# Application Architecture

## Team members:

**Arjun Vijayvargiya(2012B1A7733H) and Bikkumala Karthik(2012B4A7748H)**

**Method:** Vector Space Model

**Algorithms:**

- Porter's algorithm for stemming
- Cosine Similarity algorithm for evaluating score of a given document with respect to given query

Vector space model uses term frequencies and document frequencies to evaluate score. Each term that appears in the text document is stored in Class Dictionary with term's name and its document frequency and a linked list that contains variables of Class Posting. Class Posting contains document's name and term frequency thus each Dictionary class variable is associated with a linked list of documents and term frequencies that contained the term. All the Dictionary Class variables are stored in a linked list named words.

Vector Space model takes as input a free text query. Query terms are also tokenized and normalized after which they are stored in a linked list of Class Queryterm. Class Queryterm contains the term name and term frequency of that particular term in the given query.

List of variables of Dictionary class

| | |
|---|---|
| word=hoarsely | (1)->Judith(13), |
| word=hobbit | (2)->Hobbit(16),HobbitLarger(32), |
| word=hobbitboi | (2)->Hobbit(1),HobbitLarger(2), |

Term name (document frequency) ->document name (term frequency)

**Graphical User Interface:**