

ОНТОЛОГІЇ НА ОСНОВІ АНОТОВАНИХ ТЕКСТІВ

Карповський Д.О¹., Шинкаренко В.І².

¹УДУНТ, аспірант, Україна

²УДУНТ, д.т.н., професор, Україна

Анотація. У роботі аналізуються методи анотування текстів, які застосовуються для роботи з онтологіями. Анотування допомагає структурувати інформацію та спрощує її аналіз. Розглядаються два основні підходи: ручне та автоматичне анотування. Ручне анотування є точнішим, оскільки враховує контекст та специфіку мови, але є трудомістким і вимагає значних ресурсів. Автоматичне анотування є більш швидким, набагато швидше піддається масштабуванню, проте може допускати помилки через обмежене розуміння контексту та малий об'єм початкової бази знань. Окремо висвітлюються інструменти і технології для анотування, зокрема програмне забезпечення, що автоматизує цей процес. Визначено проблеми, пов'язані з підвищенням точності автоматичних систем та інтеграцією різних методів аналізу даних. Анотація текстів має важливе значення для розвитку онтологій, автоматичного перекладу та аналізу даних, що сприяє покращенню технологій у цих сферах. На основі анотованих текстів створюється онтологія, яка є основою для подальших досліджень за темою. За допомогою неї вирішується проблема неповноти інформації в тексті, а також виявлення та виправлення протиріч та невідповідностей.

Ключові слова. анотування текстів, онтології, природна мова, автоматичне анотування, обробка текстових даних, інформаційні системи.

Анотування текстів є ключовим етапом у процесі обробки природної мови, яке має велике значення для створення ефективних онтологій. Онтології використовуються для структурування знань у різних сферах, таких як медицина, фінанси, біологія та багато інших, де важливо зберігати точність та узгодженість даних. Вони дозволяють створювати точні моделі для автоматичної обробки інформації, а також виявляти та виправляти суперечливі або неповністю покриті сценарії. Анотації є важливим інструментом для організації текстових даних, що дозволяє покращити їх подальший аналіз, класифікацію та інтеграцію в автоматизовані системи. У роботі розглядаються основні методи анотування текстів, їхні переваги та недоліки, а також роль анотацій у побудові онтологій. Зокрема, ми

досліджуємо, як різні типи анотування можуть допомогти у покращенні точності та ефективності обробки текстових даних для подальшого виявлення й виправлення недоліків у процесах автоматичного аналізу [1].

Методами анотування тексту є «ручне анотування» та «автоматичне анотування».

Ручне анотування – це класичний метод, який полягає у залученні вузьконаправлених фахівців для додавання міток до текстів. Це може включати маркування сутностей (таких як певні об'єкти, місця, дати), а також вказування на відношення між сутностями, лексичні характеристики або навіть мовні синтаксичні структури. Зазвичай цей метод застосовується в дослідженнях, де необхідна висока точність анотацій.[2]

Автоматичне анотування здійснюється за допомогою алгоритмів машинного навчання і методів обробки природної мови (NLP). В цьому випадку використовуються попередньо навчені моделі для автоматичного виділення сутностей, класифікації тексту та виявлення відносин між елементами.

Ручне анотування краще обирати, якщо результати анотування мають бути високоточними, адже фахівці можуть врахувати всі особливості контексту, специфіку мови та граматичні особливості. Також його треба застосовувати в тому випадку, якщо отриману анотацію необхідно адаптувати до специфічних вимог тексту або завдання. Перевагою ручного анотування також є те, що для досягнення високої якості результатів не треба попередньо навчати моделі на великому обсязі попередньо анотованих даних. [3].

Автоматичне анотування є гарним кандидатом для обрання, якщо необхідно обробляти великі обсяги даних за короткий час, також його можна успішно застосовувати до великих наборів даних.

Типи анотування тексту можна поділити на «анотування сутностей», «зіставлення сутностей», «класифікація тексту», «анотування емоційного настрою». [4]

Анотування сутностей – тип анотування, який фокусується на виявленні та розпізнаванні важливих елементів у тексті, таких як імена, дати, місця чи

спеціалізовані терміни. Розпізнавання іменованих сутностей (NER) є важливим інструментом в аналізі тексту, зокрема в застосуваннях, таких як пошукові системи та чат-боти.

Зіставлення сутностей – після того як сутності були виділені, наступним кроком є їх зіставлення з відповідними даними в інших базах знань. Це дозволяє уточнити значення та контекст сутностей, наприклад, з'ясувати, чи йдеться про людину чи організацію під певним іменем.

Класифікація тексту – завдання передбачає анотування текстових фрагментів певною категорією або міткою. Наприклад, класифікація документів або текстів за темою чи емоційним настроєм. Таке анотування важливе в контексті автоматизації процесів пошуку та сортування інформації.

Анотування емоційного настрою – тип анотування займається визначенням емоцій або думок, виражених у тексті. Аналіз емоційного настрою є важливим для розуміння споживчої поведінки, соціальних медіа та інших областей, де потрібно інтерпретувати емоції, висловлені в текстах [5].

Онтології відіграють важливу роль у структуризації знань та забезпеченні їх семантичною обробкою. Анотовані тексти стають основою для створення або вдосконалення онтологій, що дозволяє значно підвищити точність автоматичного аналізу текстів [6].

Онтології допомагають:

- покращити точність аналізу: дозволяють правильно інтерпретувати анотовані дані, покращуючи результати обробки тексту;
- забезпечити інтеграцію даних: дозволяють об'єднувати дані з різних джерел та форматів, полегшуючи їх аналіз та опрацювання;
- спрощують пошук та класифікацію: завдяки структурованій організації знань, онтології полегшують пошук та класифікацію інформації у великих масивах даних.

