

Аналіз Використання Алгоритмів Виділення Ключових Слів та Сервісів ШІ для Генерування Зображення на Базі Загальної Природної Мови

Володимир Якимів
кафедра Прикладної Математики і Фізики
Українська Академія Друкарства
Львів, Україна
vldmr.yakymiv@gmail.com

Йосиф Піскозуб
кафедра Прикладної Математики і Фізики
Українська Академія Друкарства
Львів, Україна
pjozef57@gmail.com

Analysis of the Use Of Keyword Selection Algorithms and AI Services for Image Generation Based on Common Natural Language

Volodymyr Yakymiv
dept. of Applied Mathematics and Physics
Ukrainian Academy of Printing
Lviv, Ukraine
vldmr.yakymiv@gmail.com

Jozef Piskozub
dept. of Applied Mathematics and Physics
Ukrainian Academy of Printing
Lviv, Ukraine
pjozef57@gmail.com

Анотація—Проаналізовано використання алгоритмів природної мови та сервісів ШІ для отримання ключових слів та абстракцій для тексту художньої літератури. Розглянуто порівняння використання отриманих ключових та абстракцій для генерування зображень сервісами ШІ. Запропоновано наступні кроки для вдосконалення підходу.

Abstract—The use of natural language algorithms and AI services to obtain keywords and abstractions for the text of fiction is analyzed. A comparison of the use of received keys and abstractions for image generation by AI services is considered. The following steps are suggested to improve the approach.

Ключові слова— *штучний інтелект; генерування зображень; природна мова*

Keywords—*artificial intelligence; image generation; natural language;*

I. ВСТУП

Існує досить велика кількість досліджень спрямованих на пошук алгоритмів на знаходження ключових слів чи узагальнення абстракції із тексту. Такі алгоритми необхідні для узагальнення змісту та розпізнавання основної теми тексту. У результаті буде отримано найбільш використовувані та найважливіші слова та вирази у тексті.

Але чи зможуть отримані ключові слова чи абстракції бути використаними як вхідний текст для генерації

зображень у сервісах ШІ, як спосіб отримання певного графічного представлення чи узагальнення подій із тексту художньої літератури? А також чи можуть існуючі сервіси ШІ справитися краще за існуючі алгоритми та надати кращі результати?

II. ОГЛЯД ЛІТЕРАТУРИ

Перші публікації та дослідження на тему генерування зображень з використанням заданих параметрів почали з'являтися ще у 2015-2016 [1,2], хоча ще у 2007 була представлена система для генерації зображень по ключовій фразі [3]. Основна відмінність досліджуваного підходу від описаної у статті моделі [3], є використання для вибору ключових слів ШІ моделі природної мови замість алгоритму стандартної вибірки ключових слів для конспектування тексту [4,5].

Також числі проводилися чимало досліджень, які мали на меті дослідити зворотній процес: отримання параметрів чи опису об'єктів із згенерованих ШІ зображень. Основна мета цього - дослідити відношення вхідних параметрів до отриманого результату для аналізу, а саме які ключові слова або ж опис об'єктів може гарантувати отримання кращого результату [6,7].

III. МЕТОДОЛОГІЯ ДОСЛІДЖЕННЯ

Дане дослідження аналізує використання алгоритмів природної мови для пошуку ключових слів із тексту а також сервісів ШІ, на прикладі мовної моделі ChatGPT для отримання якісних та відповідних ілюстрації до вибраного тексту.

Для цього підготовлено список із 20 книжок класичної художньої літератури різних авторів. Із кожної книжки було вибрано 5 випадкових сторінок, тобто в загальному було отримано 100 тестових сторінок.

Для порівняння було обрано наступні алгоритми та підходи для отримання ключових чи абстракцій:

- TF-IDF - статистичний показник, що використовується для оцінки важливості слів у контексті документа, що є частиною колекції документів чи параграфа; використовується для отримання найбільш повторюваних чи “важливих” слів чи речень у тексті;
- WordEmbedding (WE) - алгоритм мовного моделювання в обробці природної мови (ОПМ), в яких слова або фрази зі словника перетворюють у вектори дійсних чисел;
- TextRank - це модель ранжирування тексту на основі графіків, яка підходить для пошуку ключових слів у тексті [8];
- WordNet - це алгоритм на основі семантичної схожості двох слів на основі бази WordNet;
- Co-occurrence semantics - це алгоритм із випадковою частотою впорядкованої появи двох суміжних термінів у виділеному тексті;
- використання API системи ШІ із специфічним підготовленим запитом описати сюжет поданого тексту моделі gpt-3.5-turbo;
- використання API системи ШІ із специфічним підготовленим запитом описати сюжет поданого тексту моделі gpt-4-1106-preview;

У дослідженні порівнюються декілька моделей ШІ gpt-3.5-turbo та gpt-4-1106-preview, які мають надати результат у вигляді набору ключових слів або ж абстракції.

