2008

Supporting Research in Historical Archives: Historical Information Visualization and Modeling Requirements

Katifori, Akrivi

IEEE Xplore

http://hdl.handle.net/11728/6476

Downloaded from HEPHAESTUS Repository, Neapolis University institutional repository
Abstract— The on-going progress in the area of digital libraries has lead to the beginning of a digitization effort in Historical Archives, as well. The requirements of historical research, which works with histories of entities and incomplete information, create the need for supplementary tools to support users in handling the digitized content. This work is based on a user study of historian information retrieval methods in order to create a set of tools for the context of historical archives, which will facilitate historical data storage, management and visualization.

Index Terms— historical archive, visualization, information retrieval, ontology, heuristics

I. INTRODUCTION

The recent progress in the area of digital libraries and the semantic web has lead to new ways of digitizing, organizing and presenting library material, enhanced with the incorporation of semantics. More and more organizations, libraries and document repositories opt for digitizing their material, either for internal use or for publishing it through the web. The great variety of digitized material has brought new needs and several research issues have arisen.

Digitized historical archives (HAs) could be considered as a special case of digital libraries; they have however, characteristics that differentiate them. In particular, the digitization process in the context of HAs is inherently more demanding than the equivalent in common digital libraries, mainly due to the large volume of the original material and its poor preservation state (at least for some portion of the material), as well as to the convoluted and archaic handwriting often found in documents of HAs. At the best case, keywords or other metadata (creation date, author etc) will be available [3]. Commonly, documents in a HA are fitted into a categorization scheme, which has proven to provide little or no help at all for information retrieval purposes [4] [7], as it is typically compiled by archivists to suit archiving purposes. As a result, even browsing becomes very difficult without the help of the experienced archive personnel, which mainly relies on their conceptual model of the archive, rather than on some explicit representation of knowledge about the archive content and tools offering guidance and automation for search tasks. Time-varying information and entity evolution are also frequently encountered in historical archives, and the modeling, processing and visualization of these aspects are not adequately addressed insofar [8] [9].

As HAs constitute a very important source for historical research, in this work we attempt to investigate the historical researchers’ needs and propose a set of tools to assist them in their research in digitized archives. These tools are based on an ontology [6], a construct that presents an overview of the domain related to a specific area of interest and may be used for browsing and query refinement. The proposed set of tools enables the user to visualize temporal information, indicating which data varies with the passage of time and enabling users to efficiently track entity evolution along the time axis. A set of heuristics is also made available to assist researchers in locating ontology entities that correspond to different periods of the same real-world entity.

The rest of the paper is organized as follows: Section 2 presents a user study of historian search methods, section 3 proposes a set of tools to support information retrieval and, finally, the last section concludes the paper and outlines future work.

II. A STUDY OF HISTORIAN RESEARCH METHODS

Historians and researchers collect and process historical data in order to document historical facts and produce information connecting them. Their main objective is to recreate the past through existing records and their interconnections. The collection of historical data is accomplished through methodical and comprehensive research in primary and secondary sources. Primary materials, which include the remaining records of archives, mail, books, etc of the time period of interest, are of special importance to historians as they constitute the basis for original historical research.

Tibbo in [7] presents the preliminary results of a user study concerning the way historians locate primary resource materi-