Overview: An open source document clustering and search tool

Overview is an open-source tool originally designed to help journalists find stories in large numbers of documents, by automatically sorting them based on topics and providing a fast visualization and reading interface. It has also been used for qualitative research, social media conversation analysis, legal document review, digital humanities, and more. Overview does at least three things really well:

- It finds what you don’t even know to look for.
- It shows broad trends or patterns across many documents.
- It makes exhaustive manual reading faster, when all else fails.

Search is a wonder tool when you know what you’re trying to find—and Overview includes advanced search features. It’s less useful when you start with a hunch or an anonymous tip. Or there might be many different ways to phrase what you’re looking for, or you could be struggling with poor quality material and OCR errors. By automatically sorting documents by topic, Overview gives you a fast way to see what you have.

In other cases you’re interested in broad patterns. Overview’s topic tree shows the structure of your document set at a glance, and you can lay entire folders at once to label documents according to your own category names. Then you can export these tags to create visualizations.

Rick Davies Comment: This service could be quite useful in many ways, including clustering sets of Most Significant Change (MSC) stories, or micro-narratives form SenseMaker data, or collections of Twitter tweets found via a key word search. For those interested in the details, and preferring transparency to simple cluster analysis. Overview uses the k-means clustering algorithm, which is explained broadly here. One caveat, the processing of documents can take some time, so you may want to pop out for a cup of coffee while waiting. For those into algorithms, here is a healthier critique of careless use of k-means clustering i.e. not paying attention to its assumptions about the structure of the underlying data:

It is the combination of searching using keywords, and the automatic clustering that seems to be the most useful, to me... So far. Another good feature is the ability to label clusters of interest with one or more tags.

I have uploaded 96 blog postings from my Rick on the Road blog. If you want to see how Overview hierarchically clusters these documents let me know. I then will enter your email, which will then let Overview give you access. It seems, so far, that there is no simple way of sharing access (but I am inquiring).

Rating: 0 (from 0 votes)

Research on the use and influence of evaluations: The beginnings of a list

Published on 14 January 2015 - 12:15 PM

This is intended to be the start of an accumulating list of references on the subject of evaluation use, particularly papers that review specific sets or examples of evaluations, rather than talk about the issues in a less grounded way.

2014

2012
- 1997
- 1986

Related docs:
- Improving the use of evaluation & monitoring processes and findings. Conference Report, Centre for Development Innovation, Wageningen, June 2014.
- "A fascinating framework of four areas of factors influencing use of knowledge..." - Podcast with Robert Rebec. 

http://mande.co.uk/ accessed online 27/01/2015