### Symposium ‘Digital Scholarship and the Role of the Library’
#### September 22 13.30-18.30
University Library

**Chairperson**
**Prof. dr. Hilde De Weerdt**


---

**‘Digital Scholarship and Leiden University Libraries as a Partner in Knowledge’**
**Isabel Brouwer, MA**

Isabel Brouwer studied History of Art and Latin American Studies at Leiden University. She is Subject Librarian for Latin American & Caribbean Studies and Film and Literary Studies at Leiden University Libraries. She is managing the project ‘Data- & Text Mining’.

Digital information and data offer new far-reaching possibilities for education and research, but also all kinds of new challenges in terms of use, survey techniques, accessibility, management and archiving. As a partner in knowledge for education and research, Leiden University Libraries started several projects to prepare for these challenges.

In my talk I will discuss the project Data- & Text Mining, that started in January 2014.

---

**‘Record Linkage in Medieval and Early Modern Documents’**
**Dr. Arno Knobbe**

Dr. Arno Knobbe is a senior researcher in the Leiden Institute of Advanced Computer Science (LIACS) Data Mining group, and supervisor of eight doctoral students, postdocs and scientific programmers. His topics of research include Pattern Mining, Spatio-temporal Data Mining, Digital Humanities and Multivariate Time Series. Arno Knobbe has been involved in a large number of research projects, both nationally and internationally, and has obtained funding from Dutch funding agencies NWO, STW and SURF as well as the EU. He is currently project leader of a large project on Structural Health Monitoring of a Dutch highway bridge. Arno Knobbe was involved in the ChartEx project (intercontinental funding from Digging into Data), a large Digital Humanities project involving medieval charters, and responsible for the record linkage component. In a more recent cooperation with the UK National Archives, he is working on connecting individuals mentioned in medieval and early modern documents through natural language processing and probabilistic record linkage.

In two recent projects, we have been analysing large collections of historical texts, in order to link the individuals being mentioned in the text. Based on the descriptions in the text, their context and properties of the documents, one can often recognize occurrences of people in different documents, that appear to refer to one and the same person in real life. In the projects, we have been developing automatic techniques for recognizing such links between documents, and determining the confidence of such a link, based on the evidence provided in the text. Having automated techniques allows one to investigate large archives, which is what we did for medieval documents in the ChartEx project, and for medieval and early modern texts in the Traces Through Time project, together with the UK National Archives. In this talk, I will discuss the technical choices behind our record linkage approach, and demonstrate examples of what has been achieved.
Computational Criticism: Using Quantitative Methods for the Interpretation of Poetry

Peter Verhaar, MA

Peter Verhaar studied English Literature and Book History at Leiden University and Computer Science at the Open University. He works as a lecturer at the MA Book and Digital Media Studies at Leiden University, where he teaches courses on text encoding, database theory and text mining. In addition, he works as a senior Project Manager at Leiden University Libraries, where he focuses mostly on innovations in the field of scholarly communication, such as virtual research environments, data curation and open access publishing.

Computational analyses of texts are frequently based on prior quantifications of low-level linguistic features, such as the most frequent words or occurrences of specific grammatical constructions. Arguably, analyses on the basis of such data are intellectually remote from traditional forms of literary scholarship, which generally focuses on the description and the interpretation of aspects such as meter, figures of speech, imagery or themes. This paper presents software that was built for the recognition of a wide range of such literary devices, in attempt to align traditional practices and scholarship based on data processing. The tools enables literary scholars to explore correlations between, for instance, specific figures of speech and imagery, or to identify noteworthy occurrences of such devices within text corpora.

Datasets for Researchers at the Koninklijke Bibliotheek

Steven Claeyssens, MA

Steven Claeyssens is Data Services Coordinator at the Koninklijke Bibliotheek. He studied Germanic Philology (Ghent University) and Book and Publishing Studies (Leiden University). He will defend his PhD thesis on the Dutch publishing firm De Erven F. Bohn by the end of 2014 (Leiden University).

The National Library of the Netherlands (Koninklijke Bibliotheek) has the ambition to digitise, OCR and make available for use ‘everywhere and by everyone’ all books, newspapers and periodicals printed in the Netherlands since 1470. To achieve this goal the KB works with public and private partners, such as University Libraries, Google and ProQuest. Already in 2013, 10% of this enormous task was completed, resulting in 73 million digitised pages. Many are available via the recently launched single entry point to (re)search the various text corpora simultaneously: Delpher. In my talk I will present the full text digitised collections the KB currently has and our efforts to make these large historical Dutch text corpora available as datasets by setting up a Data Service. The service provides machine readable access to the collections and enables scholars to research entire datasets using digital research tools.

