

Клочко Андрій АндрійовичНародний депутат Верховної Ради України, orcid.org/0000-0002-1691-2333

Верховна Рада України, Київ

ВПРОВАДЖЕННЯ ТЕХНОЛОГІЙ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ТЕКСТОВИХ ДОКУМЕНТІВ У СФЕРУ ТЕХНІЧНОГО РЕГУЛЮВАННЯ В БУДІВНИЦТВІ

***Анотація.** Статтю присвячено впровадженню технологій інтелектуального аналізу текстових документів у сферу технічного регулювання в будівництві України. Основну увагу в роботі спрямовано на вирішення питань автоматичного накопичення та інтелектуального аналізу нормативних документів будівельної галузі, які набувають надзвичайної актуальності у зв'язку з цифровізацією усіх сфер господарства. Висвітлено нагальні проблеми системи технічного регулювання в будівництві. Показано, що ці проблеми висують на перший план завдання підвищення швидкості та надійності опрацювання текстових документів в електронних інформаційних системах. Вирішення цього завдання передбачає розробку автоматичних систем, які здатні до інтелектуального пошуку документів в умовах невизначеності, що спричинена наявністю надлишкової текстової інформації. Проведено огляд інформаційно-пошукових систем, які використовуються для опрацювання текстових документів в електронних інформаційних ресурсах. Досліджено передумови впровадження технологій інтелектуального аналізу текстових документів у сферу технічного регулювання в будівництві України. Обґрунтовано своєчасність їх провадження. Надано основні поняття, що використовуються у процесі розроблення моделей і методів автоматичного вилучення значущої інформації з текстів. Досліджено процес інтелектуального аналізу даних та проаналізовано наявні моделі інтелектуального аналізу текстової інформації, які використовуються в різних інформаційно-пошукових системах. Запропоновано схему впровадження технологій інтелектуального аналізу текстових документів в Єдину державну електронну систему у сфері будівництва. При цьому передбачається розв'язання задачі кластеризації текстових документів покласти на штучні нейронні мережі. Розглядається можливість використання таких моделей, як глибоко структурована семантична модель і самоорганізуюча карта Кохонена. Вибір зазначених моделей ґрунтується на їх здатності до визначення міри близькості інформаційно-пошукових образів текстових документів. Практичне значення роботи вбачається в удосконаленні пошукових систем у сфері технічного регулювання в будівництві та можливості суттєво прискорити процес реструктуризації галузі будівництва і архітектури України.*

Ключові слова: автоматичний пошук; інтелектуальний аналіз; кластеризація; невизначеність; технічне регулювання

Вступ

Зростання попиту на системи автоматичного пошуку та опрацювання текстової інформації в різних сферах діяльності людини спричинило стрімкий розвиток інтелектуальних лінгвістичних систем і технологій інтелектуального аналізу текстових документів. Актуальними ці задачі є і для сфери технічного регулювання в галузі будівництва і архітектури України, оскільки інтеграція країни в європейський правовий простір спричинила значні зміни в організаційно-технічній системі «Технічне регулювання в будівництві», а велика частина інформації у цій сфері представлена текстами [1; 2].

Адаптація системи технічного регулювання в будівництві до вимог Європейського Союзу обумовлює створення додаткових документів для нормування порядку дій у разі появи невизначеності різного характеру. Внаслідок створення додаткових документів у процесі реструктуризації сфери технічного регулювання в будівництві з'являються неузгоджені та конфліктні документи (рис. 1).

Накопичення цих документів ускладнює процес технічного регулювання в цілому і може призвести до юридичних колізій чи колізій нормативних актів. Вчасне виявлення надлишкової інформації значно прискорює процес узгодження невизначеності та запобігає появі колізій, що можуть з'являтися та з'являються внаслідок реформування законодавства в частині, що стосується технічного регулювання в будівництві.

