

Contributor/Presenter:

Günter Mühlberger, Innsbruck University
Department for German Language and Literature
Digitisation and Digital Preservation

Title:

Structural metadata – a Key for Indexing Digitized Newspapers

Abstract:

Though newspapers are among the most frequently used materials in libraries, their recording in a library catalogue is rather shallow. Apart from general information about the (changing) titles and (changing) publishers of newspapers as well as missing volumes and issues the information in usual library catalogues is tenuous. In rare cases special libraries or documentation centres maintained newspaper bibliographies or news clipping archives collected articles for special topics but in general the indexing of newspapers is poor.

With the digitization of newspapers this situation changes dramatically. Now automated processes are possible such as Optical Character Recognition which provide on the one hand a full-text index for information retrieval, as well as segmentation information about newspaper pages containing physical (blocks, lines, strings) and logical (headlines, articles, by-lines, captions) units. Taking this into account it seems inevitable that current metadata schemas need to be extended towards a more detailed and fine grained level. The need for a common understanding of “what can be found in a (historical) newspaper” and “how can it be named” is therefore of eminent importance for improving access to digitized newspapers.

This paper will be a first attempt to summarize the discussion on this topic and to provide a comprehensive proposition how these questions could be answered. It will introduce several concepts with which structural metadata can be recorded, may it be automatically, manually or with the support of crowd users. The paper will also provide criteria how to apply the newly introduced concepts based on examples from several European newspapers and epochs. The paper is an outcome of the ICT-PSP Project Europeana Newspapers. A Gateway to European Newspapers.