It's All About Location: Present-day and Future Digital Cataloguing of Maps

Martijn Storms, PhD

Martijn Storms MA (Arnhem, 1978) studied human geography and spatial planning with specialisation in GIS and cartography at Utrecht University. Since 2006 he is curator of maps and atlases at Leiden University Libraries. He is also project coordinator of Koeman's Atlantes Neerlandici (the bibliography of atlases published in the Netherlands) at Brill publishers and member of the editing board of Caert-Thresoor (the Dutch magazine on the history of cartography).

Leiden University Libraries hold a collection of ca. 100,000 maps, 3,000 atlases and 25,000 topographical prints and drawings. With the addition of the collections of the Royal Tropical Institute (KIT) in 2013 and Royal Netherlands Institute of Southeast Asian and Caribbean Studies (KITLV) in 2014, the size of the collection almost doubled and the emphasis was placed more on maps of the former Dutch colonies. The colonial collection of the Royal Tropical Institute is accessible with a special map viewer. All maps have been georeferenced. With an extensive register of place names you can search for a
location in Indonesia, Suriname or the Dutch Antilles. Alternative toponyms have been added too (for instance: Jakarta, Djakarta and Batavia). By linking coordinates to these place names, the application can show all maps that include the place name in a single search. The order of results is shown from large scale to small scale. You can refine them by specific periods, map types and scales. A practical feature is the possibility to navigate from sheet to sheet within the various map series. You simply click the arrow symbols in the sideline. You will also find a clickable index for map series. Part of the maps of the UBL collection and KITLV collection has been digitised too, but only accessible via a general image database without specific geographical search options. Our challenge for the coming years is to make one entrance for all digitised maps. Maps and topographical images can always be linked to a specific location. By adding coordinates to the map (georeferencing), these maps can be linked to their geographical location and even overlaid on a modern reference map like Google Maps. Some institutions already started with this new way of cataloguing maps. See for example CartoMundi, Old Maps Online and the David Rumsey Collection.

‘Text Analysis and Visualisation: an Overview of Tools’
Peter Verhaar, MA

Peter Verhaar studied English Literature and Book History at Leiden University and Computer Science at the Open University. He works as a lecturer at the MA Book and Digital Media Studies at Leiden University, where he teaches courses on text encoding, database theory and text mining. In addition, he works as a senior Project Manager at Leiden University Libraries, where he focuses mostly on innovations in the field of scholarly communication, such as virtual research environments, data curation and open access publishing.

In recent years, a large number of tools have become available for the analysis and the visualisation of large collections of digital text. As the number of tools continues to grow, it becomes increasingly difficult for scholars to choose the applications that are most suitable for their own research. To offer guidance in this field, Leiden University Libraries has developed a list of the most significant tools for the analysis of texts, together with an overview of the various functionalities that are offered by these applications. This session discusses a website on which the matrix of tools and functionalities are presented, and explains specific text analysis tasks such as clustering, concordancing, collation and categorisation.

‘Wmatrix: a Corpus Linguistics Analysis and Comparison Tool’
Morana Lukač, PhD

Morana Lukač is a PhD student and a lecturer in Corpus Linguistics at Leiden University Centre for Linguistics (LUCL). Morana is working within the project “Bridging the Unbridgeable: linguists, prescriptivists and the general public” and the title of her thesis is “Linguistic Prescriptivism and the Media”. Before joining LUCL, Morana received MA degrees in English and Philosophy from the University of Osijek (Croatia), in English and American Studies from the University of Graz (Austria) and she worked as a lecturer at the University of Zadar (Croatia).

This workshop will focus on demonstrating the functions of Wmatrix, a web-based corpus processing environment (Rayson 2009). Wmatrix is a tool for advancing the statistical analysis of electronic corpora of linguistic data, it allows the user to run corpus linguistic tools via a web browser and it runs on any computer. Wmatrix facilitates the production of word frequency lists, statistical comparisons of those lists and concordances. By integrating part-of-speech tagging and semantic field tagging, Wmatrix allows the user to identify key grammatical categories and key semantic concepts. Whether or not you have any background in corpus linguistics, this workshop is designed to demonstrate how corpus linguistic tools allow the macroscopic analysis of texts or varieties of language to inform the microscopic level as to which linguistic features should be investigated further.

‘Working with GIS in the Humanities’
Patrick Gouw, PhD

Patrick Gouw is Subject Librarian for History & International Studies. At the Leiden University Libraries he is currently implementing a project aimed at providing future services within the field of GIS. His personal interest and involvement with the subject started when working as a researcher in Ancient History.

Geographical Information Systems (GIS) have become a common feature in today’s society. Moreover, they have also been firmly established as a valuable research tool within many academic disciplines. GIS-techniques are being used to analyze and visualize research data in a spatial matter, leading to new insights and results. This workshop aims at researchers who are not yet familiar with GIS, but want to find out whether its implementation could be beneficial to their work. This session will provide a basic introduction to GIS and will show that interesting results can be gained with relatively modest means.