

Inhaltsüberblick

Warum WR für Informationswissenschaftler?

Vektorraum IR - IRS - IS - Methodentransfer in IR

Warum mehrere/viele WR-Formate?

Elefanten-Gleichnis - No Free Lunch Theoreme - Polyrepräsentation

Warum wird WR in Zukunft wichtiger?

Welche WR- und IR-Ansätze werden behandelt?

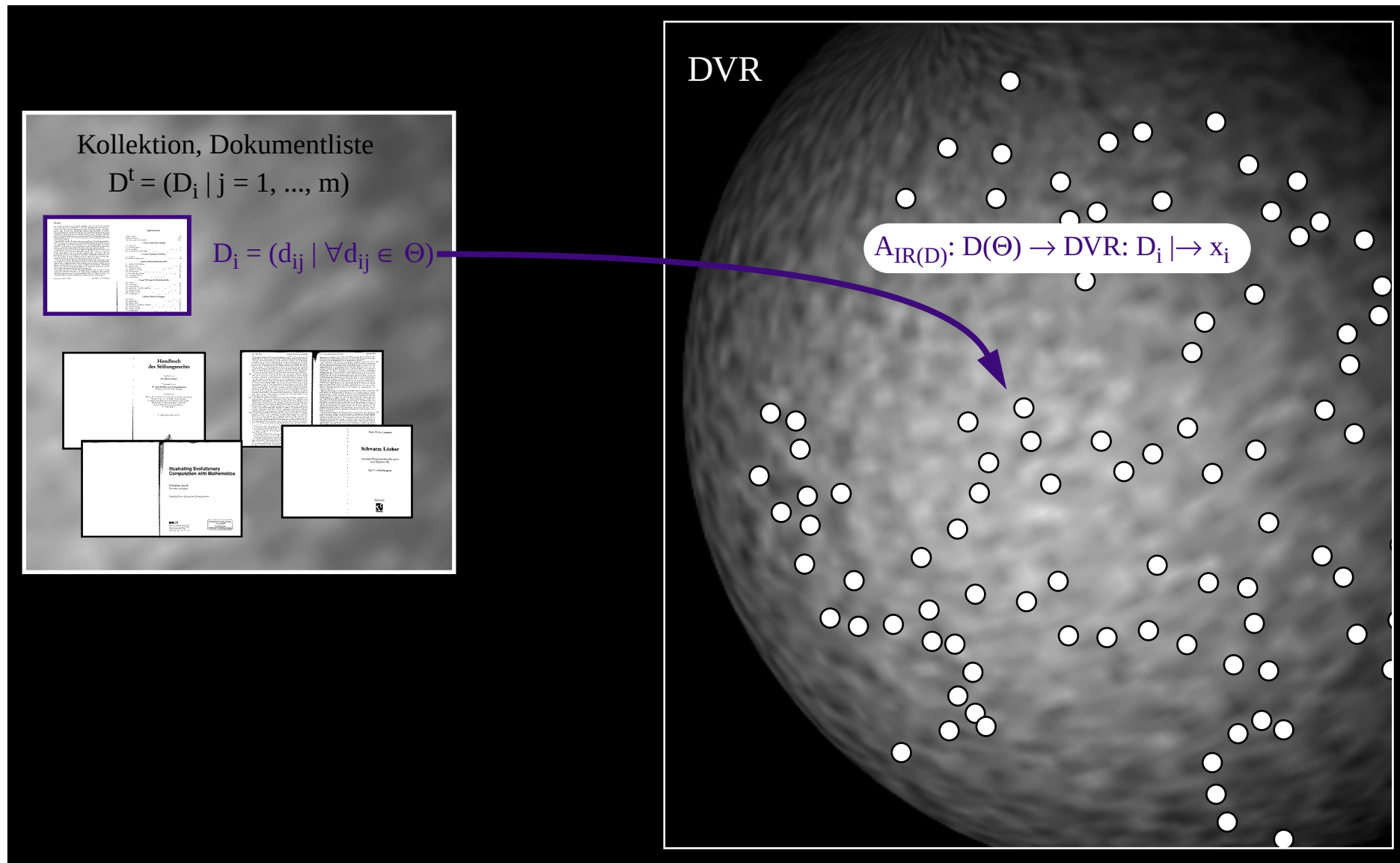
Frage:

Warum muss man sich als
Informationswissenschaftler mit
Wissensrepräsentation beschäftigen?

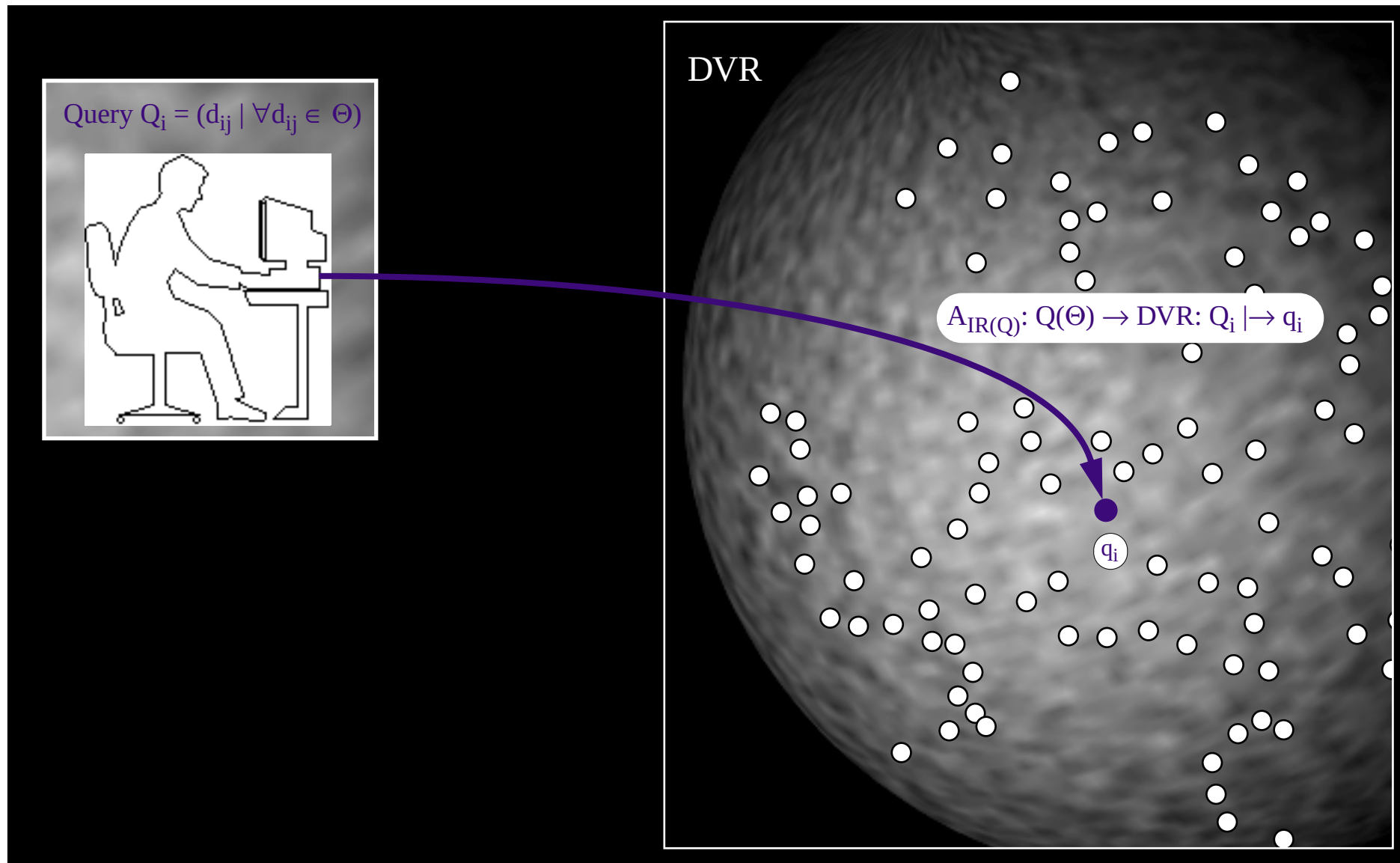
Antwort:

Repräsentation von
Informationenobjekten ist
notwendiger Bestandteil jedes
Informationssystems!

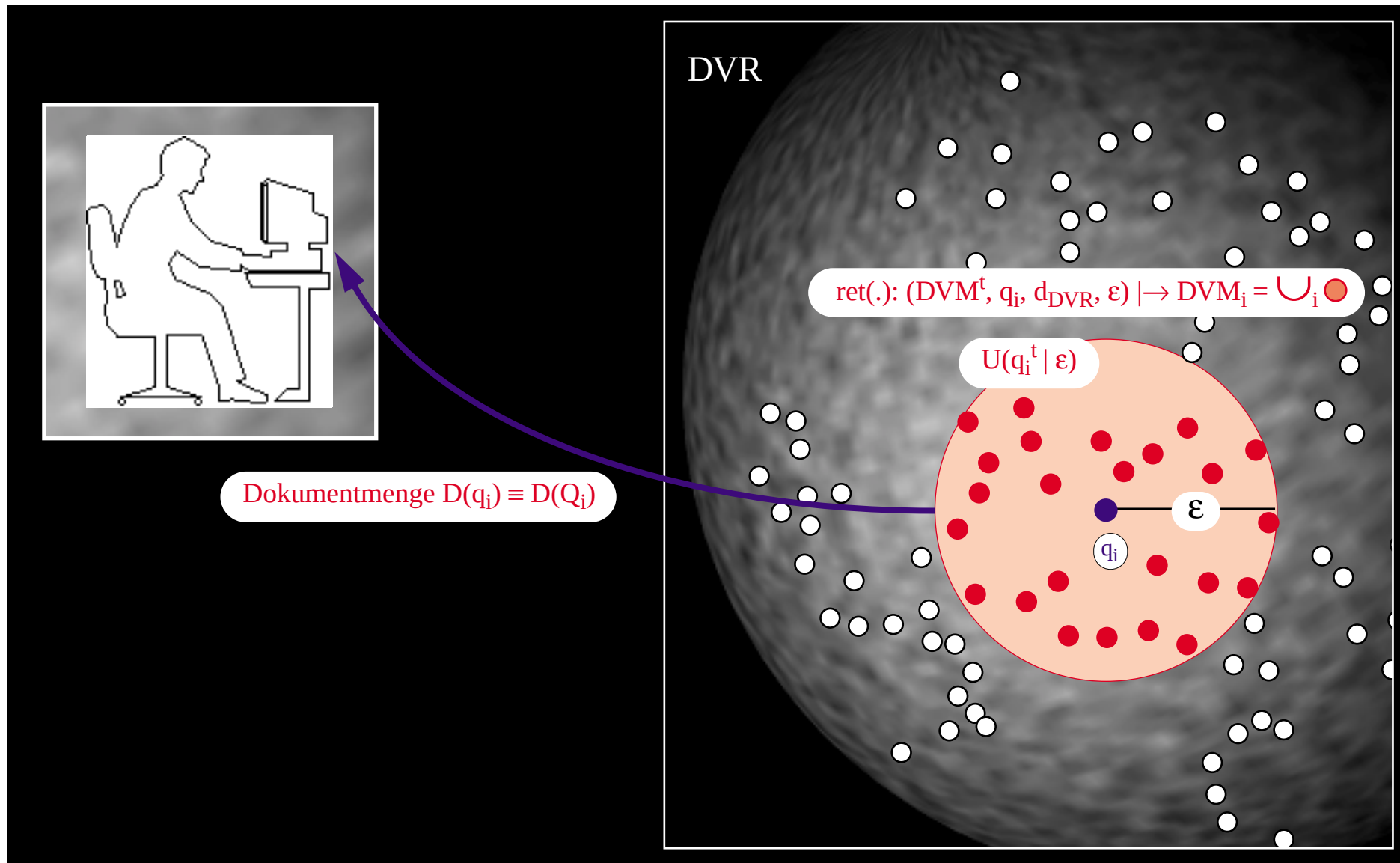
Dokument-Indexierung im Vektorraummodell (Repräsentationsaufbau)



