



**Journal of Digital Information Management Volume 19 Number 2 June 2021**

**Contents**

**Editorial**

**Research**

Retrieving and Processing Images from the Pages of a Historical Newspaper and Modeling the Text Topics-  
Gildácio J. de A. Sá, José E. B. Maia 41

Incorporating Quality Measurement into Scientific Document Retrieval -  
Nedra Ibrahim, Anja Habacha Chaibi, Henda Ben Ghézala 47

Migration Intelligence: AI and Bio-surveillance of Migration Flows-  
Giacomo Buoncompagni 59

**Book Review** 65

**Conference Notification** 66

- Second International Conference on Digital Data Processing, Tangier, Morocco
- Third International Conference on Science & Technology Metrics (STMet 2021)
- 29<sup>th</sup> FRUCT Conference: Call for Participation

51  
178

Keywords: Newspapers, Lexical Standardization, Information Retrieval, Natural Language Processing

Keywords: Processing, Image Processing, Newspaper Information Retrieval

Keywords: Review Scale: 0-6; Review Stars: 4.55; Impact Factor: 28.23%

Keywords: Retrieving the content of historical newspapers

Editors  
The original pages on paper have been worn out due to handling and in general are dry and brittle so that the photograph of the pages is the first step most used because it is the least invasive. Due to factors such as lack of standardization, text misalignment, page wear and poor quality of photographs in some cases, the OCR (Optical Character Recognition) [10] process often returns incomplete words, noises of various types and disconnected passages.

As for the natural language processing tasks, it presents other challenges. Note that a regional newspaper collection is a domain collector and as such may have a limited vocabulary resulting in searches using a broad