

Світлана Коломієць

кандидат педагогічних наук, професор
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Київ, Україна
ORCID ID 0000-0003-1070-6907
svitlanak29@gmail.com

Валерія Коломієць

студентка факультету лінгвістики
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Київ, Україна
ORCID ID 0000-0003-2863-6126
proworker11@gmail.com

МЕТОДИ АНАЛІЗУ ПАТЕНТНОЇ ДОКУМЕНТАЦІЇ

Інтелектуальний пошук патентного тексту є важливим завданням, яке потребує знання предметної області. У сучасному світі, коли зростаючі темпи технологічних інновацій вже зараз призводять до фундаментальних змін, а саме: входження в четверту промислову революцію Industry (4.0), розвиток штучного інтелекту та віртуальної економіки, пошук інформації стає все більш важливим завданням, відтак, підходи до аналізу патентного тексту потребують комплексного вивчення. У цій статті розглядаються методи аналізу патентних документів, що включають ідентифікацію теми, сегментацію тексту, вибір ознак створення кластерів та виділення підсумків. Описані методи можуть бути використані для кращого розуміння зв'язків між різними технологіями та їх пріоритетами, оцінки конкуренції та можливостей їх використання на ринку. Розглянуто також основні виклики та проблеми, пов'язані з аналізом патентної документації, такі як обмеженість обсягу даних та неоднорідність їх структури. Висновки статті можуть бути корисними не тільки для лінгвістичного аналізу патентної документації, але й для дослідників, інженерів, менеджерів та інших спеціалістів, що працюють у сфері інновацій та розвитку технологій.

Ключові слова: резюмування, виділення фраз, аналіз спільних слів, кластеризація, відображення тем.

1. ВСТУП

Патент – це інтелектуальна власність, ефективне використання якого дозволяє патентовласнику підвищити конкурентоспроможність та збільшити свою присутність у певній галузі. Однак, патенти мають цінність не лише для їхнього власника.

Патентування дозволяє компаніям отримувати конкурентні переваги за рахунок захисту своїх інноваційних розробок, що означає набуття монопольного права на використання технологій та реалізації випущеної продукції на ринку.

Актуальність цієї статті полягає у тому, що вивчення методів аналізу патентної документації є значущим аспектом лінгвістичного дослідження у теоретичному відношенні, оскільки сприяє подальшому накопиченню знань у галузі патентної документації та лінгвістики тексту. Відтак патентні документи містять важливі результати досліджень, які є цінними для певної галузі, для ділової, юридичної та політичної спільнот, а технічні деталі та впровадження нових технологій, розкривають бізнес-тенденції, надихають на нові промислові рішення та сприяють формуванню ефективної інвестиційної політики (Артамонова, Горбань, Кулинич, 2015, с. 69). Різні аспекти патентної документації плідно досліджуються у вітчизняних та іноземних наукових розвідках. Зокрема, дослідженням жанрово-стильових особливостей патенту займалися Г. Добриніна (2005), С. Коломієць, М. Шевченко (2011), В. Андрощук (2011); визначенням специфіки перекладу патентів – М. Кросс (2010), Б. Гамбург (2010), С. Лі (2009); розробкою технології навчання перекладу патентів – С. Царьова (2006) тощо.

Перспективам ринку високотехнологічних проектів на основі аналізу патентної документації присвячені роботи Лівей Чжан І та Чжихуей Лю (2020). Аналітики звертають особливу увагу на патентну літературу, що містить велику кількість технічної та ринкової інформації, а саме технічні характеристики, право власності та їх комерційна вартість, аналізують патентні документи як сегменти на рівні слів. Спочатку вони їх розбивають на ряд слів, потім вилучають прикметники, прислівники, іменники та багатослівні слова (Zhang, Liu, 2020).

