



# Integrating Knowledge Discovery into Knowledge Management

Katharina Morik, Christian Hüppe,  
Klaus Unterstein

Univ. Dortmund LS8

[www-ai.cs.uni-dortmund.de](http://www-ai.cs.uni-dortmund.de)





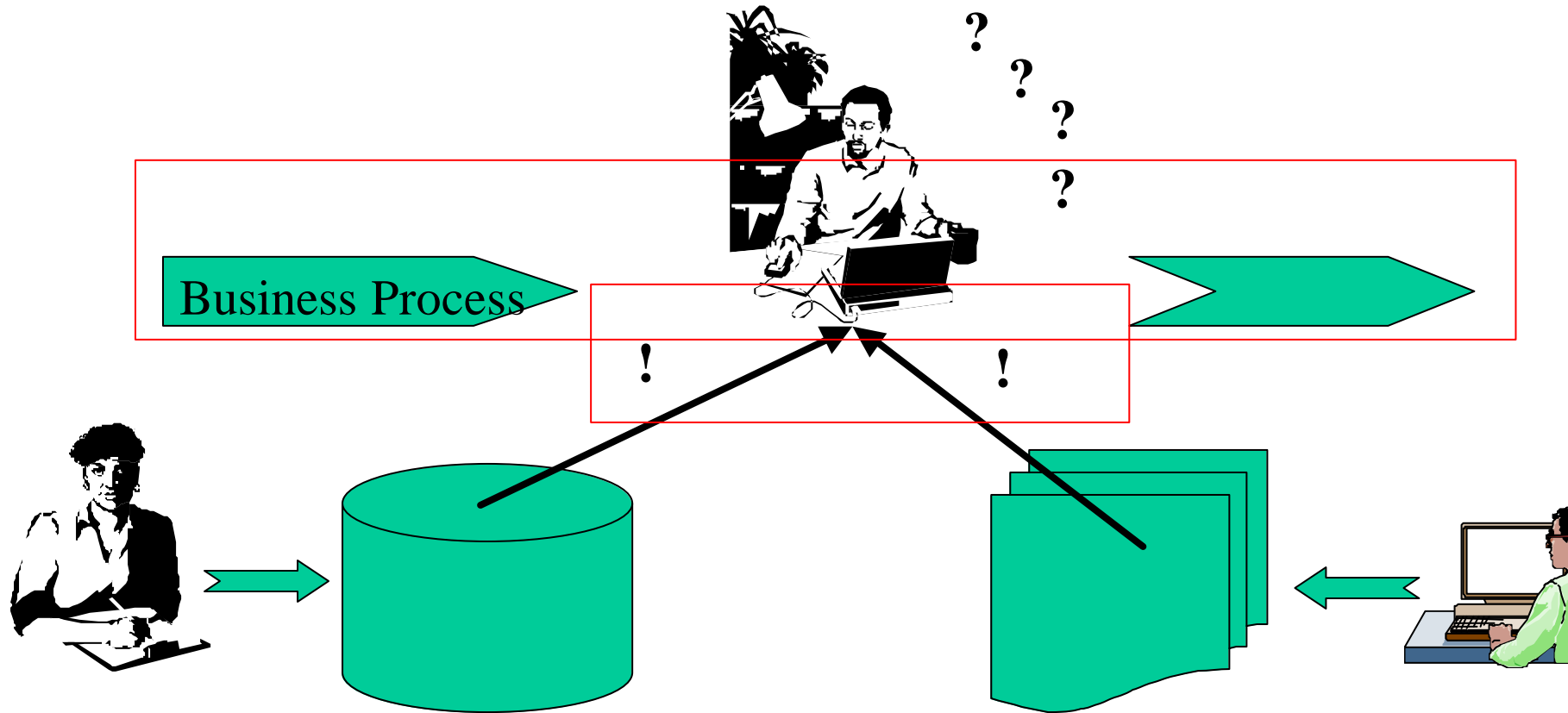
# Overview

- Integrating given data into a knowledge management system (KMS)
- System architecture of EAMS
- Integrating given document collections by learning the right retrieval function
- Integrating given databases by knowledge discovery



