

ll at,0/a  
bne at,at,20  
or v0,at,at  
or at,a0,a0  
sc at,0(,1)



Wissenschaftlerinnenkolleg  
Internettechnologien



TECHNISCHE  
UNIVERSITÄT  
WIEN

VIENNA  
UNIVERSITY OF  
TECHNOLOGY

# Using Taxonomies for Content-based Routing with Ants

Workshop on Innovations in Web Infrastructure (IWI2)

WWW2006, May 22, 2006

**Elke Michlmayr**

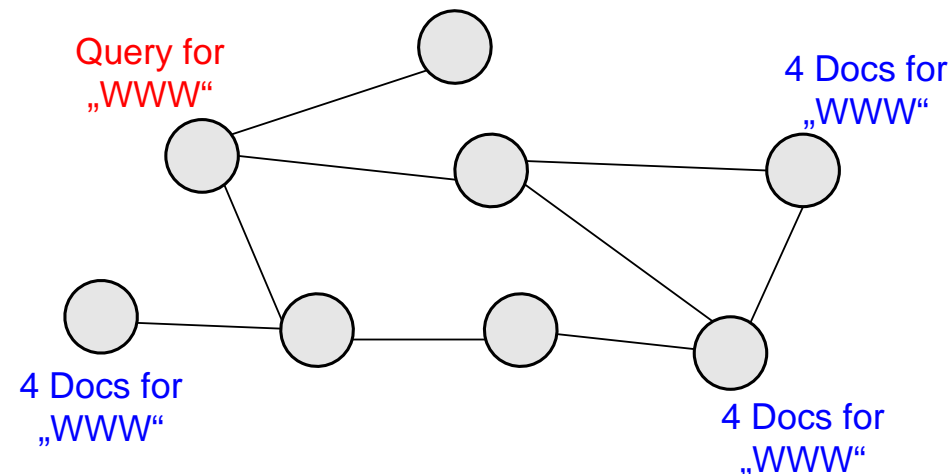
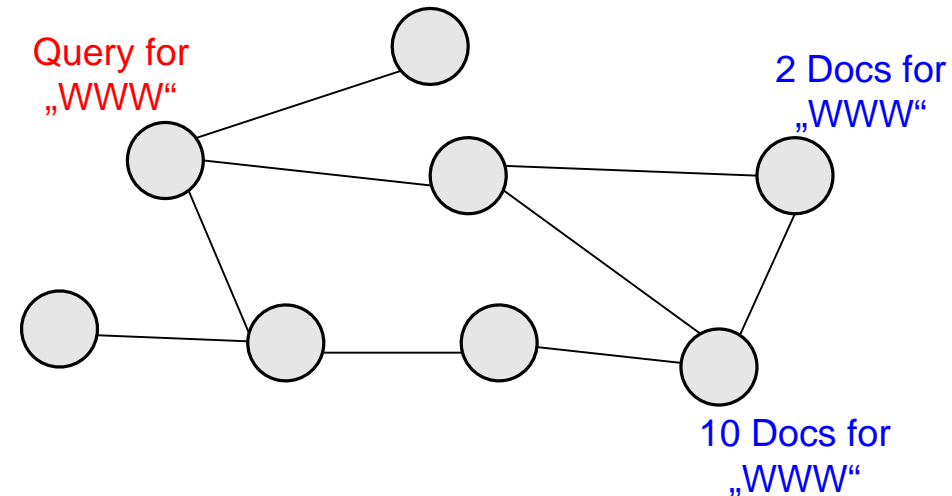
Women's Postgraduate College for Internet Technologies (WIT)

Vienna University of Technology, Austria

<http://wit.tuwien.ac.at/people/michlmayr>

# Motivation

- Ant algorithms for search in peer-to-peer networks
  - Content-based approach
  - Meta-data for annotating documents
  
- Content distribution influences performance
  - To which extent?
  - Can use of taxonomies improve performance?



# The SemAnt algorithm

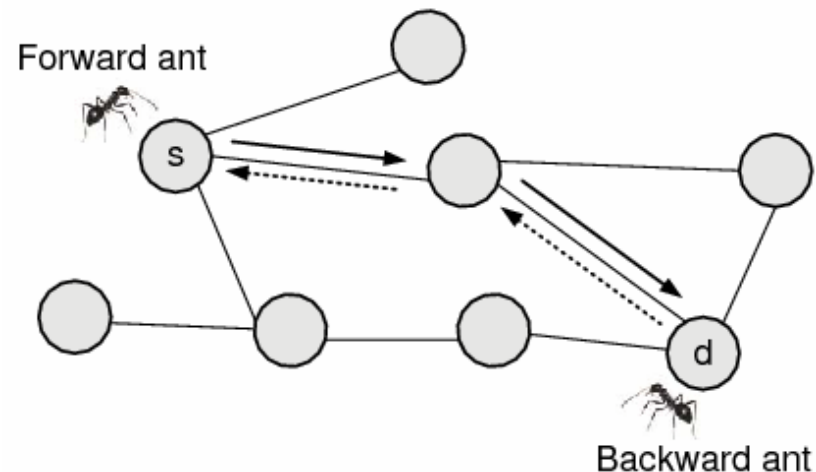
## ■ Ant algorithm for search in peer-to-peer networks

