

Web Intelligence

EXERCISE FOR LECTURE 4

1. Does the base of the logarithm used in the logarithmic scaled tf-idf scores affect the retrieved search results for a query? Why / why not?
2. Create a small example of, say, 4 documents containing terms from a vocabulary consisting of 5 different terms. Consider a query with two of these terms and evaluate scores in the tf-idf vector-space model.

When done, you may start working on the next part of your mini-project; the content-based ranking! Your constructed example can now be used for simple debugging!