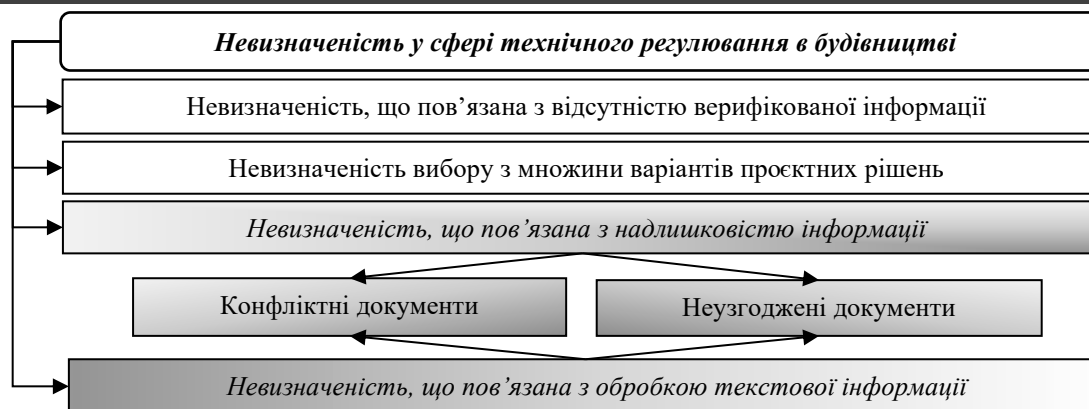


Рисунок 1 – Класифікація невизначеності за характером інформації

Одним із основних шляхів підвищення швидкості та надійності опрацювання текстових документів в електронних інформаційних ресурсах є розроблення автоматичних систем, які здатні до інтелектуального пошуку, а глобальна цифровізація усіх сфер господарства створює сприятливі умови для розв'язання цієї задачі з використанням інформаційно-пошукових систем (ІПС) [3 – 5].

Аналіз останніх досліджень і публікацій

Аналіз останніх досліджень і публікацій засвідчив, що ІПС розділяють на фактографічні та документальні залежно від характеру інформаційних ресурсів [6].

Фактографічні ІПС оперують фактами, що представлені у вигляді сутностей та властивостей реального світу, що допомагає знаходити:

- сутності із заданими властивостями;
- властивості заданих сутностей.

Документальні ІПС призначені для зберігання і пошуку документів, які містяться в електронних інформаційних системах.

Згідно з [7]:

– документом називається одиниця інформації, що використовується для реєстрації, передачі та збереження інформації про будь-який предмет, в просторі інформаційних ресурсів;

– критерієм пошуку називають необхідний компонент змісту, що являє собою опис заданих користувачем властивостей документа, за якими здійснюється пошук даних у базі даних.

У результаті опрацювання запиту користувача система видає документ чи множину документів, що задовольняють цей запит (рис. 2).

Проблеми виникають, якщо видані на запит користувача документи з вказівками до визначення одного параметра регулювання є неузгодженими чи конфліктними. Такі результати пошуку на запит користувача суттєво ускладнюють розроблення проєктної документації та проведення експертних оцінок унікальних об'єктів [4; 8].

Отже, швидке зростання обсягу інформації є однією з основних причин ускладнення пошуку потрібних документів та їх організації у структуровані за змістом сховища. Окрім того, аналіз роботи ІПС, що функціонують в різних електронних інформаційних системах, виявив проблеми, які спричинені збільшенням кількості інформації.

Серед цих проблем виокремлюються такі [9; 10]:

- неструктурована інформація становить значну частину електронних текстових документів;
- більшість технологій роботи з текстовими документами не забезпечує передачу їх змісту внаслідок відсутності семантичного індексування;
- ефективне розв'язання задачі автоматичного пошуку текстових документів потребує необхідності з кожним документом пов'язати знання, що допоможуть інтерпретувати та опрацювати дані, які зберігаються в цьому документі;
- вирішення завдання автоматичного пошуку текстових документів у нормативній базі будівництва потребує впровадження технологій інтелектуального аналізу текстових документів.

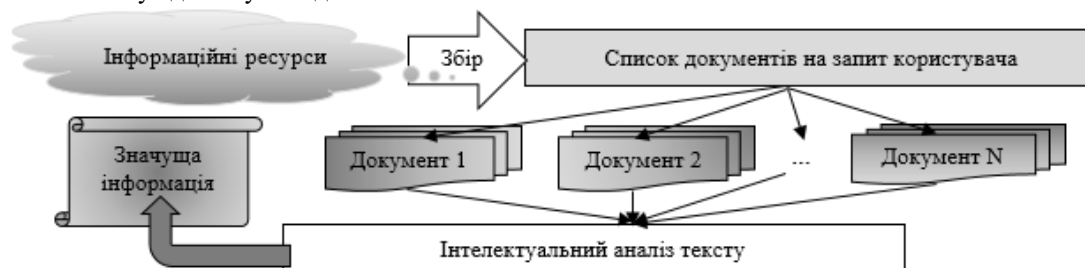


Рисунок 2 – Схема пошуку значущої інформації в електронних інформаційних ресурсах

Мета статті

Метою цієї роботи є впровадження технології інтелектуального аналізу текстових документів у сферу технічного регулювання в будівництві України.

