DESIGNING AN INTERFACE FOR MULTILINGUAL AND MULTISCRIPT DATABASE OF MEDIEVAL MANUSCRIPTS

S M Shafi*

&

Nadim Akhtar Khan**

ABSTRACT

The paper discusses the design of an interface for developing a database of 'medieval manuscripts' for accommodating data elements in multilingual and multiscript medium especially in Arabic, Urdu, and Persian using 'Unicode character sets' with Visual Basic in front end and Ms Access in back end.

KEY WORDS

Unicode, Multilingual Interface, Medieval Manuscripts.

INTRODUCTION

Museums, archives and libraries collect information sources in a variety of rare and fragile media and many institutions and libraries have undertaken initiatives to exploit the potential of digital technology for archiving, retrieving and displaying such unique and fragile materials. Medieval

* Professor, DLIS, The University of Kashmir, Srinagar-J&K 190006

e-mail:shafi_sm2000@yahoo.co.in

** Faculty, DLIS, The University of Kashmir, Srinagar-J&K 190006

manuscripts, rich sources of culture and history, are available in different formats like parchment, vellum, palm leaves, paper, birch bark, papyrus, wood, stones etc., languages and scripts in various collection sizes and conditions. The advancement in information technology have made it possible to preserve, organize and disseminate them in more economic and effective way. With fast developments in software and database management systems to design a database in various languages/scripts has become more convenient and possible. It is with this background that present paper reports designing of an interface for a multilingual database of medieval manuscripts.

PROBLEM

a) Most interfaces are usually in English language. Therefore, exploring possibility of developing an interface for non-roman language and script has become necessary.

b) The absence of interface in non-roman scripts especially 'Urdu language' creates problem for accessing, editing, retrieving information to many 'Urdu language' literate users.

c) The non-roman script languages are written and accessed from right to left but interfaces are mostly available for left to right script languages/scripts.

SCOPE

The present study, besides English language, is confined to non-roman languages particularly Arabic, Persian and Urdu languages/scripts.

OBJECTIVES

The following objectives are laid down for the study:

a) To design an interface for multilingual /multiscript database of medieval manuscripts.

b) To populate data for evaluating the interface for addition, deletion, edition and retrieval of non-roman records.

LITERATURE REVIEW

The literature surveyed identifies a number of papers on Unicode standard but a few on manuscript interfaces. **Chandraker (2004)** points out that issue of localization of library and museum databases has surfaced with the advent of digital libraries and their interoperability. He argues for Unicode as a multilingual standard and discusses the related technology available for localizing Indian language materials.**Tull & Starley (2003)** support the use of Unicode as a standard for global exchange of textual information with facility for display and searching of information. The concept is further extended by **Jilvoskey & Cunningham (2005)** from a practical and technical perspective especially in multilingual cataloguing with more emphasis on issues like storage and conversion, data input and display.**Chandraker (2002)** has highlighted the technological perspective of digitization of valuable resources like manuscripts with more emphasis on multilingual database creation. **Michos ,Stamatatos & Fakotakis (1999)** have made a case for making full advantage of Artificial Intelligence based methods for digitization by putting forward a multilingual information system consisting of an integration of new and

existing multilingual information tools with high user friendliness, low cost maintenance and improved search results.**Shafi (2004, (a))**further supports the use of modern information tools for digitization of medieval manuscripts for retrieval purposes, indulging into different techniques and aspects of architecture of digital library and its implementations.**Rampler and Calabretto(1999)** demonstrates the use of active X ,CGI and java based approach for designing a Virtual library for manuscripts. **Shafi (2004, (b))** extends the concept by highlighting the importance of digitization of manuscripts by focusing on technical and planning issues of digital process and role of metadata. **Cathro (2001)** emphasises the need for managing traditional and digital collections including texts in different scripts in an integrated manner with an aim for providing persistent access to digital information.

METHODS AND MATERIALS

Visual Basic 6.0. is selected as a front end and MS Access as a back end or data base store keeping in view their obvious and well known benefits in data processing applications associated with the use of DBMS technology and associated techniques such as entity relationship modeling. Unicode, a universal character set for expressing the world's languages and character scripts is used to enable the non-roman script as well as Bi directional (BIDI) feature.