Анотування текстів для подальшої обробки онтологій є складним завданням, що вимагає вирішення кількох проблем [7]:

- автоматизація процесу анотування: зменшення потреби в людському втручанні при збереженні високої точності;
- виявлення протиріч та неповноти документу: аналіз документу на предмет протилежних за змістом тверджень, а також не повних сценаріїв;
- обробка багатомовних текстів: необхідність створення моделей, які ефективно працюють з текстами на різних мовах;

- інтеграція анотованих даних з іншими системами: забезпечення сумісності між різними типами даних та онтологіями.

Перспективи розвитку теми включають покращення технологій машинного навчання, застосування глибоких нейронних мереж для анотування та інтеграцію цих технологій в практичні додатки [8].

Висновок. Анотування текстів є важливою складовою частиною обробки природної мови та побудови онтологій. Ручне та автоматичне анотування мають свої переваги та обмеження, але в сукупності вони дозволяють створювати точніші, більш ефективні моделі для подальшої автоматичної обробки даних. Обрання методу анотування текстів залежить від задачі та наявних ресурсів (людських, часових, матеріальних). Залучення онтологій для структурування анотованої інформації відкриває нові можливості для розвитку технологій в багатьох сферах, таких як аналіз даних, автоматичний переклад та штучний інтелект. Використання анотування та онтологій даватиме можливість виявляти протиріччя в одному початковому текстовому файлі, а також знаходити неповну інформацію, тобто ту частину певного описуваного об'єкту яка не є описаною.

ЛІТЕРАТУРА

1. Дмитраш О., Романюк А., Тимощук П. Анотування текстів для здійснення автоматичного розпізнавання іменованих сутностей. Вісник Національного університету «Львівська політехніка» – 2013. №751. С. 395–402.
2. Бісікало О.В., Назаров І.О. Огляд методів автоматичного анотування текстів. Праці ВНТУ – 2023, № 2, С. 17–21.
3. Белоіваненко М.В. Модель інтеграції семантичної анотації математичної залежності в онтологічні бази знань. Eastern-European Journal of Enterprise Technologies – 2012. №2(56). С. 54–57.
4. Литвин В., Мушаста С. Метод побудови ембедінгів ознак у задач глибинного навчання на основі онтологій. Вісник Національного університету «Львівська політехніка» - 2023. №13. С. 189–197.
5. Gruninger M., Fox M. Methodology for the Design and Evaluation of Ontologies. Proceedings of IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing – 1995. P. 10–17.
6. Munira K., Anjum M.S, The use of ontologies for effective knowledge modelling and information retrieval. Applied Computing and Informatics – 2018. №14(2). P. 116–126.

7. Карповський Д.О., Шинкаренко В. І., Перспективи рефакторингу онтологій. Information Technologies in Metallurgy and Machine building–ITMM – 2024. С. 257–259.
8. Kuropiatnyk, O., Shynkarenko, V., Zhuchyi, L. and Lyakhova, M. Geometric Fractals' Constructive-Synthesizing Models using Ontological Means. CEUR Workshop Proceeding – 2024., P. 419-431.

USING ONTOLOGIES IN TEXTS ANNOTATION

Dmytro Karpovskyi, Viktor Shynkarenko

Abstract. *The work analyzes the text annotation methods used to work with ontologies. Annotation helps to structure information and simplifies its analysis. Two main approaches are considered: manual and automatic annotation. Manual annotation is more accurate, as it considers the context and language specifics, but is laborious and requires significant resources. Automatic annotation is faster, scales much faster, but can be prone to errors due to limited understanding of the context and a small amount of initial knowledge base. Tools and technologies for annotation are separately highlighted, software that automates this process. Problems associated with increasing the accuracy of automatic systems and integrating various data analysis methods are identified. Text annotation is important for the development of ontologies, automatic translation, and data analysis, which contributes to the improvement of technologies in these areas. Ontology helps solve the problem of incomplete information in the text, as well as identify and correct contradictions and inconsistencies.*

Keywords: *text annotation, ontologies, natural language, automatic annotation, text data processing, information systems.*

REFERENCE

1. Dmytrash O., Romaniuk A., Tymoshchuk P. Annotation of texts for automatic recognition of named entities. Visnyk of Lviv Polytechnic National University - 2013. №751. P. 395–402. [in Ukrainian]
2. Bisikalo O.V., Nazarov I.A. Review of methods of automatic annotation of texts. Works of VNTU - 2023, №2, P. 17–21. [in Ukrainian]
3. Beloivanenko M.V. Model of integration of semantic annotation of mathematical dependence into ontological knowledge bases. Eastern-European Journal of Enterprise Technologies – 2012. №2(56). P. 54-57. [in Ukrainian]
4. Lytvyn V., Mushasta S. Method of constructing feature embeddings in deep learning problems based on ontologies. Visnyk of Lviv Polytechnic National University - 2023. №13. P. 189-197. [in Ukrainian]
5. Gruninger M., Fox M. Methodology for the Design and Evaluation of Ontologies. Proceedings of IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing – 1995. P. 10–17.
6. Munira K., Anjum M.S, The use of ontologies for effective knowledge modelling and information retrieval. Applied Computing and Informatics – 2018. №14(2). P. 116–126.
7. Karpovskyi D.O., Shynkarenko V.I., Prospects for refactoring ontologies. Information Technologies in Metallurgy and Machine building–ITMM – 2024. P. 257-259. [in Ukrainian]
8. Kuropiatnyk, O., Shynkarenko, V., Zhuchyi, L. and Lyakhova, M. Geometric Fractals' Constructive-Synthesizing Models using Ontological Means. CEUR Workshop Proceeding – 2024., P. 419-431.