
Knowledge Sharing in Multi-Layer P2P Networks

Paraskevi Raftopoulou & Euripides Petrakis
Dept. of Electronic & Computer Engineering
Technical University of Crete
NCC Workshop - 25/09/2005

Motivation

- Consider a **knowledge sharing** scenario
e.g., users searching for educational material, digital libraries, ...
- Various **aspects of knowledge** specific to
 - Data type
 - Data representation
 - Users
 - ...
- Various **ways to relate** data items
e.g., different similarity measures, ...

Which is the best way to exploit the various cases and provide effective query processing and data retrieval?

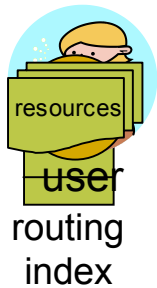
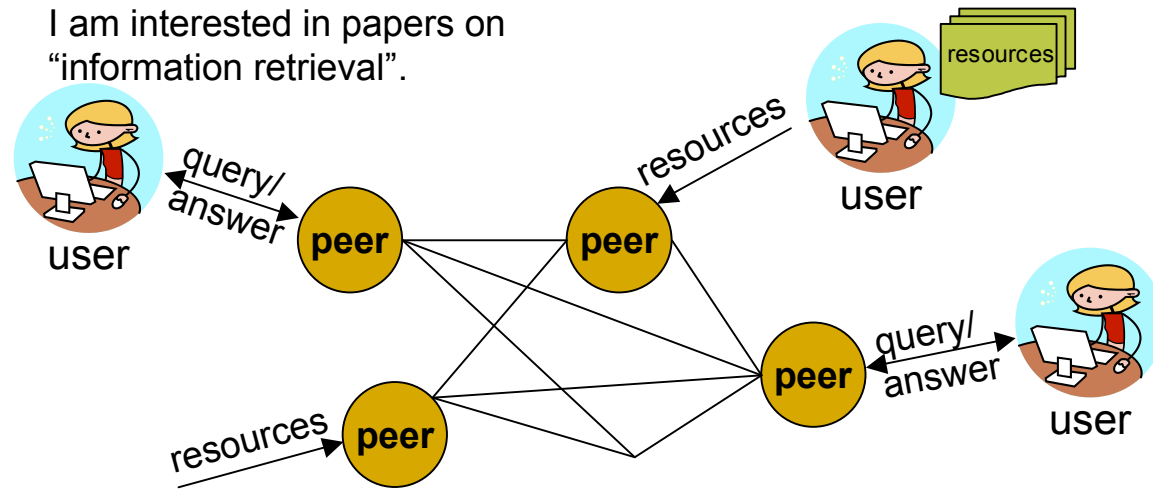
Outline

- Related research
- Scenario
- Current approach
- Multi-Layer Graph Model (MLGM)
- Multi-layer search
- Conclusions & Future work

Related research

- Information retrieval [BR99]
- P2P networks (e.g., [Sch04], [LT03], ...)
- Structured overlay networks (e.g., [CG03], [SML03], ...)
- Clustering of content [TXD03] / of peers [LWS03]

Scenario



users generate queries for information retrieval and information is downloaded from the Web
peers provide information to peers and by information downloaded from the Web

