JACOBS UNIVERSITY

Computer Networks and Distributed Systems

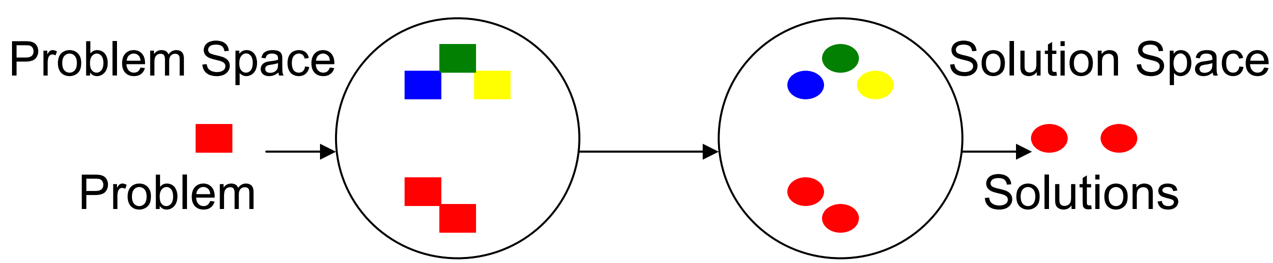
Heuristic Search using a Feedback Scheme in Unstructured Peer-to-Peer Networks

Ha Manh Tran and Jürgen Schönwälder
Computer Science, Jacobs University Bremen, Germany

International Workshop on Databases, Information Systems and Peer-to-Peer Computing
University of Vienna, 24 September 2007

DARING: Distributed Case-based Reasoning System

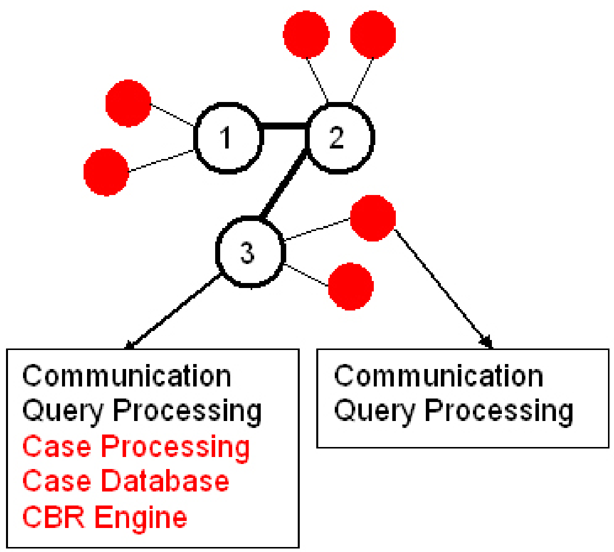
- Assisting network operators in resolving faults in large-scale, diverse communication systems
- Searching for solutions by previous experience sharing in decentralized environments
  - P2P to explore ubiquitous fault cases
  - CBR to exploit relevant fault cases



Case-based Reasoning

DARING Overview

- Unstructured P2P network overlay
  - Search performance issue
- Super peers bearing CBR engines
  - Reasonable bandwidth and power processing
- CBR engines proposing fault-matching solutions (proposed solutions)
  - Local case database and reasoning engine
  - High computational resource consumption



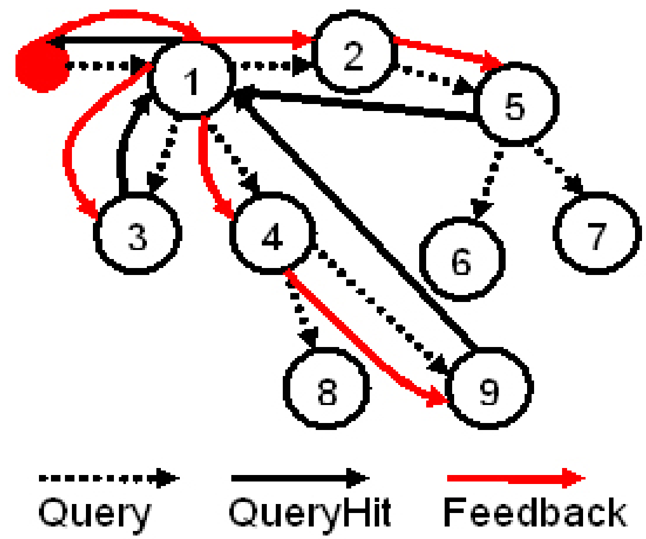
Heuristic Search in DARING

- Finding solutions:
  - Flooding the overlay is avoided
  - Proposed solutions are probably incorrect
  - Finding promising peers by looking at their previously accepted solutions

- Using a feedback scheme to announce the accepted solutions to peers

Feedback Scheme

- A querying peer verifies and accepts solutions among fault-matching solutions, then feedback the accepted solutions to specific peers.



