkovogo skladi suchasnoi angliyskoi movy pochatku XXI stolittya: anglo-ukrainskii slovnyk* [Innovations in the vocabulary of English at the beginning of the 21st century: the English-Ukrainian dictionary]. Vinnytsya, Nova Knyga, 2008. 360 p.
11. Zulinska M. O. *Afiksatsiya yak metod utvorennya angliyskikh neologizmiv sfery informatsyinyh tehnologiy* [Affixalization as a method for the formation of English neologisms in the field of information technology] *Naukovyi visnyk Volynskogo natsionalnogo universitetu imeni Lesi Ukrainky : nauk. zhurnal* [Scientific herald of Volyn National University named after Lesia Ukrainka]. Lutsk , PVV "Vezha", VNU im. Lesi Ukrainky, 2009, no. 6, pp. 362–366.
12. *Macmillan Dictionary*. Available at: <https://www.macmillan-dictionary.com> (accessed 18.10.2018).
13. *Wordspy*. Available at: <https://www.wordspy.com> (accessed 18.10.2018).
14. Duzhikova E. A. *Abbreviatsiya sravnitel'no so slovoslozheniyem na materiale sovremenno angliyskogo yazyka* [Abbreviation compared with the vocabulary of modern English]. – Moscow, Nauka, 1997. 150 p.
15. *ABBYY Lingvo*. Available at: [www.lingvo.ua](http://www.lingvo.ua) (accessed 18.10.2018).

Поступила (received) 17.11.2018

## Відомості про авторів / Сведения об авторах / About the Authors

**Недайнова Ірина Василівна (Недайнова Ирина Васильевна, Nedainova Iryna Vasylivna)** – кандидат філологічних наук, доцент, Національний технічний університет «Харківський політехнічний інститут», професор кафедри ділової іноземної мови та перекладу; ORCID: [orcid.org/0000-0003-3785-0597](https://orcid.org/0000-0003-3785-0597); e-mail: [nedafox@gmail.com](mailto:nedafox@gmail.com).

**Однолеткова Анастасія Олександрівна (Однолеткова Анастасия Александровна, Odnoletkova Anastasiya Oleksandrivna)** – Національний технічний університет «Харківський політехнічний інститут», студентка; м. Харків, Україна; e-mail: [odnoletkova@ukr.net](mailto:odnoletkova@ukr.net).

УДК 004.912

DOI: 10.20998/2227-6890.2019.01.17

**С. В. ПЕТРАСОВА, Я. Р. ГАЛКІНА, І. О. МАНУЙЛОВ, О. Р. БОРОДІНА, С. І. ШВЕЦЬ**

## МЕТОД АВТОМАТИЧНОГО ВИЗНАЧЕННЯ СЕМАНТИЧНО БЛИЗЬКИХ ФРАГМЕНТІВ НОВИНИХ ТЕКСТІВ

Складність семантичного аналізу текстової інформації, що міститься в новинних повідомленнях, визначається багатозначністю і синонімічністю, які властиві мові на всіх рівнях її представлення, що, перш за все, впливає на визначення смислової близькості мовних одиниць. Виявлення семантично близьких фрагментів текстів або перефразувань є актуальною проблемою у таких наукових напрямках як семантичний пошук інформації, видобування інформації, машинний переклад, визначення порушень авторських прав і т.п., крім того широко використовується при рерайтингу. У статті проаналізовано основні проблеми рерайтингу, зокрема перефразування синтаксичних одиниць тексту зі збереженням смислового навантаження. Розглянуто сучасні методи визначення семантичної близькості слів, вказано основні переваги та недоліки. Запропоновано метод автоматичного виявлення синонімічних фрагментів новинних текстів на основі використання WordNet та розроблених синтаксичних правил, які зберігають інформацію про граматичні характеристики слів. Перевагою даного методу є те, що аналізується як граматична структура мови, так і зміст слів. Досліджуваний корпус представлено новинними текстами інформаційного агентства Reuters, служб CNN і BBC World News. Запропонований метод ідентифікації семантично зв'язаних фрагментів тексту дозволяє виявити спільний інформаційний простір актуальних новин та може використовуватися для ефективного визначення близькості за змістом текстів в інформаційно-пошукових, експертних, аналітичних інформаційних системах. Вирішення завдання автоматичного визначення семантичної близькості може застосовуватися при автоматизованій побудові онтологій по тексту, для розширення існуючих і створення нових тезаурусів.

