

# Optimized platform for capturing metadata of historical correspondences

## Andert, Martin

[martin.andert@informatik.uni-halle.de](mailto:martin.andert@informatik.uni-halle.de)  
Computer Science Department, Martin-Luther-Universität  
Halle-Wittenberg, Germany

## Ritter, Joerg

[joerg.ritter@informatik.uni-halle.de](mailto:joerg.ritter@informatik.uni-halle.de)  
Computer Science Department, Martin-Luther-Universität  
Halle-Wittenberg, Germany

## Molitor, Paul

[paul.molitor@informatik.uni-halle.de](mailto:paul.molitor@informatik.uni-halle.de)  
Computer Science Department, Martin-Luther-Universität  
Halle-Wittenberg, Germany

Various projects around the world focus on analyzing and visualizing the correspondences and social networks of scholars (academics, writers, etc), e.g., the 'Mapping the Republic of Letters' project centered at Stanford University (Edelstein et al., 2008) and the project 'Vernetzte Korrespondenzen' (engl.: Networked Correspondences) (Burch et al., 2012) run by Trier Center for Digital Humanities, German Literary Archive Marbach, and Martin-Luther-University Halle-Wittenberg. In addition to the design and implementation of adequate tools for analyzing and visualizing the big data, an indexing scheme for the content of the letters of the corpus under consideration has to be developed and the letters have to be captured, annotated, and indexed with respect to the indexing scheme.

The capturing, annotating, and indexing process should be supported by an interactive browser based platform which allows the professional philologists (as well as 'citizen scholars' from non-academic backgrounds) to quickly enter metadata (e.g., date, name/identification of senders and addressees) in an accurate und simple manner. To meet the simplicity requirement, the platform should accept unformatted entries as far as possible; to meet the accuracy requirement, the entries have to be clear without ambiguity. The two requirements appear to be competitive and conflicting. To sort out the problem, we propose to support unformatted data entry by auto-completion based on external catalogues, e.g., the 'Deutsche

Nationalbibliografie' (DNB, 2013b) provided by the German National Library. The auto-completion tool should provide additional biographical information on the proposed scholars as soon as the philologist enters first data, to help the philologist uniquely identify scholars.

In preparation for the above mentioned 'Vernetzte Korrespondenzen' project (Burch et al., 2012) which investigates the correspondences and social networks of German scholars which went into exile during Nazi Germany, we have implemented such a browser based platform.

- The platform is generic in the sense that it can be trained on different centuries by using one or several external catalogues. In principle, each catalogue which provides a web service that offers infix search of names with the possibility of limiting the search to a specified time period can be used for training the platform. The illustrations presented in the next paragraph assume that the platform has been trained on scholars of the 18th and 19th Century.
- The platform only propounds scholars based on PND/GND-normed data together with additional accurate biographical and historical data which are extracted from existing audited catalogues. These supplementary data should allow the philologist/historian to uniquely identify the scholar. After identification the system files the scholars in a standardized manner, i.e., normed data are inserted in the metadata.
- The platform provides capturing of accurate dates by nearly unformatted entries.

As soon as the philologist enters a part of the name of the sender or the addressee, e.g., 'Joseph', the tool makes proposals of scholars of the 18th and 19th Century that match the requirement, e.g., 'Eichendorff, Joseph von', 'Hübner, Joseph Alexander von' and so on, together with their dates of birth and death, identification number of the name authority file ('Gemeinsame Normdatei') published by the German National Library (DNB, 2013a), and short vitae. Especially these secondary data are of benefit if several scholars with similar (or same) name exist. Once the philologist made her choice, the system pins the scholar's biographical data to the work space. If the person is unknown to the system which may be the case if the letter was sent, e.g., to the pharmacist next door (whose name is not specified in the letter), the philologist enters this new person into the system. The autocompletion mechanism of the platform uses these new data during future requests.