- Queries are ants
- Routing with pheromone trails
  - Trail laying and trail following
  - One type of pheromone for each keyword



## ■ SemAnt combines features from

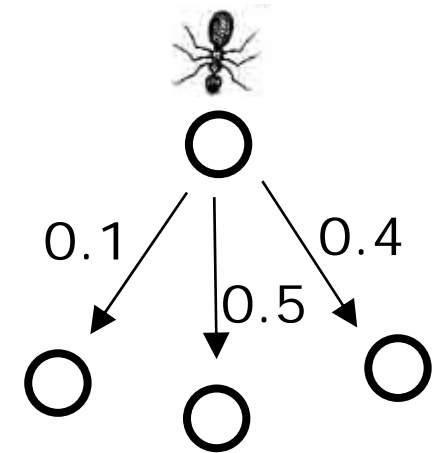
- Ant Colony System
  - For graph-based optimization problems
  - Trail laying and trail following
- AntNet
  - For load-adaptive routing of data packets
  - Forward and backward ants



# Trail following and trail laying

## ■ How do forward ants select the next peer?

- Exploiting strategy
  - Select best link
- Exploring strategy
  - Derive goodness value  $p_j$  for each link
  - Compute random value  $q$  in  $[0, 1]$
  - Select link if  $q \leq p_j$

