

# Personalized Query Routing in Peer-to-Peer Federations of Digital Libraries

Matthias Bender, Yannis Ioannidis, Donald Kossman, Henrik Nottelmann, Hans-Jörg Scheck, Gerhard Weikum, Pavel Zezula, Christian Zimmer  
**Contact Author: Gerhard Weikum, Max-Planck Institute for Informatics, weikum@mpi-inf.mpg.de**  
**Poster by: Christian Zimmer, Max-Planck Institute for Informatics, czimmer@mpi-inf.mpg.de**

## Research Objectives

The peer-to-peer (P2P) paradigm is an intriguing approach for coping with dynamically evolving federations of loosely coupled digital libraries. In addition to the libraries, user agents with powerful personalized tools may participate as peers, too. Such a P2P system promises unlimited scalability, robustness to failures, fluctuation, and load dynamics, and also much reduced vulnerability to attacks and information manipulation. Peers remain autonomous and participate in the P2P system and collaborate on behalf of user requests at their discretion. In a typical scenario, a user peer would utilize its local profile to compile an information demand and issue queries to the library peers that most suitable for the given demand. The decision about where a query should be sent is known as *query routing* or *database selection* or *resource selection*.

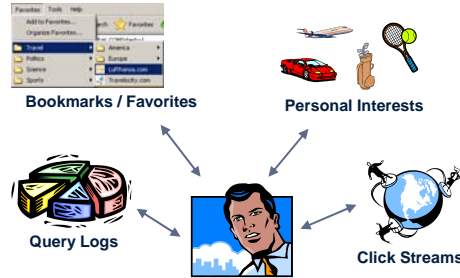
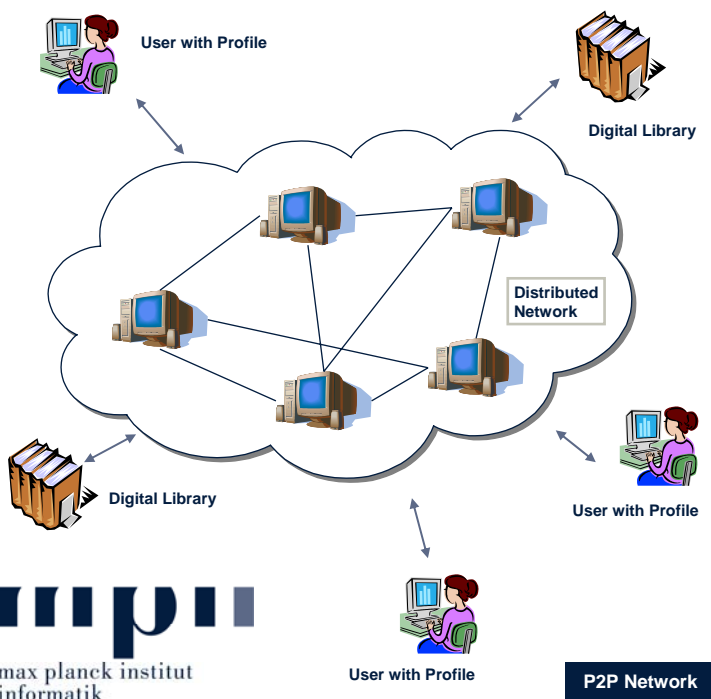
**„Leveraging the Power of Peer-to-Peer Systems for Collaborative Search“**

## New Challenges

Previous work on database selection assumed a static architecture with a fairly small a number of data sources. The P2P scenario is different:

- ✓ Peer-to-Peer is large scale and high dynamics such that the dissemination, maintenance and effective use of statistical summaries about peers' information content entails difficult questions.
- ✓ Personal Strategies for query routing based on local knowledge, query-logs, and further user-behavior with other peers provide new opportunities to improve query execution.