Аналіз окремих складових патенту також є об'єктом дослідження вітчизняних та іноземних науковців. Наприклад, розглядаються лексичні маркери офіційно-ділового стилю, архаїзми та канцеляризми у текстах патентної документації (Криворучко, 2013, с. 245); Паргаонкар (Pargaonkar, 2016) аналізує особливості поєднання спеціальної лексики в тексті патенту в рамках мови для спеціальних цілей та розглядає частотність вживання тих чи інших частин мови у тексті патенту.

Мета статті полягає у визначенні особливостей методів аналізу патентної документації.

2. МЕТОДИ ДОСЛІДЖЕННЯ

У цій статті були використані такі загальнонаукові методи дослідження як: індукція, дедукція, аналіз і синтез для уточнення теоретичних засад і формування висновків дослідження.

3. РЕЗУЛЬТАТИ Й ОБГОВОРЕННЯ

Останніми роками вибухнула патентна діяльність майже в усіх галузях техніки. Дослідники припускають, що 90-95% усіх винаходів світу можна знайти в патентних документах (Roudsari, Afshar, Lee, 2021, с. 151). Також, існує така думка, що патенти є ключем до комерційного використання винаходів (Добриніна, 2005, с. 112). Інновації вважаються рушійною силою конкуренції, економічного зростання та створення робочих місць (Crevier, 1993, с. 231). Патенти відіграють центральну роль у реєстрації інновацій і в розробці бізнес-стратегії. Сьогодні патенти стають все більш значущими для успіху ІТ бізнесу.

Патентна документація має певні переваги у порівнянні з іншими видами науково-технічних публікацій. Насамперед, вона достовірна, тобто патентна документація відображає реальні технічні рішення і не містить неперевіраних та рекламних відомостей, позбавлена нальоту клановості та політичної кон'юнктури. Новизна (як правило, світова), промислова застосовність та винахідницький рівень цих рішень підтверджуються державною експертизою (Crevier, 1993, с. 216). Крім того, у низці країн, до публікації відомостей про видачу патенту публікують і відомості про заявку на винахід (Agrawal, Imielinski, Swami, 1993). Слід також зазначити, що після публікації таких відомостей, будь-яка особа має право ознайомитися з матеріалами заявки, що публікуються і вжити відповідних дій.

Патентна документація є найбільш систематизованим і повним зібранням усіх науково-технічних досягнень людства за останні 150-200 років (Chang, Wu, Leu, 2010). Фонд патентної документації на відміну від інших видів науково-технічної літератури (журналів, збірників наукових праць, книг, монографій, дисертацій, промислових каталогів тощо) є найбільшим та систематизованим зібранням відомостей про всі охоронні винахідницькі та патентні правові технічні рішення, створені людством (Crevier, 1993, с. 221). Наскрізна нумерація патентів, яка застосовується в більшості країн, дозволяє легко упорядкувати величезні масиви документів у хронологічному порядку, а індексація відповідно до Міжнародної патентної класифікації винаходів, дозволяє полегшити пошук документів.

Патентна інформація за своїм характером є точною та оперативною (Артамонова, Горбань, Кулинич, 2015, с. 75). Відповідно до патентних законів більшості країн світу, передчасне (до подання заявки) розголошення винаходу перешкоджає видачі патенту. Тому заявник прагне зберегти інформацію про винахід у секреті якомога довше. Зазвичай, перша публікація відомостей про патент на 3–5 років випереджає інші джерела інформації (Calero, Buter, Valdés, Noyons, 2006,

с. 347). Ще більше часу минає від моменту публікації опису винаходу до появи відповідних даних у монографіях та книгах. Таким чином, патентна документація багато в чому випереджає інші види публікацій, а у багатьох випадках виключає їх.

Ще однією особливістю патентної документації є її однорідність. Ця властивість дозволяє систематизувати за єдиною системою значні масиви документів, що належать до однієї технічної галузі (Crammer, Singer, 2002, с. 266).

