

EXPERIMENT 2

AIM: To implement Boolean Retrieval Model in Jupyter Notebook.

RESOURCES REQUIRED: Jupyter Notebook, 4GB RAM and above, i5 Processor and above.

THEORY:

BOOLEAN RETRIEVAL MODEL:

In the Boolean Model for Information Retrieval, a document collection is a set of documents and an index term is the subset of documents indexed by the term itself. An index term can also be seen as a proposition which asserts whether the term is a property of a document, that is, if the term occurs in the document or, in other words, if the document is about the concept represented by the term.

The interpretation of a query is set-theoretical. In practice, a query is a Boolean expression where the set operators are the usual intersection, union and complement, and the operands are index terms. The document subsets which corresponds to the index terms of the query are combined through the set operators. The system returns the documents which belong to the subset expressed by the query.

EXAMPLE:

searching “by numbers”

! Sequence of queries driven by number of retrieved documents

- e.g. “lincoln” search of news articles
- president AND lincoln
- president AND lincoln AND NOT (automobile OR car)
- president AND lincoln AND biography AND life AND birthplace AND gettysburg AND NOT (automobile OR car)
- president AND lincoln AND (biography OR life OR birthplace OR gettysburg) AND NOT (automobile OR car)

CONCLUSION: Hence we have successfully implemented Boolean Retrieval Model.