Sharing Factual Knowledge from Research in Film and Media Studies by using the Structured Knowledge Base Wikidata

Claudia Müller-Birn (1), Adelheid Heftberger (2), Jackob Höper (3), and Niels-Oliver Walkowski (3)

(1) Freie Universität of Berlin, (2) Brandenburg Center for Media Studies (ZeM), Orcid 0000-0001-7909-6858, (3) Berlin-Brandenburg Academy of Sciences and Humanities

In certain fields of research such as medicine, biology or chemistry, scholarly communication has significantly changed over the past fifteen years. The scientific article — far from disappearing as many authors have suggested — turned into a complex multi-purpose object. In the humanities comparable examples are rare. Journals such as the Zeitschrift für digitale Geisteswissenschaften (ZfdG) or the TEI Journal offer at least TEI-XML versions of their articles. There is more than one reason for the slow development or even a certain resistance or obliviance towards developing new ways of making articles machine-readable. The main reason arguably is that many disciplines within the humanities regard factual knowledge as just one aspect of their research, while the text and the verbal reasoning constitute the bulk of the research (Drucker 2011; Owens 2011; Schöch 2013).

In our presentation we will show how the logics behind Nano-Publications, publications which contain factual knowledge linked with data about its source and context (Mons and Velterop 2009; Groth 2009), can also be realized by extracting factual knowledge from film and media studies' journal articles by using the open knowledge base Wikidata. Building on our use case, the Film and Media Studies Open Access journal Apparatus, we will present how a workflow for such a mediation as well as the underlying tool chain can look like. We will demonstrate how factual knowledge from scholarly articles can be semi-automatically extracted by using the semantic annotation software neonion, which allows creating triples in a non-technical way by annotating relevant parts of the text. The tool is tightly coupled both with Wikidata and property recommender system Snoopy (Zangerle et al. 2010). Snoopy supports the annotation process by recommending suitable and missing properties from Wikidata for specific items. We aim to automatically ingest the annotations into Wikidata without further effort and technical understanding by the creator. We believe that both workflow and toolchain are transferable to other use-cases.

Wikidata follows a formal and data-oriented approach to the representation of information. This is challenging especially for non-technical experts from the humanities, where the discursive function of text is a constitutive feature of information. These disciplines not only publish but also work on the basis of texts. Consequently, the usage of Wikidata in these disciplines needs mediation and support, to bridge the different thought systems and allow scholars to integrate factual knowledge from existing research results into Wikidata easily.

Our aim is to also explain how the publication of extracted factual knowledge can greatly benefit humanities research in areas like such as New Film History. Furthermore, we highlight how in reverse the data quality of existing collaborative projects, such as the Wikidata project, can profit from scholarly research and how opening up research results can improve scholarly communication.

Drucker, Johanna. 'Humanities Approaches to Graphical Display'. Digital Humanities Quarterly 5, no. 1 (2011). http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

Mons, Barend, and Jan Velterop. 'Nano-Publication in the E-Science Era'. In Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009), 2009. http://www.surffoundation.nl/SiteCollectionDocuments/Nano-Publication - Mons - Velterop.pdf.

Owens, Trevor. 'Defining Data for Humanists: Text, Artifact, Information or Evidence?' Journal of Digital Humanities 1, no. 1 (2011). http://journalofdigitalhumanities.org/1-1/defining-data-for-humanists-by-trevor-owens.

Schöch, Christoph. 'Big? Smart? Clean? Messy? Data in the Humanities'. Journal of Digital Humanities 2, no. 3 (2013). http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/.

Shotton, David. 'Semantic Publishing: The Coming Revolution in Scientific Journal Publishing'. Learned Publishing 22, no. 2 (2009): 85–94.

Zangerle, Eva and Gassler, Wolfgang and Specht, Günther. 'Recommending Structure in Collaborative Semistructured Information Systems'. Proceedings of the fourth ACM conference on recommender systems (ACM 2010): 261–264.