

**Student Name:** \_\_\_\_\_

**Midterm Exam**  
Information Retrieval (11-741)  
March 8, 2007

Answer all of the following questions. Each answer should be thorough, complete, and relevant. Points will be deducted for irrelevant details. Use the back of the pages if you need more room for your answer.

The exam should take you about 70 minutes to complete. The points are a clue about how much time we think each question should take to answer. We assume about 1.5 points per minute, so a 10-minute question is worth 15 points. Plan your time accordingly.

Good luck.





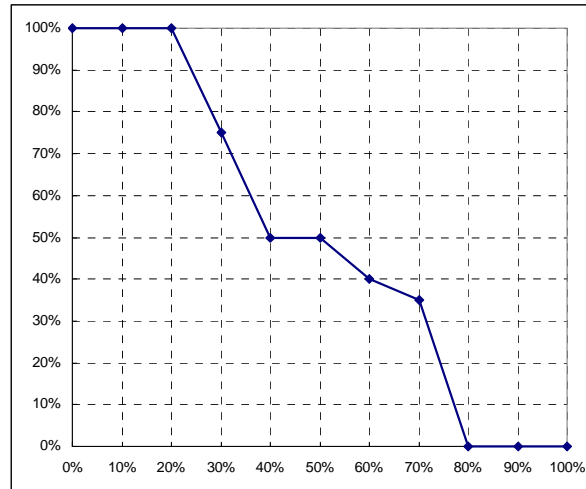


6. Hypertext Retrieval [18 points]

a. Write the pseudo-code for the HITS algorithm, including the input and output of the system.

b. Let  $A$  be the adjacency matrix of a hypertext graph with  $N$  nodes: if there is link from node  $i$  to node  $j$  then  $A_{ij}=1$ ; otherwise,  $A_{ij}=0$ . Show (prove) i) how the hub scores and authority scores converge with a sufficiently large number of iterations in HITS, ii) what those scores converge to (respectively), and iii) how the stability of the convergence is affected by the “eigen gap” of some matrix. Specify the matrixes for generating the hub scores and the authority scores, respectively, with the notation of  $U = (\vec{u}_1, \vec{u}_2, \dots, \vec{u}_N)$  as the eigenvectors of an  $N \times N$  matrix, and  $(\lambda_1, \lambda_2, \dots, \lambda_N)$  as the corresponding eigenvalues, where  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ .

7. For a particular query, an information retrieval system returned a ranked list of 20 documents that produced an interpolated precision-recall curve as shown below. Assume there are 10 relevant documents in the entire corpus.



- a. What is the precision when the system has retrieved 7 relevant documents? **(4 points)**
- b. The documents returned by the system are provided below, ordered according to the number shown above each document. Based on the provided precision-recall curve, write an **R** inside the document that is relevant. **(4 points)**

1	2	3	4	5	6	7	8	9	10
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11	12	13	14	15	16	17	18	19	20
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

- c. Calculate the Mean Average Precision (MAP)? Show your calculation. **(5 points)**
- d. Describe how Precision at 80% Recall is calculated? What is it in this case? **(5 points)**