- Upon receiving the feedback, any peer learns solutions and peers for subsequent queries

Algorithms

- Peer learning
  - Learning from feedback
  - Updating the lists of good peers (expertise values) and queries (query information)
- Peer ranking
  - Finding similar queries and corresponding peers
  - Ranking these peers
- Peer selection
  - Selecting peers from lists of good peers, recently active peers, and random peers in order
  - At least, one random peer

Similarity Function

- Learning and ranking algorithms
- Field-value pairs presentation
- Ordered Weighted Averaging [Ronald Yager 1988]

$$\mathbf{sim(q,c)=\sum w_i sim_{\sigma(i)}(q_i,c_i)}$$

q_i,c_i: field i;

w_i: weight i (a monotonic function satisfying $\sum \mathbf{w_i}=1$)

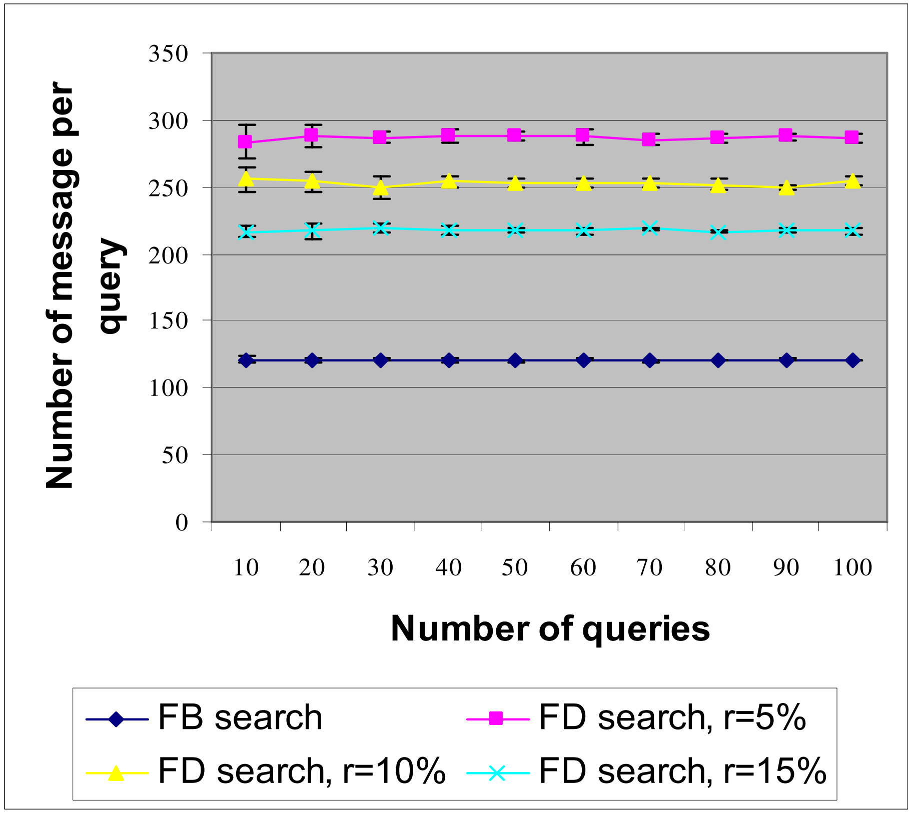
$sim_{\sigma(i)}(q_i,c_i)$: distance of q_i,c_i following a permutation $\sigma(i)$

Experiment Setup

- Gnutella network simulation
  - 100 peers
- SIMILE and CACM bibliographic data-sets
  - 35 bibtexes per peer
  - Query sets of 20%, 50% and 85% similarity
- Comparison 3 search mechanisms
  - Flooding-based search (FD, 4 neighbors)
  - Feedback-based search (FB, 3 selected neighbors)
  - Random-based search (RD, 3 random neighbors)

