

# Homework #1 Basics of IR and project teaming

---

Due February 3rd, 2014 5:00pm

## Required:

- **Problem I** (6 points): Vector space model for information retrieval
  - (1) Describe, in your own language, what is the vector space model for information retrieval? (2 points)
  - (2) What's the drawbacks of Boolean retrieval model? (2 points)
  - (3) How do you compare vector space model with Boolean retrieval model? (2 points)
- **Problem II** (4 points): TF-IDF
  - (1) Describe, in your own language, what is Term Frequency? (1 point)
  - (2) Describe, in your own language, what is Inverse Document Frequency? (1 point)
  - (3) Explain, in plain language, why TF-IDF is effective in measuring relevance? (2 points)
- **Problem III** (2 points bonus): teaming report
  - Please form a 5-6 persons team. The subsequent course projects are all going to be team projects. You are expected to present the names of the team members, and the program language(s) you plan to use in the projects. You should also indicate who will be the point of contact of your team. **The point of contact will need to email Prof. Hua a spreadsheet which list the names of all team members.**
  - Student who did not find their own team will not get the 2 point bonus.

## What to turn in?

You should make the written part of your answers in a PDF file and name it as:

**[yourfirstname]\_[yourlastname]\_HW1.pdf**

For your program, you may use any programming language. However, your submission should be the executable, the source code, and a detailed readme file on how to run it. You should have a detailed report in the written part of your homework on what you have tested and what are the results you obtained. Please make sure you packed additional dependent libraries, if any, used in your program. If your program cannot run, you lose 0.5 point automatically.

Package your PDF file with the code and supplementary Readme file in a single ZIP file as:

**[yourfirstname]\_[yourlastname]\_HW1.zip**

and please submit it through the Moodle system.

**\*\*\*\*Important Notice: Please also bring a printed copy of your answers to the class in the due day of the assignment. This is mandatory\*\*\*\***

**Grade: 10% with bonus**

Late submission policy applies universally with no exception.

If you have a compelling excuse, you must inform me at least 2 days before the due date. I don't accept excuses such as "**I am overloaded by other courses**".