Problem 1 [7 Points]. Google and Bing, two of the most popular web search engines these days, adopt different query expansion/suggestions strategies. Which one of the two search engines provides the most useful words for query suggestions? Justify your answer.

Problem 2 [7 Points]. Find an instance of misleading anchor text on the web and show the instance.

Problem 3. Given the following web graph that includes a sample Internet consisting of three web pages:

Compute the following values, assuming that the initial PageRank (PR) values for all pages, A, B, and C, are the same, i.e., \( \frac{1}{3} \).

(a) [3 Points]. \( PR(A) \)

(b) [3 Points]. \( PR(B) \)

(c) [3 Points]. \( PR(C) \)

Problem 4. Given the following equation

\[
Q_{opt} = \alpha Q + \beta \left( \frac{1}{R} \times \sum_{i \in D_R} DOC_i \right) - \gamma \left( \frac{1}{NR} \times \sum_{i \in D_{NR}} DOC_i \right)
\]

where \( \alpha, \beta, \) and \( \gamma \) are constant, \( Q_{opt} \) is an improved query vector of initial query \( Q \), \( D_R \) is the initial set of relevant documents \( R \), and \( D_{NR} \) is the initial set of non-relevant documents \( NR \) of \( Q \).

(a) [7 Points]. Under what conditions would the optimized query \( Q_{opt} \) be the same as the original query \( Q \)? In all other cases, under what conditions would \( Q_{opt} \) not be closer than \( Q \) to the centroid of relevant documents?

(b) [7 Points]. Why is positive feedback likely to be more useful than the negative feedback to an IR system. Why might using only one non-relevant document be more effective than using several?

(c) [7 Points]. Assume that a user’s initial query \( Q \) is “cheap CDs, cheap DVDs, extremely cheap CDs.” The user examines two retrieved documents of \( Q \), \( d_1 \) and \( d_2 \), and judges that \( d_1 \), with the content “CDs cheap software cheap CDs”, is relevant and \( d_2 \), with content “cheap thrills DVDs” is non-relevant. Using the equation given above, what would the revised query vector, \( Q_{opt} \), be after relevance feedback, assuming that \( \alpha = 1 \), \( \beta = 0.75 \), and \( \gamma = 0.25 \), and the weight of each component in the initial query is defined by the frequency of word occurrence?

(d) [6 Points]. Give three reasons why relevance feedback has been little used in web searches.