

# Intelligentes multimediales Retrieval

Forschungsaktivitäten in der Abteilung MIND



***GMD – Forschungszentrum  
Informationstechnik GmbH***

***Institut für Integrierte Publikations-  
und Informationssysteme***

**e-mail: [mindcore@darmstadt.gmd.de](mailto:mindcore@darmstadt.gmd.de)  
<http://www-cui.darmstadt.gmd.de/mind/>**

***Reginald Ferber***

# Überblick

## Information Retrieval

- Unsere Sicht

## Das MIRACLE Modell

## Dialogmanager

## Intelligente Retrieval Methoden

- Regelbasierte Verfahren
- Abduktives Retrieval

## Automatische Indexierung

- Dokumentstrukturen nutzen
- Inferenznetze
- Automatische Bildindexierung

## Wissensextraktion aus Datenbanken

- Knowledge Discovery
- Assoziative Query Expansion

## Mitarbeitende in der Abteilung MIND

## ***Information Retrieval***

| Nutzende   | Retrieval System                        |
|--|---|
| Anfrage formulieren  | Dokumente auswählen und anzeigen        |
| Dokumente ansehen<br>relevante Dokumente auswählen<br>Anfrage verfeinern                               | Dokumente auswählen und anzeigen        |
| Neue Dokumente mit denen der letzten Anfrage vergleichen<br>Anfrage verfeinern<br>Anfragen kombinieren | ...<br>Dokumente auswählen und ausgeben |

*Herkömmliche Retrieval Systeme überlassen den Großteil der intellektuellen Arbeit den Nutzenden*



MIND

## Unsere Sicht

Information Retrieval (IR) ist ein **Prozeß**, in dem

- Nutzende die Datenbank erkunden
- verschiedene Anfragemöglichkeiten geprüft werden
- der Informationsbedarf der Nutzenden geklärt wird
- Informationen aus der Datenbank abgerufen werden

Dieser Prozeß muß als **Dialog** zwischen den Nutzenden und dem IR-System gesehen werden.

Dabei sollte ein intelligentes IR-System mit den Nutzenden kooperieren und sie von möglichst vielen Aufgaben **entlasten**.



MIND

### Komponenten eines intelligenten IR-Systems:

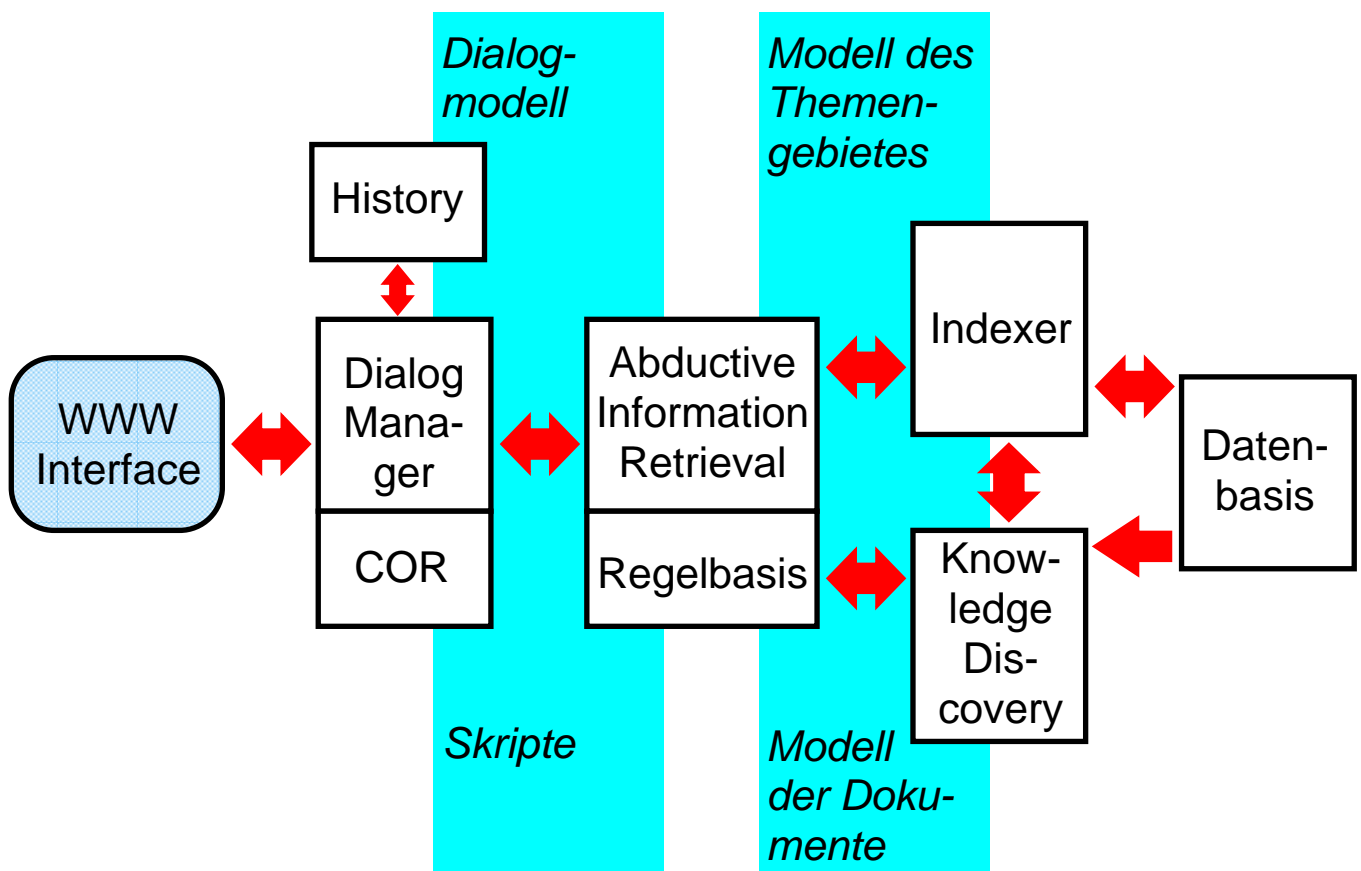
- Modell des Dialogs
- Dialogverwaltung
- Modell des Themengebietes der Datenbank
- Möglichkeit, externes Wissen zu nutzen
- Inferenzmechanismen zur Verarbeitung der Information
- Inhaltsorientierte Indexierungsmechanismen
- Methoden zum Finden und Nutzen von Regelmäßigkeiten unter den Objekten der Datenbank