Для досягнення цієї мети сформовані такі задачі:

- дослідити передумови впровадження технологій інтелектуального аналізу текстових документів у сферу технічного регулювання в будівництві;
- запропонувати схему запровадження технології інтелектуального аналізу текстових документів в Єдину державну електронну систему в будівництві (ЄДЕСБ) України.

Виклад основного матеріалу

ЄДЕСБ являє собою загальнонаціональну інформаційну систему, що призначена впорядкувати процес будівництва в Україні та зробити його прозорим і вільним від корупції через максимальну публічність інформації. Ця система запроваджується для реалізації всіх етапів будівництва, від отримання містобудівних умов і обмежень до введення об'єкта будівництва в експлуатацію. Функціонування ЄДЕСБ регулюється [10 – 12].

Користувачами ЄДЕСБ є [10]:

- органи державного архітектурно-будівельного контролю;
- місцеві органи містобудування та архітектури;
- саморегулюючі організації у сфері архітектурної діяльності та атестовані ними особи;
- проектні й експертні організації;
- органи з присвоєння адрес.

Це означає, що інформація та документація, яка необхідна для ведення будівництва, створюватиметься відповідними суб'єктами в ЄДЕСБ, що допоможе централізувати та стандартизувати дані з багатьох розрізнених джерел інформації. Проте швидкість і надійність роботи ЄДЕСБ в умовах реструктуризації галузі залежить від вчасного виявлення, узгодження конфліктних

і неузгоджених документів у таких інформаційних ресурсах, як [10]:

- електронні реєстри дозвільних документів;
- електронний реєстр містобудівних умов і обмежень;
- електронний реєстр атестованих осіб.

Глобальна цифровізація усіх сфер господарства створює сприятливі передумови для розв'язання задачі автоматичного виявлення конфліктних і неузгоджених документів нормативної бази електронних документів у будівництві і промисловості будівельних матеріалів, але потребує впровадження технологій і систем інтелектуального аналізу текстів у сферу технічного регулювання галузі [13].

Задача інтелектуального аналізу тексту полягає в автоматичному вилученні значущої інформації з недостатньо структурованих текстових даних [14; 15].

Згідно з [16 – 18] інтелектуальний аналіз тексту здійснюється в такій послідовності (рис. 3):

- збирання текстів із різних інформаційних ресурсів;
- фільтрація текстів;
- вилучення значущої інформації.

Збирання текстів із різних інформаційних ресурсів включає попереднє опрацювання текстів.

Попереднє опрацювання текстів складається з таких кроків: токенізація, видалення «стоп-слів» і визначення походження слів [16 – 18].

У процесі попереднього опрацювання виконується:

- поділ тексту на окремі слова з видаленням пустих місць і знаків пунктуації;
- видалення «стоп-слів» (сполучників, артиклів, інших незначущих слів);
- ідентифікація коренів слів, що полягає у зведенні однокорених слів до основної форми.

Для визначення походження слів найчастіше використовують алгоритм швейцара [19; 20].

Під час фільтрації тексту виконується вибірка потрібних та усунення непотрібних властивостей тексту, а також трансформація вихідної форми текстів у проміжні форми документів.