The following procedure was adopted for creating the application:

- Case studies regarding various issues related with the creation of digital Libraries of Manuscripts were thoroughly studied especially

Digital Library for Ottoman Manuscripts and C-DAC (Bangalore) and European project BAMBI.

- A thorough study for providing multilingual approach to the manuscript database was carried out and various issues concerning the design of a multilingual interface were resolved.
- Microsoft Access is used as it can take an approach based on OLEDLL, which helped to read the appropriate strings from the resource file.
- Every control in the system include text Labels, Combo-boxes, Text boxes, command buttons. Buttons with text is given a unique identifier , used as the key while linking up the actual text in the database.
- The data entry forms are created with metadata elements (fields) namely Creator, Title, Contributor, languagestyle/script, subject, Date of transcription, size, Folios, Lines, Media, Notes, Institutes, accession number, see also, sources, state.
- A user friendly front end interface is developed with proper built in message facility for interaction with user navigation, addition, deletion and modification of records made possible by providing command buttons.
- Programming for each control is written for the proper functionality. Security is ensured for data entry or manipulation so that only authenticated users can add or delete data.

- Data was added, saved and manipulated on many occasions and tested for various operations for satisfaction of the basic purpose of developing the interface.
- Search for various field items was carried by choosing the desired option and the results were analysed for their relevance.
- Connectivity with Ms Access provided easy methods of accessing various fields by binding them with the help of ADODB control. This reduced the redundant coding to greater extent.
- The application was tested for Storing and manipulating Unicode strings.
- The application was tested for handling Unicode-conformant collation, normalization, and text boundary.
- Display, print and editing of Unicode text was carried out.
- The handling of BIDI displays i.e. Right-to-left characters were tested at several occasions.
- The handling of character shaping for scripts of Arabic and Urdu was confirmed.

METADATA SET FOLLOWED

The metadata is an issue special and unique in many respects to medieval manuscript. There cant be Consensus about the elements to be included in the database, though many projects use metadata which range from TEI to MARC21.The Dublin Core Metadata Element is found useful as a kind of "Lingua Franca"supporting the discovery of information

resources from a variety of communities.Cathro **(2001)**. The present Initiative has followed Dublin Core (DCMES, 1999) which qualifies for the increasing the richness and precision of the description Baker, 2000, Sugimoto et al, 2001 **(as cited in Shafi, 2004)**

FINDINGS AND LIMITATIONS

The usefulness and uniqueness of interface is vividly clear from following seven figures. The fig 1.1 displays opening window where one can choose menus like data entry in different languages or search options to proceed further. The fig 1.2 and 1.3 demonstrate interface for data addition, edition and search facilities in English language for Medieval Manuscripts. The similar pattern has been followed for non-roman scripts, using appropriate terminology in Urdu language for data entry. The fig 1.4-1.6 displays submenus for addition, deletion and saving of records in Urdu language. The last window depicts retrieval mechanism in Urdu language.

