**HBase Frequency Index Builder**

**Introduction**

Hbase FreqIndexBuilder is an advanced WordCount program which counts the number of occurrences of each word in a given text input dataset and also stores the related document name (identification number) as HBase inverted index records. These Inverted indices for text data are built for supporting efficient searches in a huge set of text data.

Inverted Index:

For a given word, we record the count of that word in each document.

E.g.

“the” -> doc3, doc5, doc6

**Program Flow**

There is no implementation of Reducer class. We, use the default reducer class of Map Reduce framework.

Main Program:

Configuration of the Map and Reduce class. Here, reducer is set to null as we have not overridden the reducer function. Default reducer is used.

Mapper:

Input-> <key, value> corresponds to <rowkey, content>

Key is the rowkey of an HBase record related to specified URI

Content is the stored text of the URI.

Input is taken from the HBase table clueWeb09.

Output-> <word, <docId, frequency>>

Word is the word being counted

<docId, frequency> is the docId of the document in which that word occurs and its corresponding frequency count within the document.

Output is stored in HBase table clueWeb09IndexTable