

OOD

This program provides a way to efficiently find the documents associated to a certain word. It consists of four main classes, Document, which implements iterator, Stoplist, SearchEngine and CLI. SearchEngine class contains most of the functions

class Document implements Iterable

attributes:

- String name - stores the name of the Document
- String content - stores content of document

services:

- Document(String pathString) Creates a Document based on the contents of a file @param pathString, a String containing the path to reach the file
- void readContent(String filename) Reads content of a file and converts it into a String param filename, String of the path to file
- Iterator iterator() Iterates through the content of the Document using a specific delimiter pattern @return An Iterator object
- String documentName() Return the name of the document @return A String, the name of the Document
- String toString() Creates a human readable representation of the Document @return A String

class Stoplist

attributes:

- Set stoplist - stores words in stoplist file

services:

- Stoplist(String path) Creates a Stoplist, a set of stoplist words from a file @param path, a String containing the path to the file
- boolean hasStopWord(String w) Returns whether a word is a stoplist word @param w, a String containing the word to find in set @return A Boolean, true if the word is a stopword

class SearchEngine

attributes:

- HashMap<String, Set<Document>> invertedIndex - stored word key with documents associated to it
- Stoplist stoplist - stoplist collection
- boolean display - the way to display

services:

- SearchEngine(Stoplist stoplist, boolean display) Creates a SearchEngine, containing an invertedIndex hashmap @param stoplist, Stoplist with all stopwords @param display, a Boolean deciding the way to display
- void buildIndex(Document doc) Builds the Hashmap, with a String key (the word) that is associated to a set of documents it is found in @param doc, a Document to be analysed and added to invertedIndex
- Set makeClean(Document doc) Removes all stop words and punctuation from the content of a document @param doc, Document to be cleaned @return a Set of Strings containing all the words in the document
- void findQuery(String query) Finds query word in Hashmap @param query, a String containing word to be found
- void displayInvertedIndex() Displays invertedIndex Hashmap
- void display(String query, Set<Document> querydocs) Displays the documents query was found in @param query, String to be found @param querydocs, documents associated to the query
- void findMultiWord(String[] query, String querytofind) Finds all documents that query words share @param query, a String array with all the query words @param querytofind, String query
- ArrayList makeClean(String[] query) Cleans of stoplist words the string array with all the query words @return An ArrayList of strings with the cleaned query words

class CLI

attributes:

- String[] args - command line interface arguments
- Boolean display - how to display

services:

- CLI(String[] a) Creates a CLI object to read and interpret command line arguments @param a, a String array containing command line arguments

- `void usage()` Prints a usage message
- `void parse()` Interprets String arguments, parses arguments as parameters to specific functions