Спільною частиною для всіх алгоритмів отримання ключових слів складається із розділення тексту на окремі слова, видаленням пустих значень із отриманого набору слів, токенизації та POS тегування, яке визначає частину мови слова та отримання слів-кандидатів (Рис. 1).

Для отримання слів-кандидатів, з набору слів, отриманого з попередньої етапу видаляються стоп-слова з підготовленої бази NLTK. Далі алгоритм відфільтровує слова за POS тегами (частиною мови), відфільтровує унікальні слова (для всіх алгоритмів, окрім TF-IDF) а також відфільтровує короткі слова. Наступні кроки для оцінки та вибору ключових слів із отриманої матриці чи графіку векторів слів у кожному алгоритмі відрізняються.

Моделі системи ШІ gpt-3.5-turbo та gpt-4-1106-preview обрано серед доступних та найбільш продуктивних та оновлених моделей OpenAI. Для дослідження після використання списку різноманітних команд була вибрана команда “You will be provided with a block of text, and your

task is to return a short one sentence of what is happening in provided text.” та “You will be provided with a block of text, and your task is to return list of keywords (maximum 10 keywords) for provided text.” для отримання стислого та

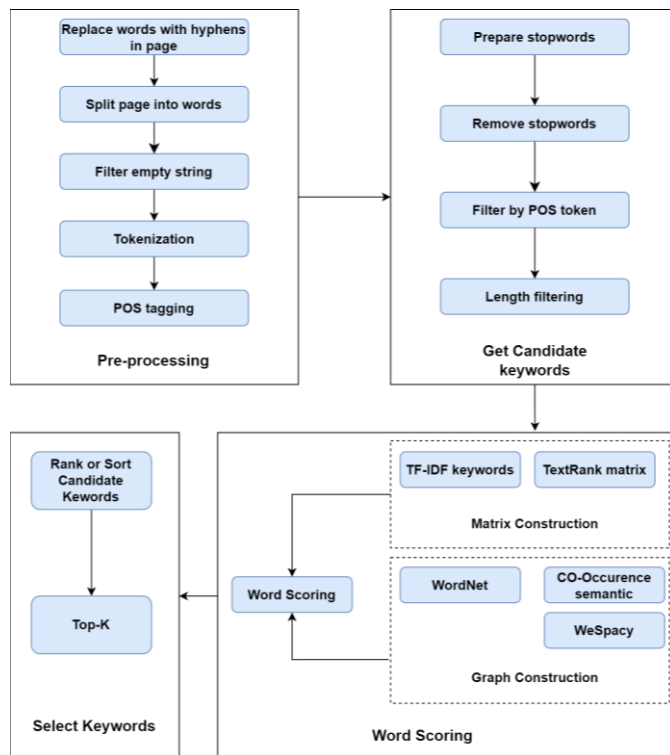


Рис. 1. Загальна схема алгоритмів підбору ключових слів

узагальненого опису подій а також списку ключових слів відповідно.

Наступним етапом дослідження є генерування зображень. На даному етапі використовуються та порівнюються дві системи ШІ, які мають відкрите API:

- Dall-e;
- Stable Diffusion;
- Midjourney.

Для оцінки отриманих ключових слів чи тексту використовується задана шкала (Таблиця. 1):

ТАБЛИЦЯ І. ШКАЛА ОЦІНКИ РЕЛЕВАНТНОСТІ КЛЮЧОВИХ СЛІВ АБО ОПИСІВ ТЕКСТУ

Текстова оцінка	Числова оцінка	Опис
Не відповідає	1-2	Надані ключові слова або резюме мають мінімальний зв'язок зі змістом тексту або зовсім не пов'язані з ним. Тема або інформація майже не збігаються.
Мало відповідає	3-4	Є деякі незначні зв'язки між ключовими словами чи резюме та текстом, але відповідність слабка. Ключові слова можуть торкатися периферійних аспектів.

