

11. Exploration tool for investigative journalists

Given a pile of documents, a journalist faces the task to dig deeply into the various layers that could yield a newsworthy story on political and other powers. Especially for investigative journalists we have developed the exploration tool FIDO (*Freedom of Information Document Overview*). FIDO supports the exploration, search and analysis of large collections of texts, such as the ones that journalists can get through Freedom of Information Act-requests (WOB in Dutch).



The journalists' starting point is the Documentcloud.org framework, where WOB-documents are stored, processed and annotated. FIDO extends this tool suite in order to support search for entities and context more intelligently. It also produces powerful visualizations of social networks, word clouds, maps and timelines.

ICT science question

How can we discover in a large set of documents the key entities? How can we find the evidence central to these entities and the contextualization of mentions of the same entities in other sources? Specific technical challenges are the heterogeneity of resources to be integrated, the multi-lingual character inherent to the real-world setting and the noise resulting from optical character recognition (OCR - used to automatically generate a digital text by scanning an analog, printed text). Unique about FIDO is that it tackles these issues 'on-the-fly', instead of via a priori analysis.

Application

Our demo shows how FIDO supports example scenarios handed in by journalists. One is inspired by an investigation of the Dutch-Swiss company Vitol, the largest independent energy trader in the world. Largely unknown to the public, Vitol emerges as a spider in the political-economic web in



Matje van de Camp
infiniti.taalmesters.nl/

COMMIT/ project
INFINITI Information retrieval for Information services

FIDO. Compared to how the initial article on Vitol was written, FIDO offers a much more effective workflow.

Alternative Application

The analysis of piles of documents is useful in many places. The type of challenge faced by investigative journalists can be likened to other intelligence activities in which the starting condition is 'a pile of documents'. Therefore, alternative application areas are business intelligence, police investigation and academic research. We have plans to do usability studies with financial analysts responsible for reputation and trend analyses at a big investment management company. FIDO could increase the effectiveness of text analytics and show its relevance for generating financial forecasts.

Nice to know

The development of the demo followed an interactive user-centric design process. Investigative and data journalists were and are being interviewed on their workflow and how FIDO should be shaped to fit this workflow.



Investigative journalists dig through more and more data to give the public what it wants: freedom of information.



Semantic search: integrating the most robust natural language and search technologies for unlocking information from text at superhuman scale and speed.



The only way to keep up with companies and governments producing increasing amounts of textual data, is to move to next-generation semantic search technologies.



Semanticizing and linking heterogeneous, multilingual sources with state-of-the-art natural language processing and information retrieval technologies.