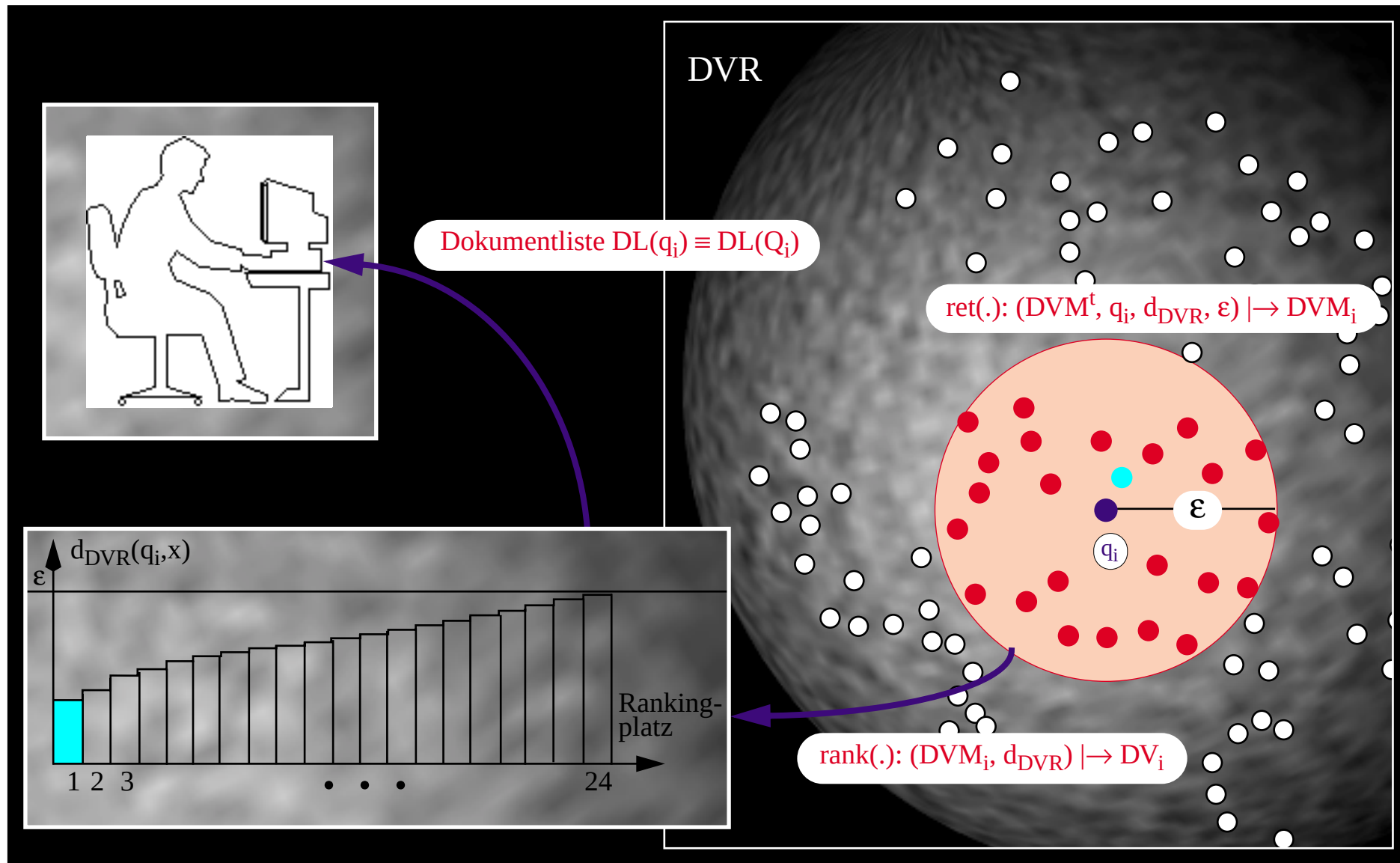
Query-Indexierung im Vektorraummodell (Repräsentationsaufbau)



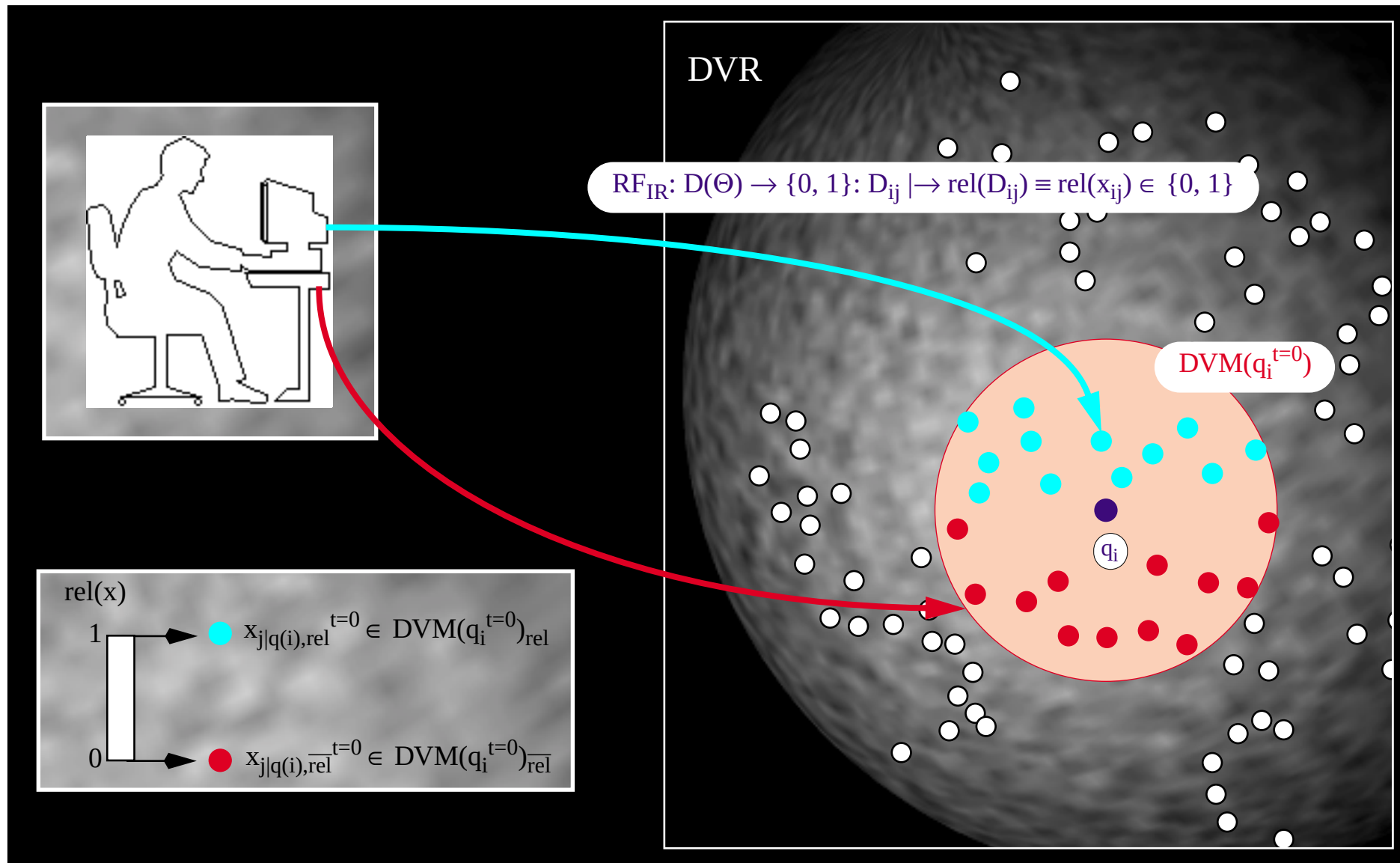
Retrievalfunktion (mit ε -Umgebung)



