CISC 879 Homework for Thursday, Sept 12:
Individually read the whole paper: "An Information Retrieval Approach to Condept Location in Source Code," Marcus et. Al. 2004.

In your 5 groups, prepare to lead class discussion on each section as designated below.
In class, each group will take turns in the "fishbowl" where they ask and discuss the answers to the questions as assigned with another group. The other group will be chosen randomly to be in the fishbowl with them for discussion. Each person in the groups in the fishbowl should participate actively in the discussion and one person in each group should be in charge of leading the discussion for one of their questions.

The rest of the class will observe and reflect on the discussion, individual participation in the discussion, and provide feedback.

You may bring notes to use during your discussion as well as your marked-up paper. You will be graded individually for your class participation and get feedback on it.

Assigned questions:
Related Work: Group 1
  - What other SE problems have IR been used for?
  - Anyone else use IR for feature/concept location? If so, who and when and how?
  - What other approaches to Feature location are there besides IR? Who created those approaches and what does IR approach contribute beyond them?
 3-3.2: Group 2
   - What are the main subproblems to be solved to prepare the corpus?
   - Explain the key aspects of LSI and VSM and compare and contrast the approaches.
   - Why do the authors say that LSI is well suited for source code?
   - What is the typical similarity measure used?
  - Compare/contrast the two different ways users can formulate queries.
 3.3,4.1: Group 3
  - What is the strategy for determining what the user should inspect in the ranked results? How is partitioning involved?
  - What is the goal of the case study – what hypothesis are they investigating?
  - What was the subject of study and how did they justify using that subject?
  - What were the threats to bias the study and how did they minimize them?
  - What did they measure and why those measures?
4.2-4.3: Group 4
  - How did they prepare the corpus? How did they justify that?
  - How did they obtain queries for their study? How did they avoid bias/maximize generalize of results?
  - How many queries and how did they justify that number?
  - What were the results presented? How did they view those results?

4.4-4.5: Group 5
- How did they automatically generate queries?
- What did they measure?
- What were the results for automatically generated queries and how do they compare with manual user queries?
- What did they compare their approach with? How do those approaches work?
- What were the results of how they compare?
- What is their interpretation of those results?

4.6-Conclusions: Group 1
- Why did they compare with grep-like search?
- What methodology did they use to implement and compare the searches?
- Summarize the main contributions
- Summarize the results of the evaluation
- What did they see as future work in 2004?