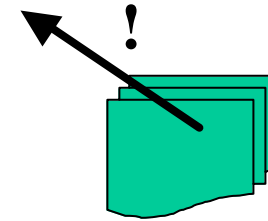


# Knowledge Management





# Integrating Given Data into KMS 1

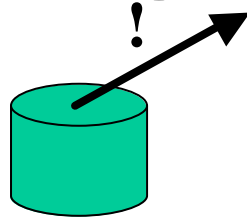


- Preparing documents for a KMS is an extra effort
- Structuring document collections according to an ontology is time-consuming, too
- Why not having the machine learn which document a user wants as the answer to his query?
  - Learning the retrieval function for each user
  - according to an ontology





## Integrating Given Data into KMS 2

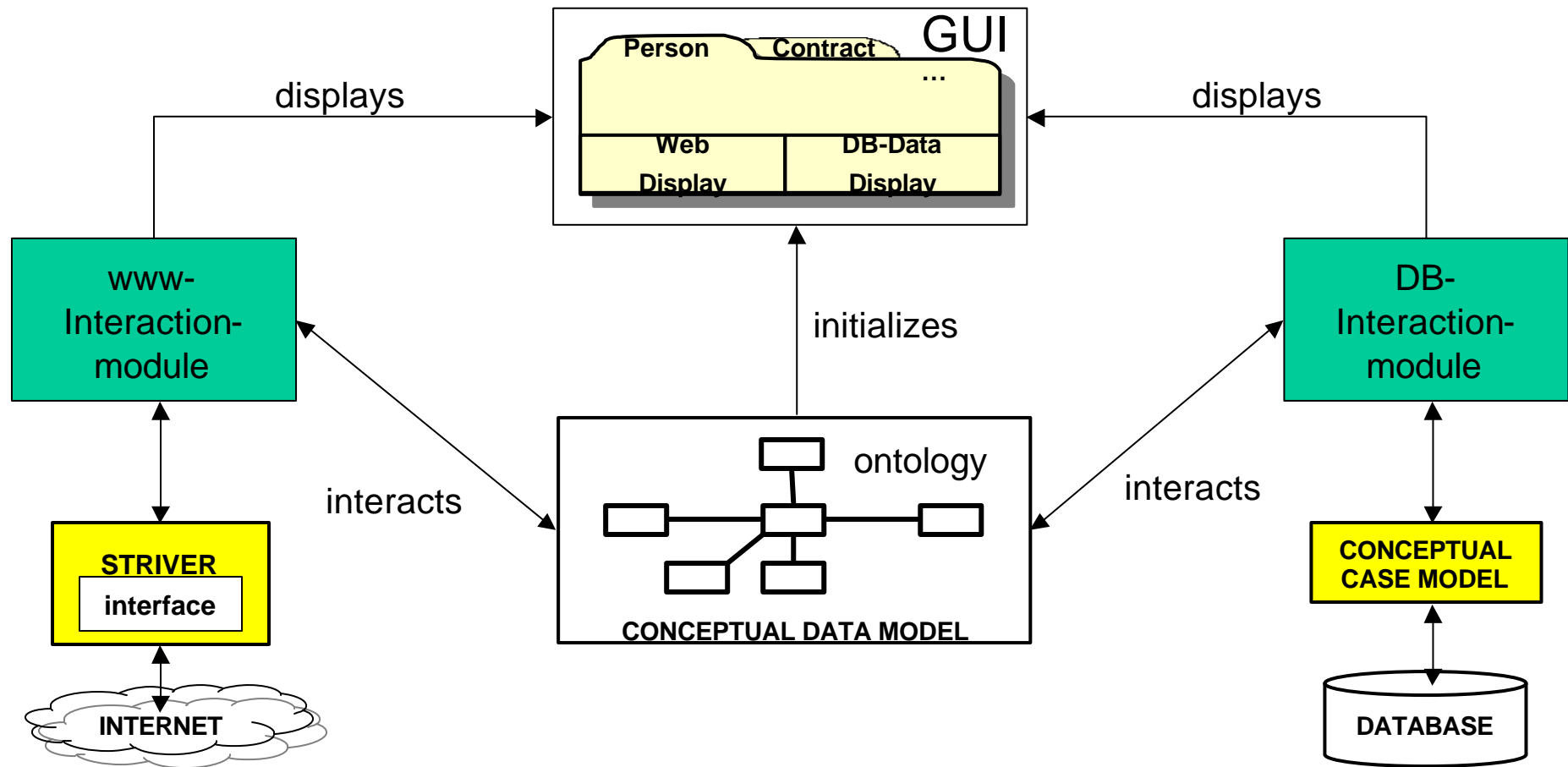


- The main data sources in organizations are databases.
- Why not using them?
  - Knowledge discovery is a high-level query language.
  - Meta-data about knowledge discovery cases can be organized according to an ontology.



