

# Towards a Benchmark Dataset for the Digital Humanities

Felix Ernst<sup>1</sup>, Nicolas Blumenröhr<sup>1</sup>

<sup>1</sup>Karlsruhe Institute of Technology, Karlsruhe, Germany

## Abstract

Applications of ontology matching such as ontology engineering, information sharing or query answering are of growing importance in the field of Digital Humanities (DH). To gain knowledge about suitable ontology matching tools for DH research, a successful evaluation of these is crucial. Unfortunately, there exist no reference alignment datasets that address DH specific requirements such as support of multiple (historic) languages, domain specific terms and an easily applicable data format. Therefore, we propose the creation of a dataset as base for a future DH OAEI track which uses knowledge bases and expert surveys as reference alignment sources. Using surveys leads to a graduated scale of term similarity and association, which makes advanced evaluation metrics possible.

## Keywords

Ontology Matching, Reference Dataset, Digital Humanities, Multilingual, FAIR Digital Objects

## 1. Introduction

Due to the growing amount of data and its increasing complexity, fields such as the Digital Humanities are now in the process of applying semantic tools, including ontologies, to gain new research insights [1]. Automatic ontology matching plays a crucial role in creating or extending ontologies [2], facilitates collaboration and reduces the workload of researchers. The Ontology Alignment Evaluation Initiative (OAEI) is a reputable source for high-quality benchmark datasets and the evaluation results of matching tools. However, the scarcity of datasets for application within the DH remains a significant obstacle in conducting reliable evaluations of ontology matchers for the DH use case. Existing OAEI tracks do not adequately address all the following specific DH requirements: (a) a wide range of (historical) languages and writing systems; (b) domain specific terms; (c) use of a data model suitable for easily creating knowledge organization systems such as SKOS. Moreover, the dataset will take account of the distinction between association (e.g. cup; coffee) and similarity (cup; mug) which is missing in multiple popular gold standards for similarity ratings [3], which limits their use for comprehensive evaluations of ontology matchers for the DH.


We propose addressing these challenges by creating a new benchmark dataset tailored to the DH requirements, with its prospective application as a newly established OAEI track. In addition, this dataset will be represented as a FAIR Digital Object (FDO) [4], a concept that


---

OM-2023: Proceedings of the 18th International Workshop on Ontology Matching, November 2023, Athens, Greece

 felix.ernst@kit.edu (F. Ernst)

 0000-0002-2102-4170 (F. Ernst); 0009-0007-0235-4995 (N. Blumenröhr)

 © 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

implements the FAIR principles [5] with a particular focus to facilitate machine-actionable tasks which makes the data ready for ML matchers and applications.

## 2. Criteria for Dataset Construction

To ensure the dataset's effectiveness, several criteria were established. The dataset will encompass both similarity and association relations, allowing for a comprehensive evaluation. Furthermore, it will strike a balance between general terms and specialized domain-specific terminology, enabling a broader applicability across different disciplines within the humanities. When establishing the ground truth, objectivity is of high importance and only the same part-of-speech classifications will be compared. Additionally, the dataset will encompass a variety of languages used within the humanities, with adequate translation of terms.

In order to make these specific data characteristics assessable also for applications outside of ontology evaluation toolkits, the information has to be represented in a standardized way and actionable by operations that can be performed on the data. This is enabled by the concept of FDOs, as a Persistent Identifier (PID) is assigned to the data, and a set of typed metadata attributes that are associated with operations, making the data machine-actionable.

## 3. Methodology

Two methods will be employed to construct the reference alignment dataset. The first method will use DH specific terms within freely available knowledge graphs such as WordNet<sup>1</sup>, GermanNet<sup>2</sup>, Wikidata<sup>3</sup> and specialized terminology resources like Loterre<sup>4</sup>. Only relations of synonyms classes such as `skos:exactMatch` will be used, disregarding relations like `skos:related` due to their subjectivity and dependency on the depth within the hierarchy. The reference alignment will be manually created using synonyms as matches between terms, including translations coming from the knowledge graphs or from language dictionaries. A benefit of this method is that many people have contributed to the knowledge graphs, which makes it less sensitive to bias in synonym classification. In addition, the dataset will have a broad domain coverage and supports a variety of (historic) languages that are relevant for the specific domains. The major drawback is that only synonyms and no associations can be covered, and discipline-specific terms of several small disciplines might be missing entirely. This is particularly regrettable since smaller disciplines are in high need of digital research tools, but often are neglected when tailoring applications of ontology matching to their research.

To address the aforementioned drawbacks, the second method for constructing a reference alignment utilizes surveys that are conducted among domain experts. They will quantify the similarity and association (while taking care of the distinct nature of both) of preselected terms coming from both the first method and from controlled vocabularies created by collaborating researchers of different domains. This approach introduces varying degrees of similarity, which

---

<sup>1</sup><https://wordnet.princeton.edu/>

<sup>2</sup><https://uni-tuebingen.de/en/142806>

<sup>3</sup><https://www.wikidata.org/>

<sup>4</sup><https://www.loterre.fr/>

can be utilized to incorporate an additional metric to the F1-score. This metric would effectively capture the nuanced nature of term matching, where a matching tool's failure to align two terms may be considered less critical if the rated similarity between the terms is categorized by the researchers as 'somewhat similar.' Conversely, if the similarity rating indicates 'very similar,' the absence of a match becomes more significant. By incorporating this refined measure, the evaluation of ontology matchers can better account for the subtleties in term matches, leading to a more comprehensive and accurate assessment of their performance. Furthermore, the surveys focusing on specific domains will yield ontologies containing highly domain-specific terms, presenting a novel challenge for ontology matching tools.

To achieve high quality, the dataset is restricted to domains in which contact to scholars is already established through ongoing DH projects, including the fields of Greek studies, Egyptology, or philology. Multilingual terms are only introduced if the confidence of translation is high, e.g. given by scholars of the respective field.

## 4. Conclusion

This abstract proposes the creation of a multilingual benchmark dataset that addresses the limitations of existing resources when applied to the DH domain. By incorporating both similarity and association relations, focusing on domain-specific terminology, and combining knowledge bases and expert surveys as sources, this dataset has the potential to contribute significantly to the evaluation of ontology matching tools. Additionally, its prospective integration as OAEI track offers opportunities for widespread evaluation and comparison. As a further stage, the proposed methodology can also be applied to other disciplines beyond the humanities, thus facilitating cross-disciplinary evaluation of matchers.

## Acknowledgments

This research is funded by the German Research Foundation (DFG)—CRC 980 Episteme in Motion, Project-ID 191249397, and the Helmholtz Metadata Collaboration Platform (HMC) and supported by the German National Research Data Infrastructure (NFDI4Ing, NFDI-MatWerk).

## References

- [1] E. Hyvönen, Using the Semantic Web in digital humanities: Shift from data publishing to data-analysis and serendipitous knowledge discovery, *Semantic Web* 11 (2020) 187–193.
- [2] J. Euzenat, P. Shvaiko, *Ontology Matching*, Springer, Berlin, Heidelberg, 2013.
- [3] F. Hill, R. Reichart, A. Korhonen, SimLex-999: Evaluating Semantic Models With (Genuine) Similarity Estimation, *Computational Linguistics* 41 (2015) 665–695.
- [4] E. Schultes, P. Wittenburg, FAIR Principles and Digital Objects: Accelerating Convergence on a Data Infrastructure, in: *Data Analytics and Management in Data Intensive Domains*, CCIS Series, Springer International Publishing, Cham, 2019, pp. 3–16.
- [5] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, et al., The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* 3 (2016) 160018.