**Ключові слова:** рерайтинг, перефразування, семантична близькість, синтаксичні правила, WordNet, корпус новинних текстів.

© С. В. Петрасова, Я. Р. Галкіна, І. О. Мануйлов, О. Р. Бородіна, С. І. Швець, 2019

**С. В. ПЕТРАСОВА, Я. Р. ГАЛКИНА, И. А. МАНУЙЛОВ, А. Р. БОРОДИНА, С. И. ШВЕЦ  
МЕТОД АВТОМАТИЧЕСКОГО ОПРЕДЕЛЕНИЯ СЕМАНТИЧЕСКИ БЛИЗКИХ  
ФРАГМЕНТОВ НОВОСТНЫХ ТЕКСТОВ**

Сложность семантического анализа текстовой информации, содержащейся в новостных сообщениях, определяется многозначностью и синонимичностью, свойственные языку на всех уровнях его представления, что, прежде всего, влияет на определение смысловой близости языковых единиц. Выявление семантически близких фрагментов текстов или перефразирований является актуальной проблемой в таких научных направлениях как семантический поиск информации, извлечение информации, машинный перевод, определение нарушений авторских прав и т.п., кроме того широко используется в рерайтинге. В статье проанализированы основные проблемы рерайтинга, в частности перефразирование синтаксических единиц текста с сохранением смысловой нагрузки. Рассмотрены современные методы определения семантической близости слов, указаны основные преимущества и недостатки. Предложен метод автоматического выявления синонимических фрагментов новостных текстов на основе использования WordNet и разработанных синтаксических правил, которые хранят информацию о грамматических характеристиках слов. Преимуществом данного метода является то, что анализируется как грамматическая структура языка, так и смысл слов. Исследуемый корпус представлен новостными текстами информационного агентства Reuters, служб CNN и BBC World News. Предложенный метод идентификации семантически связанных фрагментов текста позволяет определить общее информационное пространство актуальных новостей и может быть использован для эффективного выявления близких по смыслу текстов в информационно-поисковых, экспертных, аналитических информационных системах. Решение задачи автоматического определения семантической близости может применяться при автоматизированном построении онтологий по тексту, для расширения существующих и создание новых тезаурусов.

**Ключевые слова:** рерайтинг, перефразирование, семантическая близость, синтаксические правила, WordNet, корпус новостных текстов.

**S. V. PETRASOVA, YA. R. GALKINA, I. O. MANUILOV, O. R. BORODINA, S. I. SHVETS  
METHOD FOR AUTOMATIC IDENTIFICATION OF SEMANTICALLY SIMILAR  
FRAGMENTS OF NEWS TEXTS**

Depending on the ambiguity and synonymy at all language levels, the identification of semantic similarity of linguistic units is the challenging task of semantic analysis of text information contained in news reports. The extraction of semantically similar fragments of texts or paraphrases is an up-to-date problem in fields of science such as semantic information retrieval, information extraction, machine translation, detection of copyright infringements, etc. and is widely used in rewriting. The article analyzes the main problems of rewriting, in particular, the paraphrasing of syntactic text units keeping the sense load. The modern methods for identification of semantic similarity of words, their advantages and disadvantages are considered. Based on the use of WordNet and the developed syntactic rules that store information about the grammatical characteristics of words, a method for automatic identification of synonymous fragments of news texts is proposed. The advantage of this method is that both the grammatical structure of the language and the meaning of words (using WordNet) are analyzed. The experimental corpus is represented by news texts from Reuters news agency, BBC World News and CNN services. The proposed method for identifying semantically similar text fragments allows defining the common information space of current news and can be used to effectively identify related texts in information retrieval, expert, analytical information and rewriting systems. The automatic identification of semantic similarity could be implemented in automated construction of ontologies, in expansion of existing and creation of new thesauri.

