**Mid-term 1 Guidelines**

Topics to be covered in mid-term 1

1. Basic concepts of Information Retrieval Systems

\* Architecture of IR systems

\* Binary user judgments (i.e., relevant vs nonrelevant documents)

2. Sample IR Systems:

\* Google, Google Scholar

\* Web of Science

\* DBLP

3. Boolean retrieval models

4. System evaluation based binary user judgments

(i.e., relevant and nonrelevant documents) and

binary system results (retrieved and non-retrieved

documents).

\* recall

\* precision

\* recall-precision graph (with respect to ranked system results)

\* F measure

5. Indexing theory

\* Construction of indexing vocabulary (i.e., a set of index terms)

\* Term weighting (i.e., term frequency, term document frequency, and

inverse document frequency, etc.)

6. System modeling and evaluation based on user preferences (i.e., preference relations)

\* Weak order

\* Order preserving utility functions

\* perfect ranking

\* acceptable ranking

\* performance measure based on distance between user and system rankings.

You need to understand both conceptual understanding and computational

methods. That is, based on definitions of various notions, you should be

able to compute various quantities, such as recall, precision, recall-precision

graph, ndpm etc.

Sample questions

1. Give the basic architecture of IR systems and explain the functionality

of each model.

2. Give the definition of precision and recall measures. Explain the meaning

of these measures.

3. Suppose that for a query q there are 10 relevant documents. Consider the

following ranking produced by a system:

- + + - - + - + + - - - + - + + - - + + ...

where + denotes a relevant document and - denotes a nonrelevant document.

Draw the recall-precision graph

4. Describe the main ideas of Boolean retrieval model. Discuss potential

difficulties with Boolean model.

5. Provide an argument to show that binary user judgment is too restrictive.

6. Describe the basic ideas of inverse document frequency of a term. Give

two formulas to computing inverse document frequency.

7. Give the definition of a weak order.

8. Give the definition of an equivalence relation.

9. Why do we need that a user preference relation is a weak order. What

is the implications of a weak order.

10. Prove that if a preference relation is a weak order, then it is transitive.

11. Give the definition of a irreflective binary relation. Give a definition

of asymmetric binary relation. Prove that asymmetry implies ifflexivitity.

12. Describe the ideas of ndpm.

**Mid-term 2 Guidelines**

Topics to be covered in mid-term 2

1. Vector space model of IR

\* Document representation

\* Query representation

\* Retrieval (i.e., matching function), similarity measures

\* Binary vector space models

- similarity measures

- precision-oriented measures

- recall-oriented measures

2. Probability distribution models

\* Expected utility

\* Similarity based on entropy of distributions

3. Basic ideas of relevance feedback

\* Classical relevance feedback in VSM

\* Relevance feedback under user preferences

(Lean a query by error correction)

\* The connection between the above to approaches