

## 1. COLLABORATIVE DOCUMENT RE-RANKING

**1.1. Approach.** When we use collaborative document re-ranking, we will re-rank the documents returned by a search engine based on comparing the user to a number of most similar users and upgrade the ranking score of a document proportional to the user similarity with other users that clicked the documents. As the search engine we are using does not provide any ranking whatsoever, we will assign a ranking score based on user similarity only. The ranking score for a document  $d$  for a given user  $u_a$  is defined as:

$$r(d, u_a) = \sum_{u_b \in U} sim(u_a, u_b) \delta(u_b, d)$$

where  $U$  denotes the set of  $k$  most similar users,  $sim(u_a, u_b)$  is the function that computes the user similarity and  $\delta(u_b, d) = 1$  if user  $u_b$  clicked document  $d$  and zero otherwise. A lot of different methods can be used to compute  $sim(u_a, u_b)$ , we have chosen a language modeling approach in which we define  $sim(u_a, u_b)$  as:

$$sim(u_a, u_b) = \frac{1}{|Q_{u_a}|} \sum_{q \in Q_{u_a}} p(q|u_b)$$

where  $Q_{u_a}$  is the set of queries from user  $u_a$  and  $p(q|u)$  can be computed as:

$$p(q|u) = \prod_{q_i \in q} \frac{tf_{q_i, u} + 1}{\sum_{x \in V_u} tf_{x, u}}$$

where  $V_u$  denotes the vocabulary of user  $u$ . We can define the vocabulary in different ways. We have evaluated two methods of collaborative document re-ranking, the first one uses the Simple User Language Model (SULM) where the user's vocabulary consists only of its queries. In addition to this, we have also defined the vocabulary as the users's queries complemented by all words from the documents the user has clicked. We call this the Extended User Language Model (EULM).

**1.2. Results.** We compare our approach against a random ranking (RANDOM) and greatest hits ranking where documents are ranked according to the number of clicks they got (POP).

**1.3. Discussion.**