Abstract

Whether its consumers comparing all the available options on Amazon or data scientists analyzing datasets to find patterns and themes, users often need to explore large quantities of unstructured information beyond an individual's mental capacity to process them fully. Typically, users reduce task uncertainty by learning the unknown unknowns as they process individual pieces of information to gain deep qualitative insights. However, the cost of evaluating these insights under the global context can be high, prohibiting users to evaluate their generalizability and whether they lead to high-yield information patches. Most existing approaches either focused on aggregation techniques of unstructured data (e.g., topic modeling, review summarization and aspect extraction) or interaction techniques for exploring structured data (e.g., faceted navigation and multivariate visualizations), and do not support this process of bottom-up exploration and interpretation of unstructured online data. This thesis explores systems and interaction techniques that support users in exploring large and unstructured textual data, allowing them to both examine each piece of information to gain local insights and at the same time evaluate them under the global context. I identify and focus on two domains where addressing this issue can lead to high impact.

The first half of the thesis focuses on the domain of crowdsourced sensemaking, in which crowdworkers are limited by the scope of microtasks. The second part of the thesis focuses on supporting individual sensemaking, in which an individual explores and synthesizes online information scattered across different webpages for their own personal tasks. Through lab and field deployment user studies, I investigated the costs and benefits of the systems for supporting personal online sensemaking. This thesis makes research contributions in the design, implementation and evaluation of five interactive systems for exploring large and unstructured textual data, as well as a general design pattern for future system builders.