Sample PDF for OCR and Text Analysis

This is a sample PDF document created for the purpose of testing OCR and text analysis techniques. The document contains various sections of text that can be used to experiment with preprocessing, topic modeling, and keyword extraction.

Section 1: Introduction

Optical Character Recognition (OCR) is a technology that allows computers to extract text from images or scanned documents. It is widely used in various applications, including document digitization, data entry automation, and more.

Section 2: Applications of OCR

OCR technology has many applications, such as automatic number plate recognition, digitizing printed documents, and assisting visually impaired individuals.

Section 3: Experimentation

This document can be used to experiment with different OCR tools and techniques. You can preprocess the text, perform topic modeling, and extract keywords from the text contained within this PDF.

Text and Image PDF Example

This is an example PDF document containing text and an image with text. You can use this document to experiment with text extraction and image OCR techniques. Below is an image that contains some text as well.

Text in Image

This is an example PDF document containing text and an image with text. You can use this document to experiment with text extraction and image OCR techniques. Below is an image that contains some text as well.