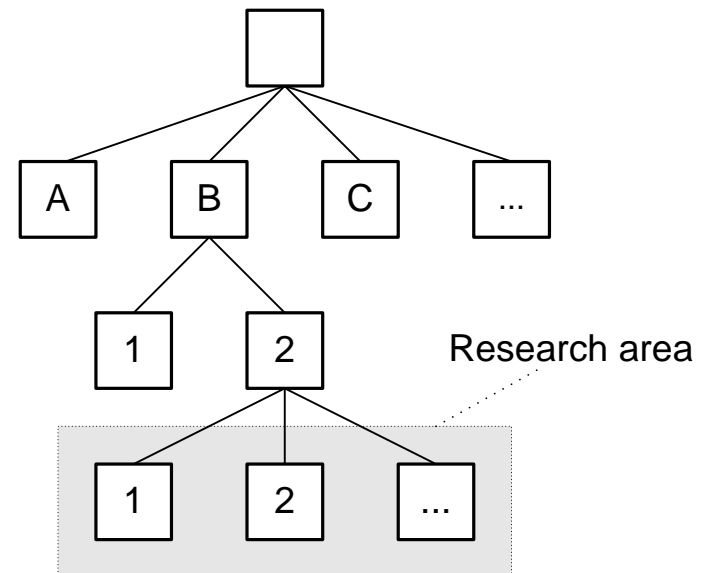


## ■ How do backward ants update pheromone trails?

- Amount depends on
  - Number of documents found
  - Number of hops

# Application scenario

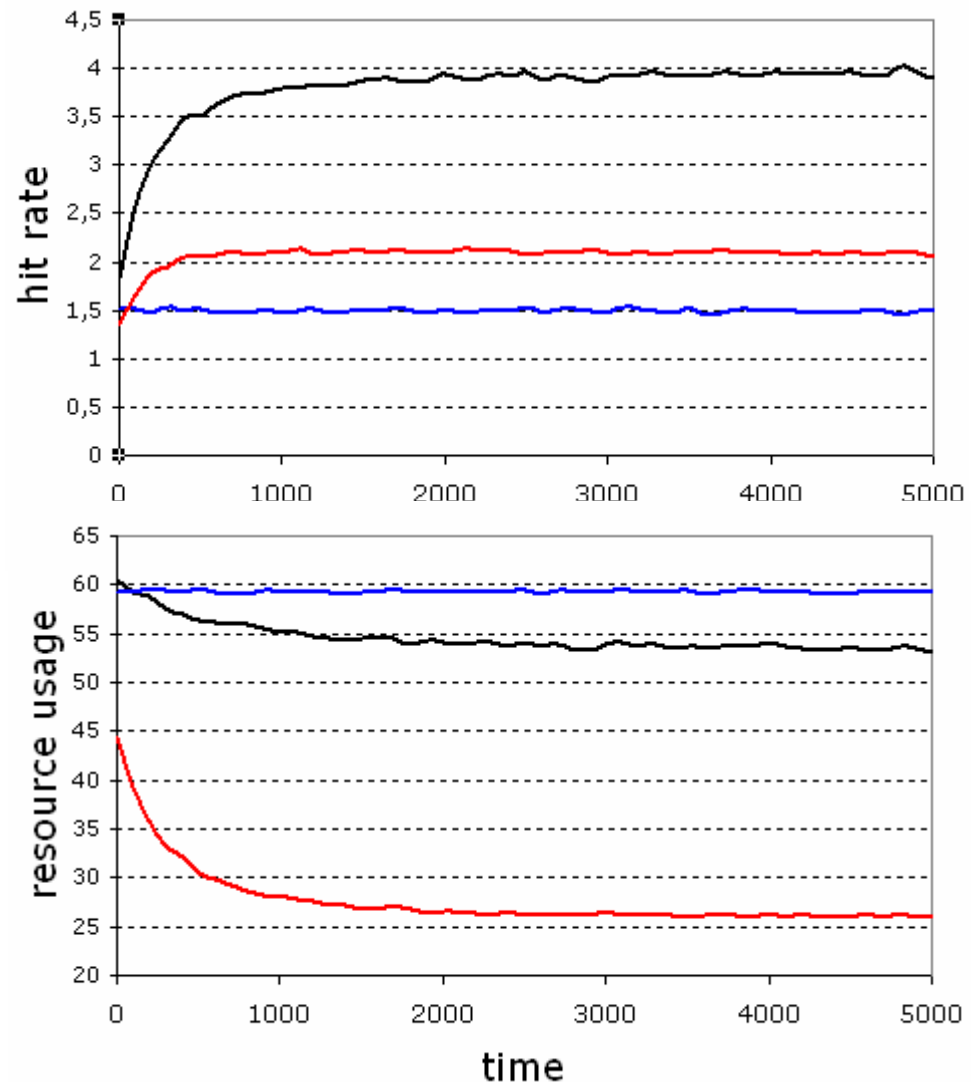
- Network topology
  - Small-world network with 1024 peers
- Test data: ACM Computing Classification System taxonomy
  - Documents annotated with one leaf topic
  - Queries consist of one leaf topic
  - Research area
    - sub-topics of a 3rd-level topic
- Query distribution is uniform
  - Probability that a peer issues a query is 0.1 per time unit
- SemAnt parameters
  - 85% exploiting strategy, TTL of 25



# Performance comparison with k-random walker

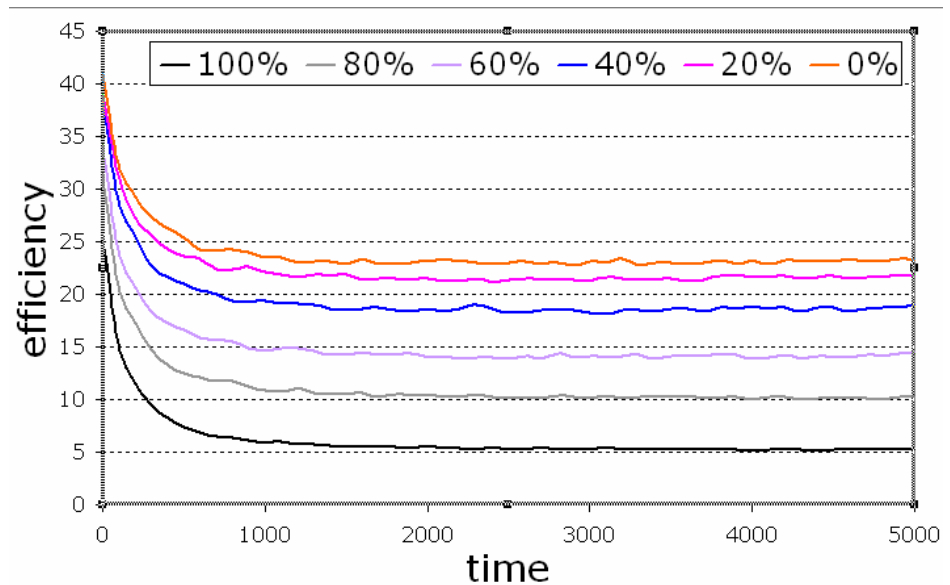
- Peer is an expert in one research area
  - 60% of its documents
  - 20% belong to another research area
  - 20% of documents are instances of random topics

