

## Character Recognition Approach Based on Ontology

Hacene Belhadef<sup>1</sup>, Aicha Eutamene<sup>1</sup>, Mohamed Khireddine Kholadi<sup>1</sup>

<sup>1</sup>Mentouri University of Constantine, Department of Computer Sciences, Road of Ain El Bey, Constantine 25017, Algeria

{hacene\_belhadef, aicha.eutamene, kholladi}@umc.edu.dz

### ABSTRACT

In this paper, we present a general description of a new character recognition process based on ontology, semantically annotated by a domain expert. Such process is based essentially on a matching step between two ontologies, the first represents a domain ontology, containing the typographical description of different characters represent an alphabet of a well definite language (Latin for example), the second ontology, describes the document in question in the form of concepts where each concept represents a grapheme located in a -defined order in the document. In this work we consider that there is a robust extraction module of primitives that will serve us to instantiate our ontology, by the different graphemes and their spatial relationships that exist in the processed document.

### KEYWORDS

character recognition, grapheme, matching, ontology, semantic annotation, typographical features.

### 1 INTRODUCTION

The general purpose of character recognition either printed or handwritten is transformed into a readable representation and exploited by a machine. The recognition process is not always easy as long as the contents of the documents may have multiple representations. In the case of printed documents, size, style (Bold, Italic... etc), and melting characters and

other factors play a crucial role in such a process. As for handwritten documents, the conditions of safeguard are often not adequate. Today a large number of books and manuscripts are preserved in museums and archives and are at risk of disappearing due to several factors such as moisture, acidity .., which requires scanning these documents in order to preserve the heritage and exploit a more efficient manner. The digitization of documents is the most effective and fast for this problem, it is to convert a document in paper form into a digital image. Transcription is another solution, but it is less used and limited to manuscripts documents not-long.

The result image of such operation of digitization is used as raw material in the recognition process, to decorticate the content and extract the necessary primitives for the identification and characters recognition, also the entire contents of the document, to use it in a lot of area such as the restoration of national heritage or world, classification, indexing and archiving.

Whatever the rich content of the documents, but this wealth is still insufficient to help the process of character recognition. In this article we are aware that one step annotation document is required to add information further helping this process to accomplish its task.

The annotation of an image through the construction of ontology constitutes the main tool for associating semantics to an image and allows the use of research

















