P2P Web Search: Make It Light, Make It Fly

Matthias Bender, Tom Crecelius, Sebastian Michel, Josiane Xavier Parreira
Max-Planck-Institut für Informatik
Saarbrücken, Germany

Jan 8, 2007
Outline

1. Motivation
2. MINERVA System Model
3. Demo
Motivation

Potential of peer-to-peer (P2P) systems
- scalable
- efficient
- resilient to failures and dynamics

P2P Web Search
- benefit from intellectual input of user community
- prevent information-resource monopolies

Key Challenge: Make it usable
- Bingo! (Focused Web crawler), MINERVA (P2P Web Search), Cloudscape (Database Backend)
  ⇒ Easy user deployment, common GUI
Architectural Model

- Peers autonomously and independently crawl the web
- build local indexes
- can execute queries on local indexes
Architectural Model

- share **per-key statistics** about local indexes
- form conceptually global **directory**
- Each directory peer responsible for randomized subset of keys
Architectural Model

- use directory to identify promising peers (Query Routing)
- send query to selected peers
- merge results appropriately
1. Step: Import Bookmark File

- Click Start button to import bookmark file
2. Step: Crawl the Web

- BINGO! fetches bookmark documents, builds classifier
- Web crawl starts automatically, can be stopped anytime
- Start Minerva
3. Step: Instantiate Minerva

- Enter nickname, network settings
- Click Create Ring (or Join Ring, if exists)
4. Step: Update Minerva Index

- **Update Index** imports current data from Bingo!, computes metadata
5. Step: Publish Metadata to directory

- Use Show Widgets menu to open Received-Posts panel
- Click Post All button
- Click Refresh button to inspect directory
6. Step: Execute Query

- Enter Query, click Execute query
- Inspect Results
7. Step: Tag Documents

- Right-click on document, *Add New Tag for URL*
- Add desired tag(s)
8. Step: Retrieve Annotations for Document

- Right-click on document, *Retrieve all Tags for URL*
- Inspected submitted tags
9. Step: Query for Annotated Documents

- Enter query as tag=value, possibly combine with keywords
- Inspect results
Thank you for your attention

Download and more details available at:
http://www.minerva-project.org