IR Practical Task-1 Group No.12

1. Document Collection: The file attribute on this class denotes the text file that was utilized to generate the document collection. A list of all the individual papers that have been produced from the file would be included in the documents attribute.

2. Document: The text attribute of this class would be used to describe the contents of the document. Additionally, each processing step that is applied to the document (stop word elimination, basic/root form reduction) would have attributes. It would also contain a title attribute, which would represent the document's title.

3. Search: The term being searched for would be represented by the search word attribute of this class. A list of all the pages that contain the search term would be included in the results attribute.

4. User Interface: This class would have an attribute that represents the user interface being used to interact with the software, either graphically or textually.

5. Retrieval Model: The type of retrieval model being used (Boolean or vector space) would be an attribute of this class. Additionally, it would have a property indicating the retrieval model implementation (linear search, inverted list, signature method, etc.).

6. Key Figures: The recall and precision of predefined test queries would be shown by attributes in this class.

The class functions are described here in further detail:

- User: Represents a system user who can communicate with the system by using a user interface.
- Query: Represents a search query that a user has entered.
- Processor: Organizes the coordination of document processing and search query execution. It communicates with the classes for Document and Index.
- Document: Represents a document in the system as a document. is composed of a title, a body of text, and a boolean flag indicating whether it has been processed. uses techniques to get rid of stop words and shorten words to their root or stem form.
- Index: Displays the inverted index that is used to search the collection of documents. contains a dictionary that links words to the texts they are used in.

The use of algorithms implemented within the Index class allows the system to accommodate various retrieval models and implementations. The processed boolean flag in the Document class allows for tracking of the document processing steps.

For predefined test queries, the important metrics Recall and Precision are automatically calculated. Users can interact with the system using a graphical or textual user interface, depending on how it is implemented.

Here are the different kinds of relationships between the classes indicated by the arrows:

1. Document Collection → Document: aggregation (because the collection is made up of individual documents)

2. Document → Search: association (because the search is performed on the document)

3. User Interface → Document Collection, Search, Retrieval Model, Key Figures: association (because the user interface interacts with all of these classes)

4. Retrieval Model → Document Collection: association (because the retrieval model operates on the document collection)

5. Key Figures → Search: association (because the key figures are determined by the search)