Fig 1.1 Window displaying menus for data entry in different languages

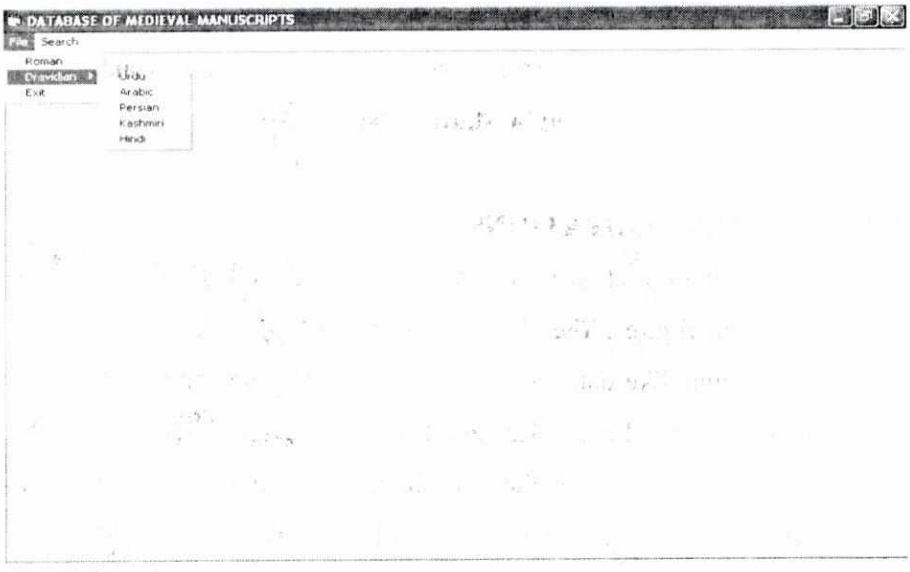


Fig 1.2 Window for data entry in English language.

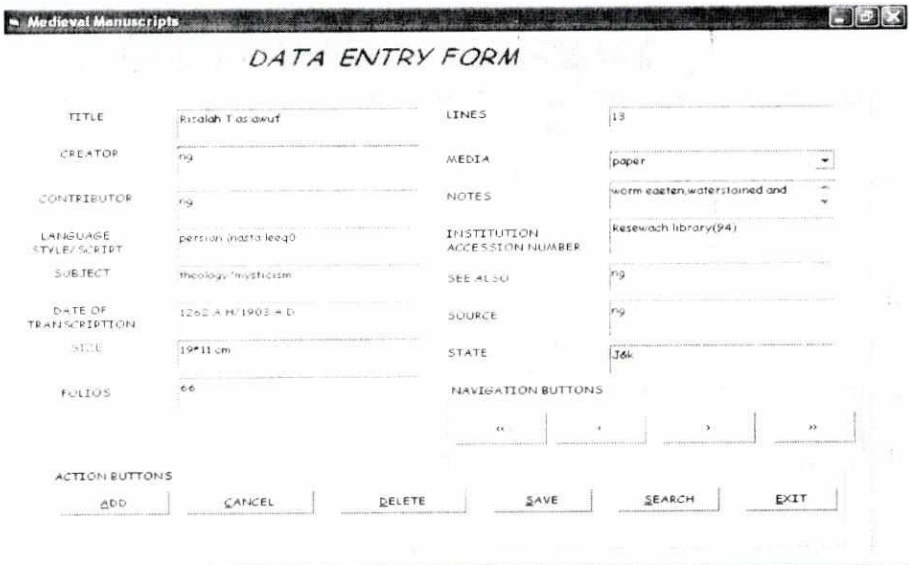


Fig 1.3 Window displaying results searched under Creator.

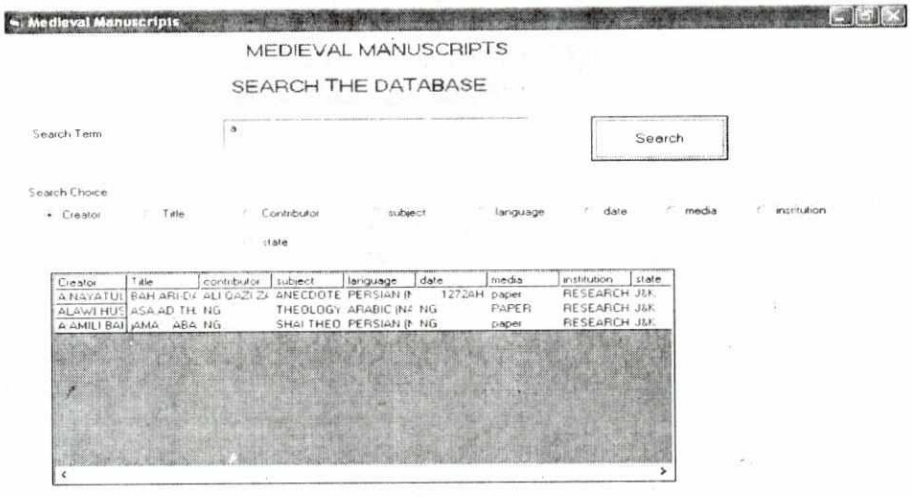


Fig 1.4 Window for Data Entry in Non Roman Languages (Urdu)

Fig 1.5 Window displaying the process of saving a record in the database.