The entry of the letter's date poses a further challenge as the determination of the date is harder than you think. For instance, a letter can be written over several days, the date may be hardly readable, it can be given by a

holiday, or it can be specified imprecisely. That's why our platform allows the philologist input the date of the letter in various formats. For instance, the letter can be dated 'early January 1798', 'Eastern 1770', 'mid-April 1770', '19/04/1770', or '15/04/1770-20/04/1770'. The conversion of the entry to an exact date or a period is automatically done by the underlying system. 'Early January 1798', e.g., is converted to the period 01/01/1798-10/01/1798, 'Eastern 1770' to the period 12/04/1770-16/04/1770, 'mid April 1770' to 11/04/1770-20/04/1770. After the date had been entered, the platform provides additional information: It runs plausibility checks, e.g., whether the date of the letter is in life of the correspondents, and provides information on historical events that took place immediately before or immediately after that date. For instance, in the case of Eastern 1770, the system would provide the information that the German evangelical theologian August Wilhelm Reinhart passed away on 17 April 1770 and Marie Antoinette got married per procurationem Dauphin Louis-Auguste in the Augustinian Church in Vienna on 19 April 1770. Especially, later on, such data should help the philologist annotate the letters. These latter data can be extracted from digital chronicles, e.g., the 'Chronik deutscher Zeitgeschichte' (Overesch et al., 1982) for adapting the platform to the project 'Vernetzte Korrespondenzen' (Burch et al., 2012).

If possible, we will distribute the tool in different digital research infrastructures for the humanities, e.g., DARIAH (DARIAH-EU, 2012), CLARIN (CLARIN, 2011).

## Funding

This work was supported by the German Federal Ministry of Education and Research (BMBF) [grant number 01UG1354C]

## References

- Edelstein, D., P. Findlen, and N. Coleman (2008). Mapping the Republic of Letters. Stanford University. <https://republicofletters.stanford.edu/> (accessed 15 February 2013)
- Burch, T., V Hildenbrandt., S. Kamzelak, P. Molitor, C. Moulin, and J. Ritter (2012). Vernetzte Korrespondenzen — . BMBF-Projekt 01UG1354. <http://kompetenzzentrum.uni-trier.de/de/projekte/projekte/briefnetzwerk/> (accessed 15 February 2013)
- Deutsche Nationalbibliothek (2013a). Homepage. [http://www.dnb.de/DE/Home/home\\_node.html](http://www.dnb.de/DE/Home/home_node.html) (accessed 15 February 2013)
- Deutsche Nationalbibliothek (2013b). Deutsche Nationalbibliografie. <http://www.dnb.de/>

EN/Service/DigitaleDienste/DNBBibliografie/dnbbibliografie\_node.html (accessed 15 February 2013)

Overesch, M., and F. W. Saal (1982). Chronik deutscher Zeitgeschichte. Droste Geschichtskalendarium. Droste Verlag 1982-1986.

DARIAH-EU (2012). Introducing DARIAH-EU. Information Brochure. <http://www.dariah.eu/> (accessed 15 February 2013)

CLARIN (2011). CLARIN ERIC STATUTES. <http://www.clarin.eu/external/index.php?page=about-clarin> (accessed 15 February 2013)

# An Interactive Interface for Text Variant Graph Models

**Andrews, Tara Lee**

tara.andrews@arts.kuleuven.be  
KU Leuven, Belgium

**Van Zundert, Joris Job**

joris.van.zundert@huygens.knaw.nl  
Huygens ING, Royal Netherlands Academy of Sciences (KNAW)

In the last three years there has been an increasing adoption of the variant graph as a suitable computer-internal data model for textual variation in tools and applications such as NMerge (Schmidt & Colomb 2009), CollateX (Dekker & Middell 2011), and Stemmaweb (Andrews & Macé 2013). The variant graph model can be visualized easily and succinctly; these visualizations become for the text researcher the first interface through which to approach the particular similarities and differences in a given text. In this short paper we consider from the theoretical perspective some hermeneutic risks inherent in the static visualization most commonly used at present, when the interface stands between the scholar and the data model (whether produced by a collation tool such as CollateX or NMerge, or generated from a published critical apparatus or text collation). The result of our application of the theory is the tool we present, meant to address these risks in part. Implemented in JavaScript<sup>1</sup>, the tool allows the user to interact with and modify the model of a particular text from within the browser window. By doing so the scholar can move beyond the 'black box' nature of visualizations produced by analysis tools and test alternate hypotheses for understanding a text.