**Main Question:** How can this potential be effectively exploited?



- Personal Infos**
- Bookmarks
  - Query Logs
  - Click Streams
  - Overlap Awareness
  - User Behavior
  - User Profile
  - Personal Interests

## Proposed Work

Information request posed by the user can vary from keyword queries to structured queries (SQL or XQuery) and may include entire sessions of searching, browsing, annotation, classification, and refinement steps.

**How to exploit the different kind of user-behavior information for the purpose of query routing in the distributed P2P federation?**

- ✓ How can and should user-behavior information be represented in a compact form, and where and how should it be made available - on demand or by proactive dissemination in a "gossiping" style?
- ✓ Once a compact representation is found, how is it used for deciding to which peers an information request should be sent? How do we choose the most suitable peers for a given query - in terms of similarity to the combination of query and user profile, and at the same time avoid that queries are redundantly sent to many peers that provide more or less the same information?
- ✓ What kinds of statistical models are most appropriate for matching user profiles and queries with the contents of peers? How expensive is their computation? How do we cope with the tradeoffs in expressiveness vs. complexity?

## Approach and Expected Outcome

- ✓ Designing models and strategies for personalized query routing and implementing the most promising approaches in a demonstrator testbed.
- ✓ The evaluation of strategies poses a critical issue and includes questions about appropriate datasets, data placements onto peers, workload generation, data and system perturbation, and quality measures of interests. This leads to the first benchmark for query routing in P2P digital libraries.
- ✓ An application-oriented demonstrator for personalized search on e-health information.

## References

- [1] Matthias Bender, Sebastian Michel, Christian Zimmer, Peter Triantafyllou, Gerhard Weikum: Improving Collection Selection with Overlap-Awareness. 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR05).
- [2] Bender M., Michel S., Zimmer Ch., Weikum G.: Leveraging the Power of Peer-to-Peer Systems for Collaborative Web Search. Digital Library Architectures. Preproceedings of the 6th Thematic Workshop of the EU Network of Excellence DELOS.
- [3] Matthias Bender, Sebastian Michel, Gerhard Weikum, Christian Zimmer: Bookmark-driven query routing in peer-to-peer web search. Proceedings of the SIGIR Workshop on Peer-to-Peer Information Retrieval: pages 46-57.
- [4] Georgia Koutrika, Yannis E. Ioannidis: Personalized Queries under a Generalized Preference Model. Int. Conf. on Data Engineering (ICDE), 2005.
- [5] Henrik Nottelmann, Norbert Fuhr. Evaluating different methods of estimating retrieval quality for resource selection. ACM SIGIR Conference on Research and Development in Information Retrieval, 2003.
- [6] David Nowak, Pavel Zezula: Indexing the Distance Using Chord: A Distributed Similarity Search Structure, DELOS Workshop on Future Digital Library Management Systems, 2005.
- [7] M. Springmann, H.-J. Scheck, H. Schuldt: Combination of Building Blocks for Similarity Search in Multimedia Electronic Patient Records (in German). Proceedings of 49th Conf. of the German Society for Medical Informatics, Biometry, and Epidemiology (GMDS), 2004.
- [8] Martin Theobald, Claus-Peter Klas: BINGO! and DAFFODIL: Personalized Exploration of Digital Libraries and Web Sources, 7th International Conference on Computer-Assisted Information Retrieval (RIAO), 2004.

**ECDL 2005**

9th European Conference on Research and Advanced Technology for Digital Libraries  
September 18-23, 2005 Vienna, Austria

Poster by  
**Christian Zimmer**



This joint research is supported by the EU Network of Excellence **DELOS**. The authors of this extended abstract are the principal investigators of the participating institutions: 1) the **Max-Planck Institute for Informatics Saarbrücken**, Germany, 2) the **National University of Athens**, Greece, 3) the **University for Health Sciences, Medical Informatics and Technology in Innsbruck**, Austria, 4) the **University of Duisburg-Essen**, Germany, 5) the **Masaryk University in Brno**, Czech Republic, 6) the **Swiss Federal Institute of Technology Zurich**, Switzerland.