

Peer-to-Peer Information Retrieval

Almer S. Tigelaar

Database Group, Faculty of Electrical
Engineering, Mathematics and Computer Science,
University of Twente, P.O. Box 217, 7500 AE,
Enschede, The Netherlands.
a.s.tigelaar@utwente.nl

THE Internet has become an integral part of our daily lives. However, the essential task of finding information is dominated by a handful of large centralised search engines. In this thesis we study an alternative to this approach. Instead of using large data centres, we propose using the machines that we all use every day: our desktop, laptop and tablet computers, to build a peer-to-peer web search engine. We provide a definition of the associated research field: peer-to-peer information retrieval. We examine what separates it from related fields, give an overview of the work done so far and provide an economic perspective on peer-to-peer search. Furthermore, we introduce our own architecture for peer-to-peer search systems, inspired by BitTorrent.

Distributing the task of providing search results for queries introduces the problem of query routing: a query needs to be sent to a peer that can provide relevant search results. We investigate how the content of peers can be represented so that queries can be directed to the best ones in terms of relevance. While cooperative peers can provide their own representation, the content of uncooperative peers can be accessed only through a search interface and thus they can not actively provide a description of themselves. We look into representing these uncooperative peers by probing their search interface to construct a representation. Finally, the capacity of the machines in peer-to-peer networks differs considerably, making it challenging to provide search results quickly. To address this, we present an approach where copies of search results for previous queries are retained at peers and used to serve future requests and show participation can be incentivised using reputations.

There are still problems to be solved before a real-world peer-to-peer web search engine can be built. This thesis provides a starting point for this ambitious goal and also provides a solid basis for reasoning about peer-to-peer information retrieval systems in general.

Thesis links:

<http://almer.tigelaar.net/2012/09/25/peer-to-peer-information-retrieval/>

<http://dx.doi.org/10.3990/1.9789036534000>

Supervisor: Peter Apers (University of Twente, NL), **Assistant Supervisor:** Djoerd Hiemstra (University of Twente, NL). **Committee:** Jamie Callan (Carnegie Mellon University, US), Fabio Crestani (Università della Svizzera Italiana, CH), Arjen de Vries (Delft University of Technology, NL), Franciska de Jong (University of Twente, NL), Dirk Heylen (University of Twente, NL), Johan Pouwelse (Delft University of Technology, NL). **Defended:** Wednesday, September 26th 2012 at the University of Twente, Enschede, The Netherlands.