Зауважимо, що патентна інформація практично не дублює відомості науково-технічного характеру, у тому числі опубліковані в непатентній літературі. Важливо також відзначити, що інформація, що міститься в описах винаходів, більш докладна, ніж в інших видах науково-технічних публікацій. Це пов'язано насамперед з вимогами патентного законодавства про повноту та надання детального опису винаходу при заповненні заявки. Крім об'єкту винаходу зазвичай описуються способи дії та найбільш ефективний варіант його практичного застосування (Zhang, Liu, 2020).

Схеми патентної класифікації використовуються для організації та індексування технічного вмісту патентних специфікацій на конкретну тему. До публікацій, патентним документам надається один або кілька класифікаційних кодів на основі текстового вмісту для тематичного аналізу та пошуку (Calero, Buter, Valdés, Noyons, 2006, с. 369)

Слід зазначити, що разом з тим, патентний документ містить десятки елементів для аналізу. Деякі з них структуровані, це означає, що вони однакові за семантикою та форматом, таким як номер патенту, заявка, дата або правонаступники; деякі з них є неструктурованими, тобто це вільні тексти різні за довжиною та змістом, наприклад назва, формула винаходу, реферат або опис винаходу.

Важливим структурним елементом патенту є опис винаходу, який можна сегментувати за областю винаходу, історією резюме та детальним описом, хоча деякі патенти можуть і не мати всіх цих сегментів (Добриніна, 2005, с. 125) Ці структуровані дані можна аналізувати за допомогою *методів інтелектуального аналізу* даних або добре налагодженої бази даних або за допомогою *засобів керування*, таких як модулі OLAP (онлайн аналітичної обробки).

Ще одним методом аналізу патентної документації є *узагальнення сегментів (сурогат документа)*, який дає стисле уявлення про документ, полегшує повторне використання проаналізованих патентів та прискорює подальшу автоматизовану обробку через меншу кількість текстових даних завдяки вилученню менш релевантних фрагментів. Не зважаючи на певні застереження щодо автоматизованої обробки текстової інформації, прості методи, що засновані на ранжуванні речень і відборі, часто дають достатню продуктивність для деяких завдань інтелектуального аналізу тексту (Zhang, Liu, 2020). Так наприклад, експерименти з патентної класифікації з використанням Naive Bayes, KNN і SVM як класифікаторів Fall показали, що навіть використовуючи лише перші 300 слів із розділів анотації, завдяки формулам винаходу та опису, можна значно підвищити ефективність пошуку інформації ніж із використання повних текстів (Crevier, 1993, с. 231).

Наступним етапом аналізу патентної документації є *виділення термінів з набору сурогатів документів*. Отримані терміни далі фільтруються за списком стоп-слів і за критеріями частоти їх використання. Мета полягає в тому, щоб отримати високоякісні умови для індексації та подальшого аналізу. Тим не менш, ідеальне індексування включає визначення основи терміну (морфологічний аналіз) і *кластеризацію термінів*, щоб терміни в різних формах, що відповідають подібним поняттям, асоціювалися разом в одному наборі (Roudsari, Afshar, Lee, 2021, с. 154). При цьому можна застосувати різні алгоритми кластеризації для визначення основних концепцій. Ці концепції можуть бути згруповані в теми, які, у свою чергу, можуть бути згруповані в категорії або домени (Cho, Kim, Lee, 2010). Тут відмінність між поняттями, темами та категоріями є не суттєвою.

Інтелектуальний аналіз тексту є досить новим методом, що був запропонований для виявлення знань із колекцій неструктурованого тексту. Зазначений метод часто розглядається як процес пошуку неявних, раніше невідомих та потенційно корисних закономірностей у великому тексті набору даних (Calero, Buter, Valdés, Noyons, 2006). Звичайна практика – це

додавати мітки до слів у документі, тоді документ у текстовому форматі може показувати ключові слова та слова-підказки, які витягуються за допомогою алгоритму аналізу тексту.

