



PRESENTER

Alex Brandsen

Background

Too much text! We have tens of thousands of excavation reports, and the current metadata search is not good enough. In response, we made **AGNES**: a full text search engine. Domain-specific entities¹ play a central role in archaeological IR, so we combine full text with named entity search. To make this work, we first need to do Named Entity Recognition.

Methods

We further pre-train BERTje with our corpus. Using all 65,000 documents, the Dutch BERTje model² is further pre-trained on excavation reports, creating **ArcheoBERTje**. We then fine-tune this model on NER training data to predict entities. We compare ArcheoBERTje with BERTje and Google's multilingual model.

Results

Higher performance, lower standard deviation

Model	Precision	Recall	F1 (Std.)	Fails
CRF Baseline	0.785	0.526	0.630 (-)	n/a
multiBERT	0.623	0.550	0.583 (0.015)	4
BERTje	0.718	0.682	0.699 (0.005)	0
ArcheoBERTje	0.743	0.729	0.735 (0.004)	0

Extra Results

Ensemble methods and domain knowledge do not improve the model. We also tried ensemble methods by majority voting or combining the predictions from the 3 BERT models into a CRF, but this did not improve performance. Adding domain knowledge (thesaurus entries) as a feature also did not lead to increased F1.

¹ Artefacts, time periods, materials, places, contexts, species

² De Vries, W., van Cranenburgh, A., Bisazza, A., Caselli, T., van Noord, G. & Nissim, M. (2019). BERTje: A Dutch BERT Model. *ArXiv preprint: arxiv.org/abs/1912.09582*



ArcheoBERTje outperforms both the multilingual and Dutch BERT models on NER in the archaeology domain, reaching an average F1 score of 0.735

CLIN31, 09-07-2021

Distribution of F1 scores over 10 runs with different seeds, for each of the 5 folds (50 runs per model)