**Keywords:** rewriting, paraphrasing, semantic similarity, syntactic rules, WordNet, corpus of news texts.

**Вступ.** Кожного дня мережа Інтернет поповнюється великою кількістю нової інформації різної направленості. У той же час багато сайтів, особливо новинних, створюють подібний за змістом текстовий контент. Проблема формування унікального тексту на основі вже існуючих статей залишається актуальною, зокрема, при ідентифікації інформаційного простору актуальних новин та подальшому їх просуванні в пошукових системах.

Одним з інструментів вирішення цієї проблеми є рерайтинг – ретельна переробка тексту, результатом якої є новий текст, що має з оригіналом спільний зміст та відрізняється лексичними засобами, семантично близькими словами та стилем написання.

Існуючі програмні реалізації, зазвичай, нездатні виконувати рерайтинг цілого тексту чи дискурсу. Зокрема, складність організації пошуку семантично близьких слів визначається рядом причин. По-перше, автору невідомо загальноприйнятої кількісної міри для визначення ступеня синонімічності значень слів.

По-друге, поняття синонімії визначено не для слів, а для значень слів, тобто синонімія нерозривно пов'язана з контекстом.

По-третє, мова – це постійно змінна субстанція, відкрита система. Слова можуть застарівати або

отримувати нові значення. Тому в напрямках науки, що активно розвиваються, спостерігається особливо активний словотвір і присвоєння нових значень словами.

Використання методу визначення семантично близьких фрагментів новинних текстів дозволить виділяти спільний простір актуальних новин, що забезпечить релевантний пошук та написання (рерайтинг) статей за схожими темами.

**Аналіз останніх досліджень і публікацій.** У сучасній лінгвістиці у зв'язку із зростаючим інтересом до дослідження смислу мовних одиниць широко застосовуються технології встановлення змістової подібності між одиницями. Формально такий механізм зводиться до синонімічної заміни елементів, розгортання структури (додавання елементів) або згортання структури (опущення елементів).

При перетворенні складного тексту в більш простий та, як результат, написанні унікальних текстів у рерайтингу використовуються наступні методики роботи.

Трансформація прямої мови у непряму. Ця техніка дозволяє залишити необхідний зміст у тексті,

але у той же час робить інформацію для пошукових систем унікальною.

Зменшення розміру тексту для його спрощення та кращого розуміння контенту.

Переробка структури тексту: переміщення абзаців тексту, зміна граматичних конструкцій речень, що крім унікальності додає до нового тексту новий стиль написання, не змінюючи його смисл [1].

При рерайтингу широко використовуються перефразування (або парафрази). В залежності від того, на якому рівні мови ведеться перефразування, в його процесі змінюється лексика, синтаксична структура, морфологічні характеристики слів, їх число і порядок. При цьому може замінюватися одне слово при збереженні всієї структури або вся структура при збереженні лексичних одиниць.

Існує декілька способів перефразування синтаксичних одиниць тексту:

- зміна граматичної структури речення, наприклад, заміна підмета і додатка;
- заміна слів з однієї частини мови в іншу, наприклад, з дієслова зробити іменник або прикметник;
- збільшення структури (додавання елементів) і навпаки;
- розбиття довгих речень на кілька менших та навпаки;
- синонімічні заміни слів/словосполучень [2].

Серед існуючих методів виявлення семантично близьких слів виділяють дві групи: статистичні методи [3–8] (табл. 1) і методи, що основані на використанні онтологій і тезаурусів [9–10].

На сьогоднішній день для визначення відношення семантичної близькості широко використовуються онтології – формальний опис термінів предметної області і відношень між ними. До достоїнств онтологій відноситься їх висока ефективність при інтеграції різнорідних джерел інформації та знань, які можуть бути розмічені за допомогою термінів онтології, незалежно від їх форматів і представлень.

Однак для вирішення багатьох практичних завдань необхідні гібридні знання, що включають не тільки понятійні, а й процедурні, евристичні знання. Тому при ідентифікації смислової близькості опис предметної області, створений за допомогою онтологій, не є вичерпним.

На ряду з онтологією часто використовують і тезаурус, як особливий різновид словників загальної або спеціальної лексики, в яких вказані семантичні відношення (синоніми, антоніми, гіпоніми, гіпероніми тощо) між лексичними одиницями.

