
Preface

The key to understanding the breathtaking development of mankind lies in the ability to objectify knowledge. Just like human knowledge has undergone changes, so have the libraries in their role as treasure chambers of said knowledge. With the advent of the digital age, traditional libraries were complemented by huge collections of digital documents. These digital libraries, or e-libraries, result from the fact that knowledge can nowadays be objectified in the form of multifaceted digital copies. Relatively speaking, we are still at the dawn of the digital age—let us not forget that the Web is only three decades old—and yet, the documents that have been collected in a digital library display an astonishing range in terms of multimedia feature diversity: traditional texts, photos, and graphics, recorded audio files (e.g., conversations, music, and sounds of all kinds), moving images (e.g. video sequences, films, or live recordings of events), to name but a few. Future developments will certainly lead to further forms and formats; it is therefore safe to say that the portfolio of digital documents will increase exponentially, and so will the number of digital libraries, consequentially.

As a result of this development, modern users are confronted with new challenges that come in the form of explosions of objectified knowledge caused by digital technology. How does one manage to find the most relevant documents about a certain topic? Who is going to assist the user with this procedure? In traditional libraries of the pre-digital era, one could address the librarian and ask for human assistance. The librarian knew the order by which the book inventory was arranged and could thus give the required information without being an expert on any of the subject matters. When looking for a similar kind of support when using a digital library, one finds first and foremost high-performance keyword search engines, mostly Google, which, despite their considerable capabilities, exhibit limitations when it comes to interacting with the user; their user interfaces have generally not been optimized for this particular task, and they usually suffer from an inability to put the user's query into a precise context.

When using such a search engine, one has to enter individual keywords and then obtains a list of digital documents as a result. These documents have previously been indexed by the search engine and have been marked as highly relevant to the query, merely due to the high occurrence of input keywords. In contrast to the assistance offered by a librarian in former times, one has to rely on the mere use of keywords, knowing that a search engine cannot differentiate between the literal meaning of a keyword and the context respectively subtext of that search item. This book examines the question whether this has to be accepted as a given fact. It describes a vision and identifies ways to develop an electronic service—the E-Librarian Service—that provides users of digital libraries with the service they are accustomed to from a time when they were dealing with a traditional librarian.

To design such an E-Librarian Service, three key components are required. Firstly, an E-Librarian Service needs to “understand” the meaning behind user needs on top of the actual meaning of the search item. In this field, there are several emerging technologies regrouped under the term “Semantic Web and Ontologies” which have the potential to develop into one of the key technologies of the 21st century. These technologies are able to give the users a feel for the importance and semantic meaning of the documents of a digital library and, at the same time, convey the message that the supporting technical systems actually “understand” their needs.

Secondly, efficient technologies are needed so that an E-Librarian Service will be able to provide the most relevant documents systematically for a search topic. The currently developing “Multimedia Information Retrieval” systems give cause for hope that it will be possible to find documents efficiently and reliably, even in digital libraries that largely exceed actual libraries in size and capacity.

Lastly, the users need a human-friendly interface for digital libraries in order to be able to enter their queries in a natural language, i.e., a language they are accustomed to as opposed to an artificial computer language. In fact, there are currently developments in the field of information technologies that make it seem realistic that natural language questions can be processed and “understood” properly by an IT system. The required technologies are referred to as “Natural Language Processing” and “Description Logics and Reasoning”.

As a matter of fact, the first implementations of such E-Librarian Services are already being used. As part of his Ph.D. thesis, one of the authors of this book, Serge Linckels, developed an E-Librarian Service called CHESt which high-school students can use to explore knowledge in the field of computer history. CHESt is able to find a clip or several clips that correspond to the queries, in this case complete questions in German or English. The underlying foundation is a digital library that consists of more than 300 video clips that were taped with the tele-TASK system that was developed by Christoph Meinel, who is this book’s second author, and his team. Research has proven that CHESt was able to retrieve videos more accurately and precisely than

the use of pure keyword search engines would have been able to do. Apart from CHESt, other E-Librarian Services have been developed and tested, such as MatES and the Lecture Butler, and despite the fact that the underlying concepts might remind us of science fiction films, research in this field shows that those E-Librarian Services could eventually become reality.

However, before we introduce the basic ideas regarding the concept and structure of E-Librarian Services and their first implementations, such as CHESt, MatES, or Lecture Butler, we are going to introduce the reader to the underlying technologies, such as “Semantic Web and Ontologies”, “Description Logics and Reasoning”, “Natural Language Processing” and “Information Retrieval”, which are of great interest in themselves. All these technologies have the potential of becoming key technologies of the 21st century and will be dealt with in individual chapters. We greatly enjoyed discovering how these technologies that are currently being developed independently from each other can be combined to create unexpected new possibilities to design services and thus provide new and astonishing results through their application. We hope that the interested reader will share the fun we had and perhaps feel inspired to become creative, too.

We would like to thank the staff of the chair “Internet Technologies and Systems” of the Hasso-Plattner-Institute in Potsdam for their contributions to this book. We would also like to thank Michel Bintener and Sabine Lang for editing, and everybody else who contributed to this book with their comments, suggestions, and reviews.

Serge Linckels and Christoph Meinel
Steinfort/Potsdam, April, 2011



<http://www.springer.com/978-3-642-17742-2>

E-Librarian Service

User-Friendly Semantic Search in Digital Libraries

Linckels, S.; Meinel, C.

2011, XVI, 212 p., Hardcover

ISBN: 978-3-642-17742-2