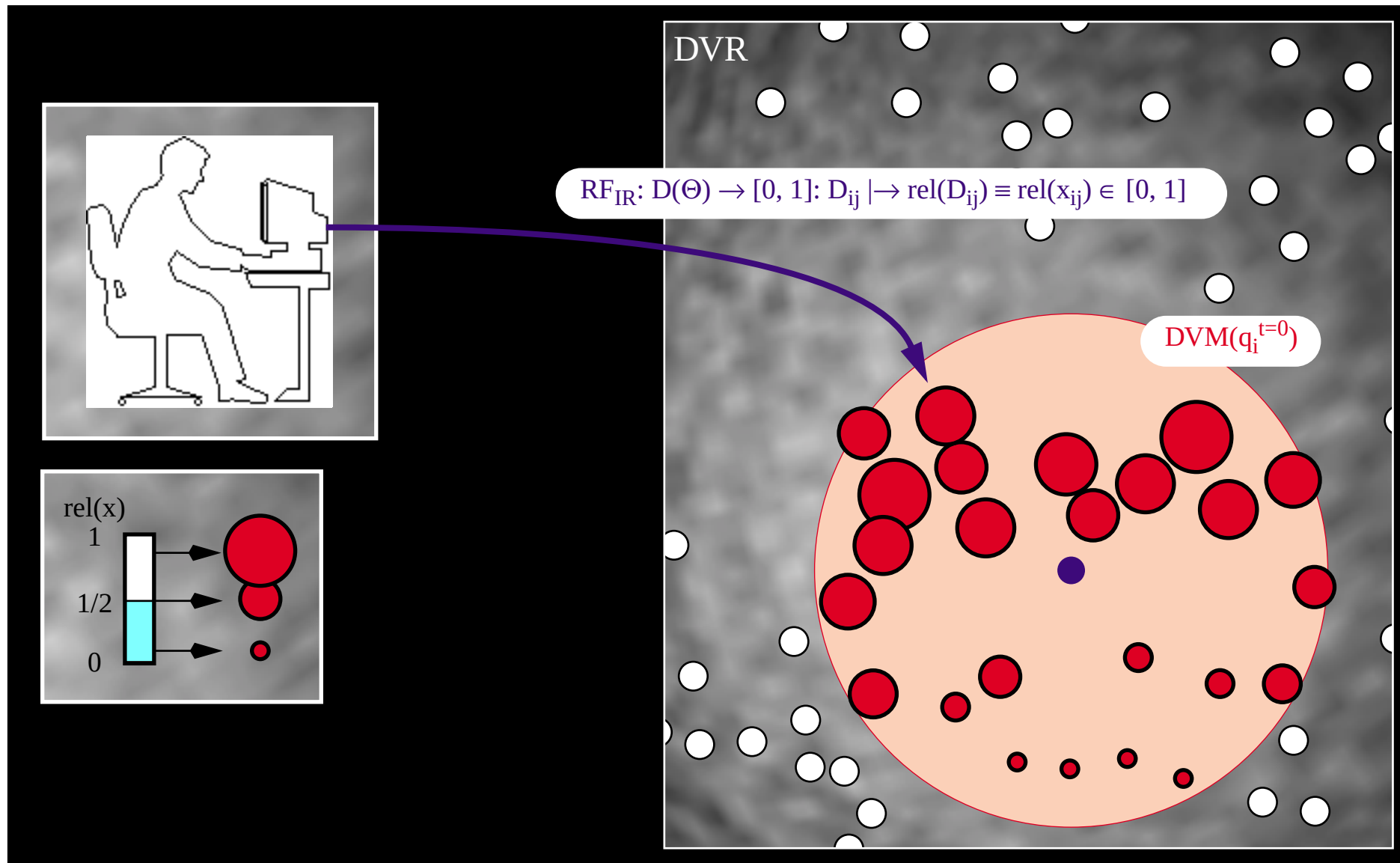
Retrieval- und Rankingfunktion (mit ε -Umgebung)



Relevanz-Feedbackfunktion (binär)



Relevanz-Feedbackfunktion (reell, normiert auf $[0,1]$)

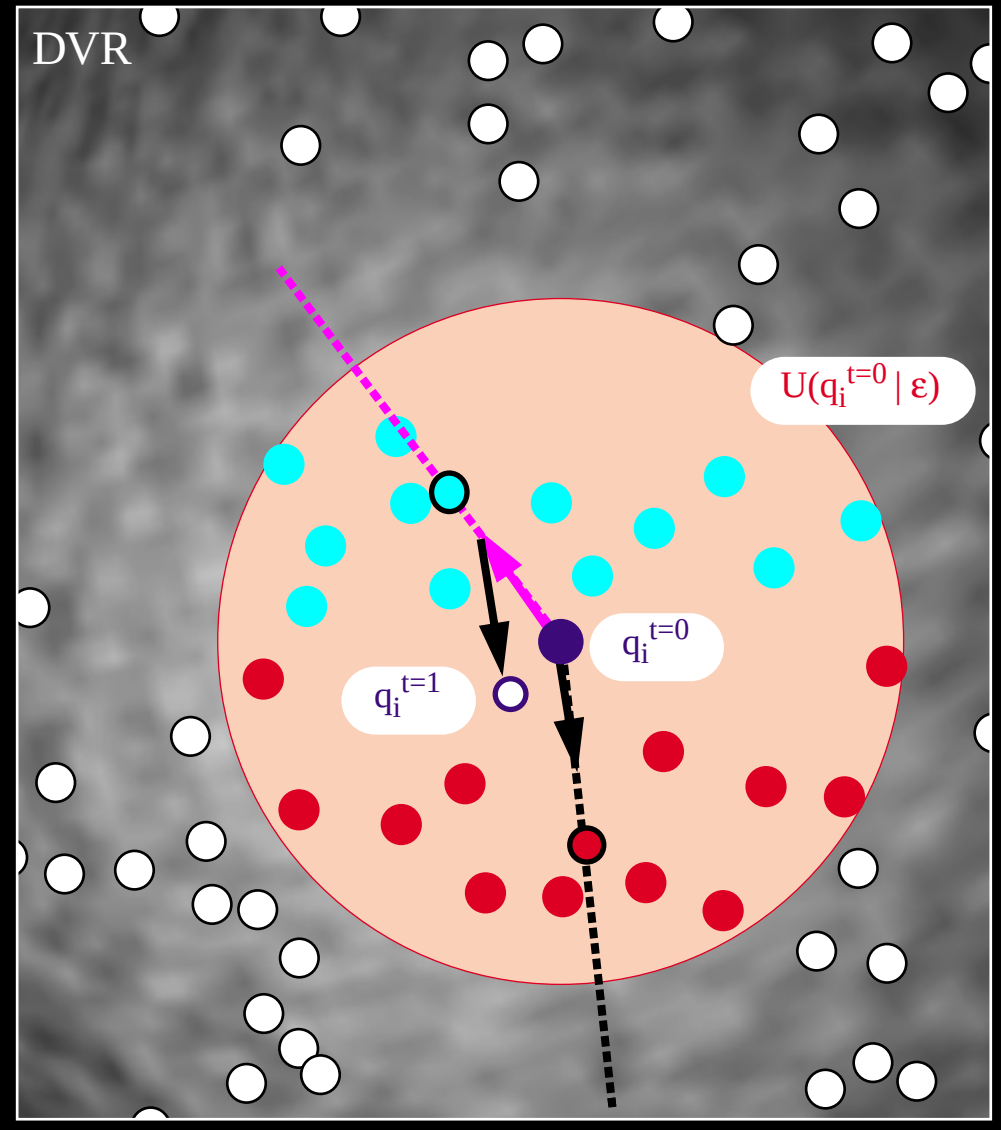


Gemischte Queryvektor-Feedback-Strategie bei binärem Relevanz-Feedback

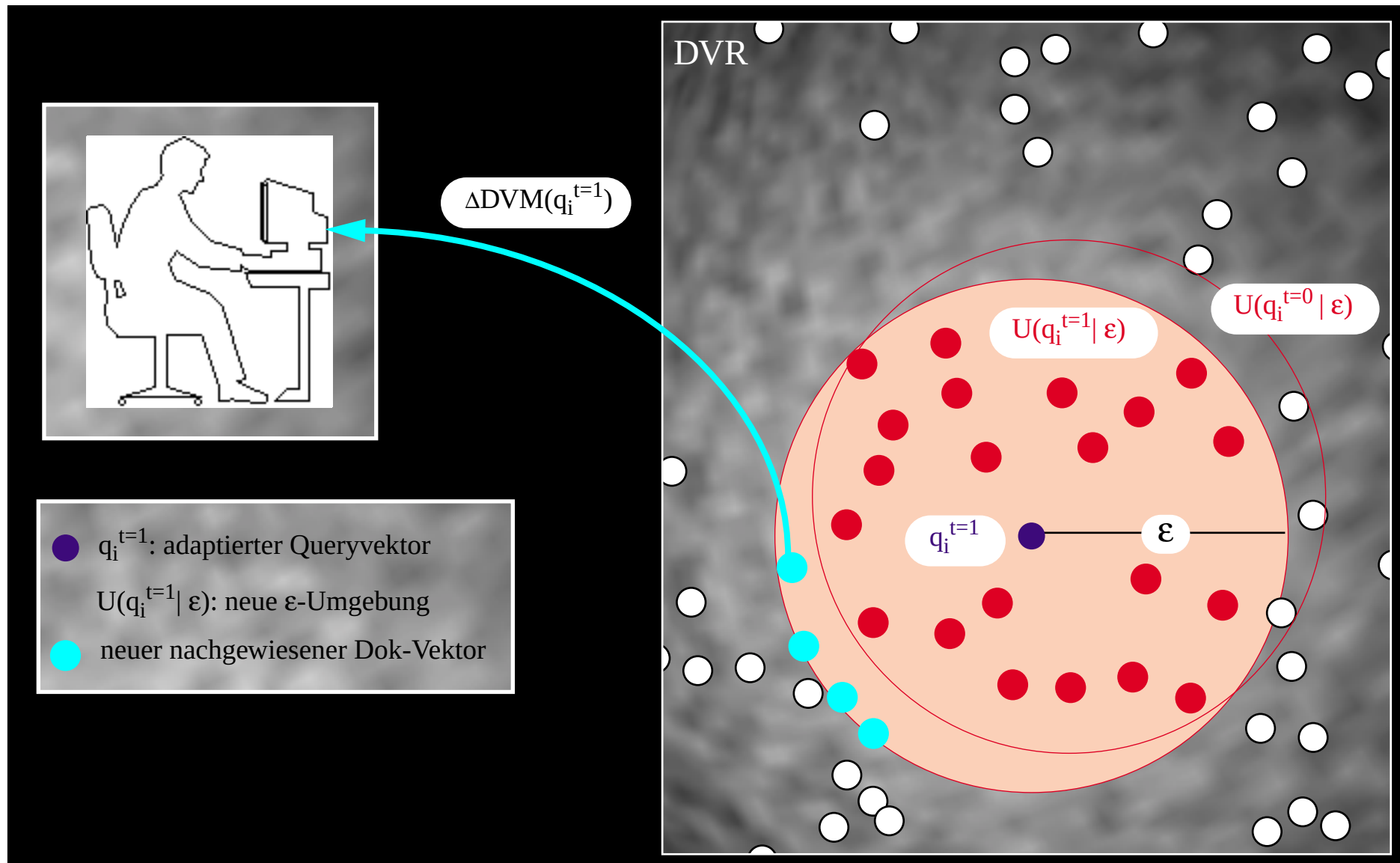


- $\bar{s}_{DVM(i,rel)}^{t=0}$: Schwerpunkt relevanter x_j
- $\bar{s}_{DVM(i,\overline{rel})}^{t=0}$: Schwerpunkt nicht relev. x_j
- ➔ positives Lernen mit Lernfaktor β
- ➔ negatives Lernen mit Lernfaktor γ
- $q_i^{t=1}$: adaptierter Queryvektor, z.B.