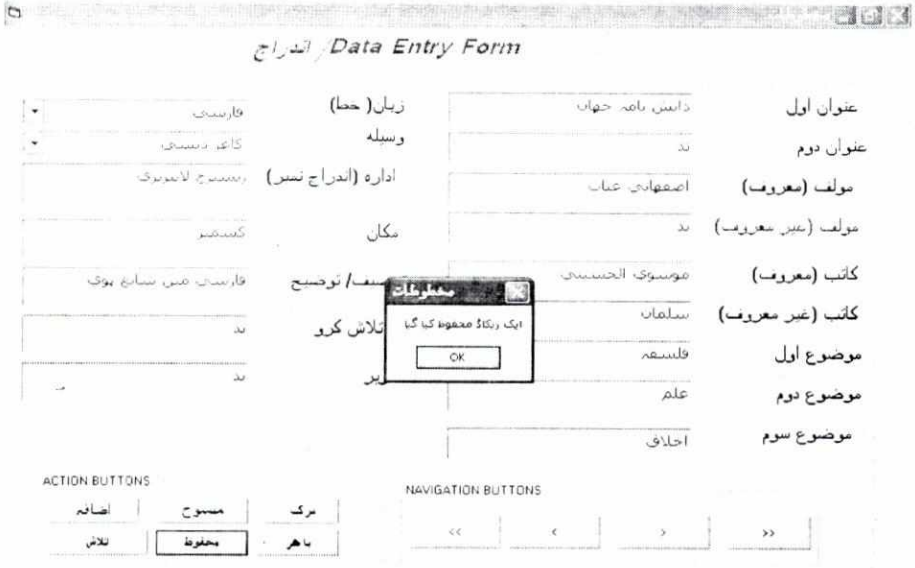


Fig 1.6 Window displaying the process of deletion of a record in the database.

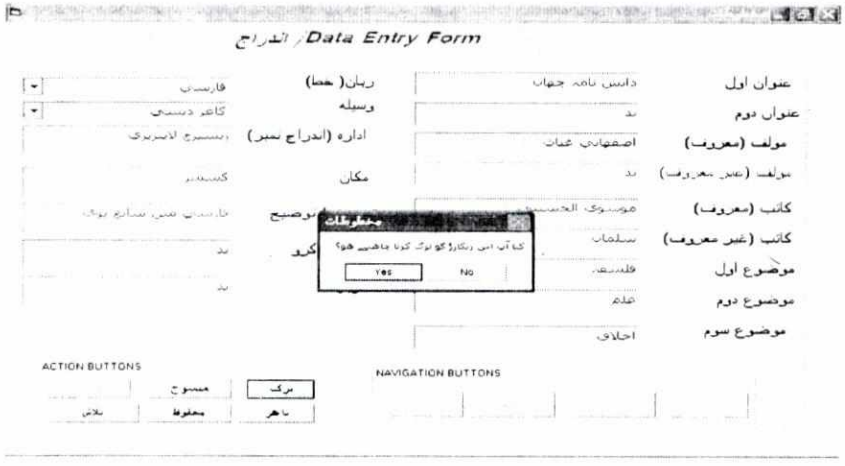
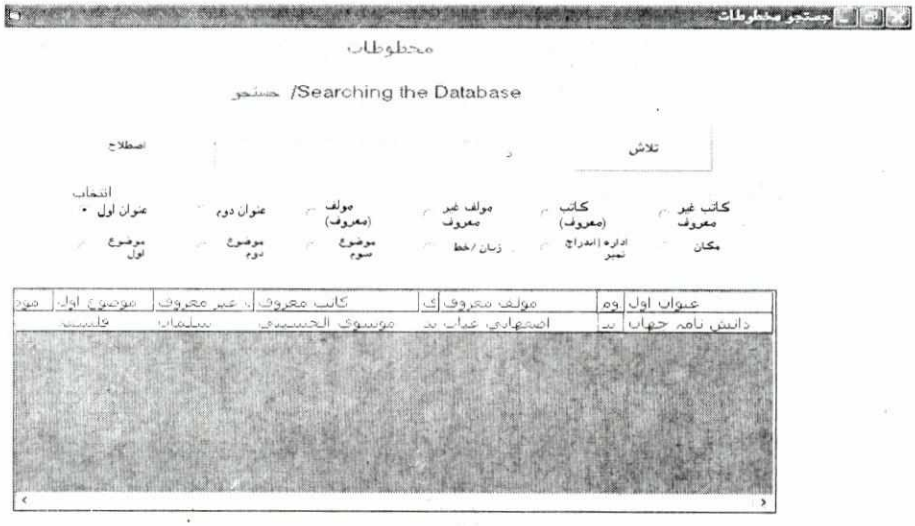