Тезаурус дозволяє виявити смисл не тільки за допомогою визначення, але і за допомогою співвіднесення слова з іншими поняттями та їх групами.

Таблиця 1 – Статистичні методи виявлення семантичної близькості слів

№	Метод	Опис	Базові джерела
1	Метод відстаней	Два слова вважаються «майже синонімами» (near synonyms), якщо в тлумачному словнику, де вони визначені: - вживаються у визначенні одного і того ж слова; - мають спільні слова у визначеннях.	Тлумачний словник
2	Формування векторних представлень слів	Семантична близькість слів відображає їх спільну зустрічальність в подібних контекстах. Формуються векторні представлення слів шляхом автоматичного виявлення статистики їхньої спільної зустрічальності з великих масивів текстової інформації.	Корпус текстів
3	Латентно-семантичний аналіз	Між окремими словами і узагальненим контекстом (реченнями, абзацами і цілими текстами), в яких вони зустрічаються, існують неявні (латентні) взаємозв'язки.	Корпус текстів
4	Метод Клайнберга	На сторінку x посилаються хаб-сторінки, які можуть мати посилання на авторитетні сторінки. На сторінку x будуть схожі ті з авторитетних сторінок, на які посилаються ті самі хаб-сторінки, які, в свою чергу, посилаються на сторінку x.	Гіпертекстово зв'язані веб-сторінки
5	Метод побудови wiki-списків семантично близьких слів	Алгоритм буде впорядкований список семантично близьких слів, ґрунтуючись на близькі за змістом статті Вікіпедії.	Тексти Вікіпедії

Одним з найбільш успішних проєктів, пов'язаних з тезаурусом, є WordNet – тезаурус або семантична мережа англійської мови, яка представляє склад і структуру лексичної мови в цілому. WordNet містить понад 150 тис. слів, організованих в більш ніж 115 тис. сінсетів. Словник складається з 4 мереж, що відповідають таким частинам мови як: іменник, дієслово, прикметник і прислівник.

WordNet групує набори слів зі схожим значенням в сінсети. WordNet містить сінсети, короткі загальні визначення до сінсетів (глоси), приклади вживань і кілька типів семантичних відношень між сінсетами.

Сінсети кожної частини мови мають власний набір семантичних відношень. Найбільш часто використовуваним відношенням між сінсетами іменників є відношення гіпонім–гіперонім, яке

встановлюється між родовими і видовими словами. Відношення між іменниками включають також відношення частини-цілого.

Між якісними прикметниками описуються відношення антонімії і подібності.

Дієслова в WordNet зв'язані відношенням гіпонім–гіперонім. Крім того, вони мають свій власний набір відношень, які не використовуються для опису іменників або прикметників, наприклад, відношення наслідку і відношення каузації [11].

Отже, аналіз методів і підходів до вирішення завдання виявлення семантичних зв'язків між словами показав, що найбільш продуктивною є комбінація формальних і неформальних методів.

**Метою дослідження** є розробка метода автоматичного виявлення семантично близьких фрагментів у корпусі новинних текстів для формування спільного простору актуальних новин.

**Матеріали і результати дослідження.** В роботі пропонується використовувати розроблені синтаксичні правила [12] для визначення словосполучень та WordNet для виявлення синонімічних слів у корпусі текстів.

Джерелом текстової інформації для корпусу є статті інформаційного агентства Reuters [13], служб CNN [14] і BBC World News [15].

Для морфологічної обробки текстів пропонується застосувати POS-тегування засобами бібліотеки NLTK мови Python.

На початковому етапі відбувається пошук кандидатів у синонімічні словосполучення в WordNet (рис. 1).

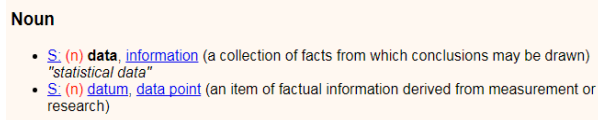


Рис. 1 – Приклад сіцнета у WordNet

Далі виконується пошук синонімічних пар у словосполученнях. Перевіряється відповідність граматичних характеристик колокатів (слів словосполучень) потенційно синонімічних словосполучень синтаксичним правилам:

$$VTr + NObjPac \quad (1)$$

$$NObjAtt + NSubAg | NSubAg + of + NObjAtt \quad (2)$$

$$NSubAg + APr | AAtt + NSubAg \quad (3)$$

Таким чином, словосполучення, граматичні характеристики яких відповідають правилам, вважаються синонімічними.

