Building an Arabic Digital Collection: The IIUM Library's Experience

Sarifah binti Abdullah

Library, International Islamic University Malaysia (IIUM)
Jalan Gombak, 53100 Kuala Lumpur, Malaysia
`sarifah@iiu.edu.my`

Abstract. The IIUM Library's mission is to be an excellent information centre providing information to meet the learning, teaching, and research needs of its customers. The Library gives special attention to Islamic studies and its information resources are predominantly in English and Arabic. To make the resources readily accessible "anytime" and "anywhere" heavy investments were made in a digital library system that implements the latest ICT advancement. The need for Arabic language capabilities has always been one of the Library's major requirements. This paper shares the Library's experience in implementing a digital library system, focusing on the challenges in handling Arabic language information resources.

1 IIUM Digital Library

The IIUM digital library is called Media on Demand (MOD) system. It comprises of the video on demand (Quick Video on Demand), content management and taxonomy management (Multimedia Object Manager (MOM)), Web application (eQUIP) and security and access management (Membership Management) systems. The MOD system runs on three IBM servers (IBM XSeries 365 with 2GB RAM, 80GB hard disk and 2X Intel Xeon 2.4 Processor).

Digital Resources: MOD acts as a repository for all types of digital material and provides public access to the digitized collections through web-based search and browsing facilities. Due copyright issues, the Library decided to focus and digitize the University's resources including theses, exam papers, videos, books, articles, manuscripts and contents pages.

Workflow: All resources are processed in accordance with the set standard workflow which are digitization (scanning, conversion to PDF format, book marking, water marking and security setting), optical character recognition (OCR) processes, metadata creation, quality checking and publishing through the web. These tasks are carried out by the Digital Library Services department.

Taxonomy: The library has decided to establish a corporate taxonomy, since we are forecasting the current digital library system as the university's repository knowledge bank. The taxonomy is being built in stages and currently we have established the body of knowledge which covers all core subjects taught in the university.

Access Management and Security: The Library has decided that only registered members are allowed to access the digital library system. A library portal has been developed by using the eQUIP system. The Membership Management system is used to control access to the digital library. For example, users are allowed to access the

metadata and first 24 pages of theses through any PCs in the campus. The full text can only be viewed through specified PCs in the Library and its branches.

2   Challenges in Handling Arabic Materials

Research and development on Arabic digital library is behind that of non-Arabic resources. The IIUM library is taking the challenge to implement its digital library which incorporates the Arabic resources.

Arabic Scripts: The Arabic scripts is different from other scripts. First, it is written from right to left while most other scripts are written from left to right. Second, Arabic text is written in many type of fonts or calligraphies such as Tsuluts, Naskhi, and etc. Third, the big difference between the nature of Arabic text and English or other languages is cursiveness. Character overlap on top of each other, either in one solid form or overhanging form. In addition, characters may vary their shapes depending on their position within a word, which implies a larger pattern classes to be recognized relating to English and other languages. It is important to point out that there are some Arabic characters, which are highly ambiguous in shape, particularly when OMNI font is considered. Typical examples are some middle shape characters of *sad*, *mim, ain*. There is considerable confusion among these characters in different fonts. Arabic human reading, however, relies on language experience for disambiguation among these characters.

Staff: To develop an Arabic digital library, it needs staff who has knowledge and expertise in the Arabic language and cataloguing and also with basic ICT skills.

Optical Character Recognition (OCR): To OCR Arabic documents, the Library is using SAKHR and Open Office software. Sakhr is the only OCR software available at the time we started the project. It was found that there were a lot of words it could not recognize. The staff has to do a lot of editing and this task is very time consuming. This problem is still not fully resolved at this point. When the RTF file is opened in windows, the character direction was not displayed from right to left. After sometime, we found the solution by opening the RTF file through Open Office. Then, the characters direction are displayed correctly from right to left. Since then, we set a procedure that all Arabic RTF files should be opened through the Open Office. The Sakhr software may support many types of fonts, but due to lack of understanding and lack of local support , the above problem is still not resolved.

Content Management System: Content management system offered by the vendor was IBM Content Manager with DB2 database. However, in the middle of project we found that the system does not sufficiently support Arabic documents as required. We could not create metadata and index Arabic materials. This problem delayed the implementation of the project. To fulfill the requirement, the vendor developed their own content management system namely Multimedia Object Manager (MOM).

3   Conclusion

The implementation of the digital library in IIUM Library is timely. The existing system needs further improvement especially in speeding up the digitization process, portal layout and incorporation of Arabic version in order to allow the Middle East community to also surf the site. We hope with the knowledge and expertise that we have gained, our University will involve us in a big way in the establishment of the University's institutional repository/memory.