

Is a graph worth a thousand words? Visual Web mining for storyunderstanding and search

Bettina Berendt
Katholieke Universiteit Leuven, Belgium
<http://www.cs.kuleuven.be/~berendt/>

Abstract:

Rich information spaces like news or blogs are full of "stories": sets of statements that evolve over time, made in fast-growing streams of documents. Even if one reads a specific source every day and/or subscribes to a selection of feeds, one may easily lose track; in addition, it is difficult to reconstruct a story already in the past. In this talk, I present the STORIES methods and tool for (a) learning an abstracted story representation from a collection of time-indexed documents; (b) visualizing it in a way that encourages users to interact and explore in order to discover temporal "story stages" depending on their interests; (c) supporting the search for documents and facts that pertain to the user-constructed story stages; and (d) navigating in document space along multiple meaningful dimensions of document similarity and relatedness. This combination provides users with more control, progressing from "surfing" the Web to "sailing" selected corpora of it, semantically in story space as well as between the underlying documents. I will conclude with an overview of answered and open questions concerning the evaluation of "story" algorithms and tools.