## 18XW86 Information Retrieval Lab

Filesystem Indexer

Project Proposal

## Team

Saketh Raman KS - 19PW26

S. Ranga Susanth Sai - 19PW30

## Objective

The objective of this project is to create an efficient file system indexing solution that allows retrieval of documents based on their content. The solution must be capable of handling frequent updates to the files in the file system by employing appropriate techniques to keep the index up to date.

## Motivation

Content based file retrieval is an important aspect of any file management system. Fast file retrieval requires the construction and maintenance of a filesystem index in order to efficiently determine the files that contain information relevant to the user’s query. The files stored in any system are often subject to continuous change, and the constructed index must account for any such changes to ensure that relevant results are retrieved during search.

## Project Plan

The implementation of the project will involve the following steps:

1. Designing and implementing a scheme to represent documents and queries.
2. Processing and converting the documents in the filesystem into the index representation.
   1. Handling text - based file formats.
   2. Parsing and tokenizing document formats.
   3. Identification of objects within images and other multimedia formats.
3. Creating a ranking function to rank retrieved documents.
4. Querying the constructed index to retrieve documents.
5. Updating the index when changes in file contents are detected.
6. Creating a user interface for the retrieval system.

## Tools and Technologies Used

* Python (programming)
* SQLite (database)
* Peewee (ORM)