ISTANBUL TECHNICAL UNIVERSITY

FACULTY OF COMPUTER AND INFORMATICS ENGINEERING

Project Description Form for Computer Project II Course

PROJECT 3A

**Assignment Date**: 27.04.2016

**Submission Date:** 18.05.2016 23:00

**Project Name: Mining Chrome Repository**

**Project Duration:** 4 week

**Project Workload (man-hour):** 40 hour/man

**Project Weekly Plan:**

1. Writing scripts to mine Chromium’s version control system
   1. Choose a specific development period, e.g. Jan 2015-Dec 2015, and extract raw commit data (what is committed, when is committed, who committed, which files are committed)
2. Identifying top developers
   1. Plot the distribution of commits in terms of each developer, e.g. How many commits each developer does, how frequently each developer makes a commit to the system
   2. Identify top developers who make 80% of the commits
3. Identifying edited file sets in commits
   1. Create a matrix in each rows represent files and columns represent developers. If the index at (i,j) in your matrix is 1, it should indicate 1 if file *i* is committed by developer *j*, or 0 otherwise.
4. Building and visualization of socio-technical network of Chrome project
   1. Draw an undirected graph in which each node represents a top developer.
   2. Draw an edge between a top developer A to another top developer B if they edited the same files. You need to use the matrix extracted in Step 3.
5. Upload your code and visualization to GitHub by creating a new project. Share the link with us.

**Keywords:** mining software repositories, Google Chrome, social and technical network

**Success Criteria:** Plots of top developers, developer-file matrix and socio-technical network should all be extracted.

**Tools:**

None. You need to write your own scripts to do this data extraction and analysis process.

**Grading:**

1. Implementation : %50
2. Results presentation: %20
3. Performance: %30

**Project Description:** Project aims to identify social and technical dependencies in Chrome project. Students are expected to extract commit-based information from version control systems of Chrome project, such as edited file sets, developers who made the commit, and co-changed files in the commits. Based on those information, a visual representation of technical dependencies and social dependencies will be done.