

CSCI 7000-IR Fall 2008 Quiz 1

Name: _____

On my honor, as a University of Colorado at Boulder student, I have neither given nor received unauthorized assistance on this work. _____.

1. **(5 points)** True/False: Jim has a high opinion of the Cuil search engine effort.
2. **(8 points)** Answer true or false for each of the following:
 - a) When stemming is used, it should be used for both indexing and query processing.
 - b) Stemming increases the size of the dictionary.
 - c) In Boolean retrieval, the use of stemming can decrease recall.
 - d) Recall/precision curves are a reasonable way to evaluate a Boolean retrieval system.
3. **(12 Points)** What is the principal *indexing* challenge faced by weblog search engines? Describe at a high level a reasonable approach to this challenge.
4. **(10 Points)** What effect does IDF weighting have on ranked retrieval results for queries that consist of a single term?
5. **(5 Points)** In the context of a tf-idf weighting why might you decide to store raw term frequencies in the postings, rather than the final tf-idf weighting for a term?
6. **(10 points)** Why might you want to store a term's postings in order of term frequency values rather than in terms of document IDs?
7. **(5 points)** What is the primary advantage of "R-precision" over the "Precision at a fixed cutoff" metric?