**Висновки.** В результаті проведеного аналізу методів виявлення семантичної близькості слів було запропоновано метод автоматичного визначення синонімічних конструкцій у корпусі новинних текстів, що дозволяє виявити спільний інформаційний простір актуальних новин. Результати дослідження можуть використовуватися для ефективного визначення близьких за змістом текстів в інформаційно-пошукових системах.

#### Список літератури

1. Колоев А.С. Рерайт как новое явление в современной журналистике. *Вестник Санкт-Петербургского университета*. Серия. Филология. 2012. Вып. 1. С. 221–226.
2. Большаков И.А. Два метода синонимического перефразирования в лингвистической стеганографии. *Труды международной конференции «Диалог 2004»*. URI: <http://www.dialog-21.ru/media/2496/bolshakov.pdf> (дата обращения 16.09.2018).
3. Senellart P., Blondel V.D. Automatic Discovery of SimilarWords. In *Survey of Text Mining II*. Springer, 2008. P. 25–44.
4. Wang T., Hirst G. Extracting Synonyms from Dictionary Definitions. *Proceedings of International Conference on Recent Advances in Natural Language processing 2009*. URI: <ftp://ftp.cs.toronto.edu/pub/gh/Wang+Hirst-RANLP-2009.pdf> (accessed 18.09.2018).
5. Мисуно И.С., Рачковский Д.А., Слипченко С.В. Векторные и распределенные представления, отражающие меру семантической связи слов. *Математические машины и системы*. 2005. № 3. С. 50–66.
6. Митрофанова О.А. Семантические расстояния: проблемы и перспективы. *Материалы XXXIV Междунар. филол. конф.* 2005. С. 59–63.
7. Kleinberg J. Authoritative sources in a hyperlinked environment. *Proc. 9th ACM-SIAM Symposium on Discrete Algorithms*, 1999. № 46 (5). P. 604–632.
8. Панченко А.И., Филиппович Ю.Н. и др. Метод и система извлечения семантических отношений из статей Википедии на основе компонентного анализа. *Материалы Междунар. науч. конф.* М.: Эйдос, 2012. С. 339–341.
9. Pantel P., Lin D. Discovering word senses from text. *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York : ACM, 2002. P. 613–619.
10. Dorow B. *A graph model for words and their meanings*. Dissertation, 2006. 187 p.
11. WordNet. URI: <https://wordnet.princeton.edu> (accessed 21.09.2018)
12. Петрасова С.В., Хайрова Н.Ф., Киселева В.Ю. Идентификация смысловой близости фрагментов текстов наукометрических баз. *International Journal of Information Technologies & Knowledge*. Bulgaria (ITHEA), 2018. Vol 12. № 1. P. 32–42.
13. Reuters. URI: <https://www.reuters.com> (accessed 10.09.2018).
14. CNN. URI: <https://edition.cnn.com> (accessed 10.09.2018).
15. BBC. URI: <https://www.bbc.com/news> (accessed 10.09.2018).

#### References (transliterated)

1. Koloev A.S. Rerajt kak novoe javlenie v sovremennoj zhurnalistike [Rewrite as a new phenomenon in modern journalism]. *Vestnik Sankt-Peterburgskogo universiteta* [Bulletin of St. Petersburg University]. 2012, issue 1, pp. 221–226.
2. Bol'shakov I.A. Dva metoda sinonimicheskogo perefrazirovaniya v lingvistichejskoj steganografii [Two methods of synonymous rephrasing in linguistic steganography]. *Trudy mezhdunarodnoj konferencii «Dialog 2004»* [Proceedings of the international conference "Dialogue 2004"]. Available at: <http://www.dialog-21.ru/media/2496/bolshakov.pdf> (accessed 16.09.2018).
3. Senellart P., Blondel V.D. Automatic Discovery of SimilarWords. In *Survey of Text Mining II*. Springer, 2008. pp. 25–44.
4. Wang T., Hirst G. Extracting Synonyms from Dictionary Definitions. *Proceedings of International Conference on Recent*