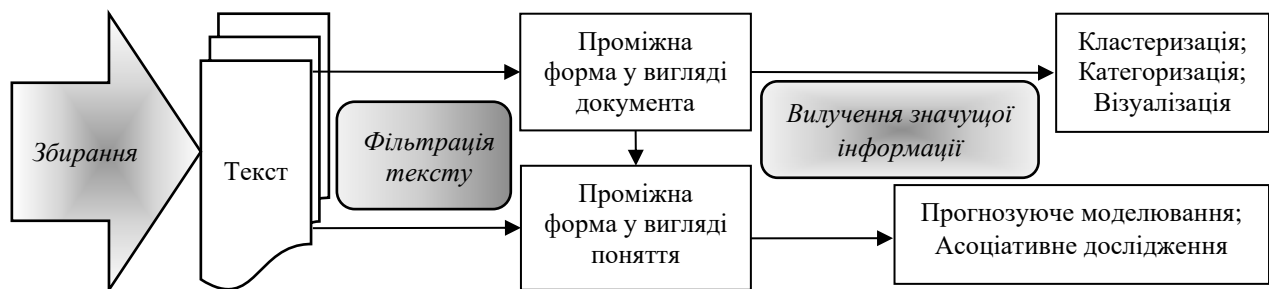


Рисунок 3 – Схема процесу інтелектуального аналізу даних

Проміжна форма може являти собою документ:

- в якому кожна сутність є іншим документом або поняттям, де кожна сутність являє собою об'єкт або набір даних з певної предметної області;
- сутність якого частково структурована (наприклад, концептуальне представлення графа);
- сутність якого структурована (наприклад, реляційне представлення даних).

Проміжна форма у вигляді документа може бути перетворена у проміжну форму у вигляді поняття шляхом вилучення релевантної інформації, що стосується об'єктів з визначеної галузі або сфери інтересів залежно від поставлених вимог. Це надає можливість виконати опрацювання знань для організації текстів за їх змістом з метою візуалізації і навігації. Аналіз проміжної форми у вигляді поняття надає можливість вилучити значущу інформацію, яка потрібна користувачеві при прийнятті рішення.

Підходи, що використовуються для створення проміжних форм, описані в [20].

Значуща інформація вилучається спеціальним програмним забезпеченням шляхом виявлення певних шаблонів (патернів) на базі статистичного опрацювання тексту або дослідження асоціацій [21].

Методи вилучення значущої інформації залежать від задач, які ставляться при інтелектуальному аналізі текстових даних.

Серед основних методів вилучення інформації:

- виділення інформації з тексту [22];
- узагальнення текстових даних [17];
- класифікація тексту [18];
- кластеризація [19];
- візуалізації документів [20].

Виділення інформації з тексту полягає у визначенні структурованих даних для отримання набору сутностей із текстових повідомлень цифрових онлайн бібліотек чи електронної пошти, виділення таблиць з тексту тощо.

Узагальнення текстових даних використовується для стислого представлення великої кількості даних. При цьому здійснюється пошук найбільш релевантних до теми речень і, як наслідок, зменшується об'єм текстових даних без втрати інформації.

Класифікацією називається процес знаходження певних ознак класу серед частково структурованих даних, під час якого створюється класифікатор на відповідних даних з визначенням міток класів.

У багатьох ІПС використовуються прості моделі пошуку, що базуються на класифікаторі. Документи і запит користувача в цих моделях представляються ідентифікаторами класів в ієрархічній структурі класифікатора. Критерієм релевантності виданого на запит документа є міра відповідності класу документа з класом запиту [6].

У складних моделях пошуку, що ґрунтуються на класифікаторі, в запиті користувача допускається вказівка декількох класів класифікатора. При цьому релевантними вважаються документи, що належать будь-якому з указаних класів або його підкласу. Така модель пошуку близька до булевої моделі. Для побудови булевих моделей використовують нечіткі імплікації. Терми булевого виразу можуть бути різними в різних варіаціях моделей пошуку. Якщо в такій моделі використовують DublinCore, то термом булевого виразу може бути задане в запиті значення міри приналежності елемента метаданих до документа. Критерієм релевантності в булевих моделях може бути істинність булевого виразу, що задається в запиті [6].

У моделях контекстного пошуку використовують представлення документа у вигляді сукупності слів і словосполучень, які зустрічаються в тексті документа, без урахування службових слів. Але застосування цих моделей потребує високої продуктивності процесорів і великого об'єму зовнішньої пам'яті прямого доступу комп'ютера, тому в досконалих системах текстового пошуку використовують векторні моделі документів і запитів. Це забезпечує кращу якість роботи ІПС і надає змогу створювати візуалізації документів [3].