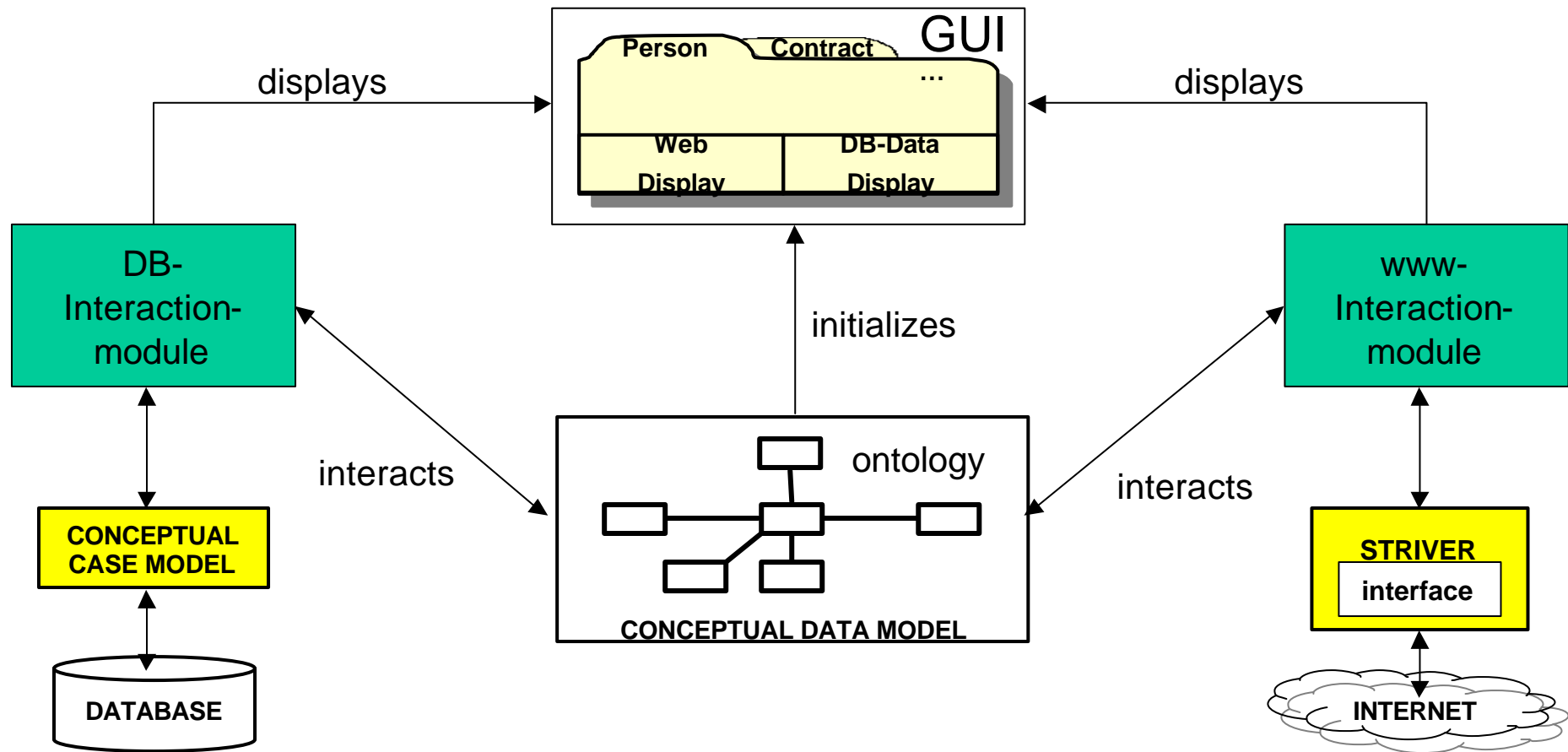


# System Architecture





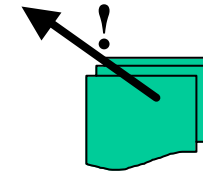
# System Architecture





# Striver: Learning a Retrieval Function

Thorsten Joachims KDD 2002

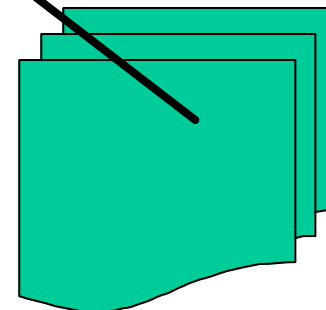


Query  $q$  ?

	Clickthrough
	$(q_1, r'_1)$ ... $(q_m, r'_m)$

$r' \in r$

Ordering  $r \in D \times D$  !



Documents  $D$   
 $\{d_1, d_2, \dots, d_n\}$







# Striver: Learning a Retrieval Function

Thorsten Joachims KDD 2002

Minimize distance between  $r'$  and learned ranking  $r\#$

	$(q_1, r'_1)$
	$l_1$ click
	$l_2$
	...
	$l_i$ click
	...
	$l_j$

$$r' \hat{=} r$$

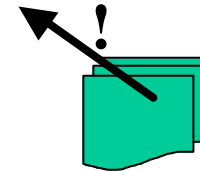
$$l_1 > l_2$$

...

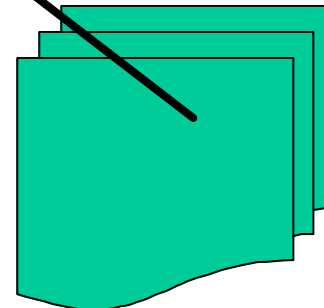
$$l_i > l_j$$



Query  $q$  ?



Ordering  $r\# \hat{=} D \times D !$



Documents  $D$   
 $\{d_1, d_2, \dots, d_n\}$





