

An Introduction to Text Mining and Visual Text Analytics

Day 1: Text Mining

1. Basic concepts of text mining: working with electronic text, Boolean model, vector-space model, latent semantic analysis.
2. Clustering: clustering on documents, hierarchical and partitioning algorithms, co-clustering.
3. Classification: text classification, Bayesian classifiers, vector space classification, support vector machines.
4. Application: applying text mining to predict Pfam families associated with proteins.

Day 2: Visual Text Analytics

1. Introduction to Visual Analytics: definition and motivation. Text visualization and interactive visualization.
2. Algorithms for information visualization: feature selection, dimensionality reduction, topic modeling, graph reduction.
3. Programming library for visual analytics: practical introduction to D3.
4. Visual analytics tools: review of different visual analytics systems.
5. Wrap-up: Take-home messages, lessons learnt and research opportunities.