Кластеризацією називається метод пошуку закономірностей, які потрібні для розбиття сукупності об'єктів на кластери, або пошуку наявних структур у даних. Кластерний аналіз являє собою сукупність методів, підходів і процедур, які розробляються для вирішення завдання формування однорідних класів (кластерів) у будь-якій предметній області [22].

Предметна область, яка розглядається в цій статті, являє собою великий динамічний масив електронної бази нормативних документів у будівництві та промисловості будівельних матеріалів (рис. 4), що суттєво уповільнює його кластеризацію експертами. Крім того, при кластеризації документів експертами є ризики, що пов'язані з людським фактором, що теж потребує розроблення технологій інтелектуального опрацювання текстів у документах сфери технічного регулювання в будівництві.

Задача автоматичної кластеризації текстових документів у загальному вигляді зводиться до оптимального (за наперед визначеним критерієм) розбиття колекції текстових документів. При цьому передбачається, що існує множина тематичних груп (кластерів), на які можна розбити цю колекцію [9].

Колекцією текстових документів називається множина текстів на природній мові. При цьому вихідними даними задачі кластеризації є не самі тексти, а їх інформаційно-пошукові образи.



Рисунок 4 – Схема запровадження технології інтелектуального аналізу текстових документів в ЄДЕСБ

Інформаційно-пошуковий образ документа являє собою багатовимірний вектор в евклідовому просторі ознак документа, що характеризує його змістовне навантаження.

Ознаки документів автоматично вилучаються з текстів згідно з вибраними моделями і методами представлення тексту, після чого об'єднуються в загальну множину.

Процес кластеризації ґрунтується на визначенні міри близькості документів. При цьому вважається, що геометрична близькість їх інформаційно-пошукових образів у просторі ознак документів всієї колекції відображає семантичну близькість оригіналів.

Для оцінювання семантичної близькості документів використовується [20]:

- значення косинуса кута;
- відстані між інформаційно-пошуковими образами документів.

Підхід, що описаний в [3; 6], ґрунтується на порівнянні семантичної близькості текстових документів за косинусом кута між векторними моделями. В цих роботах для визначення міри відповідності нормативних документів у сфері технічного регулювання в будівництві запропоновано застосовувати глибоко структуровані

семантичні моделі [23]. Проте використання цих штучних неймереж потребує їхнього навчання, що своєю чергою передбачає існування надійної вибірки для навчання.

Підхід, що використовується в [22], базується на визначенні відстані між векторами, що являють собою інформаційно-пошукові образи документів, може бути реалізованим з використанням самоорганізуючих карт Кохонена [24].

Порівняльний аналіз цих підходів є предметом подальших досліджень.

Висновки

1. Аналіз передумов впровадження технологій інтелектуального аналізу текстових документів у сферу технічного регулювання в будівництві доцільність і можливість впровадження цих технологій у документообіг ЄДЕСБ.

2. Запропоновано схему впровадження технології інтелектуального аналізу текстових документів в Єдину державну електронну систему у сфері будівництва, що передбачає застосування штучних нейронних мереж при вирішенні завдання інтелектуального аналізу нормативних документів у сфері технічного регулювання в будівництві.

