There's more to than meets the

Ineke Temming

Optimizing image retrieval - and other magic stories

Martha Larson, assistant professor in the Department of Intelligent Systems, fills us in on some of the fascinating research projects of the Delft Multimedia Information Retrieval Lab (DMIR).

Does this sound familiar? You type a search query in YouTube, you click on a few results, which are interesting but not what you are looking for, you load more suggestions that fulfill your information need even less, and after an hour of watching cool videos, you've forgotten why you went to YouTube in the first place. "This is a general problem with all social multimedia", says Martha Larson. "There's so much information available that it gets harder and harder to easily and effectively access content. In the Delft Multimedia Information Retrieval lab we are working on innovative solutions to this problem."

Multidisciplinary background
Martha, who joined TU Delft in 2008, is the classic example of a multidisciplinary researcher. She has a Bachelor's degree in Mathematics & Electrical Engineering, and a Master's in Formal Linguistics. For her PhD she spent three years in West Africa studying languages, to understand how the human brain constrains the possible structures of language. Envisioning a career in speech technology, she had a job as a computer programmer while writing her thesis. From 1999 till 2006 she worked for the Fraunhofer Institute for Intelligent Analysis and Information Systems, followed by two years on the 'MultiMatch: Multilingual/Multimodal Access to Cultural Heritage' project at the University of Amsterdam. "Here I got to know the Netherlands Institute

for Sound and Vision, which for me, as a language- and multimedia-oriented person, is one of the cathedrals of Europe!" At TU Delft Martha's research interest shifted from speech recognition techniques to social multimedia.

Video filtering by search intent
"Search results are often disappointing because existing search engines focus on the 'topic', the things that are visually depicted in an image", she continues. "When you type in the word 'cat', you get an overwhelming number of hits. Our solution is to go beyond the topic of the photo or video, and look at what the multimedia are used for. PhD candidate Christoph Kofler, who is also a Google Fellow, is developing an approach that will allow YouTube to differentiate between search intents. Currently, when you type in 'salsa', you end up with a lot of videos that don't serve your purpose. If you want to learn how to dance the salsa, you need a tutorial, but if you want to enjoy great dancing you'd better watch a dance contest. People often don't realize what their goals are, so only a minority would add a word like 'tutorial' in the search query. But in text-based search systems this only works if the person who uploaded the video has included the same word in the description."

Optimizing image retrieval
Together with MSc student Michael Riegler, Martha is working on the optimization of image retrieval by

focusing on the differences in photographic style and framing. "We exploit the fact that a mugshot is framed differently than a 'street style' fashion photo, is framed differently than a photo you take to remember an event, and so on. Each category has its own distinctive features in terms of lighting, focal point, composition, etcetera, but it's important to emphasize that we actually do not know what these are. We use crowdsourcing to discover what kind of pictures would fulfill certain information needs; or we find regularities by processing huge amounts of data from query logs, where we can see what people have clicked on. In the end it's a big data problem, and the approaches are often statistical. It can happen that we find some totally unexpected regularity which we are then able to exploit - this is the cool part about it. You just take a leap of faith that there is more to multimedia than meets the eye."

Revealing the inner structure of videos
Another problem with YouTube is that you always get the complete video, also when you're just interested in a particular part. PhD candidate Raynor Vliegendhart is developing an enhancement on existing video players that allows users to jump straight to the point where it gets interesting. Martha: "This is a research project that has to be done in academia, as YouTube's current business model requires people

multimedia eye

to stay in the network as long as possible, so that they click ads. In DMIR we aim to be very closely connected to the user's needs, which is typically 'Delft' as people here will

many more possibilities to interact with online course material, such as jumping to key parts, creating automatic summaries, hyperlinking between course videos, and so on."

sustainable, it's necessary to support people in understanding the implications of sharing personal data and in doing this safely and responsibly. "Ultimately there will be systems that will ask questions like: 'do you realize your camera has added geographic coordinates to your photo?' or that will automatically make license plates invisible. The social multimedia are great, provided that we are aware of the potential dangers to our privacy." ■

> *"In DMIR we aim to be very closely connected to the user's needs"*

always ask you: what is society getting out of it? I think online education, for example, is an area to benefit from our multimedia research. Students will have

Making social multimedia sustainable

It's Martha's conviction that in order to make social multimedia like YouTube