### Zitat:

*“It is my claim that, to design the next generation of IR systems, we will need to have a formal semantics for documents and queries. This semantic representation will interact with other types of knowledge in a controlled way, and this way is inference!”*

van Rijsbergen, ACM SIGIR, 1989 p. 81

## Das MIRACLE Modell



*MIRACLE = Multimedia concept Retrieval  
bAsed on logiCaL query Expansion*



MIND

## ***Dialogmanager***

Das Dialogmodell nutzt:

- Bisherigen Dialogverlauf als hierarchische Struktur
- Netz möglicher Dialogentwicklungen
- Skripts mit typischen Dialogen

Daraus können kontextsensitive

- Interpretationen von Dialogbeiträgen
- Hilfeangebote
- Vorschläge für weiteres Vorgehen

abgeleitet werden.

## ***Intelligente Retrieval Methoden***

### Regelbasierte Verfahren

#### *Deduktion*

|              |                                     |
|--------------|-------------------------------------|
| Regel:       | Wenn es regnet, wird die Straße naß |
| Beobachtung: | Es regnet                           |
| Folgerung:   | Die Straße wird naß                 |

#### *Induktion*

|              |  |
|--------------|--|
| Beobachtung: | Es regnet, und die Straße ist naß          |
| Beobachtung: | Gestern regnete es, und die Straße war naß |
| Beobachtung: | Neulich regnete es, und die Straße war naß |
| Regel:       | Wenn es regnet, wird die Straße naß        |

#### *Abduktion*

|              |                                     |
|--------------|-------------------------------------|
| Regel:       | Wenn es regnet, wird die Straße naß |
| Beobachtung: | Die Straße ist naß                  |
| Folgerung:   | Vielleicht regnet es                |





MIND

*Aber:*

|              |  |
|--------------|--|
| Regel:       | Wenn es regnet, wird die Straße naß                                |
| Regel:       | Wenn mein Nachbar sein Auto wäscht, wird die Straße naß            |
| Regel:       | Bei einem Rohrbruch wird die Straße naß                            |
| Regel:       | Wenn die Straßenreinigung vorbeikommt, wird die Straße naß         |
| Beobachtung: | Die Straße ist naß   |
| Folgerung:   | Vielleicht regnet es, oder mein Nachbar wäscht sein Auto, oder ... |



MIND

## Abduktives Retrieval

### Hypothesen über den Informationsbedarf der Nutzenden

Eine Anfrage ist aus dem Informationsbedarf von Nutzenden abgeleitet.