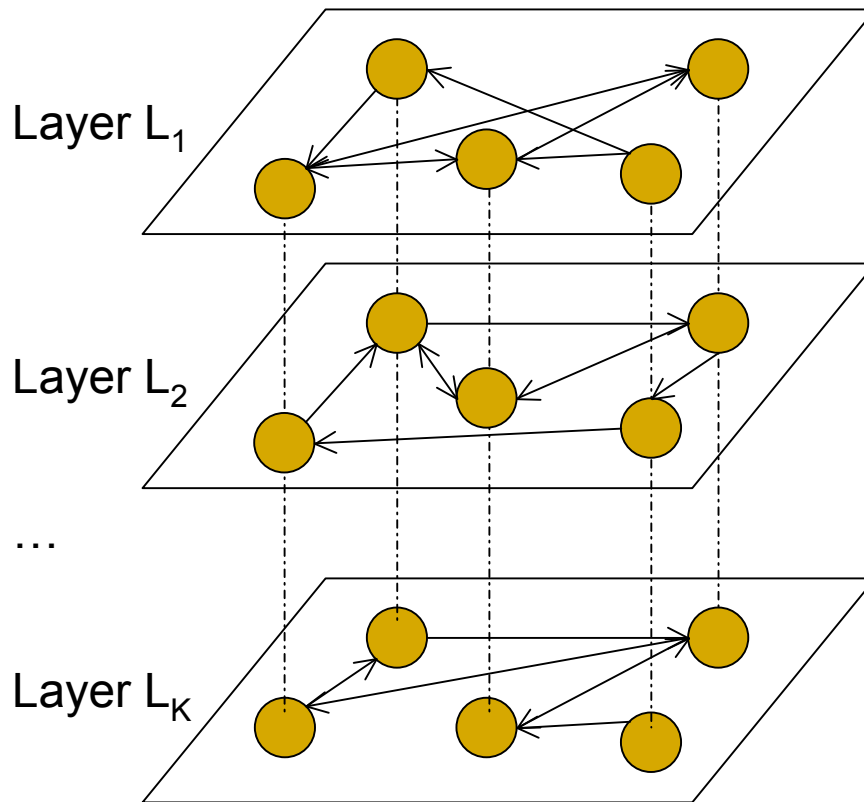
An example

- Peers contribute **documents** to the network
- Each document is represented by a **term vector**
- The **similarity** between two documents is computed according to the **VSM**
- Given a query all documents are **ranked** according to the similarity with the query

Current approach

- Apply **clustering** to the set of data
- Different **aspects of knowledge** are clustered in **separate layers**
- For a given query the **most promising cluster** at each layer is identified and searched
- The matching resources obtained from each layer are **combined** into a single answer set

MLGM



- K knowledge layers
- N peers
- A knowledge layer L_i is a directed graph $G=(V,E)$
- V corresponds to **peers**
- E corresponds to **links**
- Vertices of the various layers correspond to the **same set** of peers
- Edges depend on the **knowledge** specific to the layer

P2P Clustering

- A **distributed clustering** mechanism is needed because of large volumes of data
- Each peer organizes its resources into **clusters**
- Each cluster is represented by its **centroid**
- The centroids constitute the **description** of the peer
- Peers are organized into groups of similar interests using their descriptions and the idea of **walks**

walks: peer descriptions are forwarded in the network by some routing strategy and peers obtain information about the interests of other peers

Multi-Layer Search

- At each layer
 - query is matched against peer descriptions
 - the most promising peer clusters are identified
 - query is routed to the best matching peer cluster
 - a set of matching resources is returned
- Answers obtained from each layer are combined according to a heuristic mechanism into a single answer set

Conclusions & Future work

- Multi-layer network model for
 - efficient exploitation of different aspects of knowledge
 - effective organization of contents
 - load balancing
 - ...
- Future work
 - Identify different information aspects
 - Propose a layer and its organization specific to each information aspect
 - Implementation

Thank you!

Citations

- [BR99] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison Wesley, 1999.
- [Sch04] C. Schmitz. Self-Organization of a Small World by Topic. In *Proc. of 1st Intern. Workshop on Peer-to-Peer Knowledge Management*, Boston, MA, August 2004.
- [LT03] A. Loser and C. Tempich. On Ranking Peers in Semantic Overlay Networks. In *Proc. of 3rd Conference on Professional Knowledge Management*, 2005.
- [CG03] A. Crespo and H. Garcia-Molina. Semantic Overlay Networks for P2P Systems. Technical Report, Stanford University, May 2003.
- [SML03] I. Stoica, et al. Chord: A Scalable Peer-to-Peer Lookup Protocol for Internet Applications. *IEEE/ACM Transactions on Networking*, 11(1):17–32, February 2003.
- [TXD03] C. Tang, et al. Peer-to-Peer Information Retrieval Using Self-Organizing Semantic Overlay Networks. In *ACM SIGCOMM 2003*, Karlsruhe, Germany, August 2003.
- [LWS03] A. Loser, et al. Semantic Overlay Clusters within Super-Peer Networks. In *Proc. of Intern. Workshop on Databases, Information Systems, and P2P Computing, in 29th Intern. VLDB Conference*, Berlin, Germany, September 2003.