Список літератури

1. Terenchuk S., Belous S. (2019). Study of Ambiguity in Regulatory Framework in Construction. *Scientific Journal of ScienceRise*, 7 (60), 35–39.
2. Seryh, A. R. (2010). Technical Regulation in Construction. Analytical review of world experience. SNIP, available at: http://iceg.com.ua/wpcontent/uploads/2016/04/techno_reg_stroi_world.pdf.
3. Chernyshev D., Klochko A., Terenchuk S., Ternavska V., & Zapryvoda V. (2020). Semantic Analysis Models and Methods of the Text Information in the Building Normative Base. *International Journal of Innovative Technology and Exploring Engineering (IJTEE)*, Vol. 9, Issue-6, 1873–1879.
4. Ісасенко Д. В., Теренчук С. А. Моделювання інтелектуальної системи підтримки прийняття рішень з технічного регулювання в будівництві. *Вісник Одеської державної академії будівництва та архітектури*. 2018, № 72. С. 18–25.
5. Ісасенко Д. В., Клочко А. А., Яценко О. Ф. Аналіз проблеми цифровізації сфери технічного регулювання в будівництві *Управління розвитком складних систем*. 2020, №43. С. 91–96.
6. Klochko A., Terenchuk S., Chernyshev D., Zapryvoda V. (2020). Using Deep Structured Semantic Model to Analysis Text Documents in the Building Normative Base. 2020 IEEE International Conference on Problems of Infocommunications Science and Technology, PIC S&T 2020, 330–334.
7. Rudyuk V. (2006). Criteria for classification of electronic business documents: current trends and future schemes. *Library Science. Documentation. Informology*, 3.
8. Ісасенко Д. В. Законодавче регулювання діяльності в будівельній галузі. Особливості світового досвіду та Європейського підходу для визначення пріоритетів при формуванні життєвого середовища. *Будівельне виробництво*. 2017, №6. С. 11–15.
9. Оксанич, І. Г., Піскунов Д. М., Черниш Д. П. Інтелектуальний аналіз масиву текстових документів на основі технології Text Mining. *Системи обробки інформації*. 2013, №2. С. 139–143.
10. Єдина державна електронна система в будівельній галузі. URL: <https://thedigital.gov.ua/news/upershe-vidkritodani-pro-deklarativni-ta-dozvilni-dokumenti-z-novogo-reestru-budivelnoidiyalnosti-ta-provedeno-doslidzhennya-pro-vpliv-vidkritih-danih-namistobudivnu-galuzь>.
11. Закон України «Про регулювання містобудівної діяльності». URL: <https://zakon.rada.gov.ua/laws/show/3038-17#Text>.
12. Постанова Кабінету Міністрів «Про реалізацію експериментального проекту щодо запровадження першої черги Єдиної державної електронної системи у сфері будівництва». URL: <https://zakon.rada.gov.ua/laws/show/559-2020-%D0%BF#Text>.
13. Організація та оцифрування інформації щодо будівель та споруд включно з будівельним інформаційним моделюванням (BIM). Управління інформацією з використанням будівельного інформаційного моделювання. Частина 1. Концепції та принципи. Частина 1. Поняття та принципи (ISO 19650-1: 2018 IDT). Наказ від 18.03.2020 73 Про прийняття національних стандартів 2020 р. URL: http://online.budstandart.com/ua/catalog/docpage.html?id_doc=89571.
14. Chang, Wui Lee, Tay, Kai Meng, Lim, Chee Peng (2017). A New Evolving Tree-Based Model with Local Re-learning for Document Clustering and Visualization. *Neural Processing Letter*, 46 (2), 379–409.
15. Paltoglou, Georgios; Thelwall, Mike. Twitter, MySpace, Digg (2012). Unsupervised Sentiment Analysis in Social Media. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3 (4). P. 66.
16. Atika Mustafa, Ali Akbar, and Ahmer Sultan (2009). Knowledge Discovery using Text Mining: A Programmable Implementation on Information Extraction and Categorization. *International Journal of Multimedia and Ubiquitous Engineering*, Vol. 4, No.2, 837–848.
17. Chauhan Shrihari R., Amish Desai, (2015). A Review on Knowledge Discovery Using Text Classification Techniques in Text Mining. *International Journal of Computer Applications* (0975-8887), Vol.111, No 6.
18. Zhao Y. (2013) Analysing Twitter Data with Text Mining and Social Network Analysis in Proceedings of the 11th Australasian Data Mining and Analytics Conference (AusDM 2013), p.23.
19. Renganathan, Vinaitheerthan, (2017). Text Mining in Biomedical Domain with Emphasis on Document Clustering. *Healthcare Informatics Research*, 23 (3), 141–146.
20. Karpov I. A., Antonenko S. V. (2020). Review of text mining methods. *Current issues of automation and information technology*, Volume 24, 40–46.
21. Буряков С., Клочко А., Федоров О., Теренчук С. Аналіз проблем автоматичного пошуку веб-даних в системі технічного регулювання в будівництві. Міжнародна науково-практична конференція молодих вчених «БУД-МАЙСТЕР-КЛАС-2020», 25-27 листопада 2020. КНУБА, С. 116–117.
22. Gubin M. V. (2004) Models and methods of representation of a text document in information retrieval systems // *Scientific and technical information*, Ser. 1, 12, 12–24.
23. Xie, Z., Zeng, Z., Zhou, G. et al. (2017). Topic enhanced deep structured semantic models for knowledge base question answering. *Sci. China Inf. Sci.* 60, 110103. <https://doi.org/10.1007/s11432-017-9136-x>.
24. Kohonen T., Ageeva V. N. Self-organizing cards. Per. from English M.: BINOM. Knowledge Laboratory, 2008. 655 p.

