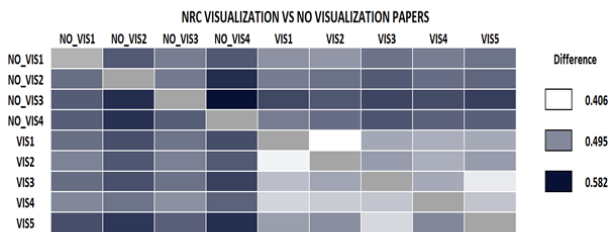


Visual Analysis of Research Paper Collections Using Normalized Relative Compression

Vazquez, Pere-Pau



The analysis of research paper collections is an interesting topic that can give insights on whether a research area is stalled in the same problems, or there is a great amount of novelty every year. Previous research has addressed similar tasks by the analysis of keywords or reference lists, with different degrees of human intervention. In this paper, we demonstrate how, with the use of Normalized Relative Compression, together with a set of automated data-processing tasks, we can successfully visually compare research articles and document collections. We also achieve very similar results with Normalized Conditional

Compression that can be applied with a regular compressor. With our approach, we can group papers of different disciplines, analyze how a conference evolves throughout the different editions, or how the profile of a researcher changes through the time. We provide a set of tests that validate our technique, and show that it behaves better for these tasks than other techniques previously proposed.