# Search String for a Web Query

The screenshot shows a web application interface with a navigation bar containing two tabs: "Kunde" and "Vertrag". The main content area displays a list of search results for insurance types. The results are as follows:

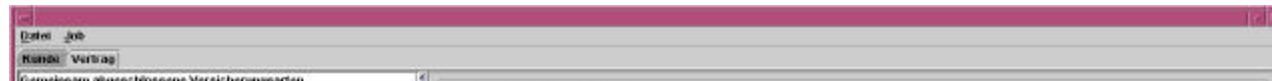
- Gemeinsam abgeschlossene Versicherungsarten
- APRIORI-Ergebnisse: Vertragsarten (Details)
- Signifikanztest: Vorgeschichte zum Rueckkauf - Vertrags
- Signifikanztest: Vorgeschichte zum Rueckkauf - Vertrags
- APRIORI-Ergebnisse: Vertragsarten
- Beliebte Versicherungen fuer bestimmte Kundengruppen
- Beliebte Versicherungen

At the bottom of the screenshot, there is a small window showing a search result for "AZKA Art Versicherung - unsere Produkte" with the URL <http://www.azka-nordstem-sei.at/produkte.html>. Below it, there is another search result for "Frankfurter Produkte" with the URL <http://www.makler-service.de/3.html>.





