Course Project Overview

**Introduction**

The course project is to give the students hands-on experience on developing some novel information retrieval and/or text mining tools. It would allow the students to potentially apply all the knowledge and skills learned in the course to solve a real-world problem. Group work is encouraged, but not required (i.e., you can have a one-person team). The maximum size of a team is 5 members to avoid challenges in efficient coordination of the work by too many team members. However, a team of a larger size is also possible subject to the approval of the instructor. A typical reason for a larger team is because the project has a very natural task division among the team members so that the need for frequent interactions and coordination of team members may be minimum despite the large size of the team. Whenever possible, collaboration of multiple project teams is strongly encouraged to minimize the amount of work of each team via expertise or resource sharing, as well as to generate “combined impact” (e.g., one team may develop a crawler that can be used by another team that develops a search engine).

More details (what to submit at each step) can be found here:

https://docs.google.com/document/d/1b-EagO17Og7\_ESj5hkP5x4EFrVPQAtBlH9YvsgEzjnY/edit?usp=sharing

**Grading criteria**

Your project will be graded based on the following weighting scheme, corresponding to three stages of work including 1) team search and formation; 2) proposal development; 3) project result submission. All team members will receive the same grade provided that every member has made sufficient effort (at least 20 hours of quality time to work on the project).

* **Team formation (5%):** Every student is required to form a team by the **beginning of** **Week 9 (Oct 17, 2022)**. The team shall need to designate one of the team members as the team leader who will be responsible for submitting and/or uploading the course project deliverables.
* **Project proposal (5%):** Each project team is required to submit a roughly one-page project proposal by the **end of** **Week 9 (Oct 23, 2022)**, and graded based on completion. Specifically, the following grades will be assigned, based on the answering of questions from the Google Doc:
  + 100: Answers all questions
  + 90: Missing one or two questions
  + 75: Missing more than two questions, significantly incomplete, cannot understand
  + 0: No proposal in repository
* **Progress report (5%):** Each project team is required to submit a short progress report by the **beginning of Week 13 (Nov 14, 2022)** listing the progress made, remaining tasks, any challenges/issues being faced, etc. It will be graded based on completion with respect to addressing the following points: (1) Progress made thus far, (2) Remaining tasks, (3)Any challenges/issues being faced. Grades will be assigned as follows:
  + 100: Addresses all points
  + 90: Missing one point
  + 75: Missing two points
  + 0: Missing all points or no submission
* **Software code submission with documentation (65%):** Due **during Week 16 (Dec 08, 2022)**, each project team will be asked to submit the produced source code with reasonable documentation. The documentation should cover both how to use the software and how the software is implemented. The 65% of the grade would be distributed as follows: **45% for source code submission**; **20% for documentation submission**. Both would be graded based on completion.
* **Software usage tutorial presentation (20%):** Due **during Week 15 (Dec 08, 2022).** 10% of the grading is based on completion and the remaining 10% is based on result of testing the software by graders

Thus, your course project work would be graded primarily based on your effort. If you have followed our guidelines and completed all the required tasks, you should receive at least 90% of the total points for the project. This is to encourage the students to pay attention to time management and set realistic goals that can actually be completed by the end of the semester. The remaining 10% is based on how well your software works; a fully functioning software would be given the whole 10%, whereas a buggy software or a software with missing functions would result in losing some of the 10% of the grade. Completion of a functioning software is emphasized also due to the potential dependency between multiple projects when they are all contributing to a larger project (e.g., one team may produce a crawler to crawl data for use by another team to build a search engine).

*The proposal, progress report and final submission will be peer-graded by you. The TAs will review all peer reviews and make sure that (1) the reviews are fair, and (2) the submissions satisfy all requirements.*

**Peer-Grading Instructions**

We'll be using peer-grading for grading project proposals, progress reports and final submissions. Later in the semester, you will receive an email from CMT inviting you as a reviewer, and we will make an announcement when the emails are sent.

Peer grading will begin soon after the submission deadlines have passed.

**The proposal peer-grading should be done between Oct 24 (12:00 am CST)-Oct 30 (11:59pm CST).**

**The progress report peer-grading should be done between Nov 15 (12:00 am CST)-Nov 20 (11:59 pm CST).**

**The final submissions peer-grading should be done between Dec 08 (12:00 am CST)-Dec 16 (11:59 pm CST).**

**Please finish the gradings on time.**

Each project will be reviewed by at least 2 students and each student will review ~2-3 projects. We will be using your comments and scores as a guide to assign the final scores to the projects. Please grade them carefully and honestly. We appreciate your help with managing the grading of such a large class. We hope the process would also be a good learning experience for you. More specific instructions will be announced and sent later.

