Vitaly Shalumov

Research fellow, Rafael, vitalysh@rafael.co.il
Rafael, P.O.Box 2250, Haifa 3102102, Israel

01/05/2024

Named Entity Recognition Dataset

To Avner Algom and the IAHLT team,

I wish to express my gratitude for the comprehensive named entity recognition (NER) dataset for the Hebrew Language introduced by IAHLT. Both the large number of high-quality annotated samples and the extension to categories not present in open-source datasets, such as "TTL," enabled the training of several high-performance transformer-based models. These models are utilized in production and contribute significantly to our entity extraction pipeline.

Additionally, I want to emphasize the openness to feedback and requests, such as additional datasets for other tasks. Apart from the Hebrew NER dataset, several other datasets are being considered for model training, demonstrating the importance of curated resources.

Best regards,

Vitaly