Abduktion liefert zu einer Anfrage Hypothesen über den Informationsbedarf der Nutzenden.

Diese können mit den Nutzenden verhandelt werden.

### Hypothesen über die Dokumente der Datenbank

Ein Dokument ist zu einer Anfrage relevant, wenn deren Inhalt aus dem Dokument abgeleitet werden kann.

Abduktion liefert zu einer Anfrage Dokumente, aus denen sie abgeleitet werden kann.



MIND

## Möglichkeiten des abduktiven Retrieval

- Integration von Wissen über das Themengebiet durch eine Regelbasis
- Nutzen einer komplexen Indexierungsstruktur
- Ableitung verschiedener Interpretationen einer Anfrage
- Verhandeln verschiedener Interpretationen der Anfrage mit Nutzenden



MIND

## ***Automatische Indexierung***

### Dokumentstrukturen nutzen

#### Textuelle Information

- Bibliographische Information in verschiedenen Feldern (Titel, Abstract, ...)
- Volltexte in elektronischen Auszeichnungssprachen z. B.:
  - SGML  
(Standard Generalized Markup Language, ISO 8879)
  - HTML  
(HyperText Markup Language, WWW)
  - TEX / Latex  
(Auszeichnungssprache im wissenschaftlichen Bereich)
  - PostScript  
(Druckerauszeichnungssprache © Adobe)
- Sammlungsspezifische Formate (DTD,...)



