# Implementation of Inverted Index using HADOOP

## Functional Requirement:

As a user using this software one should be able to find all the documents which are present in HDFS (HADOOP Distributed Files System) containing a particular word. User will be provided with a GUI which should contain text field where he will enter the word to be searched and search button. When search button is clicked user should be able to get the entire documents name which contains the particular word entered by the user. For searching inverted index algorithm must be used.

On the server side user will be the server admin. Server admin will be able to trigger inverted index algorithm whenever needed. Also this algorithm should be run every hour in the system for the new files added and updated.

## Non Functional Requirement:

* User should be able to run this software from wherever possible, that means it is not necessary that client will be on the same system where HDFS is installed.
* MapReduce should be used.
* Client should be platform independent. Hence user should able to use the software in both windows and linux based platform
* Code should be written following the clean code principals, however JUNITS are optional and can be written if time permits.
* Every module should be separately tested before performing the integration testing.
* Software should at least work on single node cluster of HADOOP.
* A proper dataset should be for testing.

## Software Requirement:

* JAVA 7
* HADOOP
* MAPREDUCE

## Hardware Requirement:

* Standard Ubuntu Machine with 4GB+ RAM and i3 or above processor.

# UseCase Diagrams

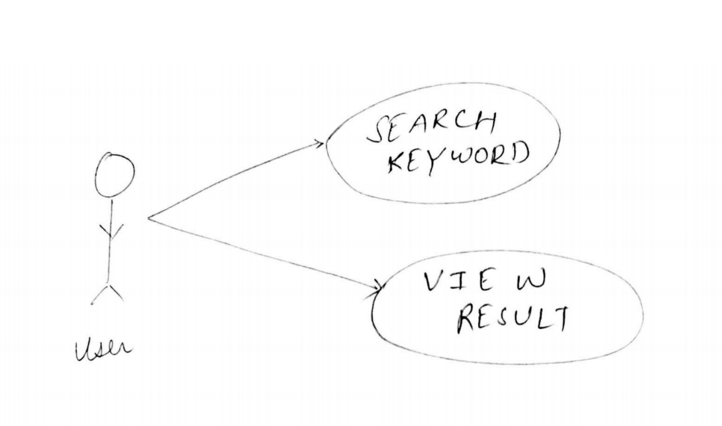


Figure User UseCase Diagram

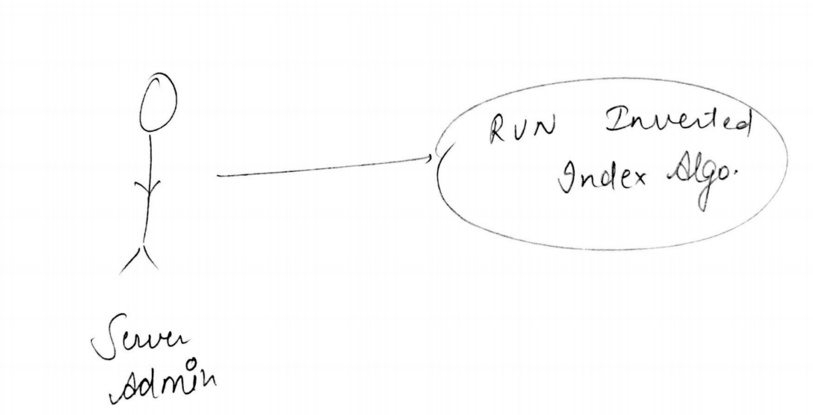


Figure Server Admin UseCase Diagram

**Note: We will try to provide to provide “and, or, not” keyword search to user if all the features are done and tested in time. And will allow user to search a document containing two words, Or will allow user to search the document containing either of the word and not will allow user to search all the documents which doesn’t contain a particular word.**