$$= \alpha * q_i^{t=0} + \beta * \bar{s}_{DVM(i,rel)}^{t=0} - \gamma * \bar{s}_{DVM(i,\overline{rel})}^{t=0}$$



Neue ε -Umgebung und neue nachgewiesene Dokumentvektoren



Information Retrieval System (IRS)

Ein (Vektorraummodell)-IRS besteht aus den Komponenten:

$IS = (A_{IR(D)}, W, Q, A_{IR(Q)}, E, \text{ret}(.), \text{rank}(.), RF_{IR}, \text{rfb}(.))$, mit

- 1) $A_{IR(D)}$ Dokument-Indexierungsfunktion als Abbildung eines Dokumentes D_i auf einen Dokumentvektor x_i .
- 2) W Menge aller internen Repräsentationen (Dokumentvektorraum DVR mit Dokumentvektormenge DVM).
- 3) Q Menge aller zugelassenen Suchfragen Q_i .
- 4) $A_{IR(Q)}$ Query-Indexierungsfunktion als Abbildung einer Anfrage Q_i auf einen Queryvektor q_i .
- 5) E Menge aller möglichen Outputmengen (P^{DVM} : Potenzmenge der DVM) bzw. Outputlisten (beim Ranking).
- 6) $\text{ret}(.)$ Retrievalfunktion als Abbildung einer indexierten Suchfrage auf eine DVM-Teilmenge (Outputmenge DVM_i).
- 7) $\text{rank}(.)$ Rankingfunktion.
- 8) RF_{IR} Relevanz-Feedbackfunktion als Abbildung von Dokumenten auf Relevanzwerte.
- 9) $\text{rfb}(.)$ Update-Funktion (z.B. Queryvektor-Adaption).

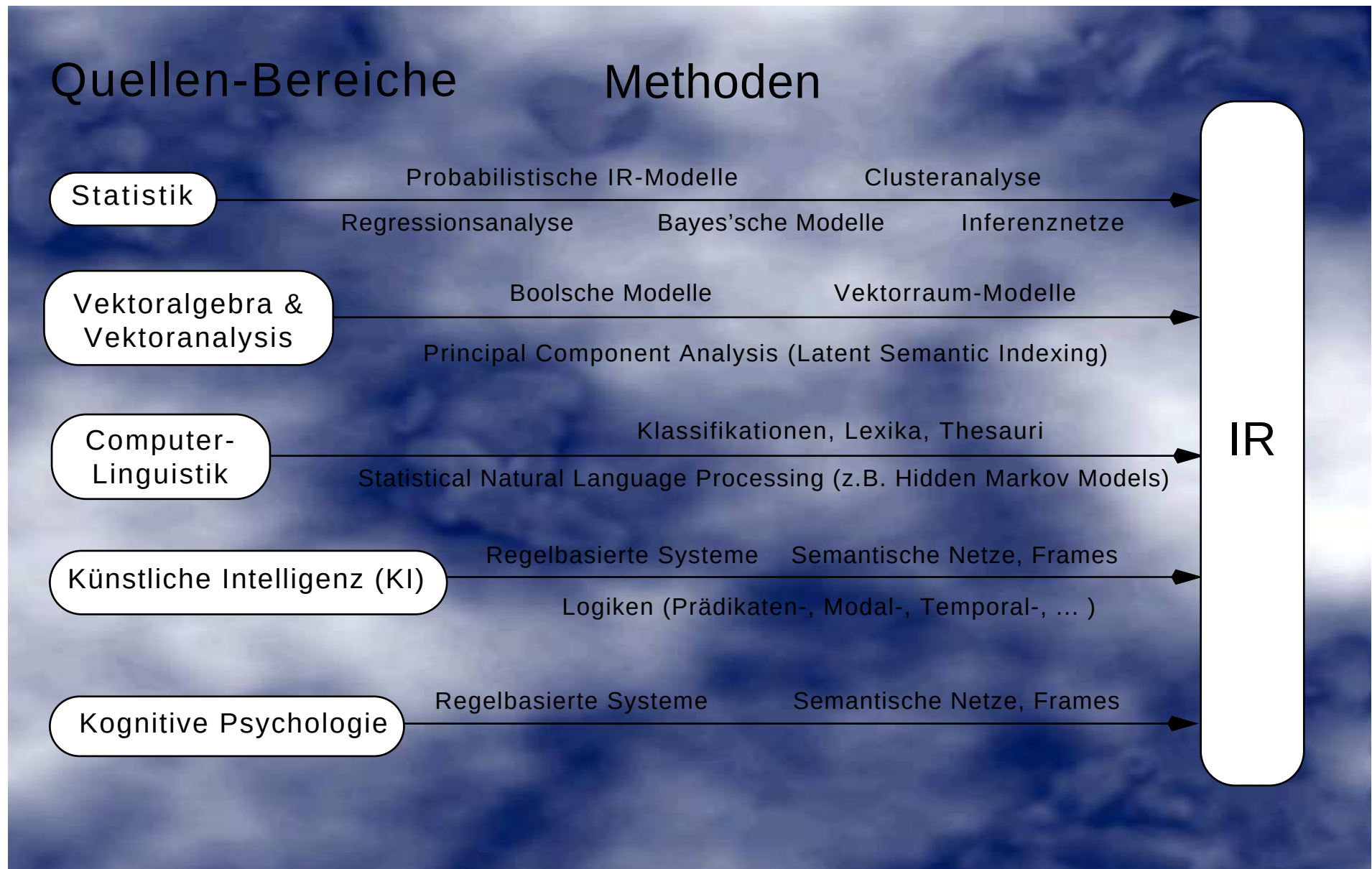
Informationssystem (IS)

Ein Informationssystem IS kann als Spezifizierung eines IRS betrachtet werden. Es besteht aus den Komponenten (Panyr (1986: 22):

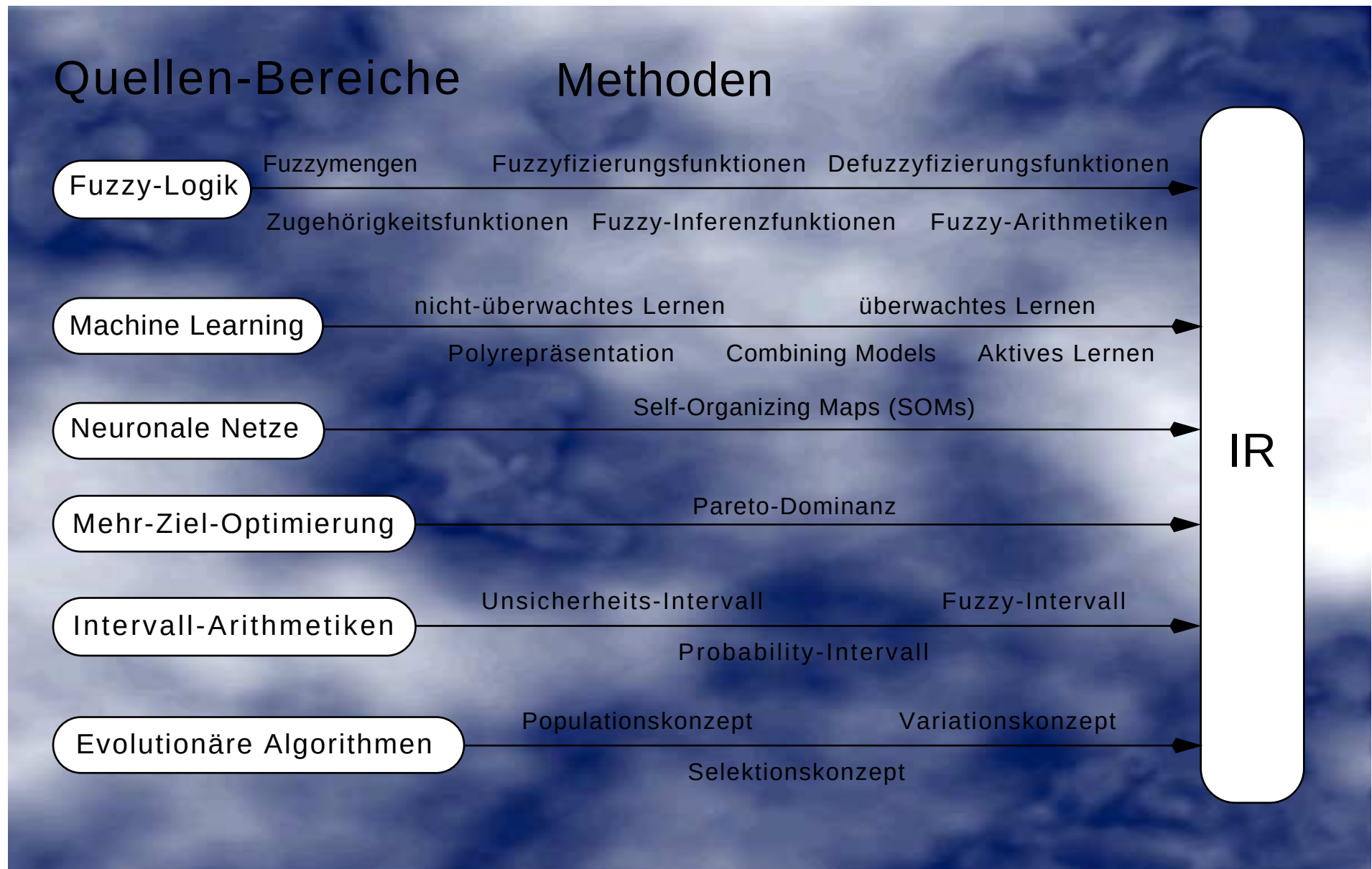
$$IS = (A, W, Q, I, E, U, D), \text{ mit}$$

- 1) **A** Inputfunktion (Erschließungsfunktion, Lernfunktion) zum Aufbau der internen Repräsentationen.
- 2) **W** Interne Repräsentationen (Wissensbasis).
- 3) **Q** Inputmenge als Menge aller zugelassenen Inputkonfigurationen (Problemformulierung, Suchfrage).
- 4) **I** Outputfunktion (Inferenzfunktion, Retrieval-, Rankingfunktion).
- 5) **E** Outputmenge als Menge aller möglichen Outputkonfigurationen (Problemlösung, Systemvorschlag).
- 6) **U** Updatefunktion der internen Repräsentationen (Relevanz-Feedback- und Adaptionfunktion).
- 7) **D** Dialogkomponente.

