ViStA: a visualization system for exploring Arabic text

Abstract

Text visualization has become a significant tool that facilitates knowledge discovery and insightful presentation of large amounts of data. This paper presents a visualization system for exploring Arabic text called ViStA. We report about the design, the implementation and some of the experiments we conducted on the system. The development of such tools assists Arabic language analysts to effectively explore, understand, and discover interesting knowledge hidden in text data. We used statistical techniques from the field of Information Retrieval to identify the relevant documents coupled with sophisticated natural language processing (NLP) tools to process the text. For text visualization, the system used a hybrid approach combining latent semantic indexing for feature selection and multidimensional scaling for dimensionality reduction. Initial results confirm the viability of using this approach to tackle the problem of Arabic text visualization and other Arabic NLP applications. © 2015 Springer Science+Business Media New York