

1 Background

Books, newspaper articles, patents, service reports, ... - large amounts of information are only available in textual form. Often these valuable resources are not used, because reading and analyzing those texts would take too much effort. With the help of a combination of automatic and visual techniques analytic tasks can be supported and significantly speeded up. Our goal is to develop approaches that enable the user to analyze a set of documents with respect to a certain analysis question without reading them.

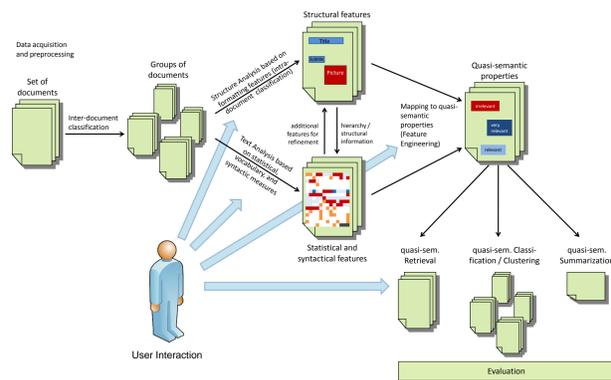


Figure 1: Research Framework for Visual Text Feature Analysis.

2 Literature Fingerprinting - Text analysis based on low-level features

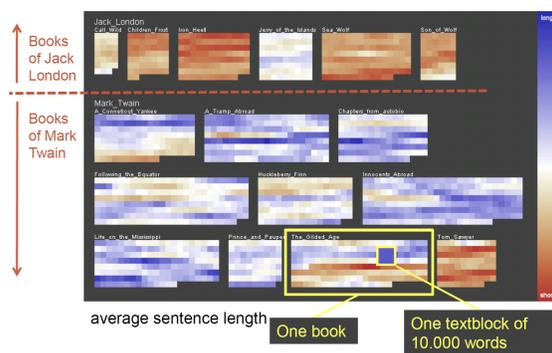


Figure 2: Literature Fingerprinting Technique. [1] Instead of calculating just a single feature value per text a sequence of feature values is extracted and presented to the user as a characteristic fingerprint. In the example above the technique is used to analyze the discrimination power of text features for authorship attribution.

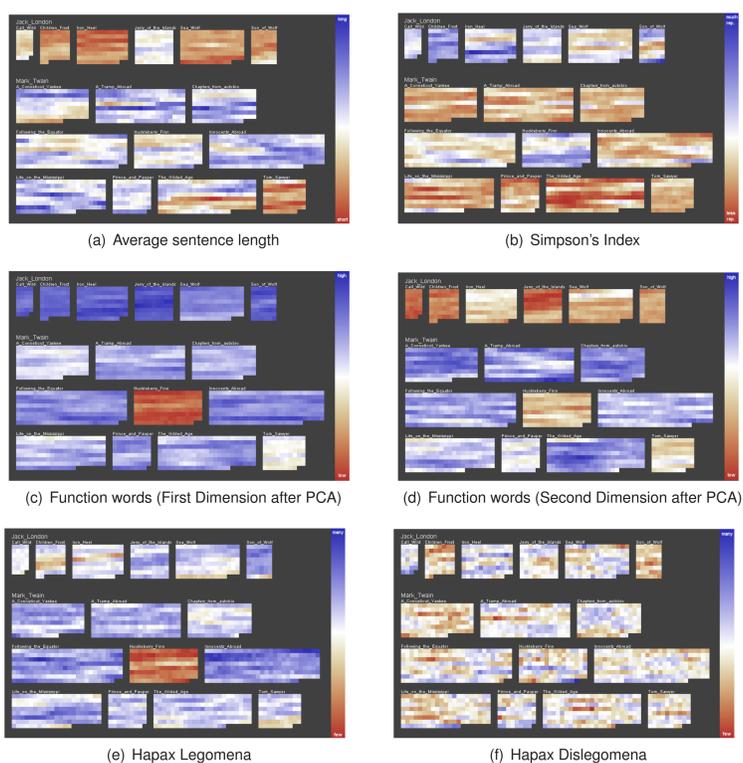


Figure 3: Different measures for authorship attribution are tested on books of Mark Twain (last three rows) and Jack London (first row).

