

## Problem Statement

Stack Exchange is a network of question-and-answer websites on topics in diverse fields, each site covering a specific topic, where questions, answers, and users are subject to a reputation award process. The reputation system allows the sites to be self-moderating.

The data files needed are attached.

You can find the metadata information here. You can also google if you have any confusions.

<https://ia800107.us.archive.org/27/items/stackexchange/readme.txt>

Create a **MapReduce** program for the questions below. You can run your code on standalone (local) Hadoop installation. The process is explained in the slides in file Hadoop Introduction. A lecture with standalone demo video is also uploaded.

1. Find the average length of comments in Comments.xml file.