Methodentransfer ins IR (1/2)



Methodentransfer ins IR (2/2)



Frage:

Warum muss man sich mit
mehreren/vielen Formen der
Wissensrepräsentation
beschäftigen?

Antwort:

Keine Repräsentationsform kann
alle Aspekte eines Phänomens
(Informationsobjektes)
repräsentieren!

Elefanten-Gleichnis (1/2) - Disjunktheit



Elefanten-Gleichnis (2/2) - Überlappungen



No free Lunch - Theorem for Search (Machine Learning)

NFL for Search

Wolpert & Macready (1996)

Kein Suchverfahren ist allen anderen Suchverfahren bezüglich allen Suchproblemen überlegen.

=> für jedes Suchproblem müssen geeignete Suchverfahren ausgewählt werden

Beziehung zwischen Suche und Repräsentation

=> jede Suche findet in einem Suchraum statt

=> jeder Suchraum ist definiert durch eine Repräsentationsart

No free Lunch - Theorem for Representation

NFL for Representation

Keine Repräsentationsart ist allen anderen
Repräsentationsarten bezüglich allen zu
repräsentierenden Phänomenen
(Informationsobjekten) überlegen.

Empirischer Beleg im IR: Zobel & Moffat (1998).

Redundanz und Polyrepräsentation

Redundanz

Mehrfache und jeweils gleiche Repräsentation eines Informationsobjektes.

Ziele: - Verbesserung von Zugriffszeiten in DBS;
- höhere Ausfallsicherheit.

Polyrepräsentation

Mehrfache und jeweils unterschiedliche Repräsentation eines Informationsobjektes.

Ziel: Effektivitätssteigerung, da bei jeder Einzelrepräsentation unterschiedliche Perspektiven fokussiert werden können, die durch ein einzelnes Modell nicht bzw. nicht in diesem Umfang repräsentierbar sind.

Polyrepräsentationsarten und Combining

Interparadigmen-Polyrepräsentation (Ingwersen (1992, 1994))

Mehrfache Repräsentation eines Informationsobjektes mit unterschiedlichen Repräsentationsarten.

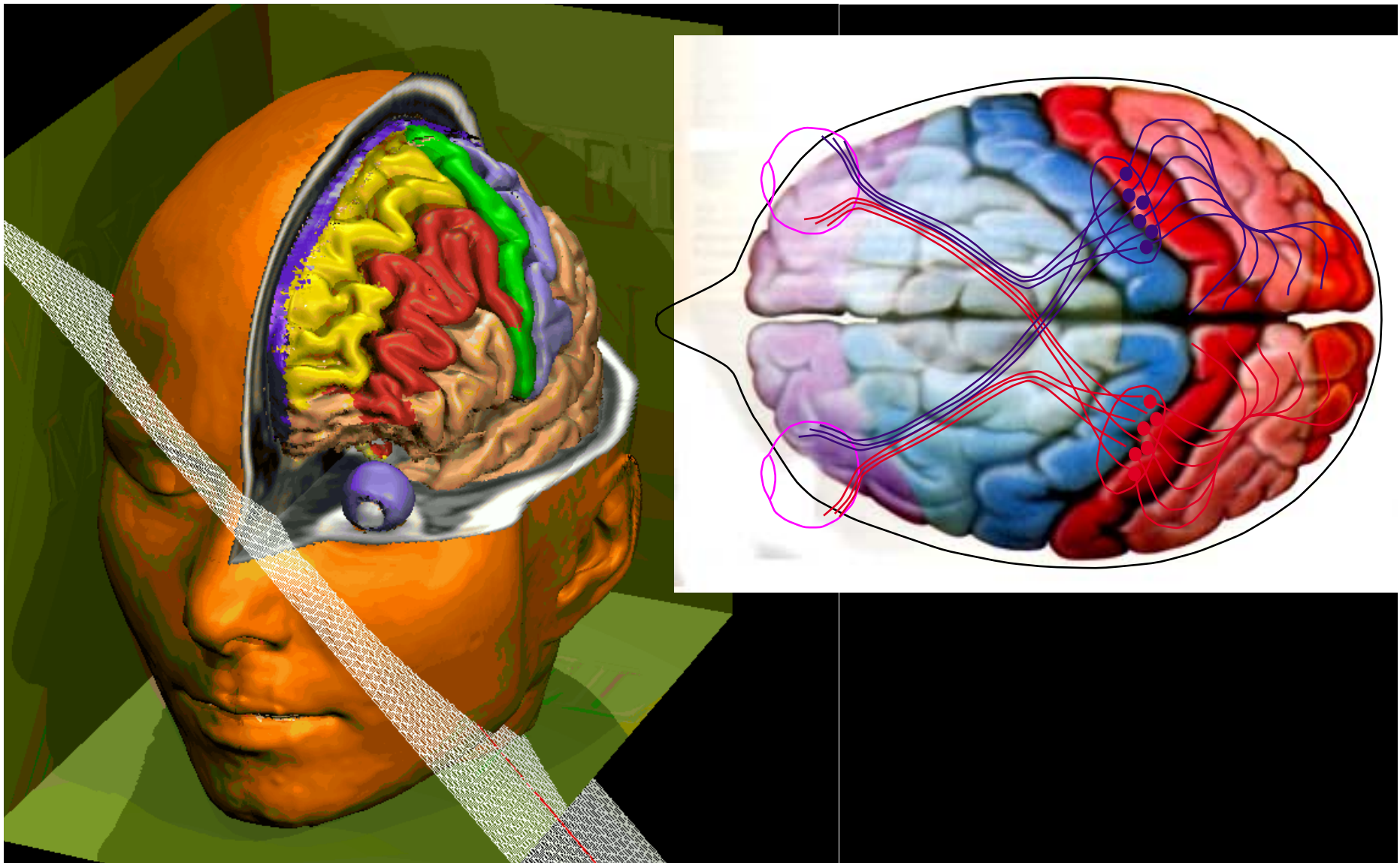
Intraparadigmen-Polyrepräsentation (Bachelier (2002))

Mehrfache Repräsentation eines Informationsobjektes mit der gleichen Repräsentationsart (mit unterschiedlichen Parametern).

Combining, Fusion

Zusammenfügen von Ergebnissen, die sich aus einzelnen Repräsentationen einer Polyrepräsentation ergeben.

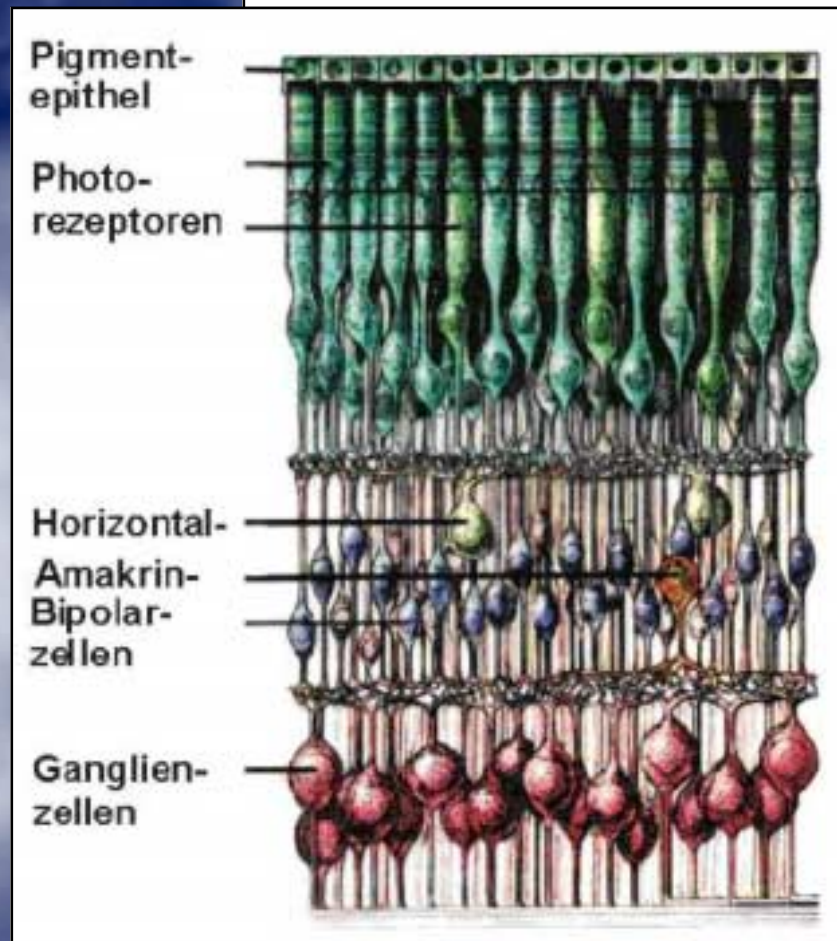
Bio: Makroskopische Input-Polyrepräsentation durch 2 Augen gleichen Typs