# Result of Web Query



Volksfürsorge Versicherung Jülich

<http://www.volksfuersorge-juelich.de/index.htm>

Grieger Mallison GMFS GmbH & Co CompactTeam Berlin KG

<http://www.compactteam.de/gr/1,1095,35004390,00.html>

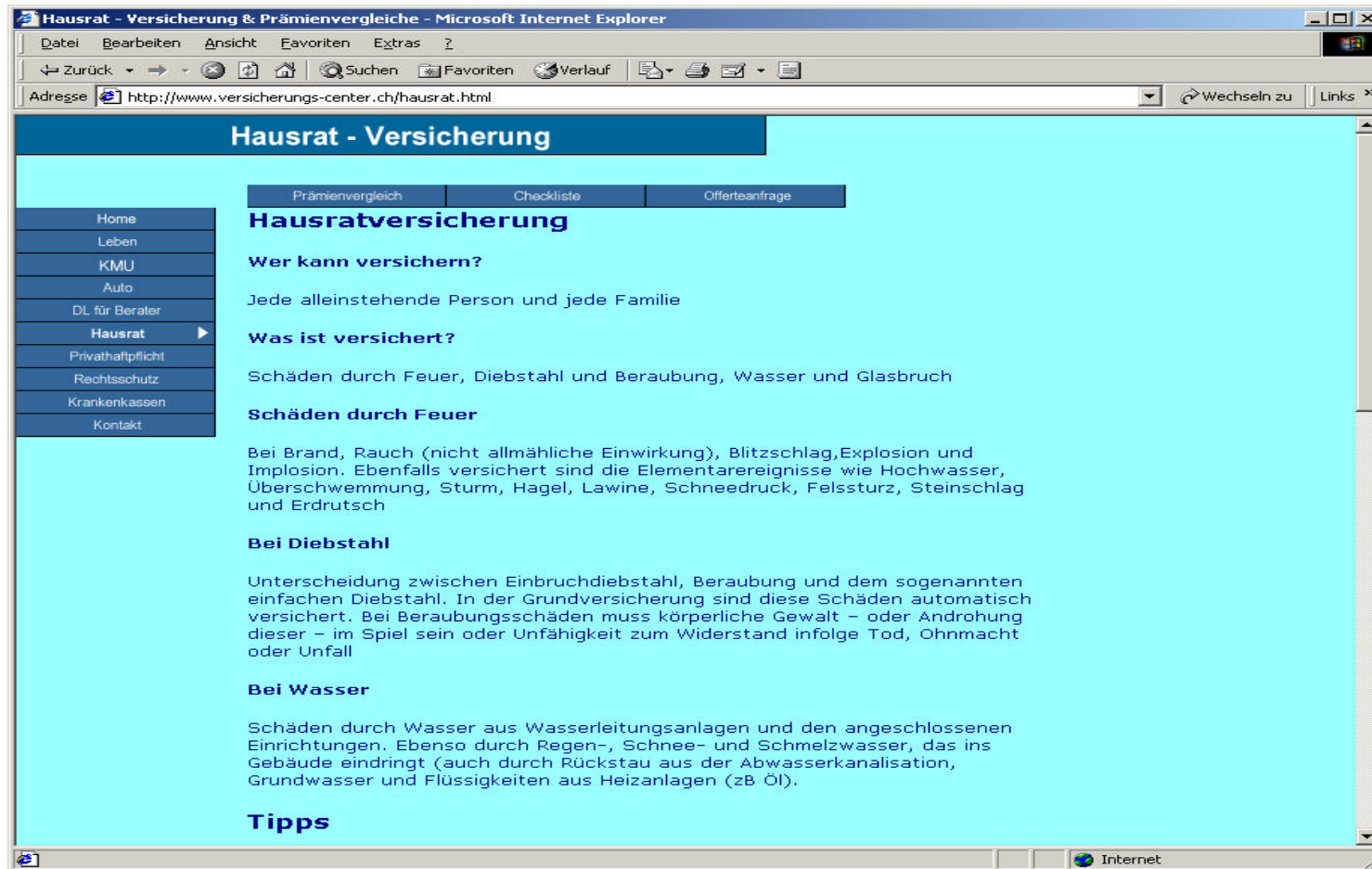
Hausrat - Versicherung Prämienvergleiche