- Advances in Natural Language processing 2009*. Available at: <ftp://ftp.cs.toronto.edu/pub/gh/Wang+Hirst-RANLP-2009.pdf> (accessed 18.09.2018).
5. Misuno I.S., Rachkovskij D.A., Slipchenko S.V. Vektornye i raspredelennye predstavlenija, otrazhajushhie meru semanticheskoy svyazi slov [Vector and distributed representations reflecting the measure of the semantic connection of words]. *Matematicheskie mashiny i sistemy* [Mathematical machines and systems]. 2005, no. 3, pp. 50–66.
  6. Mitrofanova O.A. Semanticheskie rasstojanija: problemy i perspektivy [Semantic distances: problems and prospects]. *Materialy XXXIV Mezhdunar. filol. konf* [Proceedings of XXXIV Intern. philol. conf.]. 2005, pp. 59–63.
  7. Kleinberg J. Authoritative sources in a hyperlinked environment. *Proc. 9th ACM-SIAM Symposium on Discrete Algorithms*, 1999, no. 46 (5), pp. 604–632.
  8. Panchenko A.I., Filippovich Ju.N. et al. Metod i sistema izvlechenija semanticheskikh otnoshenij iz statej Vikipedii na osnove komponentnogo analiza [Method and system for extracting semantic relations from Wikipedia articles based on component analysis]. *Materialy Mezhdunar. nauch. konf.* [Proceedings of the Intern. scientific conf]. Moscow: Eidos, 2012, pp. 339–341.
  9. Pantel P., Lin D. Discovering word senses from text. *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2002, pp. 613–619.
  10. Dorow B. *A graph model for words and their meanings*. Dissertation, 2006. 187 p.
  11. WordNet. Available at: <https://wordnet.princeton.edu> (accessed 21.09.2018).
  12. Petrasova S., Khairova N., Kysilova V. Identification of Semantic Similarity of Text Fragments in Scientometric Bases. *International Journal of Information Technologies & Knowledge*. Bulgaria (ITHEA), 2018, vol. 12, no. 1, pp. 32–42.
  13. Reuters. Available at: <https://www.reuters.com> (accessed 10.09.2018).
  14. CNN. Available at: <https://edition.cnn.com> (accessed 10.09.2018).
  15. BBC. Available at: <https://www.bbc.com/news> (accessed 10.09.2018).

Надійшла (received) 16.11.2018

#### Відомості про авторів / Сведения об авторах / About the Authors

**Петрасова Світлана Валентинівна (Петрасова Светлана Валентиновна, Petrasova Svitlana Valentynivna)** – кандидат технічних наук, 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](mailto:svetapetrasova@gmail.com).

**Галкіна Яна Романівна (Галкина Яна Романовна, Galkina Yana Romanivna)** – Національний технічний університет «Харківський політехнічний інститут», магістрант факультету соціально-гуманітарних технологій; м. Харків, Україна; e-mail: [galkina230797@gmail.com](mailto:galkina230797@gmail.com).

**Мануйлов Ілля Олександрович (Мануйлов Илья Александрович, Manuilov Illia Oleksandrovych)** – Національний технічний університет «Харківський політехнічний інститут», магістрант факультету соціально-гуманітарних технологій; м. Харків, Україна; e-mail: [banger@ukr.net](mailto:banger@ukr.net).

**Бородіна Олександра Русланівна (Бородина Александра Руслановна, Borodina Olexandra Ruslanivna)** – Національний технічний університет «Харківський політехнічний інститут», магістрант факультету соціально-гуманітарних технологій; м. Харків, Україна; e-mail: [formalino0@yandex.ua](mailto:formalino0@yandex.ua).

**Швець Софія Ігорівна (Швец София Игоревна, Shvets Sofia Igorivna)** – Національний технічний університет «Харківський політехнічний інститут», студент факультету соціально-гуманітарних технологій; м. Харків, Україна; e-mail: [sonyazno2018@ukr.net](mailto:sonyazno2018@ukr.net).