Зауважимо, інтелектуальний аналіз тексту широко використовується і для дослідження складних взаємозв'язків між патентними документами та для отримання цінної інформації з великих обсягів неструктурованого тексту.

Зазначається, що аналіз складається з кількох етапів. Першим етапом є збір текстів і попередня обробка. Вибирається потрібна область патенту і відповідні патентні документи зібрані в електронному текстовому форматі. На другому етапі, сирі патентні документи перетворюються на структуровані дані. Оскільки оригінали документів виражені у форматі природної мови, вони повинні бути перетворені в структуровані дані для аналізу та подальшого використання (Crevier, 1993, с. 236).

Стосовно патентного аналізу, текст-майнінг використовується як інструмент обробки даних і вилучення ключових слів, що уможливило вимірювання подібності між патентами (Crevier, 1993, с. 243).

На сьогоднішній день, науковцями запроваджено різноманітні методи представлення інформації про бізнес, технології або ринок колекції документів, включаючи експертний аналіз патенту або статистики підрахунку документів, аналіз класу патентів, бібліотеку графічних методів, такі як аналіз спільного входження, включаючи аналіз спільних слів або співцитат, планування технологій, та методи штучного інтелекту, такі як генетичні алгоритми, байєсівські методи навчання, моделі Маркова та подібне.

Базуючись на методиці аналізу патентних документів Лівей Чжан І та Чжихуей Лю (Zhang, Liu, 2020), яка включає 15 кроків ми виокремили такі основні етапи аналітичного процесу, що включають:

Підготовчий етап:

- 1) вибір та підготовка набору документів для аналізу;
- 2) підготовка до електронного аналізу документів, включення вибраних розділів документа в базовий сумісний формат.

Виконавчий етап:

- 1) формування списку ключових слів аналізу з одного або кількох розділів документа у початковий список слів;
- 2) сортування отриманого списку слів за частотою після видалення дублікатів або несуттєвих слів та стандартизації словоформ;
- 3) формування кореляційної матриці першого слова з використанням початкового списку слів.

Завершальний етап:

- 1) підрахунок частоти, з якою знайдена пара слів у збірнику вибраних документів;
- 2) формування першої колекції технологічних тем за вибраною парою слів;
- 3) формування додаткової матриці кореляції слів із слова у збірнику першої технології теми;
- 4) формування додаткового технологічного збірника тем асоціювання пар слів з однієї або кількох перших технологічних тем (Zhang, Liu, 2020).

Для встановлення значущості патенту, його креативності та інноваційності застосовують *аналіз патентного цитування*, який виявляє зв'язки між компанією-власником та іншими гравцями, що цитують патенти. Доцільно розглянути аналіз спільного цитування, який визначає частоту появи двох посилань разом у посиланнях на статтю наукового журналу (Crevier, 1993, с. 233). Так у своїх працях Краммер К. описує спільне цитування як метод аналізу, в якому є гіперпосилання на веб-сторінки, що розглядаються як посилання для створення індексу веб-сторінки (Crammer, Singer, 2002, с. 268).

Важливим в аналізі патентів є *аналіз основних тем та спільних слів*. Головне завдання аналізу спільних слів полягає в тому, що використовувані пари спільних слів, що вказують на основні теми, які є у наборі документів (Crammer, Singer, 2002, с. 270). Аналіз спільних слів

можна здійснити за допомогою простої база даних і карти зв'язків, що можуть виявити загальну структуру досліджуваної області.

Для ідентифікації зображень або військових цілей використовувалися *методи штучного інтелекту*. Методи штучного інтелекту, зокрема моделі Маркова, використовувалися в записі голосу, розпізнавання почерку та ідентифікації генетичної послідовності (Crammer, Singer, 2002, с. 273).

4. ВИСНОВКИ І НАПРЯМИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У ході дослідження було з'ясовано основні характеристики патенту та роль патентної документації як фіксатора технологічних модернізацій та інновацій, що відіграє значущу роль в пошуку та обробці великих обсягів структурованих і неструктурованих даних.

