

## 2018 Ph.D. Program Qualify Examination – Information Retrieval

1. (25%) Please explain how to construct an *inverted file* for indexing scheme. What is the block addressing for search?
2. (25%) Please explain the concept of “Word Frequency vs. Resolving Power” for an information retrieval system in terms of word frequency and word rank.
3. (25%) Please describe the detailed procedure of how to obtain the 11-point *precision* and *recall rates* curve. Given you have a test data set consists of  $n$  documents for the purpose of information retrieval, and you can illustrate how to draw the curve from the retrieving results.
4. (25%) What is the TFIDF weighting scheme? Please give at least 3 variant TFIDF functions and explain the difference.