

# Decentralized Indexing in a Peer-to-Peer Search Engine

Supervised by Prof. Dr.-Ing Gerhard Weikum

Begin of studies: October 2003

## INTRODUCTION

The Peer-to-Peer (P2P) approach allows handling huge amounts of data in a decentralized and self-organizing way. All peers are equal and the functionality of the whole system is shared among the participants. P2P systems have considerable advantages:

- no single point of failure
- load balancing
- high scalability
- self-organization
- autonomy

These characteristics offer potential benefits for building a search engine in a P2P environment. Additionally, a P2P search engine can possibly exploit the intellectual input of a large user community.

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

## DESIGN FUNDAMENTALS

The Peer-to-Peer (P2P) search engine combines different design decisions which influence the properties of the whole system:

- Every participant manages its own local index containing document information; every peer works as a local search engine.
- Peers share their local indexes (or specific fragments) by posting meta-information into the P2P network, thus effectively forming a large global, but decentralized directory.
- The global directory is maintained as a distributed hash table (DHT). This underlying routing system (for example a Chord ring) performs a mapping from keys to locations in an entirely distributed manner and ensures a high scalability (figure 1).

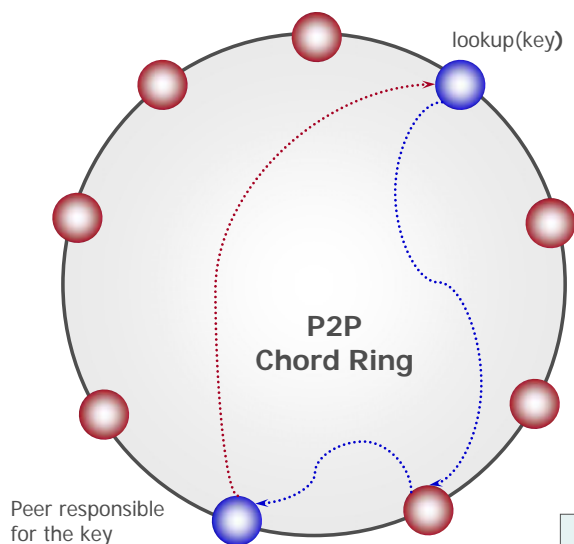


Figure 1

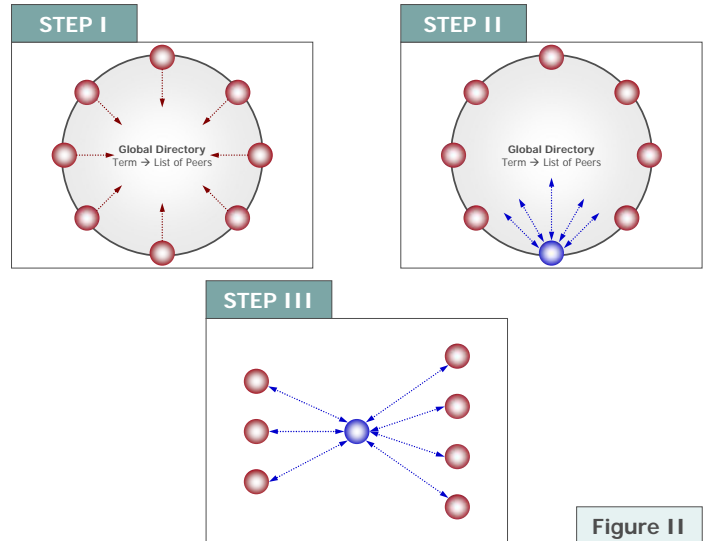


Figure II

## SEARCH ARCHITECTURE

Searching in the Peer-to-Peer environment is divided into different stages (figure II):

- Every peer posts meta-information about its local index to build up the conceptually global but physically distributed directory.
- To process a query consisting of several terms, a peer asks the peers responsible for a certain term of the query and collects lists of peers having information about the terms.
- The requesting peer merges the peerlists and contacts the best peers with the full query. The combination of the individual results produce the final result of the query.

## FUTURE WORK & THESIS GOAL

Future work will address the implementation of a prototype allowing the evaluation and comparison of different strategies:

- How to build up the global directory? What kind of QoS measures improve the peer selection?
- How to compute global authority scores or idf values in a P2P environment?
- Can user-specific bookmarks of web pages enhance collaborative search?
- How to compute relative quality measures in a decentralized manner (e.g. MI value)?

## REFERENCES & PROJECTS

- [1] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan. Chord: A scalable peer-to-peer lookup service for Internet applications. Technical Report TR-819, MIT, March 2001.
- [2] Cuenca-Acuna, F. M., Peery, C., Martin, R. P., und Nguyen, T.D.: PlanetP: Using Gossiping to Build Content Addressable Peer-to-Peer Information Sharing Communities.
- [3] 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.
- [4] Bender M., Michel S., Zimmer Ch., Weikum G.: Bookmark-driven Query Routing in Peer-to-Peer Web Search. Proceedings of the SIGIR Workshop on Peer-to-Peer Information Retrieval.

**DELIS** Dynamically Evolving, Large Scale Information Systems (EU)

**P2E2** Peer-to-Peer Enterprise Environment (BMBF)