INFO/CS 4300: Language and Information
Rocchio's Method for Relevance Feedback

Name: _____

NetID: _____   Left colleague NetID: _____   Right colleague NetID: _____

This is Rocchio's query update rule for relevance feedback:

$$\vec{q}_1 = \alpha \cdot \vec{q}_0 + \beta \cdot \frac{1}{|D_r|} \sum_{d \in D_r} \vec{d} - \gamma \cdot \frac{1}{|D_{nr}|} \sum_{d \in D_{nr}} \vec{d}$$

In the above, $\vec{q}_0$ is the initial query, $\vec{q}_1$ is the updated query, D_r is the set of relevant documents and D_{nr} is the set of non-relevant documents. If, after the update, there are negative weights in $\vec{q}_1$, they are set to 0.

Consider the terms:

- t_1 = kardashian

- t_2 = champion

- t_3 = olympics

and the initial query $\vec{q}_0 = (1, 1, 0)$.

The user gave relevance feedback for three documents:

Relevant.

- $\vec{d}_1 = (1, 2, 1)$

- $\vec{d}_2 = (3, 2, 1)$

Non-relevant.

- $\vec{d}_3 = (6, 0, 0)$

Let $\alpha = \beta = \gamma = 1$. What would be the new query according to Rocchio's update rule? Explain the intuition behind the new query and why it might be better than the query initially given by the user. (Assume that the IR system is using a collection of documents about the Kardashians.)