

VIDI-Video: Interactive semantic video search with a large thesaurus of machine-learned audio-visual concepts

In this demonstration, we highlight the interactive search engine of the VIDI-Video project that facilitates access to broadcast video content at a semantic level. Most commercial video search engines such as YouTube, Blinkx, and Baidu, provide access to video based on text, as this is still the easiest way for a user to describe an information need. The labels of these search engines are based on the filename, surrounding text, social tags, closed captions, or a speech transcript. Text-based video search using speech transcripts has proven itself especially effective for segment-level retrieval from (English) broadcast news, interviews, political speeches, and video blogs featuring talking heads. A video search method based on just speech transcripts results in disappointing retrieval performance, however, when the audiovisual content is neither mentioned, nor properly reflected in the associated text. In addition, when the videos originate from non-English speaking countries, such as France, Italy, or the Netherlands, querying the content becomes much harder as robust automatic speech recognition results and their accurate machine translations are difficult to achieve. To allow for segment-level retrieval of visual content, the core of the VIDI-Video semantic video search engine is a thesaurus of automatically detected semantic concepts, like a kitchen, an airplane, or people walking in a desert. We will show how such a thesaurus of concepts can be exploited for state-of-the-art access to video at a semantic level. In addition, we will exhibit novel browsers that present retrieval results using advanced visualizations. Taken together, the interactive search engine provides users with semantic access to broadcast video archives.

<http://www.vidivideo.info/>