Дещо відповідає	5-6	Є помірний рівень релевантності. Ключові слова або резюме охоплюють деякі аспекти тексту, але є помітні прогалини або відмінності.
Помірно відповідає	7-8	Ключові слова або резюме добре узгоджуються з текстом, фіксуючи основні ідеї та теми. Однак можуть бути деякі нюанси або деталі, які не відображені ідеально.
Дуже відповідає	9	Ключові слова або короткий виклад точно відповідають змісту тексту. Вони ефективно охоплюють основні моменти та теми, лише з незначними варіаціями.
Точно відповідає	10	Ключові слова або резюме ідеально описують текст. Кожна важлива деталь, тема, нюанс точно відображені.

Для оцінки відповідності зображень до тексту використовувалася така ж сама шкала.

IV. РЕЗУЛЬТАТИ

Ключовою особливістю даного дослідження, є те що в даному дослідженні аналізуються не просто наскільки отримані ключові слова відповідають аналізованому тексту, а наскільки вони придатні для подальшого використання як запит для генерації зображень.

Отримані результати показали, що абстракції чи ключові слова згенеровані сервісом III ChatGPT може надати такі абстракції чи ключові слова з тексту, які моделі генерування зображень зможуть краще опрацювати та надати ілюстрації близькі до тексту твору. Тільки алгоритми CO Occurrence та дещо гірше TextRank серед алгоритмів показали кращі результати, які в деяких випадках співмірні з результатами використання сервісів III.

В цілому можна помітити схожу тенденцію опрацювання зображень між різними сервісами генерування зображень (Рис. 2). Також на графіку (Рис. 3)

можна помітити, що немає схожості у результатах середньої оцінки за книгами. Іншими словами, твори деяких авторів даються складніше для абстрагування, а згенеровані зображення не надто відповідають аналізованому тексту.

При аналізі дані, було також помітно, що Stable Diffusion у багатьох випадках додає більше осіб у згенерованому зображенні, ніж це описується у вхідних даних. Помітна також залежність, коли отримані опис чи ключові слова дуже погано відповідали вхідному тексту, то Stable Diffusion не міг взагалі згенерувати нічого, було отримано просто чорне зображення, або не отримано зображення взагалі.

V. ВИСНОВКИ

Після аналізу отриманих результатів можна зробити висновки, що описаний підхід із використанням сервісами III справляються з поставленою задачею, та можуть надати сервісам генерування зображень відповідний опис тексту.

Проте згенеровані абстракції містять занадто загальний опис подій на сторінці і сервісам III важко зосередитися над загальною картиною. У подальшому дослідженні варто спробувати:

- використати даний підхід для окремих абзаців;
- використати сервіси III виключно для вибірки ключових слів (обмеженої кількості);
- змінити задану команду таким чином, що III спробував розділити окремі сцени зі сторінки чи виокремити окремі події зі сторінки тексту, та згенерувати зображення для кожної з них.

Також, оцінка відповідності чи схожості вибраних слів чи згенерованих абстракції та зображень доволі суб'єктивна, адже будь-який читач може по-своєму, уявляти та сприймати прочитаний текст, проте завданням даного дослідження є не абсолютно точно вгадати і передати сформоване уявлення про задану сторінку до того як це уявляє читач, а сформувати певний напрям образів і об'єктів для формування у читача кращого сприйняття твору.