MIND

## Andere Strukturelemente

### HTML-Elemente

- Buttons (Punkte, “New”, ...)
- Markierungen (Linien, Hintergrund)
- Links

## Nichttextuelle Informationen

### Bilder

### Graphiken



MIND

## Inferenznetze

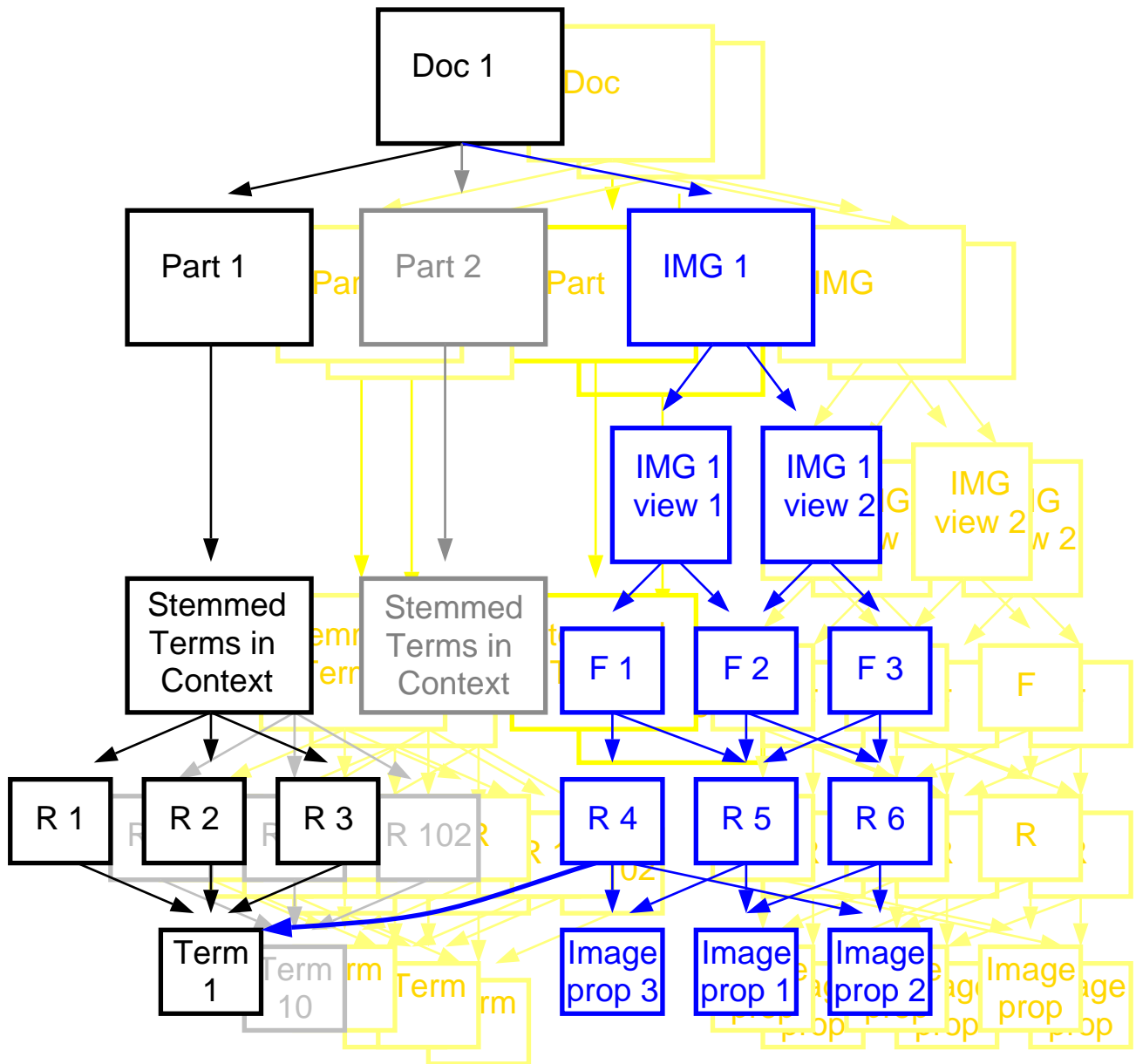
Inferenznetze arbeiten mit

- gewichteten Schätzwerten für Relevanz auf
- verschiedenen parallelen Pfaden

um eine **Rangordnung** der Dokumente bezüglich einer Anfrage zu berechnen.

Themengebietsspezifische Regeln erlauben

- zusätzliche Gewichtung spezifischer Information
- genaue Spezifizierung einer Anfrage



*MAGIC = Multimedia-based Automatic-  
Generation of Indexes and Clusters  
R: Regeln, F: Feature – Detektoren*



MIND

## Automatische Bildindexierung

### Problem:

Bilder haben keine **atomaren Inhaltseinheiten**, die einfach **automatisch** erfaßt werden können (wie Wörter in Texten).

### Ansatz:

Einer Sammlung von Bildern werden einfache Inhaltskategorien wie

- Licht (künstlich oder natürlich)
- Anteil von Schatten
- Art des Bildes (Foto, Gemälde, Zeichnung, ...)
- Bildobjekte (natürlich, künstlich, ...)

manuell zugeordnet.

Für die selben Bilder werden aus der Pixeldarstellung automatisch Werte wie

- fraktale Dimension
- Farbhistogramme
- Anteil der Pixel in Konturlinien
- Texturwerte

berechnet.

Diese beiden Wertegruppen werden miteinander in Beziehung gesetzt, um die Inhaltszuordnungen möglichst gut aus den Pixelwerten ableiten zu können.





MIND

## ***Wissensextraktion aus Datenbanken***

### Knowledge Discovery

Zusammenhänge und Regeln in Massendaten finden.

Beispiele:

- Fachgebietsspezifische assoziative Thesauern
- Assoziative Query Expansion
- Statistische Regeln



MIND

## Assoziative Query Expansion

Anfragen können durch inhaltlich verwandte Terme erweitert werden.

Dabei werden Inhalte durch assoziative Beziehungen zwischen Termen erschlossen.

Inhaltliche Assoziationen können automatisch aus dem gemeinsamen Auftreten von Termen in Dokumentsammlungen gewonnen werden.

### Zitat:

*“Objects once experienced together tend to become associated in the imagination, so that when any one of them is thought of, the others are likely to be thought of also”*

(James, 1890, Vol. 1, p. 561).

