CSCI318- Software Engineering Practices & Principles
Spring Session 2017

Group project description

Total 45 marks

Note: In students' assignments and project reports, if any part (including sentences, figures/diagrams, tables, definitions, descriptions, and so on) are copied from other people's work (including both published and unpublished papers, reports, Web articles, etc), or if any idea is from other people, such work or people must be acknowledged explicitly. If you directly "copy and paste" sentences from other people's work to your work, you shall use quotation marks to quote the copied sentences and make a citation immediately, or else it is plagiarism.

1. Size of a group: Students will work on the project in groups of three to five.
2. Progress report: Each group is required to give a 3-minute presentation to the class on their progress, in the **second lecture of Week 4**. Immediately before the presentation, submit a hardcopy of your PPT slides to the lecturer. All team members are required to attend the presentation. The progress report should cover the following points:
i. Every team member's name, email address and student number.
ii. What is your progress so far?
iii. Did you encounter any problems?
iv. Your future plan and timetable.

3. Submission of final report: Submit a **folder** including the following items: (1) a hardcopy final report that addresses the marking criteria; and (2) a CD/DVD/USB storage including a softcopy of your final report, code, executables, presentation slides, and a readme file. You can also include additional data, figures, screenshots, etc, in the storage.

Print every team member's name, student number, and login name in the cover of your report. The cover should include all team members' signatures. Also indicate each team member's individual contribution in the cover, as explained below.

4. Assessment of individual team members:
Each team member is assessed individually based on his/her individual contribution. Basically, the "individual contribution" of each team member is assessed by all the other team members in terms of "contributed", "very little", and "almost no contribution". For a team member who has "**contributed**", he/she will receive **100**% of the group marks; for a member who contributed "**very little**", he/she will receive **50**% of the group marks; for students who made "**almost no contribution**", he/she will receive **0** marks for the group project. Details of individual contributions for each team member should be printed on the cover of the final project report. In special situations, the above criteria may be adjusted by your lecturer.

5. Marking criteria (total **45 marks**)
(1) Progress report (second lecture of **Week 4**): **1 mark**.
(2) Final presentation of your group project: **1 mark**.
The final presentation will be conducted starting from the **first lecture of Week 13** (and may extend to the second lecture of the same week). Submit a hardcopy of your presentation slides to the lecturer immediately ***before*** your group's presentation. The duration for each group is about 10 minutes plus a short Q/A session. It is *not* compulsory for all team members to speak, but all team

members must stand in front of the class during their group's presentation. If a student is absent from his/her group presentation, the student will receive 0 presentation marks.

(3) In the first lecture of week 13:

a. Did you submit a hardcopy of your presentation slides to the lecturer immediately ***before*** your group's presentation? **(1 mark)**

b. Did you use a folder to enclose all materials when submitting the final report? (**1 mark**)

(4) Technical value and writing of the final project report (submission due at the start of the **first lecture of week 13**): **41 marks**.

Project title: *Automatic Evaluation of Machine Translation Software*

1. Marking criteria (41 marks)

Section 1. (**6 marks**) A comprehensive literature study

First, study the concept of *metamorphic testing* by reading the research papers available in Moodle under "Readings". Read further papers on this topic if necessary.

Review at least 50 articles, which can be refereed research articles, books, and good quality websites (including Wikipedia), on the topic of *automatic evaluation of machine translation*. As an example, the website http://www.mt-archive.info/srch/eval-10.htm lists many evaluation measures published since 2010. In addition, review available tools (the more the better) for automatic evaluation of machine translation.

Summarize the key research problem and the key contribution of each paper, and the features of each tool. For each paper and each tool, explain whether and how the idea of metamorphic testing could be applied to enhance the technology.

Section 2. (**35 marks**) Complete the following tasks in this section.

Section 2.1. (5 marks)
Propose ideas to apply *metamorphic testing* to the evaluation of machine learning software.

Section 2.2. (10 marks)
Implement the ideas that you proposed in Section 2.1 into a software testing tool. Your code can be based on any existing software components such as open source software and web services --- but you must acknowledge them explicitly.

Section 2.3. (20 marks).
Apply the tool that you developed to test machine translation applications (such as Google Translate, Bing Translator, and WeChat translate). Report your results.

Marking criteria for Section 2.3

(i) How many machine translation applications did you test? (should be at least two.)

(ii) How many languages did you test? (should include at least English and Swedish.)

(iii) How did you generate test cases and how many test cases did you use? (test case generation approach must be justified and a large number of test cases are needed.)

(iv) How comprehensive and convincing are your results and conclusion?

Example:
Applications under test: Google Translate and Bing Translator
Languages: two-way translation between English 100 other languages offered by the translation software. These 100 languages include both popular languages (such as Chinese, Spanish, Arabic) and unpopular languages in order to assess the reliability of the applications under different operational profiles.
Metamorphic relation: Round-trip translation.
Evaluation metric: Precision in terms of the top 50 Google search results (for Google translation) and Bing search results (for Bing Translator).
Test case generation: Using sentences taken from Wikipedia articles.
Objective: To assess the reliability of the translation software between English and other different languages.