

Pengembangan Alat Ekstraksi Informasi pada Teks Berita dengan Large Language Model dan Knowledge Graph = Development of Information Extraction Tools for News Texts with Large Language Model and Knowledge Graph

Fransisco William Sudianto, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920553294&lokasi=lokal>

Abstrak

Perkembangan Large Language Model (LLM) terjadi secara cepat dan mengalami kemajuan yang signifikan. Hal ini mendorong penggunaan dan pemanfaatan LLM pada berbagai bidang. Disisilain, Knowledge Graph (KG) menyediakan cara yang terstruktur dan bermakna untuk menyimpan informasi. KG sudah banyak digunakan secara luas di berbagai aplikasi, seperti mesin pencari, sistem rekomendasi, dan sistem penjawab pertanyaan. Salah satu pemanfaatan LLM dan KG yang masih jarang adalah pada bidang jurnalistik, khususnya untuk menganalisis dan memvisualisasikan berita. Penelitian ini bertujuan untuk mengembangkan alat ekstraksi informasi yang efisien, akurat, dan interaktif untuk menganalisis teks berita menggunakan pendekatan gabungan antara LLM dan KG. Metode ini menggabungkan keunggulan kedua teknik tersebut untuk meningkatkan pemahaman dan ekstraksi informasi dari teks berita yang kompleks. Tujuannya adalah agar pembaca dapat memahami informasi yang terdapat pada teks berita dengan lebih interaktif. Penulis memanfaatkan LLM yang telah terlatih secara luas dalam memahami dan menghasilkan teks untuk mengidentifikasi informasi penting dalam teks berita, seperti entitas, sentimen, kutipan, relasi antar entitas, dan unsur 5W1H (Who, What, Where, When, Why, How), urutan kronologis kejadian, dan hubungan bagian-keseluruhan (mereology) dalam teks berita. Untuk mengekstraksi informasi tersebut, prompt dimodifikasi dengan menggunakan pendekatan one-shot-prompting untuk memberikan konteks dan contoh kepada LLM dalam memahami teks berita. Kemudian, informasi yang diekstraksi di visualisasikan dalam bentuk KG yang merepresentasikan pengetahuan terstruktur tentang entitas dan hubungannya didalam teks. Selain itu, penelitian melibatkan pembuatan sebuah website yang akan menyediakan antarmuka untuk sistem agar pengguna dapat melakukan analisis teks berita secara langsung dan interaktif. Evaluasi utama yang dilakukan pada penelitian ini adalah mengukur akurasi jawaban yang dihasilkan oleh LLM pada setiap bagian informasi yang diekstraksi dan bagaimana visualisasi KG yang baik untuk informasi yang didapat. Penelitian ini menunjukkan bahwa LLM mampu mengekstraksi informasi yang diinginkan dengan cukup akurat dan visualisasi KG dapat menyajikan informasi dengan lebih interaktif dan mudah dimengerti. Penelitian ini telah menunjukkan bahwa LLM dan KG dapat dimanfaatkan sebagai alat ekstraksi dan visualisasi informasi yang ada pada teks berita.

.....The development of Large Language Model (LLM) is happening rapidly and has made significant progress. This encourages the use and utilization of LLM in various fields. On the other hand, Knowledge Graph (KG) provides a structured and meaningful way to store information. KG has been widely used in various applications, such as search engines, recommendation systems, and question answering systems. One of the uses of LLM and KG that is still rare is in the field of journalism, especially for analyzing and visualizing news. This study aims to develop an efficient, accurate, and interactive information extraction tool for analyzing news texts using a combined approach between LLM and KG. This method combines the advantages of both techniques to improve understanding and information extraction from complex news

texts. The goal is for readers to understand the information contained in the news text more interactively. The author uses LLM which has been widely trained in understanding and producing text to identify important information in news texts, such as entities, sentiments, quotes, relationships between entities, and 5W1H elements (Who, What, Where, When, Why, How), chronological sequence of events, and part-whole relationships (mereology) in news texts. To extract the information, the prompt is modified using a one-shot-prompting approach to provide context and examples to LLM in understanding the news text. Then, the extracted information is visualized in the form of KG which represents structured knowledge about entities and their relationships in the text. In addition, the study involves the creation of a website that will provide an interface for the system so that users can analyze news texts directly and interactively. The main evaluation carried out in this study is to measure the accuracy of the answers generated by LLM on each part of the extracted information and how good KG visualization is for the information obtained. This study shows that LLM is able to extract the desired information quite accurately and KG visualization can present information more interactively and easily understood. This study has shown that LLM and KG can be used as tools for extracting and visualizing information in news texts.