<http://www.versicherungs-center.ch/hausrat.html>





# Web document





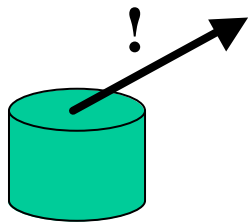
# Learning a Retrieval Function

- New version of support vector machine for ranking (Thorsten Joachims 2002).
- Optimizes given retrieval functions.
- Automatically adapts to users (tasks).
- Can be applied to the intranet without preparation.
- Inspection of the learned function shows that the weights of words make sense!





# Knowledge Discovery as High-level Query Language to Databases



- Ontological concepts:
  - Person,
  - Contract
- Query types:
  - Frequencies of attributes
  - Segmentation (subgroups)
  - Correlation of attributes
  - Classification
- Algorithms (operators):
  - Statistical stored procedures
  - Data cube
  - APRIORI
  - C4.5
  - mySVM
- Preprocessing chain





# KDD Query -- already executed job

Hier werden die Merkmale Geschlecht, Altersgruppe und Berufsgruppe der Kunden analysiert. Das Ergebnis ist ein dreidimensionaler Data Cube.

Job	Status	Dauer	Fertigstellung
Job 1000000659	fertig	0	30.06.2002, 00:00 Uhr
Job 1000000752	fertig	0	06.07.2002, 00:00 Uhr











maennlich	>65 Jahre (Berliner)	Berufsgruppe 2	6200
maennlich	=65 Jahre (Berliner)	Berufsgruppe 3	570
weiblich	>65 Jahre (Berliner)	Berufsgruppe 1	46





# KDD Result

Berufsgruppe, Geschlecht: männlich, Altersgruppe: 40-65 Jahr

	Berufsgruppe 1, 8082 (15.5%)
	Berufsgruppe 2, 1013 (1.9%)
	Berufsgruppe 3, 2103 (4.0%)
	Berufsgruppe 4, 31 (0.1%)
	Berufsgruppe 5, 6103 (11.7%)
	Berufsgruppe 6, 3040 (5.8%)
	Berufsgruppe 7, 7558 (14.5%)
	Berufsgruppe 8, 5606 (10.7%)
	Berufsgruppe 9, 12237 (23.4%)
	Berufsgruppe 10, 5150 (9.8%)







# KDD Result



sex	age group	profession	quantity
male	0-22 years	profession group 1	67
male	0-22 years	profession group 2	4373
male	0-22 years	profession group 3	1967
male	0-22 years	profession group 4	3