Стаття надійшла до редколегії 02.09.2021

Klochko Andriy

People`s Deputy of Verkhovna Rada of Ukraine, orcid.org/0000-0002-1691-2333

Verkhovna Rada of Ukraine, Kyiv

INTRODUCTION OF INTELLECTUAL ANALYSIS TECHNOLOGIES OF TEXT DOCUMENTS INTO FIELD OF TECHNICAL REGULATION IN CONSTRUCTION

Abstract. The article is devoted to the introduction of intellectual analyses technology of text documents into the field of technical regulation in Ukraine construction. The main attention in the paper is directed on the decision of questions of automatic collection and intellectual analysis of construction branch`s normative documents. These issues are becoming extremely important in connection with the digitalization of all sectors of the economy. Urgent problems of the technical regulation system in construction are highlighted. It is shown that these problems bring to the fore the task of increasing the speed and reliability of processing text documents in electronic information systems. The solution to this problem involves the development of automatic systems that are capable of intelligent document search in uncertainty conditions that caused by the presence of redundant textual information. The overview of information retrieval systems used to processing text documents in electronic information resources is conducted. Preconditions of introduction of intellectual analysis technologies of text documents into the technical regulation sphere in Ukraine construction are investigated. The timeliness of the technology implementation substantiated. The basic concepts used in the models and methods development of automatic extraction of meaningful information from texts are given. Process of data mining is studied and Models of textual information mining are analyzed that used in different information retrieval systems. Scheme of introduction of intellectual analysis technology of text documents into the Unified state electronic system in the field of construction is offered. Solution of clustering of text documents problem apply artificial neural networks is supposed. The possibility of using such models as Deep Structured Semantic Model and Self-Organizing Map is considered. The choice of these models is based on their ability to determine the proximity degree of information retrieval images of text documents. Work practical significance is seen in the improvement of search engines in the field of technical regulation in construction and the ability to significantly accelerate the process of restructuring into the field of technical regulation in Ukraine construction.

Keywords: automatic search, intelligent analysis, clustering, uncertainty, technical regulation

References

1. Terenchuk, S., Belous, S. (2019). Study of Ambiguity in Regulatory Framework in Construction. *Scientific Journal of ScienceRise*, 7 (60), 35–39.
1. Seryh, A. R. (2010). Technical Regulation in Construction. Analytical review of world experience. SNIP, available at: http://iceg.com.ua/wpcontent/uploads/2016/04/techno_reg_stroi_world.pdf.
2. Chernyshev, D., Klochko, A., Terenchuk, S., Ternavska, V. & Zapryvoda, V. (2020). Semantic Analysis Models and Methods of the Text Information in the Building Normative Base. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 9, 6, 1873–1879.
3. Isaenko, D. V., Terenchuk, S. A. (2018). Modeling of intelligent decision support system for technical regulation in construction. *Bulletin of the Odessa State Academy of Civil Engineering and Architecture*, 72, 18–25.
4. Isaenko, D., Klochko, A. & Yashenko, O. (2020). Digitalization problem analysis of the sphere of technical regulation in construction. *Management of Development of Complex Systems*, 43, 91–96.
5. Klochko, A., Terenchuk, S., Chernyshev, D., Zapryvoda, V. (2020). Using Deep Structured Semantic Model to Analysis Text Documents in the Building Normative Base. 2020 IEEE International Conference on Problems of Infocommunications Science and Technology, PIC S&T 2020, 330–334.
6. Rudyuk, V. (2006). Criteria for classification of electronic business documents: current trends and future schemes. *Library Science. Documentation. Informology*, 3.
7. Isaenko, D. (2017). Legislative regulation of activities in the construction industry. Features of the world experience and the European approach for definition of priorities at formation of an environment. *Construction production*, 63, 11–15.
8. Oksanich, I. G., Piskunov, D. M., Chernysh, D. P. (2013). Intelligent analysis of an array of text documents based on text mining technology. *Information Processing Systems*, 2 (109), 139–143.
9. The only state electronic system in the construction industry. – [Electronic resource]. Access mode: <https://thedigital.gov.ua/news/upershe-vidkrito-dani-pro-deklarativni-ta-dozvilni-dokumenti-z-novogo-reestru-budivelnoi-diyalnosti-ta-provedeno-doslidzhennya-pro-vpliv-vidkritih-danih-na-mistobudivnu-galuzь>.
10. Law of Ukraine "On regulation of urban planning activities". – [Electronic resource]. Access mode: <https://zakon.rada.gov.ua/laws/show/3038-17#Text>.
11. Draft of the Cabinet of Ministers of Ukraine "On the implementation of a pilot project for the introduction of the first stage of the Unified State Electronic System in the field of construction." – [Electronic resource]. Access mode: <https://zakon.rada.gov.ua/laws/show/559-2020-%D0%BF#Text>.
12. Organization and digitization of information on buildings and structures including construction information modeling (BIM), Information management using construction information modeling. Part 1. Concepts and principles (ISO 19650-1: 2018 IDT). Order of 18.03.2020 73 On the adoption of national standards, 2020, [online] Available: http://online.budstandart.com.ua/catalog/docpage.html?id_doc=89571.

