NAME: ZAMORA, JELBERT C. PROGRAM: BSCS – III
INSTRUCTOR: PACOT, MARK PHIL COURSE & SECTION: CSC 198 – BC1

DELIBERABLE #1
Select a research article related to sample Computer Science thesis research topics using Google Scholar. Share your personal understanding of the article in a response of at least 200 words.

Article Title	OCR based Document Archiving and Indexing using PyTesseract: A Record Management System for DSWD Caraga, Philippines
Authors	Jaymer M. Jayoma Elbert S. Moyon Edsel Matt O. Morales
Personal Comprehension	The article discusses the implementation of OCR technology and the PyTesseract library to automate the classification and indexing of records in the Department of Social Worker and Development (DSWD) Caraga in the Philippines. The main goal is to improve the management of paper-based records by digitizing them and making the text content easily accessible for efficient record keeping and retrieval. The process starts with scanning paper-based documents and then extracting the text using PyTesseract, an open-source Python library for OCR. This extracted text is then integrated into a system built with Django and MySQL for indexing and archiving, making it easier to organize, store, and retrieve records. The system aims to streamline the preservation of permanent and valuable papers, making them secure and easily accessible for future reference. Digitizing paper-based records offers several benefits, including easier preservation, access, and management of documents. It also reduces the physical storage space required for paper records and minimizes the risk of data loss due to natural calamities such as fire or flood. Additionally, the system ensures the accuracy of text extraction and indexing through expert-assisted classification and the use of an N-Grams based Linear Classifier Model for categorizing records into different classes. The system also includes a search facility for users to filter and access documents based on specific criteria. In summary, the article highlights the use of OCR technology, specifically the PyTesseract library, to digitize paper-based records and automate their classification and indexing, making it more efficient and reliable for organizations like DSWD Caraga.