Аналіз вітчизняної та зарубіжної наукової літератури дозволив виокремити такі методи інтелектуального аналізу, як резюмування, узагальнення сегментів (сурогат документа), кластеризацію, відображення тем.

Бібліометричний метод – це метод інтелектуального аналізу, який базується на цитатах, посиланнях, словах чи групах слів, що знайдені в документах і включає публікацію підрахунків патентів, аналіз цитувань та слів і співцитат.

Базуючись на зарубіжній методиці аналізу патентних документів нами було виокремлено 3 етапи аналізу патентної документації: підготовчий (відбір матеріалу для аналізу), виконавчий (формування списку ключових слів та кореляційної матриці першого слова) та завершальний (підрахунок частоти, з якою знайдена пара слів). Перспективу перекладацького аспекту у подальшому дослідженні методів патентного аналізу ми вбачаємо у аналізі різних форм представлення структури знань, зокрема, тематичної карти.

СПИСОК ЛІТЕРАТУРИ

- Андрощук В.К., Федулова Л.І., Хаустов Г.О. *Інтелектуальна власність у національній інноваційній системі*. Ін-т економіки та прогнозування НАН України. 2011. № 10. С. 201—216.
- Артамонова Н.О., Горбань А.Е., Кулинич Г.В. *Наукометричний аналіз засобів наукових комунікацій в онкології*. Український радіологічний журнал. 2015. Т. 15, № 1. С. 68—72.
- Добриніна Г.П. *Патентна інформація та документація*. Патентні дослідження: Конспект лекцій. - Ко Ін-т. інтел. власн. і права", 2005. -123 с
- Інформаційні і патентні дослідження на етапах планування і виконання науководослідних робіт з проблем медицини: Метод. рекомендації / Український центр наукової медичної інформації та патентноліцензійної роботи. — Київ; 1999. — 23 с.
- Патентні дослідження: Метод. рекомендації / Державне патентне відомство України. — Київ, 1999. — 262 с.
- Дерді Е. Т. *Своєрідність мови патентів як проблема перекладу*. Наукові записки Національного університету «Острозька академія». Серія «Філологічна». 2010. Вип. 16. С. 100–104.
- Коломієць С. С., Шевченко М. В. *Жанрово-стильова домінанта у відтворенні англомовного опису винаходу у патентній літературі мовою перекладу*. Національний університет «Острозька академія». Наукові записки. Серія «Філологічна» Випуск 20. Острого, 2011. – С. 281-290. URL http://www.irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe
- Криворучко І. М., Носенко Ю. М., Вербова О. В., Сінельник Л. М. *Патентно-ліцензійна діяльність в контексті інноваційно-інвестиційного розвитку НААН*. Вісник Центру наукового забезпечення АПВ Харківської області. 2013. Вип. 14. С. 240–247.
- Царьова С. О. *Переклад патентної документації США*. Харків: ХПІ, 2006. 32 с.
- H. Roudsari, J. Afshar, S. Lee and W. Lee, "Comparison and Analysis of Embedding Methods for Patent Documents (2021) IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju Island, Korea (South), 2021, pp. 152-155, doi: 10.1109/BigComp51126.2021.00037.
- Agrawal, R., Imielinski, T., & Swami, A. (1993), *Mining association rules between sets of items in large databases*. In Proceedings of the 1993 ACM-SIGMOD International Conference on Management of Data, New York, USA.
- Borgatti, S. P., Everett, M. G., & Freeman, L. C. (1999) UCINET 6.0 Version 1.00, Harvard: Analytic Technologies Publishers.
- Bottou, L., & Vapnik, V. (1992) Local learning algorithms. *Neural Computation*, 4, pp 888-900.
- Calero, C., Buter, R., Valdés, C. C., & Noyons, E. (2006) *How to identify research groups using publication analysis: an example in the field of nanotechnology*. *Scientometrics*, 66, pp 365-376.
- Chang, P. L., Wu, C. C., & Leu, H. J. (2010) *Using patent analyses to monitor the technological trends in an emerging field of technology: a case of carbon nanotube field emission display*. *Scientometrics*, 82, pp 5-19.
- Cho, Y. S., Kim, J. Y., & Lee, H. S. (2010) Efficient Viterbi scoring architecture for HMM-based speech recognition systems. *Electronic Letters*, 28, pp 2338-2340.