13. Chang, Wui Lee, Tay, Kai Meng, Lim, Chee Peng. (2017). A New Evolving Tree-Based Model with Local Re-learning for Document Clustering and Visualization. *Neural Processing Letter*, 46 (2), 379–409.
14. Paltoglou, Georgios; Thelwall, Mike. Twitter, MySpace, Digg. (2012). Unsupervised Sentiment Analysis in Social Media. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3 (4), 66.
15. Atika, Mustafa, Ali, Akbar and Ahmer, Sultan. (2009). Knowledge Discovery using Text Mining: A Programmable Implementation on Information Extraction and Categorization. *International Journal of Multimedia and Ubiquitous Engineering*, 4, 2, 837–848.
16. Chauhan, Shrihari R., Amish, Desai. (2015). A Review on Knowledge Discovery Using Text Classification Techniques in Text Mining. *International Journal of Computer Applications (0975-8887)*, 111, 6.
17. Zhao, Y. (2013). Analysing Twitter Data with Text Mining and Social Network Analysis in Proceedings of the 11th Australasian Data Mining and Analytics Conference (AusDM 2013), p.23.
18. Renganathan, Vinaitheerthan. (2017). Text Mining in Biomedical Domain with Emphasis on Document Clustering. *Healthcare Informatics Research*, 23 (3), 141–146.
19. Karpov, I. A., Antonenko, S. V. (2020). Review of text mining methods. *Current issues of automation and information technology*, 24, 40–46.
20. Buryakov, S., Klochko, A., Fedorov, O. & Terenchuk, S. (2020). Analysis of problems of automatic search of web data in the electronic database of normative documents in construction and building materials industry // BMC-2020 – International Scientific-Practical Conference of young scientists "Build-Master-Class-2020", November 2020, Kyiv, Ukraine.
21. Gubin, M. V. (2004). Models and methods of representation of a text document in information retrieval systems. *Scientific and technical information*, 1, 12, 12–24.
22. Xie, Z., Zeng, Z., Zhou, G. et al. (2017). Topic enhanced deep structured semantic models for knowledge base question answering. *Sci. China Inf. Sci.*, 60, 110103. <https://doi.org/10.1007/s11432-017-9136-x>.
23. Kohonen, T., Ageeva, V. N. (2008). Self-organizing cards. Per. from English M.: BINOM. Knowledge Laboratory, 655.

Посилання на публікацію

- APA Klochko, A. (2021). Introduction of Intellectual Analysis Technologies of Text Documents into Field of Technical Regulation in Construction. *Management of Development of Complex Systems*, 47, 63–70, [dx.doi.org/10.32347/2412-9933.2021.47.63-70](https://doi.org/10.32347/2412-9933.2021.47.63-70).
- ДСТУ Ключко А. А. Впровадження технологій інтелектуального аналізу текстових документів у сферу технічного регулювання в будівництві. *Управління розвитком складних систем*. Київ, 2021. № 47. С. 63 – 70, [dx.doi.org/10.32347/2412-9933.2021.47.63-70](https://doi.org/10.32347/2412-9933.2021.47.63-70).