Fig 1.7 Window displaying results searched in Urdu language.



The design of the multilingual interface for medieval manuscripts exhibits the exploitation of the opportunities offered by the Information Technology without any constraints for the users with few limitations. However, the application software can be further modified using My SQL as backend and Java or ASP or VB.net as front end that will allow the system to handle bulk of data in a convenient manner.

The prototype digital interface is a unique as it handles BIDI display i.e. Right-to-left characters as application softwares running on real environment are still uncommon or unknown. Thus the project has enough scope for development on a wider canvas to make it more useful,

effective, integrated and available for a variety of systems operative in different cultural and educational settings.

REFERENCES

Cathro, Warchick S. (2001).Visions of Fundamental change in libraries and librarianship for Asia Pacific. *Library Review*, 50(7/8), 334-342.Retrieved May 14, 2007, from http://www.emeraldinsight.com/Insight/viewPDF.jsp?Filename=html/Output/Publishd/EmeraldFullTextArticle/Pdf/0350500701.pdf

Chandraker, Rajesh. (2002).Multi-Script bibliographic database: an Indian perspective. *Online Information Review*, 26(4), 246-251.Retrieved April 13 ,2007, from http://www.emeraldinsight.com/1468-4527.htm

Chandraker, Rajesh. (2004). Unicode as a multilingual standard with respect to Indian languages. *The Electronic Library*, 22(5), 422-424.Retrieved April 7 2007, from http://www.emeraldinsight.com/0264-0473.htm

Jilovsky, Cathie., & Cunningham, Andrew. (2005). Multilingual Cataloguing: practical and technical issues. Retrieved April 23, 2007, from http://www.ifla-stockholm2005.se/pdf/IFLA%20Multicultu%C9005%20CJ%20&%20AC.pdf

Michos, Stephanos., Stamatatos, Efstathios.,& Fakotakis, Nikos.(1999).Supporting Multilinguality in Library Automation Systems Using AI Tools. *Applied Artificial Intelligence*,13(7),679-704.Retrieved May 12, 2007, from http://www.wcl.ee.upatras.gr/ai/papers/michos3.pdf

Rumpler, Beatrice., & Calabretto, Sylvie.(1999).Virtual library for ancient manuscripts. *Library Review,* 48(4), 181-186.Retrieved April 30, 2007, from http://www.emeraldinsight.com/10.1108/00242539910276460

Shafi, S.M. (a) (2004).Digitization Perspective of Medieval Manuscripts. Presented at 2nd Convention PLANNER - 2004, Manipur Uni., Imphal, 4-5 November, 2004.Retrieved April 12, 2007, from http://dspace.inflibnet.ac.in/bitstream/1944/415/1/04Planner_1.pdf.

Shafi, S.M. (b) (2004).Makhtootat: Multilingual and Multiscript Digital Library of Medieval Manuscripts Initiative Presented at International Conference on Digital Libraries. (pp 442-448).

Tull, Laura., & Straley, Dona. (2003).Unicode: Support for multiple Languages at the Ohio State University Libraries. *Library Hi Tech,* 21(4), 440-450.Retrieved April 15, 2007, from http://www.emeraldinsight.com/0737-8831.htm