

Homework Assignment 5

Homework Assignment 5 Please note that even though the homework assignments are optional, you're still highly encouraged to answer them, as they will help you prepare for your final exam. Solutions can be sent by email at ghammad@ifis.cs.tu-bs.de or can be given back to us at the lecture.

1. Clustering

Exercise 1.1 what does the Cluster Hypothesis say? How can it be exploited for information retrieval tasks?

Exercise 1.2 what is the idea behind Scatter-Gather?

Exercise 1.3 how can one determine whether a clustering is "good"?

Exercise 1.4 what is the idea underlying the k-means clustering algorithm? Using your own words, briefly explain how the algorithm works.

Exercise 1.5 what is a dendrogram and how can it be used in clustering tasks?

2. Relevance feedback

Exercise 2.1 explain in your own words the pseudo relevance feedback? What are the advantages and the drawbacks of this approach?

3. Document classification

Exercise 3.1 what are typical applications of classification algorithms in information retrieval?

Exercise 3.2 what is the difference between supervised classification and unsupervised classification? Illustrate those approaches with examples.

Exercise 3.3 how do you evaluate the quality of a classifier?

Exercise 3.4 briefly explain how Naive Bayes works. Why is it called naive? Why do we need some smoothing?

Exercise 3.5 what is adaptive boosting, what is it used for, and how does it work? Use your own words.