2004 Paper 7 Question 12

Information Retrieval

(*a*) What role does stemming play in automatic indexing? [4 marks]

(*b*) Briefly describe the principles behind the Porter Stemmer. [5 marks]

(*c*) One extreme form of "stemming" is the mapping of all words of a certain prefix length onto one term. What effect will this prefix mixing have if the prefix length is extremely short, for example only two letters long? Compare this with a situation with more linguistically motivated stemming. [5 marks]

(*d*) Consider a query with four relevant documents, and a ranked IR system which returns them in the following order ("X" represents a relevant document, "–" represents an irrelevant document).

Rank	Relevance
1	X
2	X
3	–
4	–
5	–
6	–
7	–
8	–
9	–
10	X
11	X

Calculate this system's mean precision at seen documents and 11-point average precision. [6 marks]