Leider kein Webergebnis verfügbar

maennlich	23-39 Jahre (Young Professionals)	Berufsgruppe 6	2175
maennlich	23-39 Jahre (Young Professionals)	Berufsgruppe 7	2626
maennlich	23-39 Jahre (Young Professionals)	Berufsgruppe 8	2197
maennlich	23-39 Jahre (Young Professionals)	Berufsgruppe 9	6761
maennlich	23-39 Jahre (Young Professionals)	Berufsgruppe 10	3183
maennlich	23-39 Jahre (Young Professionals)		23627
maennlich	40-65 Jahre (Geldverdienend)	Berufsgruppe 1	8062
maennlich	40-65 Jahre (Geldverdienend)	Berufsgruppe 2	1013
maennlich	40-65 Jahre (Geldverdienend)	Berufsgruppe 3	2103
maennlich	40-65 Jahre (Geldverdienend)	Berufsgruppe 4	31
maennlich	40-65 Jahre (Geldverdienend)	Berufsgruppe 5	6103
maennlich	40-65 Jahre (Geldverdienend)	Berufsgruppe 6	3040
maennlich	40-65 Jahre (Geldverdienend)	Berufsgruppe 7	7558
maennlich	40-65 Jahre (Geldverdienend)	Berufsgruppe 8	6005
maennlich	40-65 Jahre (Geldverdienend)	Berufsgruppe 9	12237
maennlich	40-65 Jahre (Geldverdienend)	Berufsgruppe 10	6463
maennlich	40-65 Jahre (Geldverdienend)		52233
maennlich	>65 Jahre (Rentner)	Berufsgruppe 1	665
maennlich	>65 Jahre (Rentner)	Berufsgruppe 2	6200
maennlich	>65 Jahre (Rentner)	Berufsgruppe 3	570
maennlich	>65 Jahre (Rentner)		46





# KDD Query -- creating a new job

Datei Job  
Kunde Vertrag  
Leide Kundenebenen

Eingabe der Parameter

Eingabe der Parameter zum zugehoerigen Job

Kundengruppe

Anzahl

Job starten

Berufsgruppe 2, Altersgruppe 0-22 Jahren (Kinder/Jugendliche)

Berufsgruppe 2, Altersgruppe 0-22 Jahren (Kinder/Jugendliche)

Berufsgruppe 2, Versicherungs-Zahler

Berufsgruppe 2, Altersgruppe >65 Jahren (Rentner), Rolle Versicherter

Berufsgruppe 2, Altersgruppe >65 Jahren (Rentner), ERWERBSART 3

Berufsgruppe 2, Altersgruppe >65 Jahren (Rentner), Versicherungs-Zahler

Berufsgruppe 2, Altersgruppe >65 Jahren (Rentner), Versicherungs-Halter

Berufsgruppe 2, Rolle Versicherter, Altersgruppe 0-22 Jahren (Kinder/Jugendliche)

Berufsgruppe 2, Altersgruppe 0-22 Jahren (Kinder/Jugendliche), Familienstand ledig

Omega 48

<http://www.statistik.uni-dortmund.de/fachscha/Omega/Omega48.htm>

maennlich	40-65 Jahre (Geldverdiener)	Berufsgruppe 3	2103
maennlich	40-65 Jahre (Geldverdiener)	Berufsgruppe 4	31
maennlich	40-65 Jahre (Geldverdiener)	Berufsgruppe 5	6103
maennlich	40-65 Jahre (Geldverdiener)	Berufsgruppe 6	3040
maennlich	40-65 Jahre (Geldverdiener)	Berufsgruppe 7	7558
maennlich	40-65 Jahre (Geldverdiener)	Berufsgruppe 8	5005
maennlich	40-65 Jahre (Geldverdiener)	Berufsgruppe 9	12237
maennlich	40-65 Jahre (Geldverdiener)	Berufsgruppe 10	6463
maennlich	40-65 Jahre (Geldverdiener)		52233
maennlich	>65 Jahre (Rentner)	Berufsgruppe 1	555
maennlich	>65 Jahre (Rentner)	Berufsgruppe 2	5200
maennlich	>65 Jahre (Rentner)	Berufsgruppe 3	570
maennlich	>65 Jahre (Rentner)	Berufsgruppe 4	45





# Mining Mart for Knowledge Management

- Making existing sources (databases) available to users – a case answers a high-level question
- The conceptual model (ontology) eases the integration with other services of a knowledge management system (e.g., web navigation).
- The conceptual model and the cases create the GUI for the EAMS user.

