UiB Informatics

Department Seminar

Lost in Publications?

How to Find Your Way in 50 Million Scientific Documents by Machine Learning and Interactive Intent Modelling

Jaakko Peltonen

COIN & HIIT, Aalto University
& University of Tampere

Thursday, 2016-08-18, 14h15

Store Auditorium, 2nd floor,
Høyteknologisenteret (data blokk)

Abstract

Before one can analyze relevant data one must first find it from among the huge amount of available data. The most common example is information seeking from large masses of documents, whether it is from the general web or from large collections. Often the search is exploratory, and it may be hard to formulate a good query, but traditional information retrieval systems do not sufficiently help users to improve unsatisfactory results. In this talk I discuss an improved system for exploratory search, where users are given power to direct their search by interacting visually with a model of their search intent.

My talk concentrates on a particular domain: information seeking of scientific documents. Finding relevant documents is a common task for researchers, who must navigate big data to keep up to date with ongoing research and place their own work in context. Current scientific knowledge includes more than 50 million published articles -- among such a huge mass of data, how can a system help a researcher find relevant documents in their field?

We introduce SciNet, an interactive search system that anticipates the user's search intents by estimating them from the user's interaction with the interface. The estimated intents are visualized on an intent radar, a radial layout that organizes potential intents as directions in the information space. The system assists users to direct their search by allowing feedback to be targeted on keywords representing the potential intents. Users can provide feedback by moving the keywords on the intent radar. The system then learns and visualizes improved estimates and corresponding documents. The resulting user models are explicit open user models curated by the user during the interactive information seeking. SciNet has been shown to significantly improve users' task performance and the quality of retrieved information without compromising task execution time. We also show how user models learned in SciNet can be used to help cold-start recommendation in another system, the CoMeT talk management system, by cross-system user model transfer across the systems.