

OMAR - An online database for oriental manuscripts

Matthias Brückner^(*) and *Paul-Thomas Kandzia*^(#)

^(*) Oriental Seminar, Freiburg University, Germany

E-mail: m.brueckner@cyberfatwa.de

^(#) Institute of Computer Science, Freiburg University, Germany

E-mail: kandzia@informatik.uni-freiburg.de

ABSTRACT

OMAR (Oriental Manuscript Resource) [11] is a database which provides extensive search facilities for oriental manuscripts. It is currently based on the collections of the Oriental Seminar at the University of Freiburg which comprise about 2.500 manuscripts. Images of the manuscripts are stored in OMAR together with structured data containing information about bibliographic aspects, the author, the title etc.

KEYWORDS: database, orient, manuscripts

INTRODUCTION

Over the last decade information technology, especially databases and the internet, has become an efficient and indispensable tool for scientific work. Nowadays, in technology related disciplines, e.g. Computer Science, huge amounts of information exist (often purely) in digital form. However, in the Humanities the potential of information and communication technology is not exhausted by far. In the following we focus on the handling of texts and manuscripts in Islam Science. Here, the advantages of digital resources and access via the internet are obvious, since the community is dispersed all over the world, and copying and distributing their "working

material" is not an easy task at all. Still, it is a quite usual practice that researchers travel far in order to examine a certain book. Furthermore, often appointments are necessary to get the permission to lay hands on a valuable worm-eaten loose manuscript. Surprisingly, though several library catalogues are accessible online, there is no digital resource which provides extensive search facilities of the metadata (that is catalogue-like data about the documents in the library) as well as for the texts or manuscripts themselves. Our project is a first step in order to change that situation: We propose the prototypical core of a database which stores images of historical Arabic manuscripts together with the scientific information known about them. As a special feature there is a flexible and comfortable Web-interface. Furthermore, we point out some very specific problems, which differentiate the database modeling from well-known areas as (occidental) business databases.

ONLINE RESOURCES OF ISLAM SCIENCE

In the 1990s Turkey, Tunisia and Iran already had been connected to the Internet. The impact of the Internet on the Orient grew bigger just since the

mid-nineties. Now, with Saudi-Arabia and Syria getting connected, most of the Islamic Orient is online [10], and most of the scientists of Middle East Studies and Islam Sciences are using the Internet, too. So, how about the mass of Oriental manuscripts lying in the sand waiting for discovery? Digitalisation in this area is still exceptional, but of growing interest.

The American University in Cairo Library is providing a catalogue which could be displayed in Arabic with the help of Unicode and MS Arabic Support¹, however, it is not possible to get access to the texts. The database of one of the most famous fatwa²-online-services can only be searched by title, keyword, mufti, and topic [6]. The database of another important fatwa-service is just searchable by number, date and keyword [7]. The latter services belong to the main resources of modern authentic Islamic texts online [2, 3]. Unfortunately, those texts are rather short. The Qur'an and Hadithantologies³ which can be searched by keyword are frequently found online, e.g. at *Musalman* [8]. The German *Orient-Institute* is providing a database with 1.380 texts [4], but these are only secondary texts about modern Oriental history. The largest recent book-collecting activity in Germany is the *Sondersammelgebiet Vorderer Orient/Nordafrika* (Near East/North

Africa), which moved from Tübingen to Halle. Both collections can be accessed by a virtual library catalogue [13]. Again, there is no access to the texts. The same applies to the *Manuscripts' Catalogue* of the *British Library* [1]. Using a search machine with keyword "Arabic language", 204 resources were listed when searching, but none of them provides direct access to the texts. A small collection with some pictures of texts is accessible at the homepage of the *Islamic medical manuscripts at the [US] National Library of Medicine* [9]. In summary, one can say that surprisingly few Arabic texts are searchable and accessible through the Internet, and that the advantages of database technology seem to be not yet fully recognized by the community. The services which provide texts are limited in their scope, e.g. fatwa-online-services, and limited in their search facilities. For the time being, there is no database which allows for direct access to full texts through enhanced search-facilities and is open to a broad scope of Arabic texts.

THE OMAR PROJECT

OMAR, a joint project of the Oriental Seminary and the Institute of Computer Science, Freiburg University, has begun to close the gap described above. Starting point is the collection of the Oriental Seminary in Freiburg, consisting of about 240 original manuscripts of various origins and about 2.300 Mauritanian manuscripts [12] as microfilms. These comprise about 279.000 pages altogether. This material and knowledge about it is successively put into a database and made public over the Web. Already, a prototype is working, which we will describe in the following.

¹ MS Arabic Support is a package of arabic True Type Fonts, which is quite common in use for arabic websites.

² Islamic legal ruling, which consists of a question posed by a layman and an answer given by an expert.