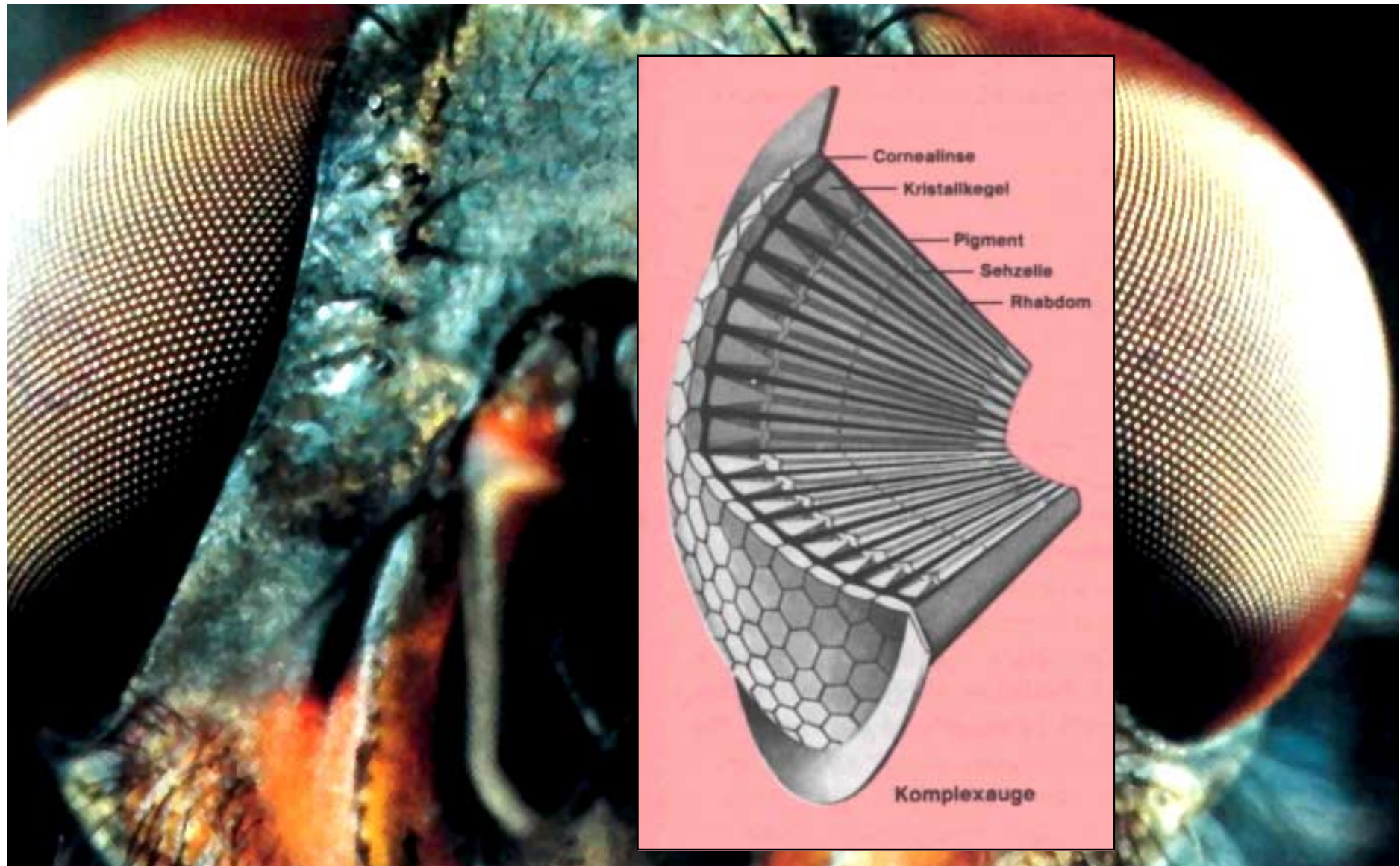
Beispiel einer fehlenden makroskopischen Input-Polyrepräsentation



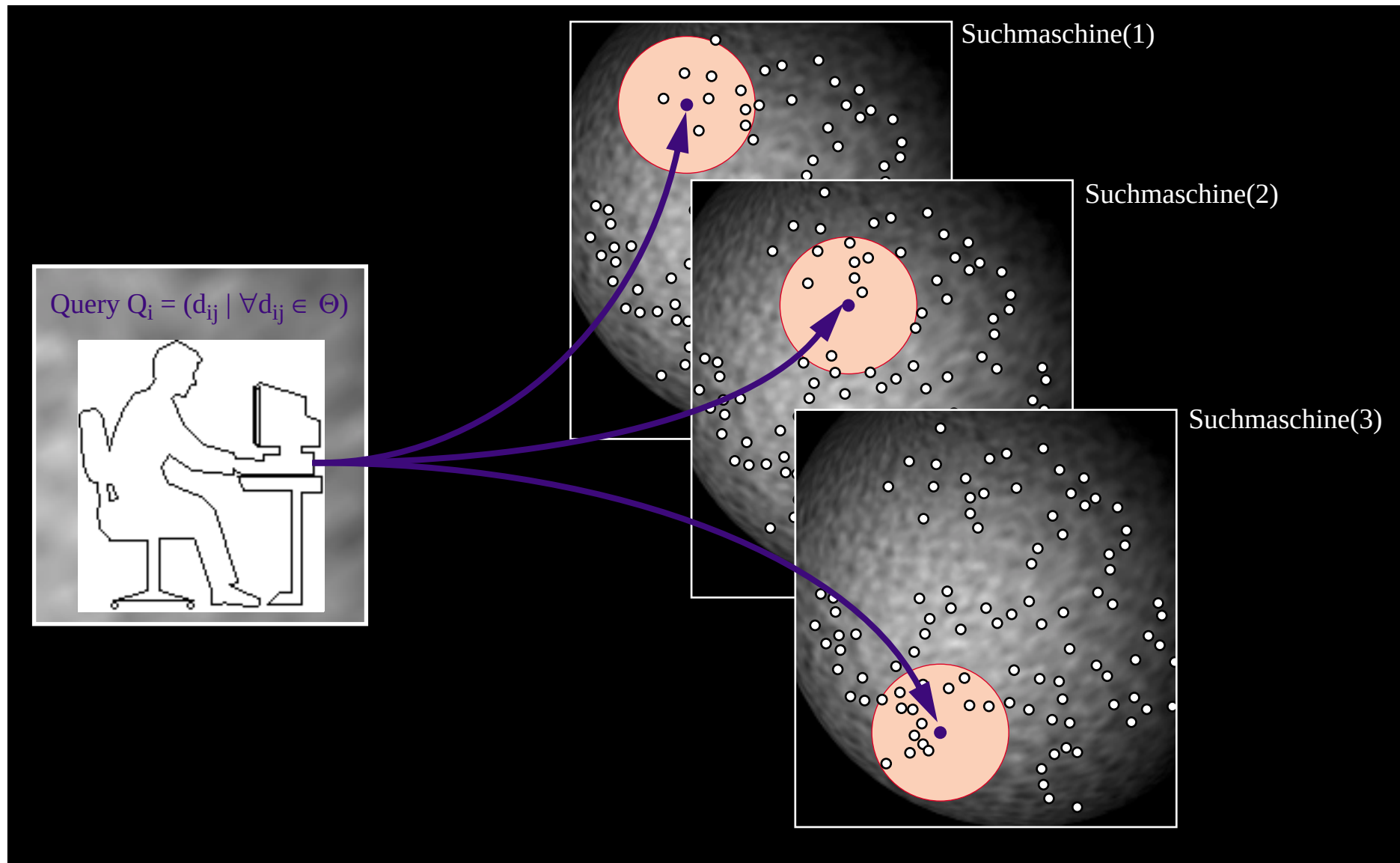
Trotzdem mikroskopische Input-(Intra- u. Inter-)Polyrepräsentation



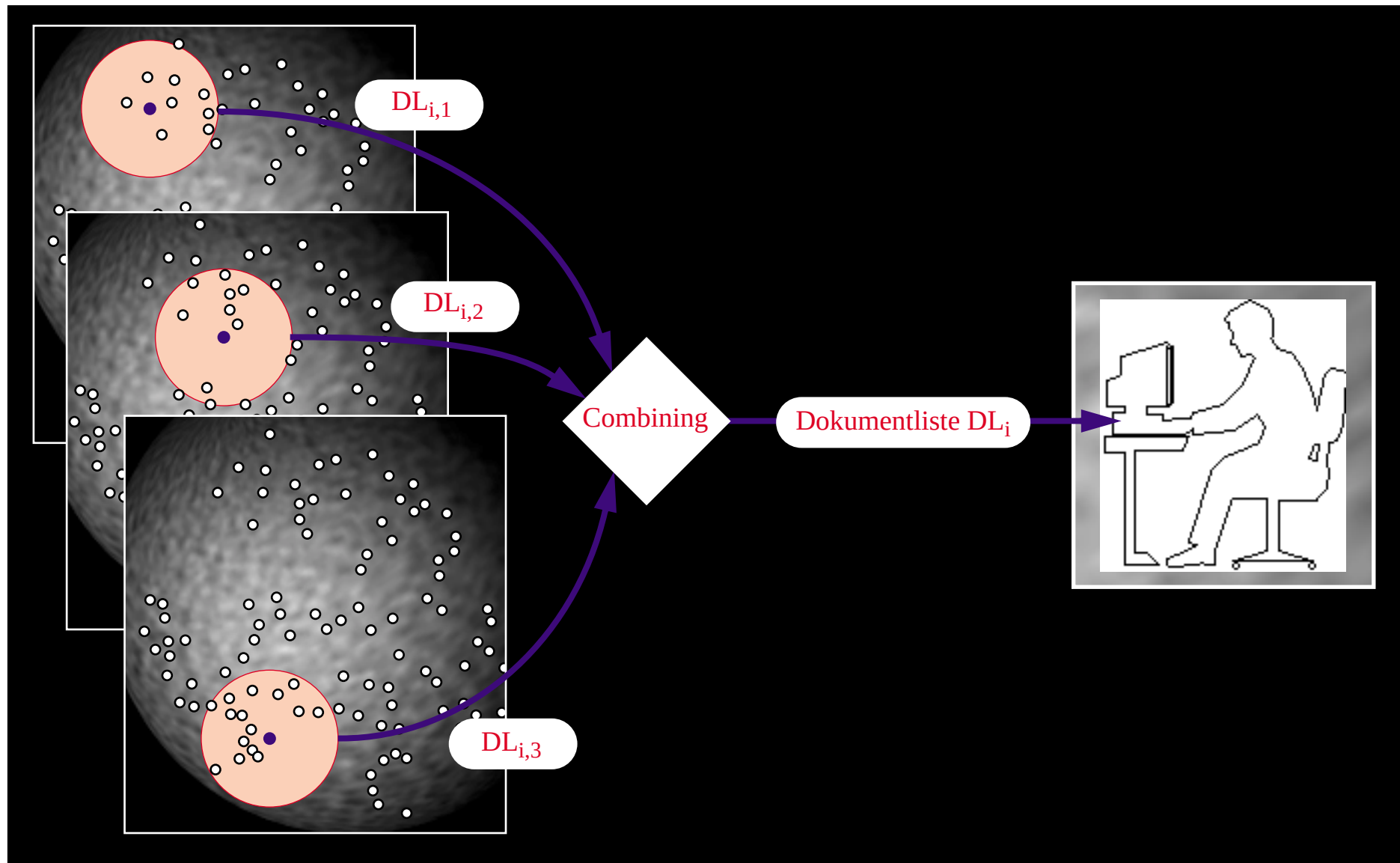
Bio: Komplexauge als mikroskopische Input-Polyrepräsentation