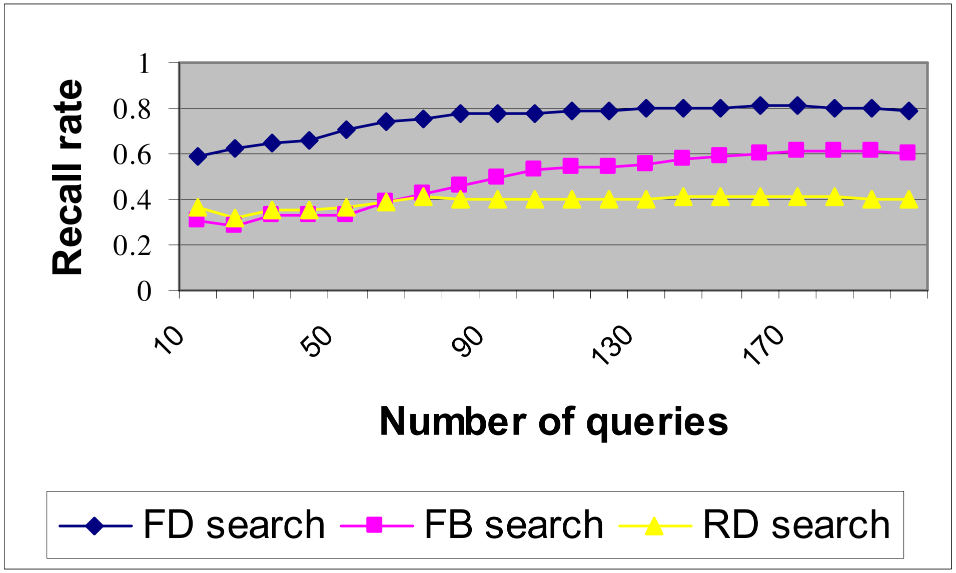
Scheme Evaluation (1)

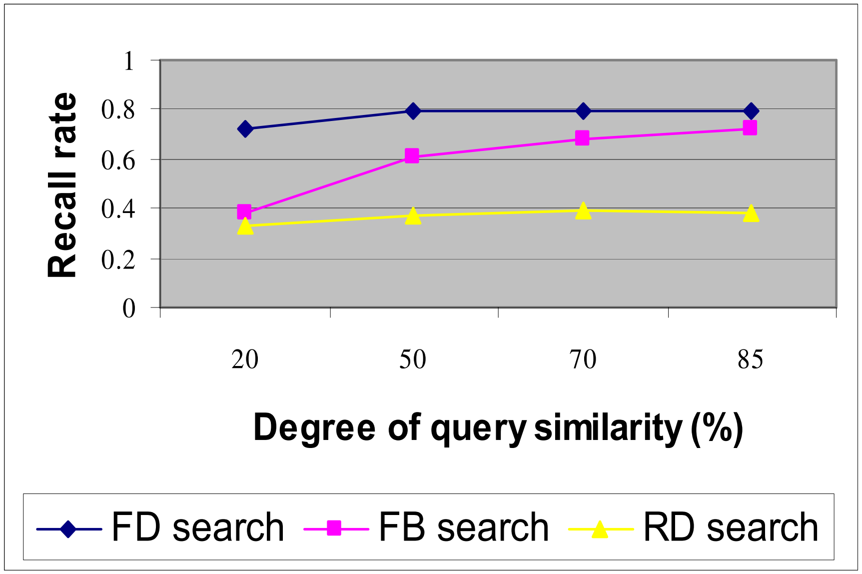
- Number of message per query
  - Reducing the cost of CBR processing
  - Reducing 1 selected neighbor
  - Cutting down 65% messages



Scheme Evaluation (2)

- Recall rate of retrieved bibtexes
  - Increasing the efficiency of the search mechanism
  - Reaching 77% of the recall rate limit with the query set of 50% similarity





Standing Issues

- Resource and query representation
  - Influencing similarity functions
  - Influencing peer learning and ranking processes
- Realistic fault data-sets
  - More complex than bibliographic data-sets
  - Case processing required

Thank you and Questions