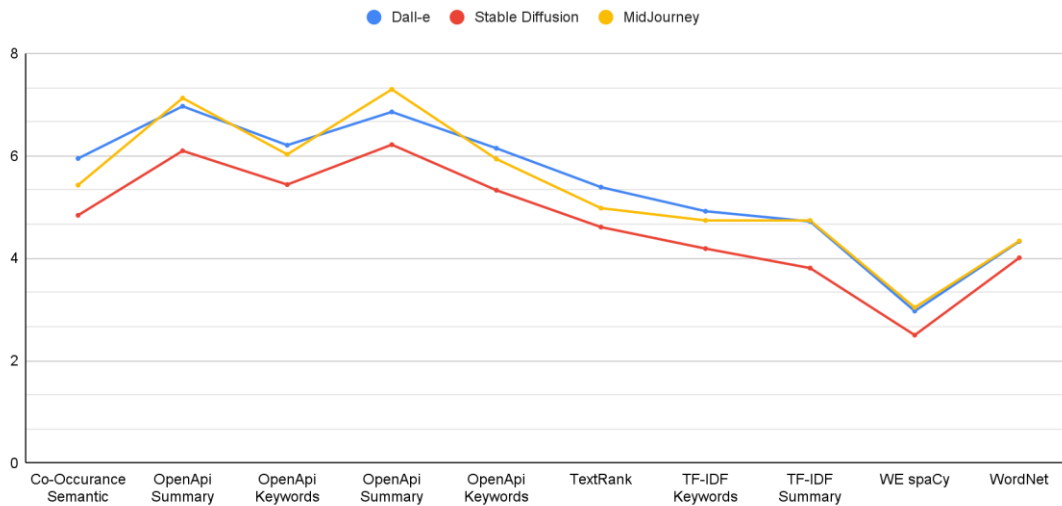


Рис. 2. Порівняння середньої оцінки між Dall-e, Stable Diffusion і MidJourney за алгоритмом

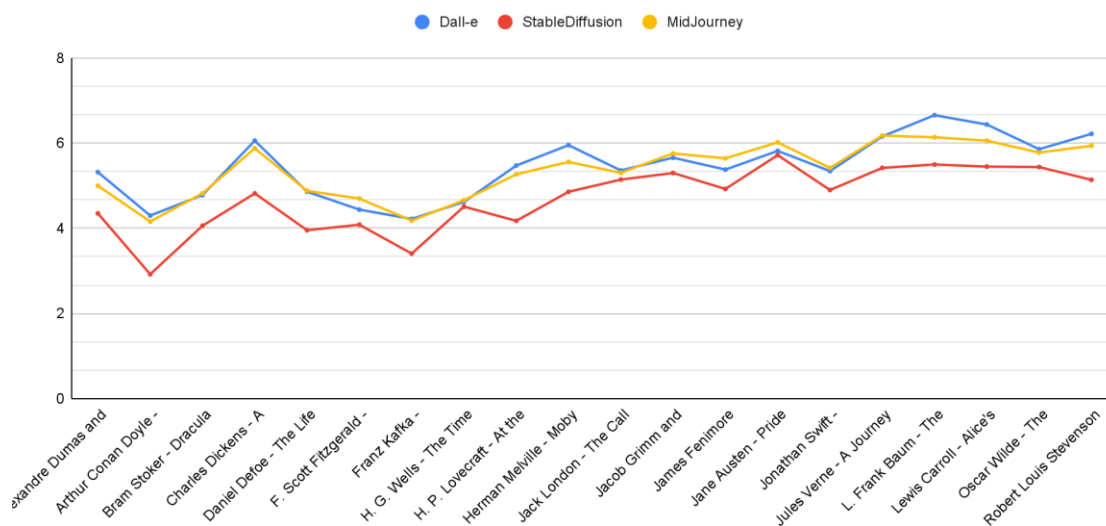


Рис. 3. Порівняння середньої оцінки між Dall-e, Stable Diffusion і MidJourney за книгами

ЛІТЕРАТУРА REFERENCES

- [1] Elman Mansimov, Emilio Parisotto, Jimmy Ba, and Ruslan Salakhutdinov, "Generating Images from Captions with Attention", ICLR, 2016, <https://doi.org/10.48550/arXiv.1511.02793>.
- [2] Reed Scott, Akata Zeynep, Logeswaran Lajanugen, Schiele Bernt, Lee Honglak, "Generative Adversarial Text to Image Synthesis", International Conference on Machine Learning, 2016, <https://doi.org/10.48550/arXiv.1605.05396>.
- [3] Zhu Xiaojin, Goldberg Andre, w B. Eldawy Mohamed, Dyer Charles R, Strock Bradley, "A text-to-picture synthesis system for augmenting communication", Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, pp. 1590–1595, 2007.
- [4] Mihalcea Rada, Tarau Paul, "TextRank: Bringing Order into Text", 2004.
- [5] Turney Peter, "Learning Algorithms for Keyphrase Extraction", Information Retrieval, vol. 2, pp.
- [6] Shen Xinyue, Qu Yiting, Backes Michael, Zhang Yang, "Prompt Stealing Attacks Against Text-to-Image Generation Models", 2023, <https://doi.org/10.48550/arXiv.2302.09923>.
- [7] Xu Xingqian, Guo Jiayi, Wang Zhangyang, Huang Gao, Essa Irfan, Shi Humphrey, "Prompt-Free Diffusion: Taking "Text" out of Text-to-Image Diffusion Models.", Pre-print, 2023.
- [8] Mihalcea, Rada and Paul Tarau, "TextRank: Bringing order into text", Proc. Conf. on Empirical Methods Natural Lang. Process., pp. 404-411, 2004.