UNIVERSITY OF OSLO

Faculty of Mathematics and Natural Sciences

Exam in INF3800/INF4800 Search Technology
Day of exam: June 8th, 2015
Exam hours: 14:30-18:30 (4 hours)
This examination paper consists of XXX page(s)
Appendices: None
Permitted materials: None

Make sure that your copy of this examination paper is complete before answering.

You can answer in Norwegian or English. Please use the language that you are most comfortable with.

STEMMING

a) False. Stemming can increase the retrieved set without increasing the number of relevant documents.
b) True. Stemming can only increase the retrieved set, which means increased or unchanged recall.
c) False. Stemming decreases the size of the vocabulary.
d) False. The same processing should be applied to documents and queries to ensure matching terms.

RELEVANCY EVALUATION

a) Average precisions:

   ariana grande: (1 + 2/3 + 3/6 + 4/9 + 5/10) / 5 = 28/45 = 0.622
   trine skei grande: (1/2 + 2/5 + 3/7) / 3 = 31/70 = 0.443
   jono el grande: (1 + 2/4 + 3/10) / 3 = 3/5 = 0.6

   MAP = (28/45 + 31/70 + 3/5) / 3 = 1049/1890 = 0.555
b) DCG = 4 + (16/log(2) + 8/log(4) + 2/log(8)) = 2 + 16/1 + 8/2 + 2/3 = 74/3 = 24.667
   Base of log doesn't matter, but rank positions is such that a choice of base 2 is obvious.
c) Ideal ranking = DCABXXX… Compute DCG for this to obtain normalization factor:
   16 + (8/log(2) + 4/log(3) + 2/log(2)) = 27.523
   Don't expect the students to evaluate the expression numerically without a calculator, but they should be able to write down the expression itself.

SIMILARITY FUN

a) See page 113 in textbook.
b) Yes, it is. Proof, where v and w are two different document vectors:

   sum[(q_i – w_i)^2] =
   sum[(q_i)^2] – 2*sum[(q_i*w_i)] + sum[(w_i)^2] =
   2*(1 - sum[(q_i*w_i)]) because q and w are unit vectors

   Thus we have:

   sum[(q_i – v_i)^2] < sum[(q_i – w_i)^2]
   ⇔
   2*(1 - sum[(q_i*v_i)]) < 2*(1 - sum[(q_i*w_i)])
   ⇔
   sum[(q_i*v_i)]) > sum[(q_i*w_i)])

LINK ANALYSIS

a) See Section 21.2, pp. 424 in textbook.

b) Note that 1/6 + ½ = 2/3, and that 1/6 + ¼ = 5/12. We then have P as in Equation (21.3), pp. 428 in textbook.
c) See Section 21.2.2, pp. 427-429 in textbook. Should show/mention iteration until convergence, and eigenvectors.