

# Distributed Algorithms - Hadoop Assignment

This project is an implementation of searching in sentences in a distributed file system, HDFS.

This project is built via maven and uses Hadoop v2.6.0. An **artifact/jar** file has been provided which can directly be run on a system with hadoop already installed.

## How to execute the jar file

```
hadoop jar dcassign02.jar Search dcassign02/input dcassign02/search/output  
politics
```

In the above command, the third argument is the query whereas the first and second argument are input and output directories HDFS respectively.

## Working

The mapper looks for the query in each news headline and passes it to the reducer if the query is found in the sentence.

The reducer then simply writes all the sentences in the output directory.