# Reconstruction of damaged documents and detail prediction

## Niwethushan Sritharan

Dissertation submitted in partial fulfilment of the requirements for the BEng (Hons) in Software Engineering degree

**Department of Computing**

**Informatics Institute of Technology, Sri Lanka**

**In collaboration with**

**University of Westminster, UK**

**2021**

# Abstract

Preservation of Information has been a part of human life from early age to now. Man kind has used various methods of preservation for years and this method have been evolving from time to time. Documentation of Information has been the most prominent methods used till today. Physical documentation was done using various paper, papyrus and parchments and these were highly used to write down information. These kind of physical documentation has a high tendency to get damaged, destroyed or lose its value by various means.

Most of the Torn or Shredded documents can easily be matched or reconstructed if its small in size or number but when its damaged to extend you can't match or when its in in bulk mixed with other pierces human brain can't easily sort this out. Reconstruction of documents will prevent loss of information and misleading of information.

For years' manual reconstruction was there and then the virtual or artificial reconstruction was introduced which focuses on edge and shape detection corner matching and pattern analyzing. The contents of the shredded or torn document wasn't taken into consideration.

The solution provided for the reconstruction is analyzing of content of torn document and classifying it as groups and using matching them or using OCR technology to retrieve the relevant information. Further to predict information where unclear documents were provided. As to achieve this documents with different shapes and sizes were classified and used. The proposed system focuses on Machine Learning, Nature Language Processing and OCR technologies for implementation.