Universiteit Leiden
The Netherlands

Digging in documents: using text mining to access the hidden knowledge in Dutch archaeological excavation reports
Brandsen, A.

Propositions

accompanying the dissertation

Digging in Documents
Using Text Mining to Access the Hidden Knowledge in Dutch Archaeological Excavation Reports

by

Alex Brandsen

1. Using the pairwise averaged F1 score for the inter annotator agreement of named entity annotation is more useful than the traditionally used Cohens Kappa.

2. While BERT models trained on out-domain data can lead to decent performance, further pre-training BERT models with more domain- and language-specific data leads to higher performance in the Dutch archaeology domain.

3. Agile development, user centred design, and thorough system evaluation leads to better software, and should be used more in archaeology and the humanities.

4. Without text mining and other data science techniques, it will be impossible to tame the large influx of archaeological data, leading to incomplete and possibly biased research.

5. Uncritically accepting a machine learning outcome is unwise.

6. Open science with free open source software published in open access journals leads to more transparent, more accessible, and overall better research.

7. GPU usage for Deep Learning in this research produced CO_2 output equivalent to a flight from Amsterdam to Prague. Due to this, less computationally expensive methods are preferable, except in cases where deep learning substantially outperforms.

8. For interdisciplinary research between archaeology and computer science, a digital archaeologist is needed to facilitate information flow between the two fields.

9. Total liberation of all human and non-human animals is required for a just society.

10. When your only tool is a trebuchet, every problem looks like a siege.