

Pengembangan Language Model KEPLER untuk Bahasa Indonesia = Development of KEPLER Language Model for the Indonesian Language

Inigo Ramli, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20525251&lokasi=lokal>

Abstrak

Model bahasa dapat mempelajari struktur suatu bahasa manusia menggunakan korpus yang tidak terstruktur. Namun, model bahasa secara umum belum dapat mempelajari pengetahuan faktual yang direpresentasikan oleh knowledge graph. Terdapat beberapa usaha untuk membuat model bahasa yang dapat mempelajari pengetahuan faktual seperti KEPLER. Sayangnya, belum terdapat penelitian yang komprehensif mengenai integrasi pengetahuan faktual terhadap pelatihan model bahasa Indonesia. Penelitian ini mengajukan model bahasa Indonesia baru bernama IndoKEPLER yang melatih model bahasa Indonesia yang sudah ada dengan korpus Wikipedia Bahasa Indonesia dan memanfaatkan pengetahuan faktual dari Wikidata. Selain itu, penelitian ini juga mengajukan metode knowledge probing baru untuk menguji pemahaman faktual suatu model bahasa Indonesia. Hasil eksperimen penelitian ini menunjukkan bahwa pelatihan model IndoKEPLER dapat meningkatkan pemahaman faktual suatu model bahasa Indonesia.

.....Pretrained language models have the ability to learn the structural representation of a natural language by processing unstructured textual data. However, the current language model design lacks the ability to learn factual knowledge from knowledge graphs. Several attempts have been made to address this issue, such as the development of KEPLER. Unfortunately, such knowledge enhanced language model is not yet available for the Indonesian language. In this experiment, we propose IndoKEPLER: a pretrained language model trained using Wikipedia Bahasa Indonesia and Wikidata. We also create a new knowledge probing benchmark named IndoLAMA to test the ability of a language model to recall factual knowledge. This experiment shows that IndoKEPLER has a higher ability to recall factual knowledge compared to the text encoder it's based on.