Short Problem Set (Module 1)

1. [25%] Approximately how many distinct words are there in the English language? Briefly justify your response.

2. [25%] Give a specific example of a word where case-folding (*i.e,* lower-casing text) can cause an IR system to make an error that would not normally occur if case distinctions were retained. Briefly explain the error.

3. [25%] Problem 1.9 (IIR, pg. 13). When the authors say "optimal", take that to mean the ordering of processing query terms that requires the least computation to produce the correct result set of documents.

4. [25%] Problem 2.8 (IIR, pg. 41).