

IndoGrad: Adaptasi Skema Winograd Bahasa Indonesia untuk Resolusi Kata Ganti yang Membutuhkan Penalaran = IndoGrad: Indonesian Adaptation of Winograd Schema Challenge on Coreference Resolution with Reasoning

Mirza Alim Mutasodirin, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20524287&lokasi=lokal>

Abstrak

Penelitian pada pengolahan bahasa manusia tentang resolusi kata ganti yang membutuhkan penalaran menjadi sangat penting agar mesin mampu menyelesaikan tugas dengan tingkat kesulitan yang lebih tinggi dari resolusi kata ganti biasa. Mesin dituntut untuk mampu menyelesaikan tugas tersebut yang membutuhkan penalaran seperti yang dimiliki otak manusia. Penelitian tentang ini sudah berjalan selama satu dekade terakhir pada Bahasa Inggris, yang disebut dengan the Winograd Schema Challenge (WSC). Namun, sepanjang pencarian kami, belum ditemukan sama sekali penelitian tentang ini pada Bahasa Indonesia. Kami menginisiasi penelitian tentang WSC pada Bahasa Indonesia dengan membangun dataset baru yang diadaptasi dari dataset WSC berbahasa Inggris yang sudah ada. Dataset baru ini diberi nama IndoGrad (Indonesian Winograd). IndoGrad memiliki 1.134 data latih, 284 data validasi, dan 318 data uji dengan format cloze-style. Untuk menguji kelayakan data ujinya, IndoGrad diujikan kepada tiga orang manusia pemegang gelar sarjana dan disimpulkan bahwa data ujinya bisa dijawab oleh manusia dengan akurasi tinggi. Performa manusia secara Full-Agreement yang didapatkan adalah 94,0% akurasi dan secara Majority-Agreement adalah 97,8% akurasi. Kelayakan data latihnya disimpulkan dari bisanya data latih dipelajari oleh model sehingga mendapatkan training accuracy mendekati 100%. Dua belas pretrained models berbasis BERT diuji untuk mengukur performa mereka terhadap dataset ini. Performa deep learning model terbaik yang didapatkan adalah 62,58% akurasi oleh IndoBERT-Large dan 68,86% akurasi oleh XLM-RoBERTa-Large. Hasil ini masih jauh dari performa manusia dan perlu penelitian lebih lanjut di masa depan.

.....Natural Language Processing (NLP) study on coreference resolution with commonsense reasoning becomes very important to make machines capable of tackling high-difficulty coreference resolution. Machines are required to complete the task that needs reasoning, like the human brain. Study on this topic has been running for the last decade on English, named the Winograd Schema Challenge (WSC). However, as far as our search goes, we did not find any study on this in Indonesian. We initiate the first study about WSC in Indonesian by building a new dataset adapted from the previously available English WSC dataset. This new dataset is named as IndoGrad (Indonesian Winograd). IndoGrad has 1,134 training data, 284 validation data, and 318 testing data in cloze-style format. To determine the feasibility of the testing data, IndoGrad was tested on three humans holding bachelor's degrees and it was concluded that the testing data could be answered by humans with high accuracy. The human performance achieved are 94,0% Full-Agreement Accuracy and 97,8% Majority-Agreement Accuracy. The feasibility of the training data is concluded from its ability to be studied by the model so that the training accuracy is close to 100%. Twelve BERT-based pretrained models were tested to measure their performance against this dataset. The best deep learning model performance achieved are 62,58% accuracy by IndoBERT-Large and 68,86% accuracy by XLM-RoBERTa-Large. This result is far from human performance and it needs further study in the future.