- Crammer, K., & Singer, Y. (2002) On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning*, 2, pp 265-292.
- Crevier, D. (1993) *AI: The Tumultuous History of the Search for Artificial Intelligence*, New York: Basic Books.
- Cross, R., Borgatti, S. P., & Parker, A. (2001) Beyond answers: dimensions of the advice network. *Social Networks*, 23, pp 215-235.
- Cross M. *Literal Translation of Patents*. 2010. № 1 (19). 267 p.
- Hamburg C. Bruce. *Translating Patents: Issues in Prosecution*, 2010. 320 p.
- Lee S, Yoon B, Lee C, Park J. Business planning based on technological capabilities: Patent analysis for technology-driven roadmapping. *Technological Forecasting and Social Change*. 2009; 76(6): 769– 786. <https://doi.org/10.1016/j.techfore.2009.01.003>
- Pargaonkar Y. R. *Leveraging patent landscape analysis and IP competitive intelligence for competitive advantage*. *World Patent Information*. 2016. Vol. 45. P. 10–20.
- Yanghong, L., Runhua, T., (2007) in *IFIP International Federation for Information Processing, Volume 250, Trends in Computer Aided Innovation*, ed. Leon-Rovira, N., (Boston: Springer), pp. 89-96.
- Zhang L, Liu Z (2020) Research on technology prospect risk of high-tech projects based on patent analysis. *PLoS ONE* 15(10): e0240050. <https://doi.org/10.1371/journal.pone.0240050>