**Instructions**

**1. Form a team (5%, due beginning of Week 9, Oct 17, 2022)**

By the beginning of Week 9, you should form teams. Groups of more than 5 members are highly discouraged. The team also needs to designate a member of the team as team leader. Your team leader will be responsible for submitting and/or uploading the course project deliverables such as: 1) this form, 2) the proposal, 3) the presentation, 4) the code, etc.

**2. Pick a topic & Write a proposal (5%, due end of Week 9, Oct 23 2022)**

**Picking a project topic**

Students will be able to choose their own topic or select a topic from a list of suggested topics which we will provide by Week 4.

**Writing a project proposal**

Each project team is required to write a one-page proposal before you actually go in depth on a topic. The proposal is due end of Week 9.

In the proposal, you should include the names and email addresses of all the team members. One member must be designated as the project coordinator/leader for the team, so please make sure to indicate that. The project coordinator would be responsible for the coordination of the project work by the team and also communication with the instructor or TA when the team needs help.

Detailed instructions about the required content of the proposal will be provided by Week 4, including a short set of questions that need to be answered in the proposal. As long as those questions are addressed (wherever applicable), the proposal does not have to be very long. A couple of sentences for each question would be sufficient.

**3.** **Peer-review project proposals (due Middle of Week 10, Oct 30 2022)**

Each student will review the progress reports of ~1-2 groups and provide feedback/suggestions. TAs will go through the peer assessments and make sure that the proposals satisfies all requirements listed in the Google Doc.

**4. Work on the project**

You should try to reuse any existing tools as much as possible so as to minimize the amount of work without sacrificing your goal. Discuss any problems or issues with your teammates or classmates on Campuswire and leverage Campuswire to collaborate with each other. Consistent with our course policy, we strongly encourage you to help each other in all the course work so as to maximize your gain of new knowledge and skills while minimizing your work as much as possible. We will do our best to help you as well. Consider documenting your work regularly. This way, you will already have a lot of things written down by the end of the semester.

**5. Submit progress report (5%, due beginning of Week 13, Nov 14 2022)**

Each team must submit a short report detailing 1) Progress made thus far, 2) Remaining tasks, 3) Any challenges/issues being faced. The graders, TAs and instructor may provide help and suggestions based on the report. Continue working on the project until the final submission.

**6. Peer-review progress reports (due end of Week 13, Nov 20 2022)**

Each student will review the progress reports of ~1-2 groups and provide feedback/suggestions.

**7. Software code submission with documentation (65%, due middle of Week 16, Dec 08 2022)**

Each team must submit the software code produced for the project along with a written documentation. The documentation should consist of the following elements: 1) An overview of the function of the code (i.e., what it does and what it can be used for). 2) Documentation of how the software is implemented with sufficient detail so that others can have a basic understanding of your code for future extension or any further improvement. 3) Documentation of the usage of the software including either documentation of usages of APIs or detailed instructions on how to install and run a software, whichever is applicable. 4) Brief description of contribution of each team member in case of a multi-person team. Note that **if you are in a team, it is your responsibility to figure out how to contribute to your group project, so you will need to act proactively and in a timely manner if your group coordinator has not assigned a task to you.** There will be no opportunity to make up for any task that you failed to accomplish. **In general, all the members of a team will get the same grade for the project unless the documentation submission indicates that some member(s) only superficially participated in the project without doing much actual work; in that case, we will discount the grade.** Everyone is expected to spend at least 20 hours to seriously work on your course project as a minimum, not including the time spent for preparing the documentation.

The 65% of the grade would be distributed as follows: **45% for source code submission**; **20% for documentation submission.** The 20% for the documentation submission includes 5% for overview of functions, 10% for implementation documentation, 5% for usage documentation. There is no strict length requirement for the documentation.

**8. Software usage tutorial presentation (20%, due middle of Week 16, Dec 08 2022)**

At the end of the semester, every project team will be asked to submit a short tutorial presentation (e.g., a voiced ppt presentation) to explain how the developed software is to be used. The presentation must include (1) sufficient instructions on how to install the software if applicable, (2) sufficient instructions on how to use the software, and (3) at least one example of use case so as to allow a grader to use the provided use case to test the software. There is no strict length requirement for this video submission, but you should target at 5~10 minutes. A presentation shorter than 5 minutes is unlikely detailed enough to help users understand how to use the software, whereas a longer video than 10 minutes might be too long for impatient users. However, feel free to produce a longer presentation if needed.

The tutorial presentation would be graded based on

1) completion of the presentation (10%); and

2) result of testing the software by graders (10%).

If the software passes the test (i.e., is working as expected), full points will be given; otherwise, points will be deducted from the 10% allocated to the “result of testing the software by graders.”

**9. Peer-review project code, documentation, and presentations (due Dec 16 2022)**

Each student will review the final project submissions of ~2-3 groups and provide feedback/suggestions.