Assoziationen zu "Frankfurt", berechnet mit dem *IMAGINE* System auf der Basis eines Kunstlexikons.

|                  |                      |
|------------------|----------------------|
| Frankfurt: 58.00 | Kupferstichkab: 8.29 |
| Städel: 57.17    | Hollstein: 8.12      |
| Schadow: 12.89   | Jakob: 8.07          |
| Karlsruhe: 12.06 | Mander: 7.96         |
| Mainz: 11.61     | Neue: 7.85           |
| Goethe: 11.47    | Mauritshuis: 7.79    |
| Philipp: 11.08   | Regensburg: 7.79     |
| Staatsgal: 11.05 | Ksthalle: 7.78       |
| Residenz: 10.24  | Heidelberg: 7.73     |
| Meister: 10.02   | Kassel: 7.65         |
| Behrens: 9.67    | Staatl: 7.61         |
| Samml: 9.56      | Breslau: 7.61        |
| Nazi: 9.25       | Stuttgart: 7.60      |
| Lessing: 9.08    | Theodor: 7.59        |
| Bonn: 9.06       | Cologne: 7.57        |
| Bremen: 9.00     | Essen: 7.48          |
| Mannheim: 9.00   | Kunstakademie: 7.44  |
| Staatliche: 8.86 | Adriaen: 7.18        |
| Gerhard: 8.84    | Mengs: 7.12          |
| Ksthist: 8.69    | Nuremberg: 7.12      |
| Gropius: 8.56    | Hochschule: 7.10     |
| Gemäldegal: 8.40 | Ludwig: 7.00         |
| Landesmus: 8.35  | Christoph: 6.97      |

*IMAGINE = Interaction Merger for Associations  
Gained by Inspection of Numerous Exemplars*

*Assoziationen zu "Frankfurt", auf der Basis eines Korpus aus Meldungen der Agentur Reuter von 1987*

|                    |                     |
|--------------------|---------------------|
| Frankfurt: 266.00  | Porsche: 29.64      |
| Mueller: 106.67    | Hamburg: 29.63      |
| Murphy: 100.00     | Babcock: 29.56      |
| Schmidt: 88.89     | Bayer: 29.56        |
| Zurich: 73.33      | Ciba-Geigy: 29.56   |
| Karl: 69.48        | GmbH: 28.84         |
| Otto: 68.17        | Indosuez: 28.07     |
| Poehl: 68.17       | London-based: 27.36 |
| Deutsche: 62.34    | Milan: 27.36        |
| Volkswagen: 49.38  | Hahn: 26.67         |
| Gerhard: 45.04     | Hesse: 26.67        |
| Keating: 44.44     | Carl: 25.33         |
| Audi: 44.33        | Chris: 25.33        |
| Stoltenberg: 43.05 | Easter: 25.33       |
| Commerzbank: 41.56 | Anthony: 23.47      |
| Louvre: 41.48      | German: 23.28       |
| Schlesinger: 38.11 | Amsterdam: 21.71    |
| Bundesbank: 37.57  | Hans: 20.52         |
| Kangyo: 33.33      | Collins: 20.46      |
| Alcan: 33.25       | Williams: 20.00     |
| Bang: 33.25        | Helmut: 19.75       |
| Dresdner: 31.92    | Kohl: 19.75         |
| Dai-Ichi: 31.29    | Visa: 19.05         |
| Bonn: 30.40        | Berlin: 19.00       |

*Assoziationen zu "Iraq", berechnet auf der Basis eines Kunstlexikons.*

|                     |                    |
|---------------------|--------------------|
| Iraq: 169.00        | Shapur: 46.11      |
| Ubaid: 159.06       | Dynastic: 44.09    |
| Baghdad: 131.76     | Achaemenid: 43.46  |
| Uruk: 128.76        | Persepolis: 40.24  |
| Ur: 125.54          | Elam: 37.56        |
| Sumerian: 106.41    | Syria: 35.21       |
| Akkadian: 87.88     | Damascus: 34.67    |
| Mesopotamian: 84.50 | Iran: 34.49        |
| Euphrates: 84.50    | Anyang: 33.83      |
| Mosul: 84.50        | Afghanistan: 32.95 |
| Elamites: 78.87     | Ahmad: 32.19       |
| Babylon: 72.43      | Anatolia: 31.71    |
| Babylonia: 71.16    | Mari: 31.69        |
| Mesopotamia: 69.69  | Dura: 31.69        |
| Fatimid: 67.60      | Hasanlu: 30.73     |
| Babylonian: 65.95   | Assyrian: 30.49    |
| Parthians: 65.72    | Cyrus: 29.82       |
| Tepe: 58.09         | Assyria: 28.19     |
| Parthian: 58.09     | Aleppo: 28.17      |
| Luristan: 56.33     | Koran: 27.05       |
| Susa: 52.81         | Pergamonmus: 26.41 |
| Sasanian: 52.81     | Islam: 26.00       |
| Persians: 50.70     | Arab: 25.47        |
| Elamite: 46.94      | Iranian: 24.41     |