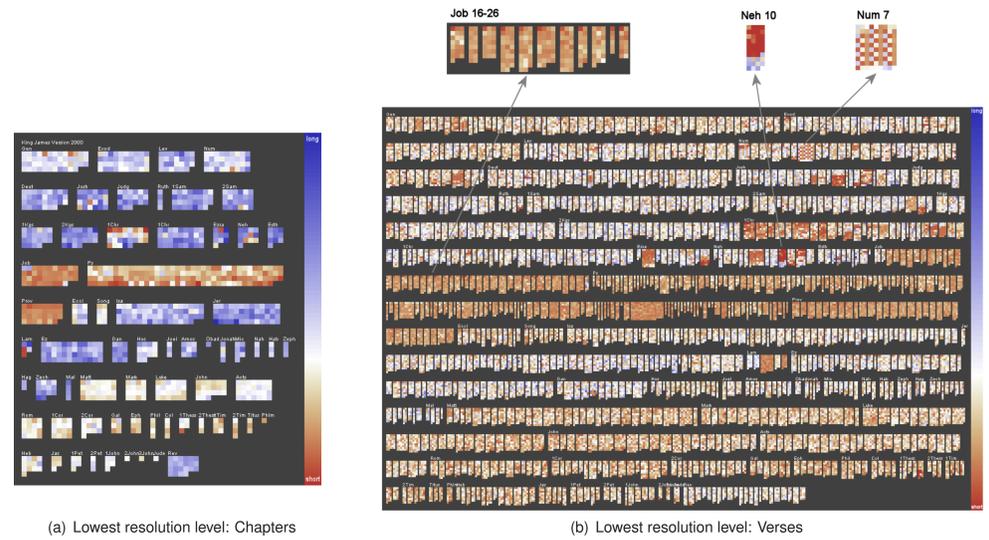


Figure 4: Visual analysis of the bible. The fingerprints are enhanced with structural information that can be used to display the text on different levels of resolution. Feature: Average verse length.

3 Future Work

- Text analysis based on quasi-semantic properties
- In cooperation with Hendrik Strobel (GK stipend, LS Deussen) and Christian Rohrdantz (GK master research student, LS Keim): Analyzing the rise and fall of topics in conference publications and the interdependencies of the conferences with respect to the topics discussed.

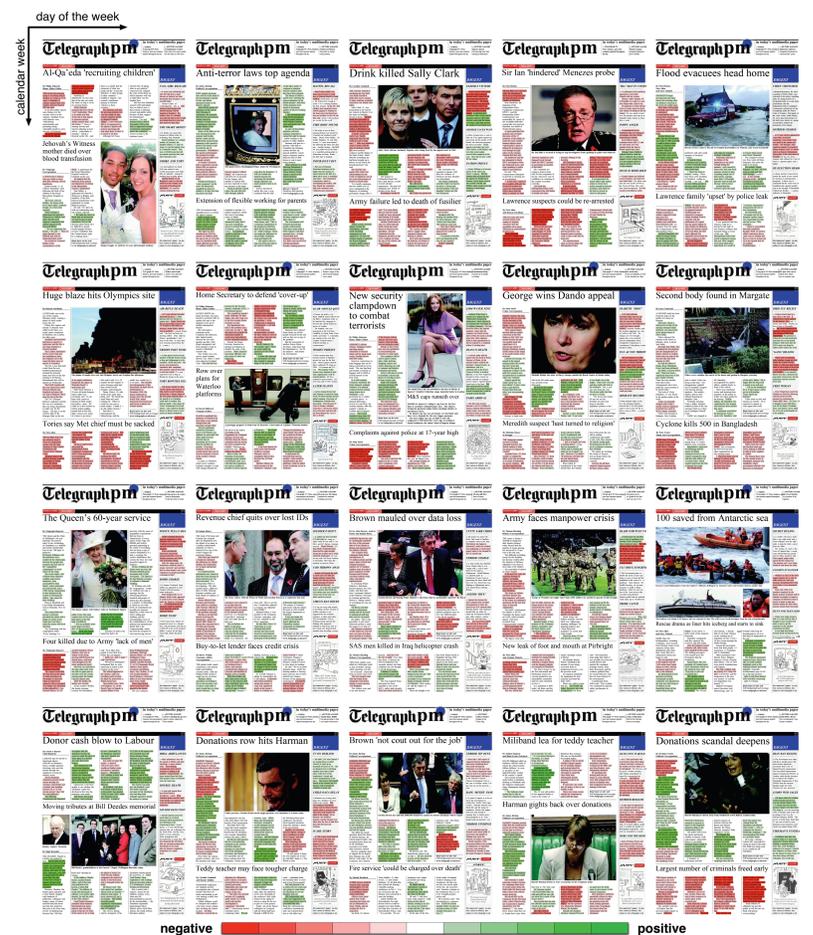


Figure 5: Title pages of *The Telegraph* in November 2007. Sentences with positive statements are highlighted in green, the ones with negative statements in red, respectively.

References

- [1] Daniel A. Keim and Daniela Oelke. Literature fingerprinting: A new method for visual literary analysis. In *Proceedings of the IEEE Symposium on Visual Analytics and Technology (VAST 2007)*, pages 115–122, 2007.
- [2] Daniela Oelke, Peter Bak, Daniel A. Keim, Mark Last, and Guy Danon. Visual evaluation of text features for document summarization and analysis. In *Proceedings of the IEEE Symposium on Visual Analytics and Technology (VAST 2008) - to appear*, 2008.