



Thijs Westerveld

Searching for the best search engine

Searching for text documents in a single language is one thing. Searching for pictures, audio and video fragments – multilingual if possible – is a different kettle of fish. Thijs Westerveld has been concerned with all of these aspects for several years now. He is currently trying to find a way to search for pictures within large document collections or on the web. Not just by using the caption, but by comparing image characteristics, too. How do you find a photograph of Mount Everest or the Eiffel Tower in summer?

“Language analysis forms the basis for our information retrieval methods. For text retrieval, this is quite clear. But in the case of audio or video fragments, there has to be an additional way to describe what you see or hear. By using speech recognition and by reading subtitles, you can make transcriptions allowing you to search for the right fragment. The search program then takes you directly to the fragment you want, and not just to the movie in which the fragment can be found. For this purpose, our group developed a good speech recognition system.”

“I now try to search for images in a smart way. Usually you use the captions, the file name or the description that goes with the picture. But the picture contains information in itself: as a human being, you recognise a great deal without reading the text. I therefore draw up a statistical description of the image, which is then combined with text models. Let’s say you are searching for a picture of Mount Everest. If you’re lucky, you might find a ‘mounteverest.jpg’ somewhere, but what if the name doesn’t relate to the content? Combining both types of information will enable a better cross-modal search. You will be able to search for pictures with a certain atmosphere and with certain colours and patterns. However, combining both worlds involves a lot of mathematics.”

“As the amount of information is growing, search demands are getting more sophisticated by the day. We have to fine-tune our methods continuously. These work in an increasingly cross-modal way: obtaining the information using all the resources we have. Needless to say, we look at search engines for the Web. But that’s not the only thing: companies now have large digital archives with texts, but also with complete movies. We often co-operate with broadcasting companies, for example. We are currently working on the digitisation of the Dutch ‘Polygoon News’, which used to be shown in the cinema. As this material dates back many decades, you have problems recognising some of the words or phrases and the quality varies, etc. Furthermore, we are making a daily transcription of the broadcasts of

Examples of current projects:

- Multimedia Indexing and Searching Environment (MUMIS; EU Fifth Framework Program)
- Digital Media Warehousing (DMW; Telematics Institute)
- Secure Multimedia Retrieval (SUMMER; Ministry of Economic Affairs)
- Multimodal Meeting Manager (M4; EU Fifth Framework Program)

the major news network in the Netherlands. We use a great deal of daily newspaper material for the necessary textual input.”

“It is possible to test the performance of your search engine at the annual TREC conference in the US. They send a large document collection in advance, so everyone does the same test. Some years ago, our natural language search method performed quite well and we expect the video search to do the same. Getting a ‘TREC record’ is not the ultimate goal, of course, but it is a very nice result to take home.”