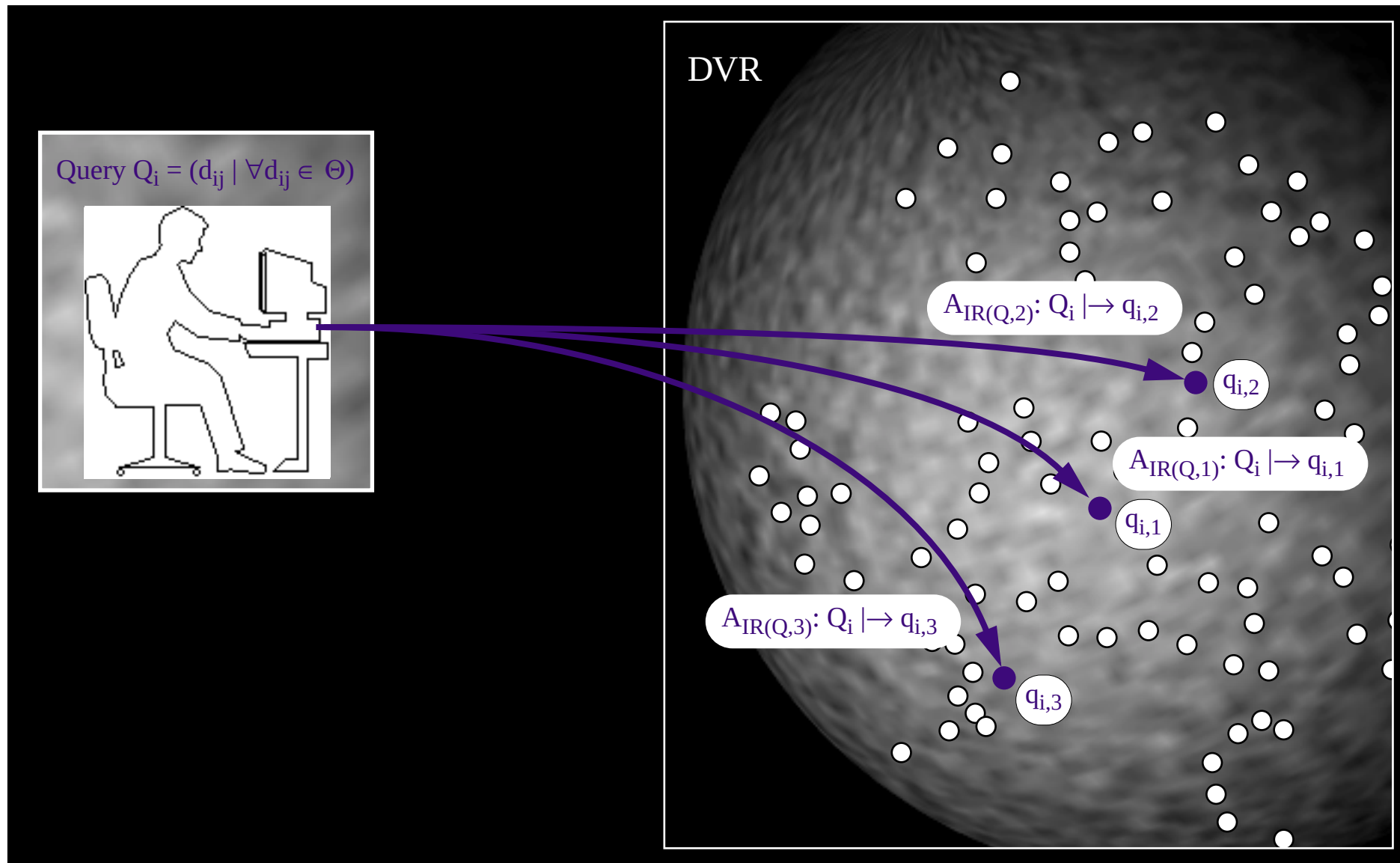
InfoWis: Metasuchmaschinen (= Dokument-Polyrepräsentation) 1/2



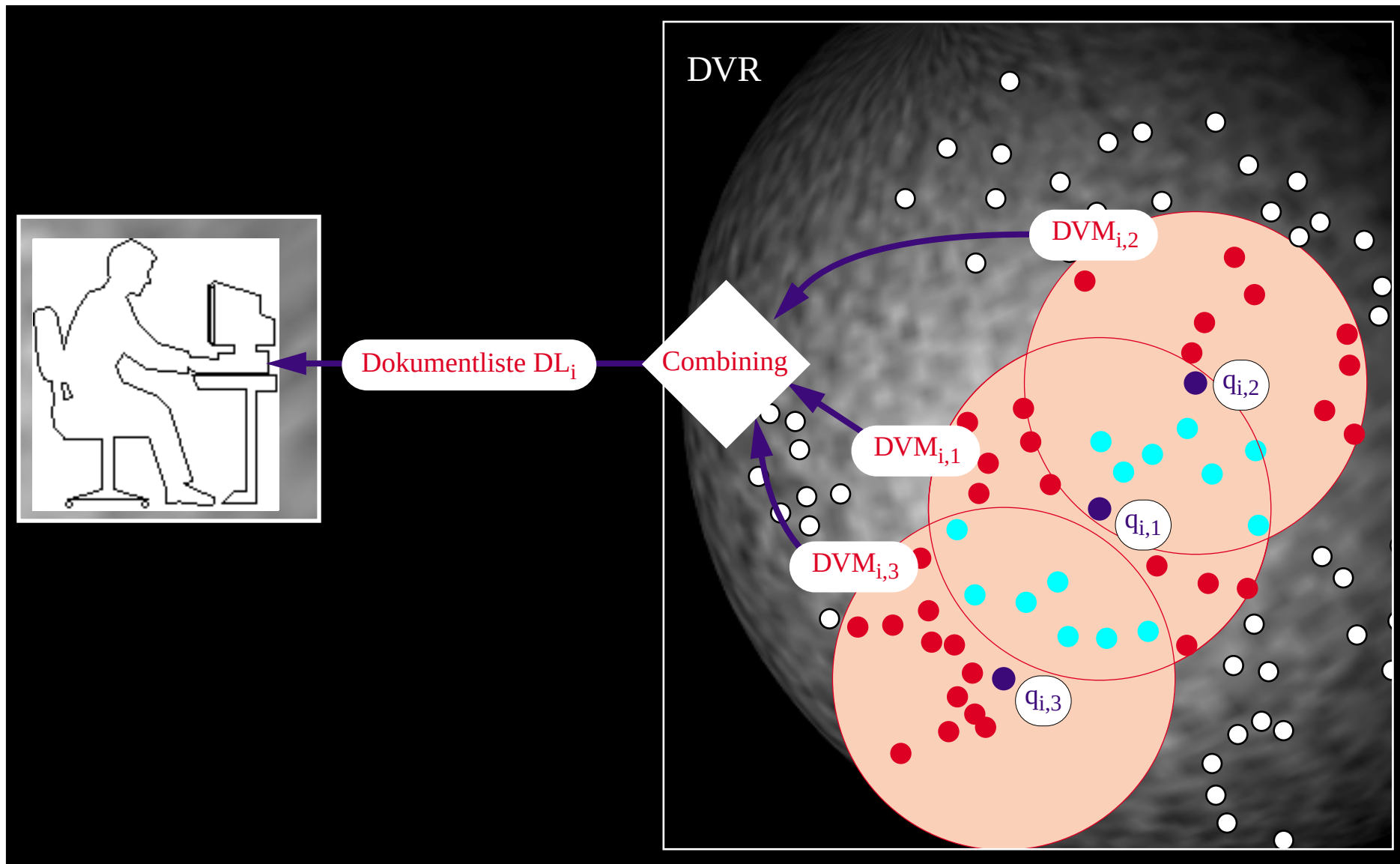
InfoWis: Metasuchmaschinen (= Dokument-Polyrepräsentation) 2/2



Queryvektor-Polyrepräsentation durch Indexierungsfkt-Polyrepräsentation



Combining der Ergebnis-Dokumentvektormengen



Frage:

Warum wird die Beschäftigung mit
Wissensrepräsentation in Zukunft
immer wichtiger?

Antwort:

Semantic Web!

Semantic Web 1/2: Ontologien und Inferenz

Semantic Web

Vision eines zukünftigen Webs, mit Technologien für eine maschinelle Verarbeitung semantischer Inhalte