³ Hadiths are traditions of the prophet Muhammad and the second important source of law after the Qur'an.

aspects. Marginal notes might tell historical stories. Important information can be concluded from the formal aspects of the manuscript, e.g. by knowing which paper or ink-colour has been manufactured and used in which time and place. Many such bibliographical aspects as well as certain relationships between persons, e.g. teacher-pupil or father-son, have been included in the database.

Altogether, the relational schema of OMAR is fairly complex, comprising 55 tables. Taking into account the growing user experience, extensions will follow in the future.

Technical Issues

The prototype has been realised with Oracle 8i on a Sun Ultra 5, using PHP 4 for the web interface [5]. Three types of users are distinguished: guests, normal users and administrators. A guest views the database as a black box, getting access via several search-facilities. It is also possible to display lists of the manuscripts ordered by title, author, keyword, signature or basmala⁴. Normal users may view schema information and even use arbitrary SQL-expressions for searching, however they have no writing permission. The latter is left to administrators.

Arabic Text Values. The values of some attributes (e.g. basmala) must be capable to contain Arabic text. However, a database has to support alternative modes of those values, as Original Arabic, transliterated Arabic

(using an international alphabet, also available as True-Type font), and an ASCII-representation. At present, OMAR uses transliterated Arabic and the conversion of transliterated Arabic according to TUSTEP (Tübinger System von Textverarbeitungs-Programmen)⁵.

Sorting. Of course, a database often has to output sorted lists. However, what is the meaning of sorted, if transliterated Arabic text values are concerned? For the time being, OMAR uses the alphabetical order of the internal ASCII-code. Of course, this is a somewhat arbitrary solution. On the other hand, there is no standard sorting order for transliterated Arabic, so this question is still open.

CONCLUSION AND OUTLOOK

At present, the metadata about the manuscripts in Freiburg have been imported into the database. The import of the images of the manuscripts will follow soon. Then, OMAR will start its work under real world conditions for German Islam scientists. The growing experience will be used to enhance the data model and search facilities. However, in order to be useful for an international community, a multilingual approach is essential. In particular, besides English, Arabic should be supported. Since the MS Arabic Support fonts seemingly are emerging as a standard, the Arabic extension will be based on it. During the work on OMAR we found quite few activities in this area. We hope that our project helps to raise a fruitful discussion on database and web technology for Islam Science.

⁴ Basmalas are innovative Arabic blessings at the beginning of a book, which allows to determine it. Often, there is no title written on Arabic manuscripts.

⁵ TUSTEP is a mighty word-processing-system based on FORTRAN.

REFERENCES

1. British Library
http://molcat.bl.uk/msscat/INDEX.ASP
2. Brückner M. *Cyberfatwa*,
http://www.cyberfatwa.de
3. Brückner M. Der Mufti im Netz.
In: Lohlker R., et al. *Islam im Internet. Neue Formen der Religion im Cyberspace* Deutsches Orient-Institut, Hamburg, 2000, CD-ROM, Version 1.0
4. Deutsches Orient Institut
http://www.deutsches-orientinstitut-datenbank.de
5. Fleig D. *Eine Datenbank zur Verwaltung orientalischer Handschriften*,
http://sansibar.informatik.uni-freiburg.de/fleig/diplom/html/index.html
6. Islam Online *http://www.islam-online.net/fatwa/arabic/searchFatwa.asp*
7. Islam - Question & Answer
http://www.islam-qa.com/search.shtml
8. Musalman
http://www.musalman.com/search
9. National Library of Medicine
http://www.nlm.nih.gov/hmd/arabic/arabichome.html
10. Network Startup Resource Center
http://www.nsrc.org/
11. OMAR (Oriental Manuscript Resource) *http://www.uni-freiburg.de/islam/db/*
12. Rebstock U. *Sammlung arabischer Handschriften aus Mauretanien*, Otto Harrassowitz, Wiesbaden, 1989
13. Virtueller Katalog Sondersammelgebiet Vorderer Orient/Nordafrika,
http://www.ubka.uni-karlsruhe.de/hylib/vk_ssg_vo.html

ABOUT THE AUTHORS

Matthias Brückner worked from 1996 to 2000 as system-administrator in the Oriental Seminar at the Freiburg University. Since 1999 he builds up an information website about fatwas (<http://www.cyberfatwa.de>). In 2000 he received the Master of Arts (M.A.) in Islam Science with a work about the distribution of fatwas about alcohol in the internet. Currently, he is working at the OMAR-database,
E-mail: m.brueckner@cyberfatwa.de

Dr. Paul-Thomas Kandzia has done research in the area of data models and object-oriented deductive databases. Currently, he is working as project manager of VIROR, the Virtual University of the Upper Rhine Valley (<http://www.viror.de>). VIROR is taking first steps to enhance the traditional lecture room scenario with more flexible computer supported settings,
E-mail: kandzia@informatik.uni-freiburg.de