

Wasfi G. Al-Khatib, S.A. Shahab, and **Sabri A. Mahmoud**

“Digital Library Framework for Arabic Manuscripts”,

The proceedings of the fifth ACS/IEEE International Conference on Computer Systems and Applications (AICCSA 2007), Amman, Jordan.

Key words: Arabic manuscripts, Digital libraries, Framework, similarity matching, word identification.

Digital Library Framework for Arabic Manuscripts

Wasfi G. Al-Khatib, S.A. Shahab, Sabri A. Mahmoud
Information and Computer Science Department
King Fahd University of Petroleum & Minerals
Dhahran 31261, Saudi Arabia
{wasfi,sadnans,smasaad}@ccse.kfupm.edu.sa

Abstract

Handwritten Arabic manuscripts have been highly underutilized owing to the fact that, to our knowledge, there is no system or interface in existence, utilizing which user can browse or search the desired information. Given vast content of these manuscripts, it is important to develop indexing systems that support content-based retrieval from historical manuscripts. In this paper, we propose a “Digital library framework for Arabic Manuscripts”. A prototype system has been implemented which supports preprocessing of document images. To enhance their quality, feature extraction of the user identified words for similarity matching purposes is carried out. Utilizing relevance feedback from the user, the system indexes the manuscript pages for later efficient retrieval. The prototype system has been tested and encouraging results were achieved.

ries will soon become a reality and play a major role in all aspects of research and education.

2. Challenges and Requirements

The scope of the proposed framework is illustrated by the following example. We consider the case of locating a saying of Prophet Muhammad, known as a *Hadith* in Arabic, in an old manuscript, as seen in Figure 1 which shows two pages scanned from one of the six books of Hadith, “*Sunan Ibn Majah*”. These pages, downloaded