*Assoziationen zu "Iraq", auf der Basis eines Korpus aus  
Meldungen der Agentur Reuter von 1987*

|                  |                   |
|------------------|-------------------|
| Iraq: 174.00     | Gulf: 34.69       |
| Basra: 168.39    | Arabia: 34.12     |
| Baghdad: 157.16  | Weinberger: 32.97 |
| Iraqi: 141.68    | Arabian: 32.37    |
| Iranians: 133.85 | Libya: 30.29      |
| Silkworm: 107.88 | Pentagon: 29.85   |
| Tehran: 92.90    | Lebanon: 29.03    |
| Hormuz: 79.09    | Saudi: 27.77      |
| Iranian: 78.30   | Sea: 26.32        |
| Lukman: 69.68    | Petrobras: 24.90  |
| Iran: 66.86      | Opec: 23.23       |
| Mideast: 63.87   | Prince: 23.23     |
| Sheikh: 61.50    | BC-u: 22.70       |
| Kuwait: 55.43    | Carlos: 22.70     |
| Brent: 55.36     | Nigeria: 22.69    |
| Kuwaiti: 55.32   | Murphy: 21.77     |
| Abu: 52.20       | Ocean: 21.77      |
| Strait: 51.61    | Ali: 21.75        |
| Qatar: 46.09     | Navy: 20.90       |
| Vienna: 38.71    | Oman: 20.50       |
| Jordan: 37.86    | Beirut: 20.47     |
| Emirates: 37.83  | Margoshes: 19.36  |
| Caspar: 37.62    | Egypt: 19.33      |
| Saudis: 37.33    | Honduras: 18.32   |
| Arab: 35.19      | Javier: 16.60     |

*Assoziationen zu "Darmstadt", berechnet auf der Basis eines Kunstlexikons.*

|                     |                          |
|---------------------|--------------------------|
| Darmstadt: 236.00   | Troost: 23.60            |
| Hess: 138.40        | Haus: 23.60              |
| Rottmann: 64.55     | Bel: 22.48               |
| Behrens: 53.29      | Kunsthalle: 22.20        |
| Bellotto: 52.44     | Lenbach: 21.52           |
| Kauffman: 47.20     | Krefeld: 21.50           |
| Riemerschmid: 44.25 | Vroom: 21.45             |
| Landesmus: 43.42    | Wallis: 21.45            |
| Baden: 39.47        | Häring: 21.45            |
| Dumonstier: 39.33   | Technische: 21.00        |
| Slevogt: 33.83      | Nietzsche: 20.52         |
| Dillis: 33.71       | Kunstakademie: 20.17     |
| Feuerbach: 31.57    | Secessionist: 19.73      |
| Geddes: 30.85       | Lichtenstein: 19.73      |
| Begas: 29.60        | Merz: 19.73              |
| Goodhue: 29.60      | Becker: 19.67            |
| Schenck: 29.50      | Worms: 19.67             |
| Marees: 26.30       | Piazzetta: 19.67         |
| Münter: 26.22       | Koch: 19.27              |
| Keller: 26.22       | Hagen: 18.93             |
| Werkbund: 24.00     | Göttingen: 18.93         |
| Unwin: 23.68        | Kunstaussstellung: 18.93 |
| Vitebsk: 23.68      | Speyer: 18.20            |
| Vallotton: 23.67    | Linz: 18.20              |



MIND

## ***Mitarbeitende in der Abteilung MIND***

e-mail: [mindcore@darmstadt.gmd.de](mailto:mindcore@darmstadt.gmd.de)  
oder: [nachname@darmstadt.gmd.de](mailto:nachname@darmstadt.gmd.de)  
<http://www-cui.darmstadt.gmd.de/mind/>  
Stand auf der Infobase: **4.1 B 31**

Dr. Ulrich Thiel

Abteilungsleiter, Information Retrieval, Dialog Management (Mi)

Dr. Reginald Ferber

Knowledge Discovery, Information Retrieval (Mi, Do)

Dr. Jon Atle Gulla

Dialog Management, Explanation Generation (Do)

Said Kutschekmanesch

Automated Indexing (Do)

Barbara Lutes

Thesauri, Terminology (Di, Mi)

Adrian Müller

Logical Retrieval, Abductive Reasoning (Di)

Dr. Adelheit Stein

Dialog Modeling, Multimodal Interfaces (Mi, Do)