

Recognition based Web information retrieval systems: An applied analytical study

The noticeable increase in the amount of the digital text images which are available on the web, their growing importance and use by researchers in all fields normally receive much attention in the Arabic digital circles, whether concerning forming groups out of these resources or the users' interest in searching for and using them.

The importance of the study springs from the fact that it might be the first academic study in Egypt that attempts to comprehensively handle the subject of digital text images recognition, as a media means , attempting to discover their nature and characteristics, how to process and organize them, how to perform a research and retrieval process based on concept and meaning, and also attempting to set a visualization of the features that should be provided by any system for retrieving and recognizing the digital text images information on the web.

The study consists of an introductory and five chapters organized as follows:

- **Chapter one:** manuscript processing in the web content.
- **Chapter two:** web information recognition.
- **Chapter three:** recognition-based information systems.
- **Chapter four:** recognition search tools in web environment.
- **Chapter five:** a proposed visualization of recognition-based text research systems in the Arabic digital content.

The study is followed by a number of results and recommendations, the most important of which is revealing how the Arabic research mechanisms show the weakness of text image recognition systems when dealing with

Arabic. The study also shows the insufficiency of Arabic lexicon both grammatically and semantically, as it should include all the linguistic data; phonetic, syntactic, grammatical, and semantic data because these data are necessary in processing grammar.

The study provides a number of features that should be included in text images recognition systems to be sufficient when used later for processing, organizing, researching retrieving, managing and having the text images available.