A Multifunctional Historical Document Research System

Eva Dyllong, University of Duisburg-Essen

In this talk, the key components of a multifunctional historical document re-
search system are discussed. An ongoing project which aims at creating a repre-
sentative corpus of documents that reflect the impact of the German philosopher
Friedrich Nietzsche in the period 1865-1945 forms the case study for the research
system.

The realisation of the system includes several working fields: the collection of
literature assets, the choice of digitization software supporting German black
letter fonts, the design and implementation of a database, the development
of fuzzy techniques for searching on documents with a non-standard spelling,
the preparation of a web-based communication interface, the implementation
of communication, annotation and visualisation tools, and the design of a user-
friendly system interface adapted for heterogeneous user group ranging from
interested amateurs to experts. Furthermore, the literature archive should uti-
lize library-oriented data standards for archival storage.

The data flow and data formats of the discussed multifunctional historical doc-
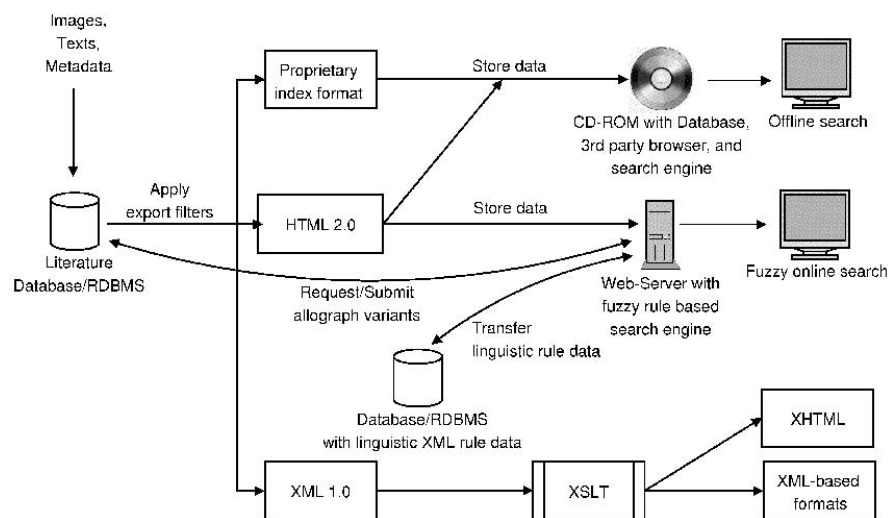ument research system is depicted in Figure 1.

Figure 1: The system architecture.

The literature sources for this project are acquired from archives and libraries,
where they are digitized and discussed via a mobile web-based expert-assisted
system. After the pertinence of the sources has been evaluated by experts,
the selected documents are completely scanned in a resolution appropriate for
optical character recognition (OCR). Finally, the scanned images, their cor-

responding text bodies, and the metadata are added to a relational database management system (RDBMS) (Biella et al 2006). Database export is achieved by modular server-based export filters, which can be customized for the desired output formats. These export filters include support for a proprietary index file format, HTML2.0, and XML1.0. Using XSL transformation (XSLT) on the exported XML data set, the content can be transformed into other XML-based markup languages such as XHTML.

In (Biella et al 2005) we described assets that enable teachers and experts to coordinate mobile literature research carried out by students or researchers in remote libraries and archives, where they digitize texts or images with a digital camera or scanner. The files are transferred to a central document server, from which they can be accessed by the tutors, who coordinate the process, advise further activities and evaluate the pertinence of the sources.

