

**DOCUMENT MANAGEMENT SYSTEM USING OPTICAL CHARACTER RECOGNITION,  
CLUSTERING, WATERMARKING AND QR CODING ALGORITHMS**

A THESIS  
presented to  
The School of Graduate Studies  
Central Philippine University  
Jaro, Iloilo City, Philippines

In Partial Fulfillment  
of the Requirements for the Degree of  
**MASTER OF SCIENCE IN COMPUTER SCIENCE**

By

**KEITH C. CENSORO**

January 2021

# **DOCUMENT MANAGEMENT SYSTEM USING OPTICAL CHARACTER RECOGNITION, CLUSTERING, WATERMARKING AND QR CODING ALGORITHMS**

Censoro, K.C.

## **ABSTRACT**

The Document Management System is a stand-alone desktop application that provides the employees of the Department of Budget and Management Regional Office VI a service that would secure the circulars and memorandum documents. The Document Management System will accept an official document sent by the Central office to be scanned into the system thru a scanning module that will use Optical Character Recognition (OCR). The captured image of the scanned document is converted to text and clustered into supervised keywords to facilitate searching using the Term Frequency-Inverse Document Frequency (TF-IDF) Algorithm. The scanned document will then undergo authentication by imprinting a watermark image and a QR Code using the watermarking algorithm using Text Brush Embedding and a QR Code Model 2 matrix code. The Evolutionary Prototyping model was used throughout the system development process. It has phases where system evaluation is made directly by the evaluators and the system is refined based on feedbacks which fastened the system process.

Employees of the Department of Budget and Management Regional Office VI evaluated the system and based on feedback, the scanned documents that underwent the authentication process was very secured as the watermark and QR code is imprinted on the document accurately.

The features of the system was able to generate a scanned document with watermark and QR Code imprint and was scanned via the verifier that accepted the document as an authenticated and secured copy generated by the system.