Faculty of History
University of Warsaw

PEOPLE
PLACES
EVENTS

Explaining (representations) of a heap of pebbles. Disambiguation in digital humanities

Workshop

Call for Participation

When? November 4, 2022
Where? Online

Increasing human reliance on computational methods and techniques of learning, reasoning, and explanation have consequences for digital humanities. With many representations of artefacts and sources, older source critical approaches and newer digital source criticism approaches, face numerous challenges. In this workshop, we invite researchers engaged with the digital humanities to present, exchange and discuss their different perspectives on disambiguation in the digital humanities.

We would like to consider this workshop in the process of transitioning historical research from traditional media-bound sources to flexible digital representations. Artificial intelligence, machine learning, hermeneutics, algorithms, geospatial analysis all offer different possibilities. They are becoming more and more common in research. The workshop should consider their advantages, like rapid results of massive computing. They have, however, disadvantages, or - to keep the positive spirits - challenges for the high expectations of interpretative research processes. Data quality as its structure, credibility, and interoperability or readability (cf. Cai, Zhu 2015) in relationship to interpretative uncertainty become an important, but still underdefined reference. One important element needs to be added here: "for AI, not only human brains, if possible". Several

questions arise related to data and visual representation in the digital disambiguation process:

- How to make analysed phenomena structured, credible, readable; simply: clearer?
- How to make the complex and fuzzy processes, events, ideas, objects, places less ambiguous?
- How to assess what and how resources and data are ambiguous and needs to be included into the disambiguation process?
- What level of disambiguation is sufficient for explaining particular phenomena?
- How do researchers determine a level of disambiguation sufficient for the data quality and research purposes?

These questions for the workshop are not new. Already a rich collection of possible responses from various disciplines offer guides for researchers (i.e. Beelen et al. 2021, Pedersen 2022, Cruz 2018, Agirre, Rigau 1996). We would like to take up the matter of disambiguation in a workshop, address these questions to continue the discussion and to contribute to it with some insights. We would like to make these questions the starting point for the workshop discussions. While the organisers' scope is in geography, cartography and history, we would like to broaden the workshop discussion to all scholars working in the digital humanities regardless of backgrounds and research fields.

At the workshop, we aim to advance discussions arising in very old dilemmas. The heap of pebbles mentioned in the title of this Call for Participation originates in an ancient philosophical paradox introduced by Eubulides from Millet: how many grains or pebbles create a heap or a pile? (cf. Krukowski 2021, p. 37). During the workshop, within the academic discussion, we would like to determine conditions for testing a hypothesis as follows: "machine learning can enhance cognitive processes in humanities".

The online workshop we are organising on **the 4th of November, 2022** will have:

- moderated discussion on above mentioned and other questions,
- key speaker lecture delivered by Miroslaw Krukowski,
- intermissions,
- friendly atmosphere.

Participation is free of charge. Your contribution to the discussion is of great value. We do not plan any post-workshop publication beyond the workshop website with abstract and contact information.

Please, do not hesitate contacting Dr Wiesława Duży in any issue connected with the workshop: w.duzy@uw.edu.pl.

Schedule:

10:00 Welcome and short introduction
10:15 Key speaker:
Mirosław Krukowski, On the imperfection and vagueness of certain geographical concepts
11:15 Discussion
12:00 Break
12:15 Presentations of projects by guests
13:30 Lunch break
14:30 Plenary moderated discussion with instant notes
16:30 Break
17:00 Plenary moderated discussion with instant notes
18:00 Summary and final remarks
18:30 End of workshop

Cai, Zhu 2015: Cai, L. and Zhu, Y., 2015. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. _Data Science Journal_, 14, p.2. DOI: http://doi.org/10.5334/dsj-2015-002.

Krukowski 2021: Mirosław Krukowski, Nieostrość w modelowaniu kartograficznym, Lublin 2021, ISSN: 978-83-227-9478-4.

Pedersen 2022: T. Pedersen, A baseline methodology for word sense disambiguation. In Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics, 2022, pages 126-135.

Cruz 2018: Manuel Márquez Cruz, A lexicographical model based on the predicative framework theory (functional grammar) for sense disambiguation. An application to Latin author dictionaries, *Digital*

Scholarship in the Humanities, Volume 33, Issue 2, June 2018, Pages 362–373, https://doi.org/10.1093/llc/fqx037

Beelen et al. 2021: Kaspar Beelen, Federico Nanni, Mariona Coll Ardanuy, Kasra Hosseini, Giorgia Tolfo, Barbara McGillivray, When Time Makes Sense: A Historically-Aware Approach to Targeted Sense Disambiguation, Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pages 2751–2761.

Agirre, Rigau 1996: Eneko Agirre, German Rigau, Word Sense Disambiguation Using Conceptual Density, COLING '96: Proceedings of the 16th conference on Computational linguistics - Volume 1, 1996 Pages 16–22, https://doi.org/10.3115/992628.992635

Ferdinando Di Martino, Vincenzo Loia, Salvatore Sessa, and Michele Giordano [in:] Fuzzy Modelling with Spatial Information for Geographic Problems, eds. Frederick E. Petry, Vincent B. Robinson, Maria A. Cobb, Berlin-Heidelberg 2020, Springer-Verlag, pp. 185-208.

NAWA CHAIR

The project is co-financed by the Polish National Agency for Academic Exchange within the NAWA Chair programme; project number is: PPN/PRO/2020/1/00004).