The current field of activity involves the implementation of various annotation levels that can be used for both the mobile web-based and the CD-ROM version of the digital literature archive system. Figure 2 shows an excerpt from the annotation tool.
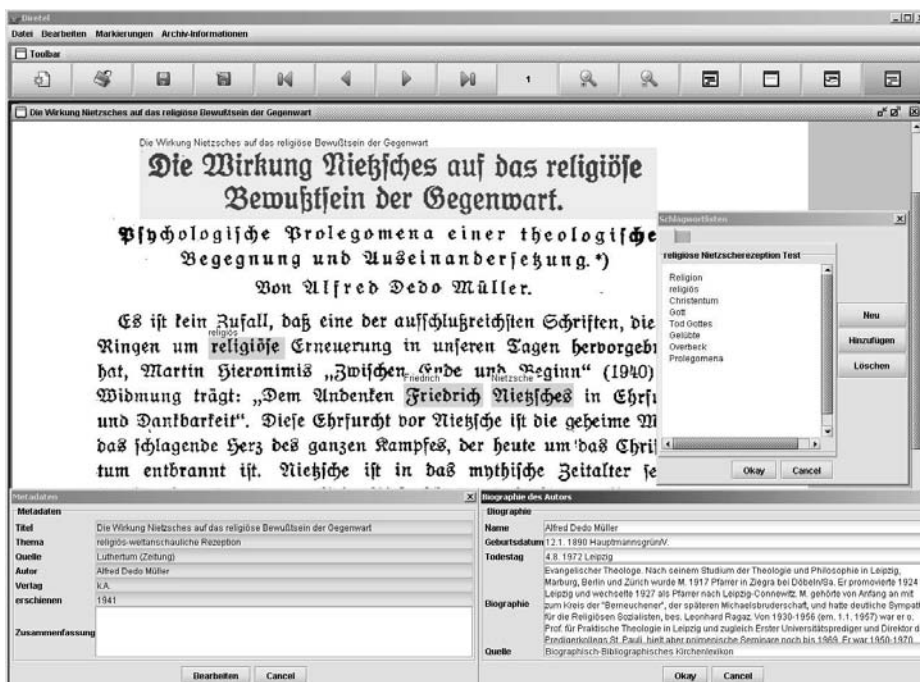

Figure 2: Annotation tool.

Moreover, within the scope of the ongoing project "Nietzsche-Archive" an appropriate visualization of the database queries is intended. To find out, for example, in which year during a particular time period the database includes

most of the documents were written, one can simply read this information from a convenient chart (see Figure 3).
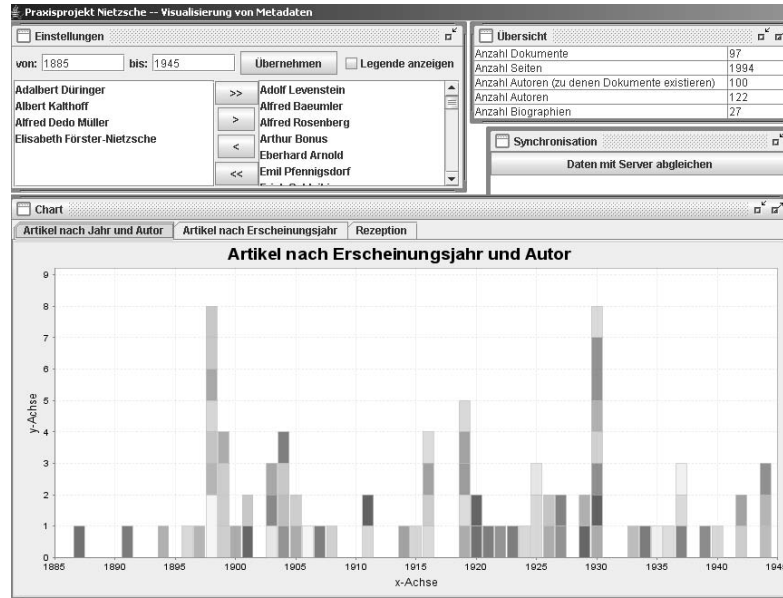

Figure 3: Visualization of metadata.

However, to complete the Nietzsche database with some context information, such as the circulation of a literary periodical, is an ambitious scheme that will require several years.

References:

[**1**] Biella, D., Dyllong, E., Luther, W., and Pilz, Th. (2005). *An On-line Literature Research System with Rule-Based Search*, In Proceedings of the 4th European Conference on e-Learning (ECEL2005), 10-11 November 2005, Amsterdam, Netherlands, pp. 67-76, 2005.

[**2**] Biella, D., Dyllong, E., Kaiser, H., Luther, W., and Mittmann Th. (2006). *Wege zur digitalen Erfassung der Nachwirkung Nietzsches in Deutschland von 1865-1945. Ein Arbeitsbericht zum Duisburger Retrodigitalisierungsprojekt* In: Zur unterirdischen Wirkung von Dynamit - Vom Umgang Nietzsches mit Büchern zum Umgang mit Nietzsches Büchern. M. Knoche, J. H. Ulbricht, J. Weber Hrsg. Harrassowitz Verlag Wiesbaden, pp. 187-199, 2006.