REFERENCES

- Androschuk V.K., Fedulova L.I., Haustov G.O. *Intellectual property in the national innovation system*. Institute of Economics and Forecasting of the National Academy of Sciences of Ukraine. 2011. No. 10. P. 201—216.
- Artamonova N.O., Horban A.E., Kulnych G.V. *Scientometric analysis of means of scientific communication in oncology*. *Ukrainian Radiological Journal*. 2015. Vol. 15, No. 1. P. 68—72
- Dobrynina G.P. *Patent information and documentation*. Patent research: Synopsis of lectures. - Co. Inst. intel own and rights", 2005. -123 p
- Information and patent research at the stages of planning and implementation of research works on medical problems: Method. recommendations / Ukrainian center of scientific medical information and patent licensing work. - Kyiv; 1999. — 23 p.
- Patent research: Method. recommendations / State Patent Office of Ukraine. — Kyiv, 1999. — 262 p.
- Durdy E. T. *The peculiarity of the language of patents as a problem of translation*. Scientific notes of the National University "Ostroh Academy". "Philological" series. 2010. Issue 16. P. 100–104.
- Kolomiets S. S., Shevchenko M. V. Genre-style dominant in the reproduction of the English-language description of the invention in the patent literature in the translated language. URL http://www.irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe
- Kryvoruchko I. M., Nosenko Yu. M., Verbova O. V., Sinelnik L. M. Patent and licensing activity in the context of innovation and investment development of NAAS. *Bulletin of the Center for Scientific Support of APV of Kharkiv Region*. 2013. Issue 14. P. 240–247.
- Tsareva S. O. Translation of US patent documentation. Kharkiv: KhPI, 2006. 32 p.
- H. Roudsari, J. Afshar, S. Lee and W. Lee, "Comparison and Analysis of Embedding Methods for Patent Documents (2021) *IEEE International Conference on Big Data and Smart Computing (BigComp)*, Jeju Island, Korea (South), 2021, pp. 152-155, doi: 10.1109/BigComp51126.2021.00037.
- Agrawal, R., Imielinski, T., & Swami, A. (1993), *Mining association rules between sets of items in large databases*. In *Proceedings of the 1993 ACM-SIGMOD International Conference on Management of Data*, New York, USA.
- Borgatti, S. P., Everett, M. G., & Freeman, L. C. (1999) *UCINET 6.0 Version 1.00*, Harvard: Analytic Technologies Publishers.
- Bottou, L., & Vapnik, V. (1992) Local learning algorithms. *Neural Computation*, 4, pp 888-900.
- Calero, C., Buter, R., Valdés, C. C., & Noyons, E. (2006) *How to identify research groups using publication analysis: an example in the field of nanotechnology*. *Scientometrics*, 66, pp 365-376.
- Chang, P. L., Wu, C. C., & Leu, H. J. (2010) *Using patent analyses to monitor the technological trends in an emerging field of technology: a case of carbon nanotube field emission display*. *Scientometrics*, 82, pp 5-19.
- Cho, Y. S., Kim, J. Y., & Lee, H. S. (2010) Efficient Viterbi scoring architecture for HMM-based speech recognition systems. *Electronic Letters*, 28, pp 2338-2340.
- Crammer, K., & Singer, Y. (2002) On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning*, 2, pp 265-292.
- Crevier, D. (1993) *AI: The Tumultuous History of the Search for Artificial Intelligence*, New York: Basic Books.
- Cross, R., Borgatti, S. P., & Parker, A. (2001) Beyond answers: dimensions of the advice network. *Social Networks*, 23, pp 215-235.
- Cross M. *Literal Translation of Patents*. 2010. № 1 (19). 267 p.
- Hamburg C. Bruce. *Translating Patents: Issues in Prosecution*, 2010. 320 p.
- Lee S, Yoon B, Lee C, Park J. Business planning based on technological capabilities: Patent analysis for technology-driven roadmapping. *Technological Forecasting and Social Change*. 2009; 76(6): 769– 786. <https://doi.org/10.1016/j.techfore.2009.01.003>
- Pargaonkar Y. R. *Leveraging patent landscape analysis and IP competitive intelligence for competitive advantage*. *World Patent Information*. 2016. Vol. 45. P. 10–20.

- Yanghong, L., Runhua, T., (2007) in IFIP International Federation for Information Processing, Volume 250, Trends in Computer Aided Innovation, ed. Leon-Rovira, N., (Boston: Springer), pp. 89-96.
- Zhang L, Liu Z (2020) Research on technology prospect risk of high-tech projects based on patent analysis. PLoS ONE 15(10): e0240050. <https://doi.org/10.1371/journal.pone.0240050>

Svitlana Kolomiets, Valeriia Kolomiets. Methods of patent documentation analysis. Intellectual search of patent text is an important task that requires knowledge of the subject area. In today's world, when the growing pace of technological innovation is already leading to fundamental changes, with entering the fourth industrial revolution Industry (4.0), the development of artificial intelligence and the virtual economy, the analysis of patent documentation is becoming an increasingly important task, therefore, traditional approaches to text analysis are insufficient for comprehensive study of the patent text. This article discusses methods of patent document analysis, including topic identification, text segmentation, information display, feature selection, clustering, and summary selection, etc. The described methods can be used to better understand the relationships between different technologies and their priorities and to assess competition and opportunities for their use in the market. Also considered are the main challenges and problems associated with the analysis of patent documentation, such as the limited amount of data and the heterogeneity of their structure. The article's conclusions can be useful not only for the linguistic analysis of patent documentation, but also for researchers, engineers, managers and other specialists working in the field of innovation and technology development.

Keywords: summarizing, phrase selection, analysis of common words, clustering, display of topics.