Arabic Text Recognition System (ATRS)

Prepared by:
Bushra Al-Zaher
Fatemah Al-Awami
Mariam Al-Jumah

Supervised by:
Dr. Feryal Haj Hassan

Project submitted in partial fulfillment for the degree of B.SC. in computer science
2006-2007      1427-1428
Abstract

Arabic text recognition has received attention as a research field for the last five decades. Due to the cursiveness nature of Arabic script and the variety of character shapes which increases the difficulty of the recognition process, it is still an open field for research. We present this project that lays the groundwork for the development of Arabic text recognition system. The system deals with unifont Arabic characters. Developing the system goes through five stages: acquiring text image through a scanner, segmenting the text, pre-processing the character images, extracting features and classification. We developed two methods to recognize the text. The first method is a statistical approach for recognition which is called moment based method. The second method is a structural approach for recognition which is called chain code based method. These methods used in this project were based in main features (moment or chain code) and additional feature (number of holes, ratio…). The main features are used in the first level classifier and the additional features are used in the second level classifier. The system, also, provide an options for correcting the errors of rejected characters and another option for editing the recognized text.