

# Domain Specific Retrieval Experiments with MIMOR at the University of Hildesheim

**René Hackl, Ralph Kölle,  
Thomas Mandl, Christa Womser-Hacker**

Information Science

Universität Hildesheim  
Marienburger Platz 22 - 31131 Hildesheim

Germany

`{mandl,womser}@uni-hildesheim.de`

# Overview



- Fusion in Information Retrieval
- The MIMOR Model
- Participation
- Outlook

# Fusion in ad-hoc Retrieval

Many studies especially within TREC show

- that the quality of the best retrieval systems is similar
- that the overlap between the results is not very large
- ➔ that fusion of the results of several systems can improve the overall performance

# Committee Machines in Machine Learning

- Fusion in IR corresponds to committee machines
- Opinions or results of several experts are combined





# Fusion in ad-hoc Retrieval

- Several fusion approaches have been developed
- How do we create one ranked list out of several ranked lists?
  - Different, but similar from fusion problem in multilingual retrieval
- Fusion in meta search engines is different
  - heterogeneous sources

# Fusion

Document  
corpus

IR System 1

IR System 2

Ranked  
Document  
List

doc weight  
doc weight  
doc weight

Ranked  
Document  
List

doc weight  
doc weight  
doc weight

Fused  
Document  
List

doc  $f(\text{weight1}, \text{weight2})$   
doc  $f(\text{weight1}, \text{weight2})$   
doc  $f(\text{weight1}, \text{weight2})$

sum

min

max

f



# Other results of TREC

- ...
- Relevance feedback is a very successful strategy

# The MIMOR Model

Combines fusion and Relevance Feedback

- linear combination
- each individual system has a weight
- weights are adapted based on relevance feedback

Goal:

High  
System  
Relevance

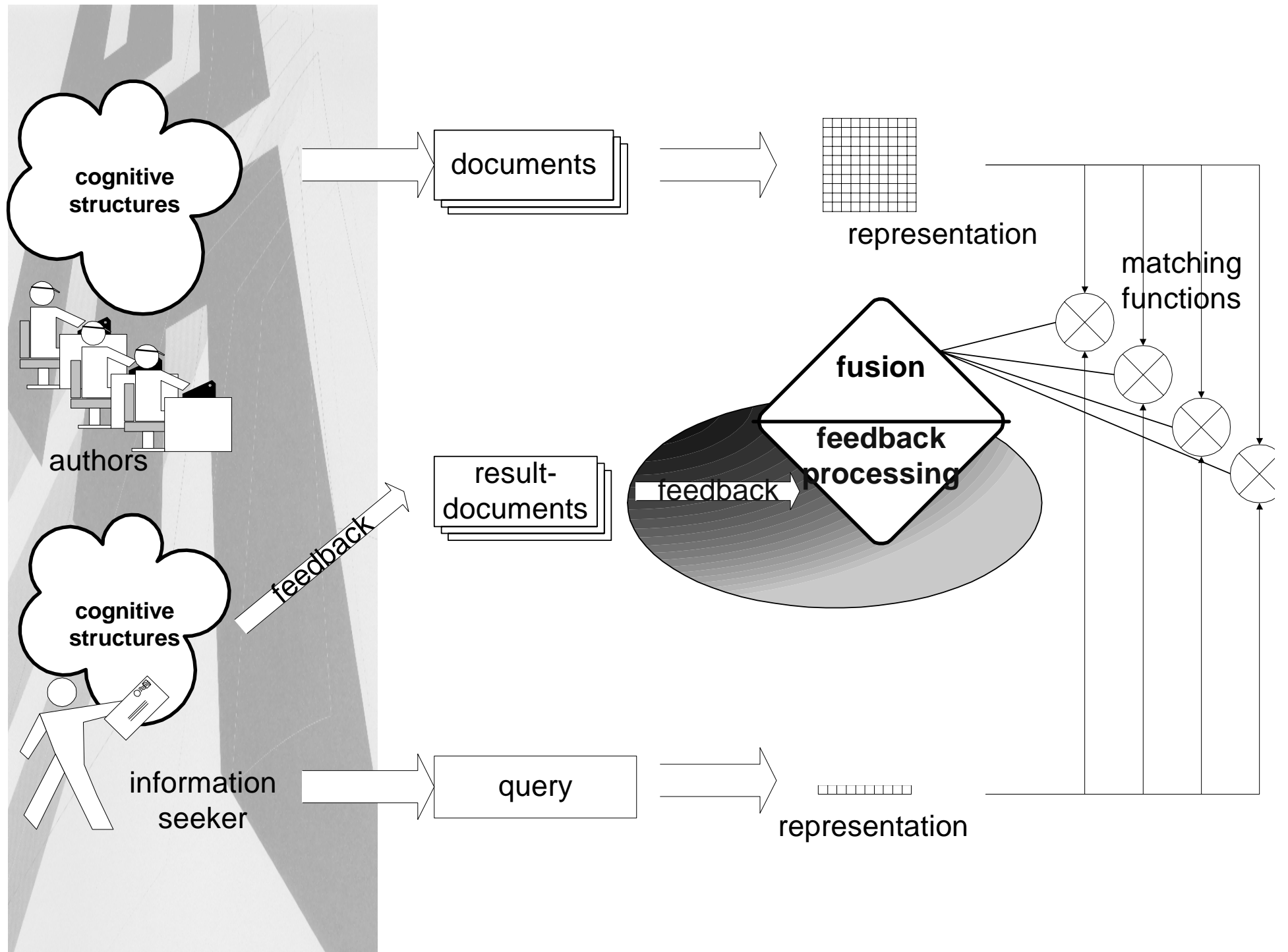
+

Positive  
Relevance  
Feedback



Increase weight of a system





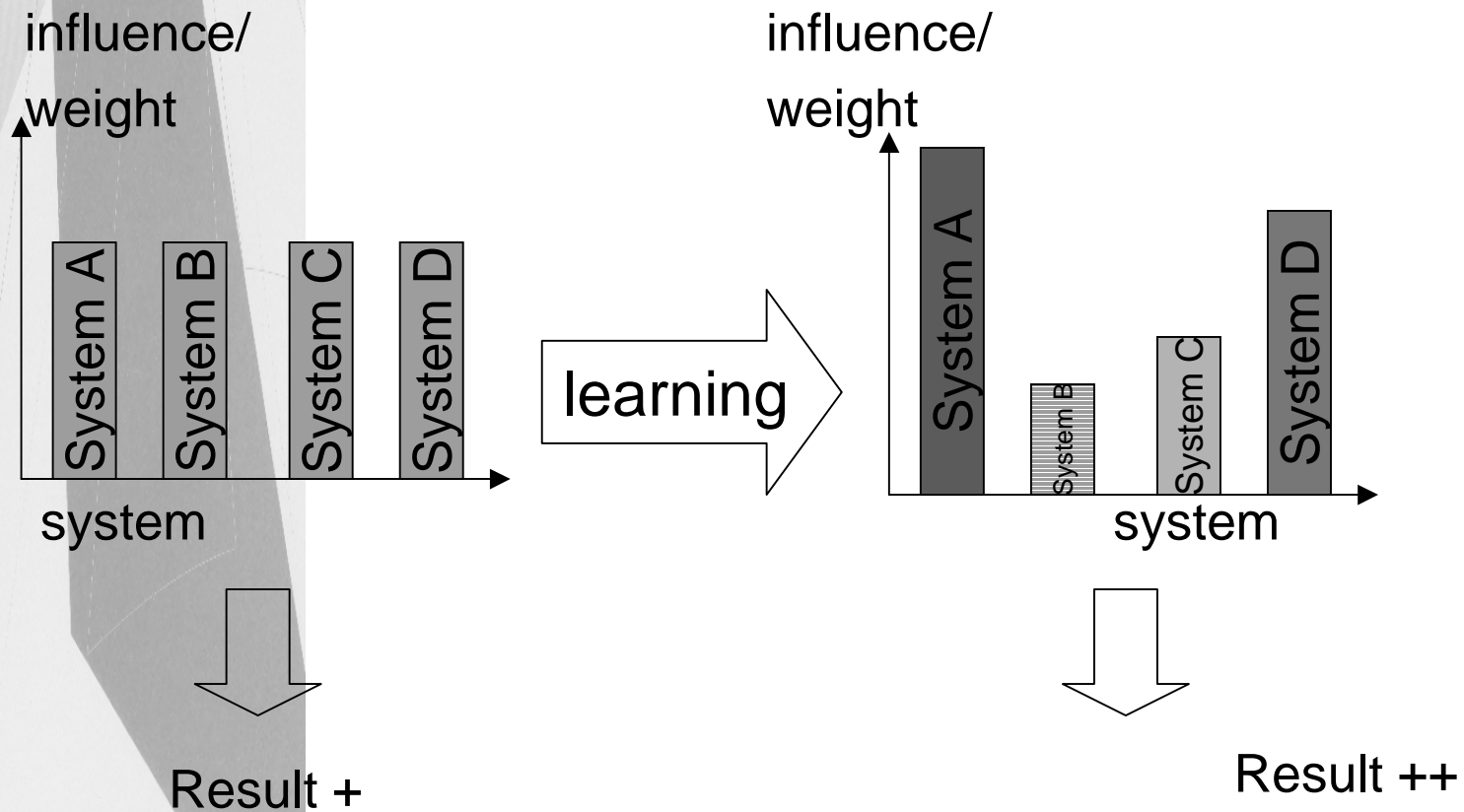
# Calculation of RSV in MIMOR

Weighted sum of single RSV

$$RSV_j = \frac{\sum_{i=1}^N (\omega_i RSV_{ij})}{N}$$

$RSV_{ij}$       *Retrieval Status Value of  
System  $i$  for Document  $j$*

# The Learning Process





# Learning in MIMOR

$$\omega_i = \omega_i + (\varepsilon RSV_{ij} R_j)$$

$\varepsilon$       *Learning rate*

$R_j$       *Relevance-Judgement for Document*

Intellectual supervision of weight  
vector possible



# Participation in CLEF 2002

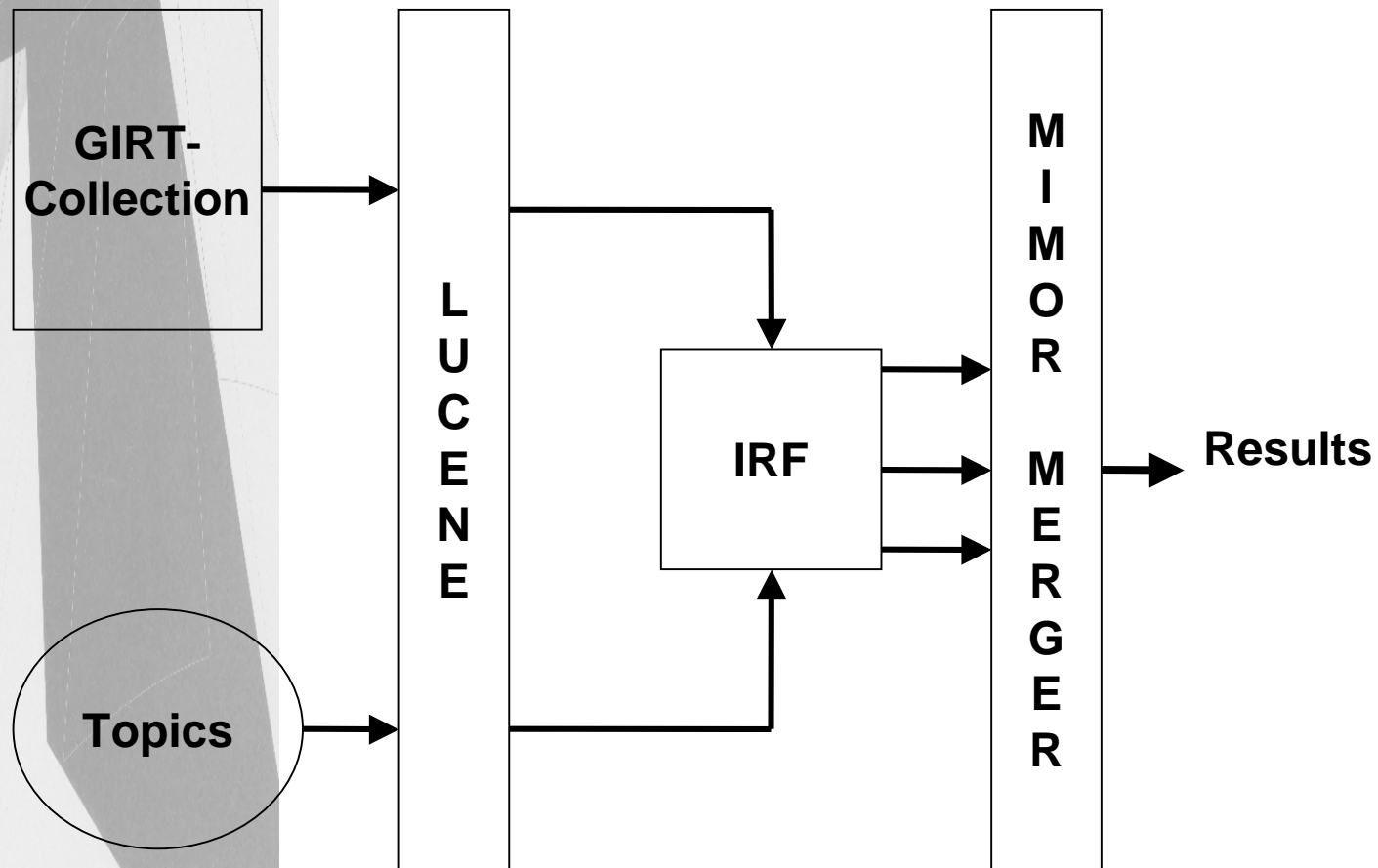
- GIRT track
  - monolingual
  - German social science data
- MIMOR
  - Fusion and optimization implemented in JAVA
  - basic retrieval systems: irf from NIST



# Participation in CLEF 2002

- MIMOR
  - LUCENE for linguistic pre-processing
  - three different results were obtained by different parameter settings in irf
  - static optimization

# Sequence of operations





# Participation in CLEF 2002

- Results
  - satisfying for a first try





# Acknowledgements

- Thanks to NIST for providing the source code of the IRF package
- Thanks to students in Hildesheim for implementing part of MIMOR in their course work



# Outlook

We plan to

- integrate more retrieval systems (suggestions or contributions welcome)
  - invest more effort in the optimization
  - participate in the multilingual track
- 