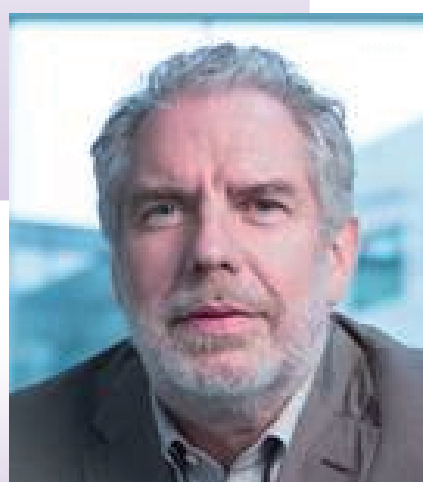


『Tsinghua Information Forum #100』 —Samsung #4』



■ **Title:** Data Overload: What Can We Do?

■ **Speaker :** Prof. Dr. Peter M.G. Apers

■ **Time :** 2013-4-19, 14:30-15:30pm

■ **Venue:** East Main Building #10-101

Abstract :

ICT is changing the world. It is changing both individuals and our society. One very intriguing change is that everybody has become both a consumer and a producer. Before, the production of books and music went through a whole procedure to guarantee quality. Nowadays everyone is putting Youtube videos on the internet, regardless of the quality. People no longer accept to pay perceived high prices. Many companies and organizations have to rethink their business model. All traditional institutions have to rethink what their added value is. Also the way we behave is changing. Many people no longer read a book or watch a movie without continuously checking their mail or social media. Furthermore, Google gives us the feeling that we have access to all the information in the world, resulting in delaying decisions because obtaining more data may lead to perceived better decisions. All this results in a number of issues that require attention:

- the amount of data is increasing at a tremendous speed, making us actually close to blind;
- the average quality of the data is going down, many data is contradicting each other, can we still trust the data on the internet?
- is it possible to extract reliable knowledge from the Internet?

During the talk, we will discuss a couple of research projects from our group. The research goal of the group is to design algorithms to provide high quality and highly relevant information along the following lines:

- Filtering or searching in a decentralized way and for specific groups. The idea is to involve producers of data in the search. The challenge is to achieve high quality results in a time-efficient way. Furthermore, children search in a complete different way than adults.
- Disclosing semi-structured data, for example from the Deep Web or from logs. Many reliable data sources are hidden in databases hidden behind web services. The challenge is to include these in a simple and uniform way.
- Providing transparency by keeping track of how the data presented is obtained. More and more decisions are taken based on data of which the way it is obtained is not clear. Data provenance keeps track on how data is obtained. The challenge is to do this in a memory-efficient way.
- Extracting knowledge from large data sets from various sources. The challenges are to link data from different sources based on content, location, time, etc. and to do it in a time-efficient way.

Biography :

Since 1985 Peter M.G. Apers is a full professor in Computer Science at the University of Twente, the Netherlands. In this role he is involved in a large number of externally funded projects. After being a member of the Executive Board of the University of Twente (responsible for Research and Innovation), he is currently Scientific Director of the Centre for Telematics and Information Technology (<http://www.ctit.utwente.nl>) of the University of Twente. CTIT, with more than 400 researchers, has a focus on ICT as a technology and on ICT applications.

At the national level he is involved in formulating the ICT-policy: in the past as chairman of the Computer Science section of NWO, member of the ICT Taskforce, member of the ICT Forum, and the board of Exact Sciences of NWO, as member of the Advisory Board of ICTRegie and now as scientific leader of the ICT Roadmap. Together with others he has taken the initiative to found NIRICT (Netherlands Institute for Research on ICT), the ICT institute of the three Dutch Universities of Technologies.

Since 2005 he is chairman of the Technology Foundation STW.

Organizer: School of Information Science and Technology

Contact: Ling Feng (62773581)