Cluster Rank Demo Harness

Philip Robinson

Oregon Health Sciences University

June 24, 2017

Abstract

It is often the case that initial query compositions result in frequent restarts as the user negotiates with their retrieval system. This is likely a product of unfortunate query formulations or choice of ranking algorithm. Our proposed retrieval system encourages diversity in displayed documents by introducing an unsupervised clustering step before displaying results. The clusters are then presented to the user with their documents ranked independent of each group. We do this by clearly seperating the retrieval process into the three steps `relevance`, `clustering`, and `ranking`, then allow the user to recurse this process on a cluster (rather than restarting their query). Additionally, we propose a simple method to compare results against varying quality tfidf queries. Our final product is a demo harness that abstracts these steps, so that others may easily produce and reproduce prototypes against their own corpora.