

# Good Applications for Crummy Entity Linkers? The Case of Corpus Selection in Digital Humanities

**Alex Olieman** [alex@olieman.net](mailto:alex@olieman.net)  
University of Amsterdam

**Kaspar Beelen** [kasparvonbeelen@gmail.com](mailto:kasparvonbeelen@gmail.com)  
University of Amsterdam

**Jaap Kamps** [kamps@uva.nl](mailto:kamps@uva.nl)  
University of Amsterdam

**Milan van Lange** [m.van.lange@niod.knaw.nl](mailto:m.van.lange@niod.knaw.nl)  
NIOD

## Keywords

Entity Linking   Real-World Applications   Semantically-Enhanced Search   Interactive Information Retrieval  
Corpus Selection   Digital Humanities

---

## Abstract

Over the last decade we have made great progress in entity linking (EL) systems, but performance may vary depending on the context and, arguably, there are even principled limitations preventing a "perfect" EL system. This also suggests that there may be applications for which current "imperfect" EL is already very useful, and makes finding the "right" application as important as building the "right" EL system. We investigate the Digital Humanities use case, where scholars spend a considerable amount of time selecting relevant source texts. We developed WideNet; a semantically-enhanced search tool which leverages the strengths of (imperfect) EL without getting in the way of its expert users. We evaluate this tool in two historical case-studies aiming to collect a set of references to historical periods in parliamentary debates from the last two decades; the first targeted the Dutch Golden Age, and the second World War II. The case-studies conclude with a critical reflection on the utility of WideNet for this kind of research, after which we outline how such a real-world application can help to improve EL technology in general.