- Comparison of
  - **k-random walker**
  - SemAnt maxResults (Ants use maximum TTL)
  - **SemAnt minResources** (Ants stop after first result)



# Influence of content distribution

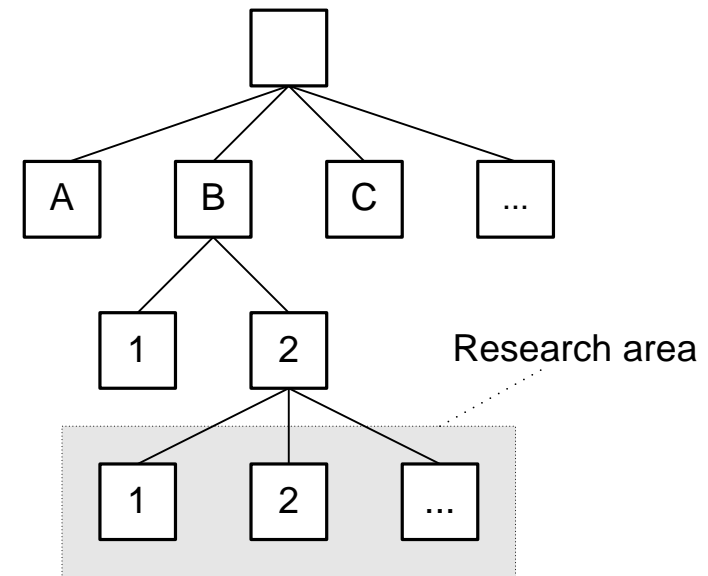
- Decrease coherence of content ...
  - Peers store 100%, 80%, 60%, 40%, 20%, 0% of their documents related to one research area
  - Remaining documents are instances of random topics



- ... and performance decreases as well!
- Can information from the taxonomy be used for enhancing it?
  - How?

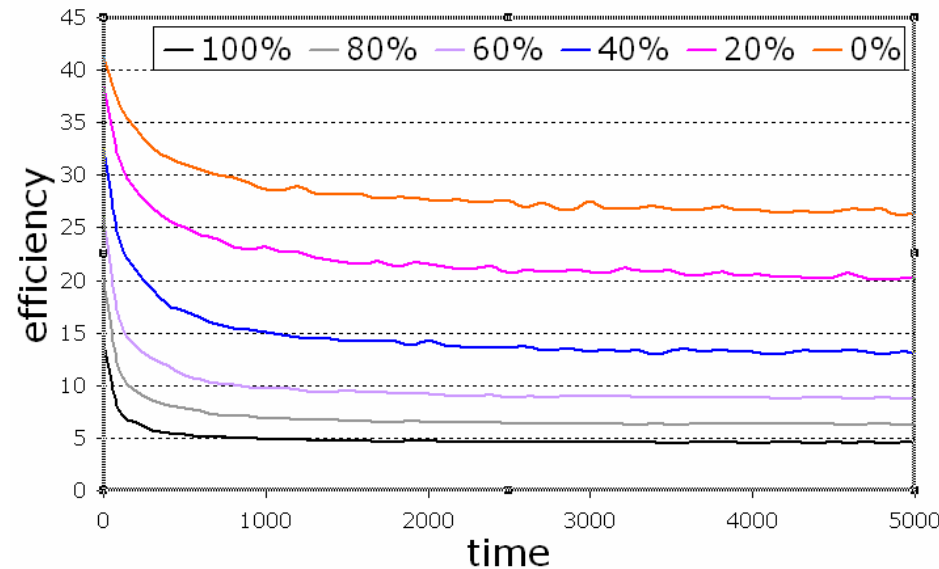
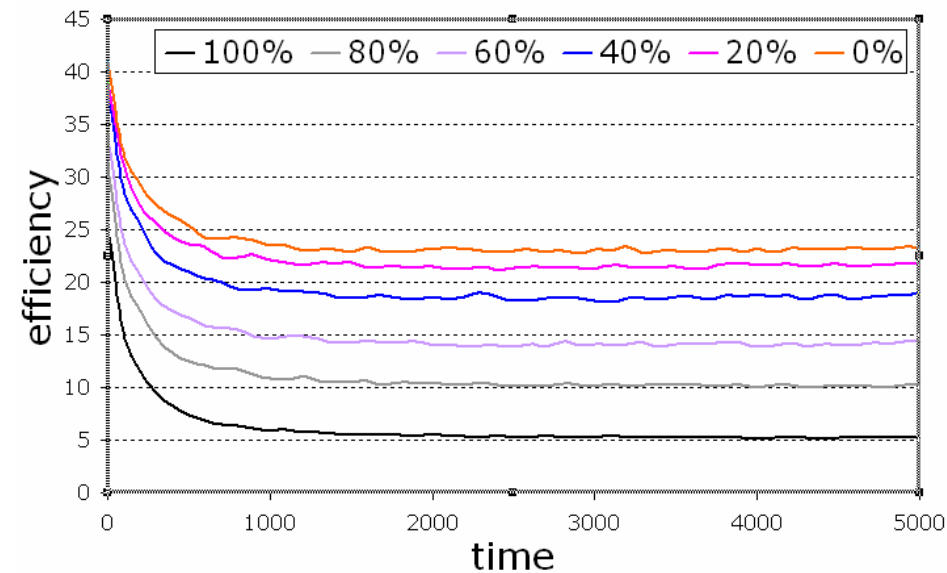
# Using taxonomies

