Readme text file explains brief about the project and gives the minimum system requirements that are needed to run this project.

System Requirements

============================================

Java version : JDK 1.7.0\_91 or higher

Hadoop version is 1.2.1 or higher

Hive version 1.0.0 or higher

Project Introduction

============================================

Finding top 10 influential papers among all the papers from the DBLP citation network. Here we used weighted pagerank algorithm to compute the page ranks. The PublicationMapper and PublicationReducer programs are used to extract the input from the input file to generate the link graph which is later used in the computations of hive queries.

Project Development

============================================

The following java files and hql files are part of this project.

PublicationDriver.java –This is the main class in this project,where we can set the output and input format of classes.

PublicationMapper.java-The mapper class which filters the unused paperID’s from the input file like a paper with no reference ID’s,etc and emits paperID along with paper title as key and reference ID as value.

PublicationReducer.java—Here the reducer class is used to aggregate all the reference ID’s of a paper and emitting it as a value and the paperID with paper title as key.

Linkgraph\_hqlfile1.hql –This hql file is used to create the link graph.

Pagerank\_hqlfile2.hql—This hql file is used to perform the 10 iterations to find the top 10 pages which has highest rank.