- Pheromone trails also for higher-level topics
  - Backward ants drop same amount as for topic itself
  - Forward ants consider these trails for exploiting strategy
  
- How many higher-level topics?
  - Useful only for direct super-topic
  
- To which extent?
  - Trail for topic: 100%
  - Trail for super-topic: 25%





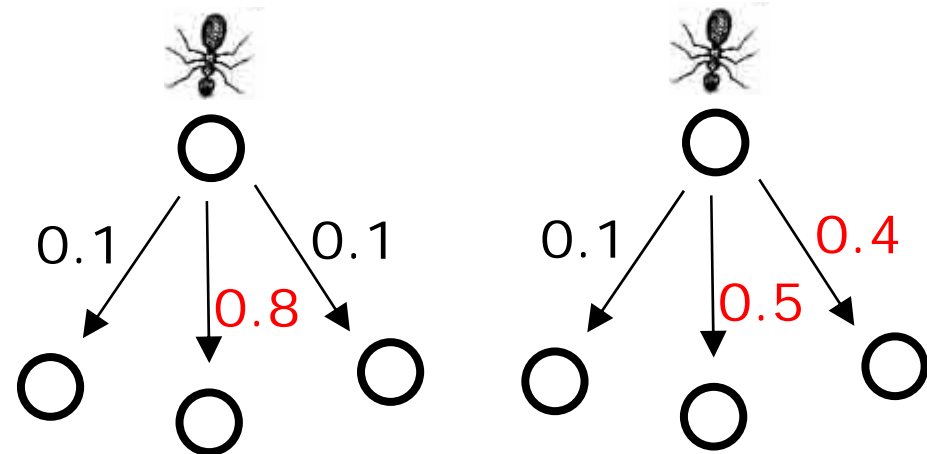
# Improvement by using underlying taxonomy



- 100%: Same, but converges faster
- 60%, 80%, 40%: Significant improvement (~40%, ~38%, ~30%)
- 20%: Slight improvement after 2500 time units
- 0%: Worse
- The lower the coherence, the longer it takes to converge

# Conclusion

- Content-based query routing in unstructured peer-to-peer networks based on the ant metaphor
  - Successful queries leave small traces in the network
  - Heuristics use traces to direct subsequent queries
  - => Ants build a map of the content distribution
  
- Contribution of this paper
  - Impact of content distribution
  - Taxonomies for improvement
  
- Future work
  - "Conditional flooding"



# References

- M. Dorigo, L. M. Gambardella. Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman Problem, 1997.
- G. Di Caro, M. Dorigo. AntNet: Distributed Stigmergy Control for Communications Networks, 1998.
- M. Dorigo, G. D. Caro. The Ant Colony Optimization Meta-Heuristic, 1999.
- M. Guntsch and M. Middendorf. Pheromone Modification Strategies for Ant Algorithms Applied to Dynamic TSP, 2001.
- O. Babaoglu, H. Meling, and A. Montresor. Anthill: A Framework for the Development of Agent-Based Peer-to-Peer Systems, 2002.
- B. Yang, H. Garcia-Molina. Improving Search in Peer-to-Peer Networks, 2002.
- V. Kalogeraki, D. Gunopulos, D. Zeinalipour-Yazti. A Local Search Mechanism for Peer-to-Peer Networks, 2002.
- D. Tsoumakos, N. Roussopoulos. Adaptive Probabilistic Search for Peer-to-Peer Networks, 2003.
- B. F. Cooper. Guiding Queries to Information Sources with InfoBeacons, 2004.
- C. Tempich, S. Staab, A. Wranik. REMINDIN': Semantic Query Routing in Peer-to-Peer Networks based on Social Metaphors, 2004.
- S. Rhea, D. Geels, T. Roscoe, J. Kubiatowicz. Handling Churn in a DHT, 2004