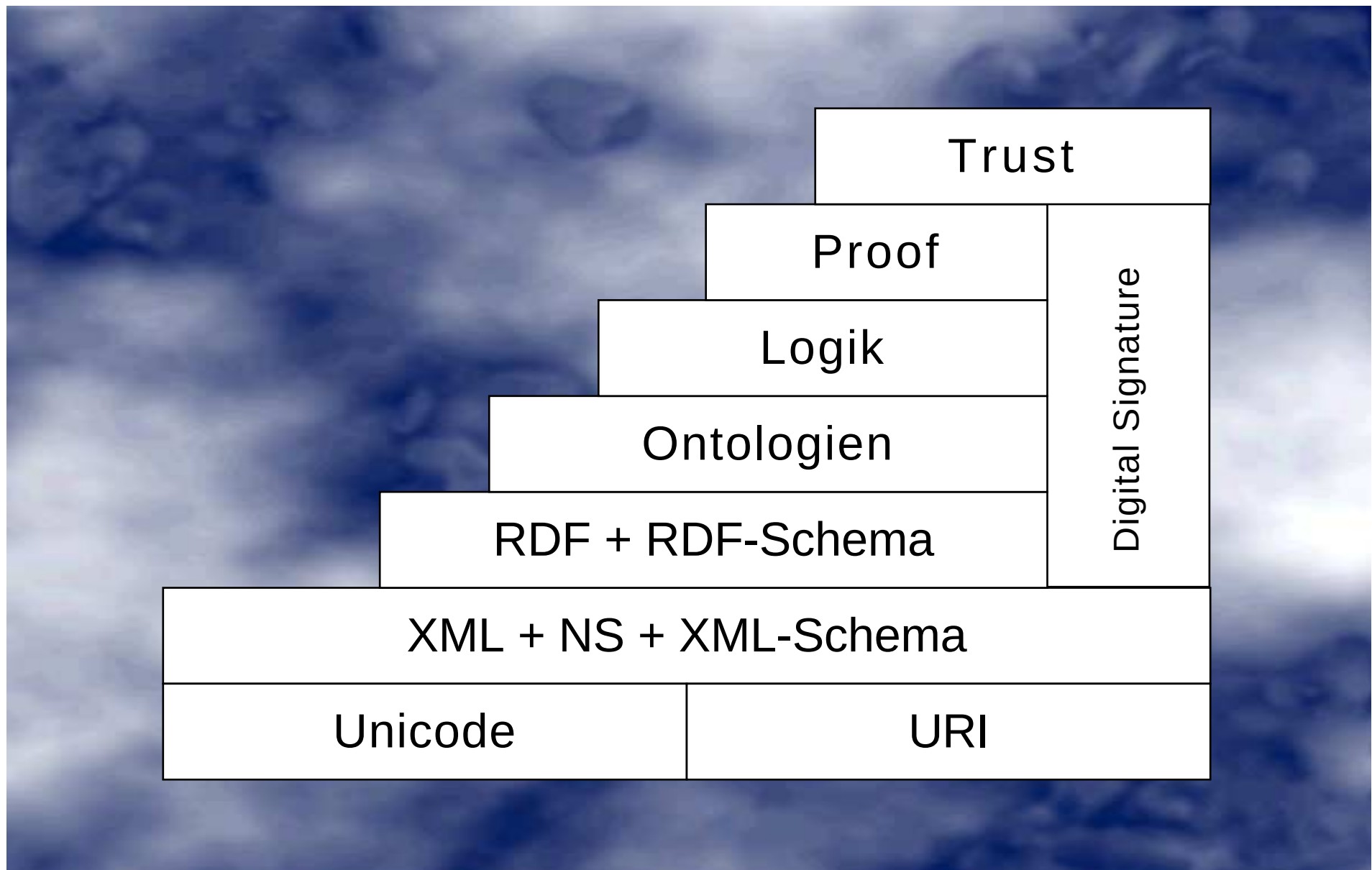
Ontologie

Formale, explizite Spezifikation einer gemeinsam verwendeten Konzeptualisierung eines Phänomenbereiches

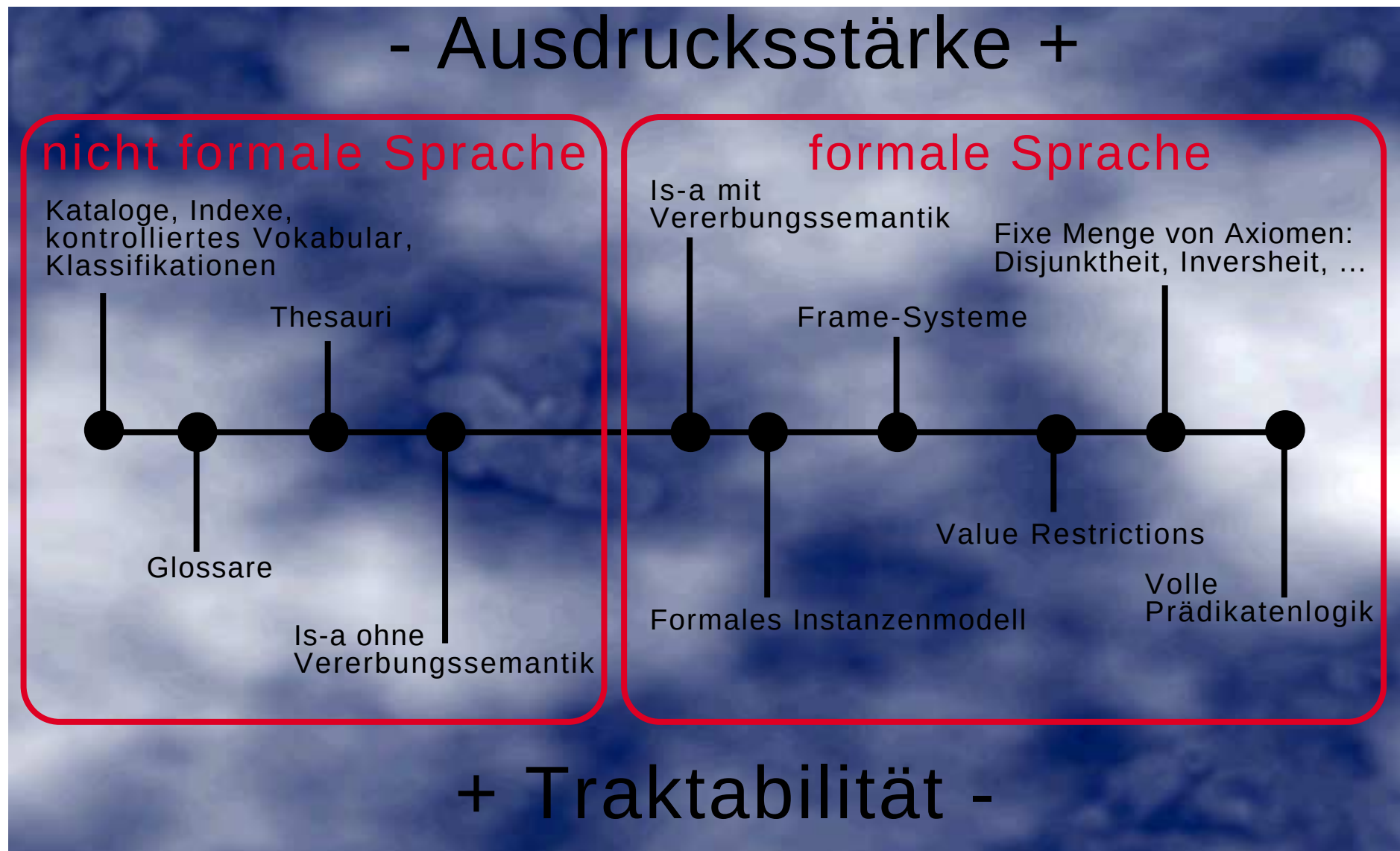
Inferenzstrategien

Ableitungsverfahren von neuem Wissen aus einer Wissensbasis oder Ontologie

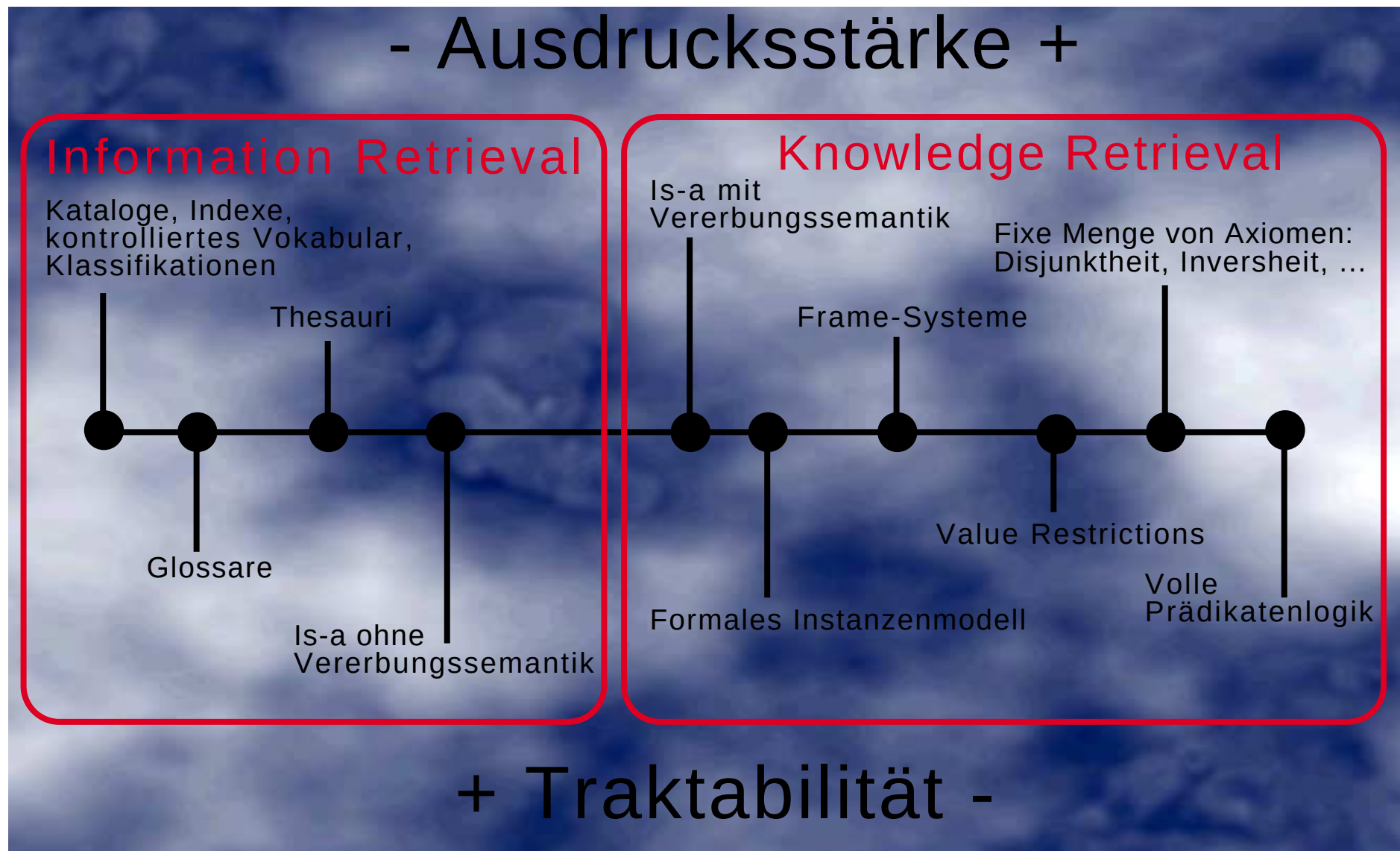
Semantic Web 2/2: Schichtmodell (s.a. Berners-Lee, Tim: [Semantic Web](#). 2000)



Ontologien als Spektrum der Formalisierbarkeit



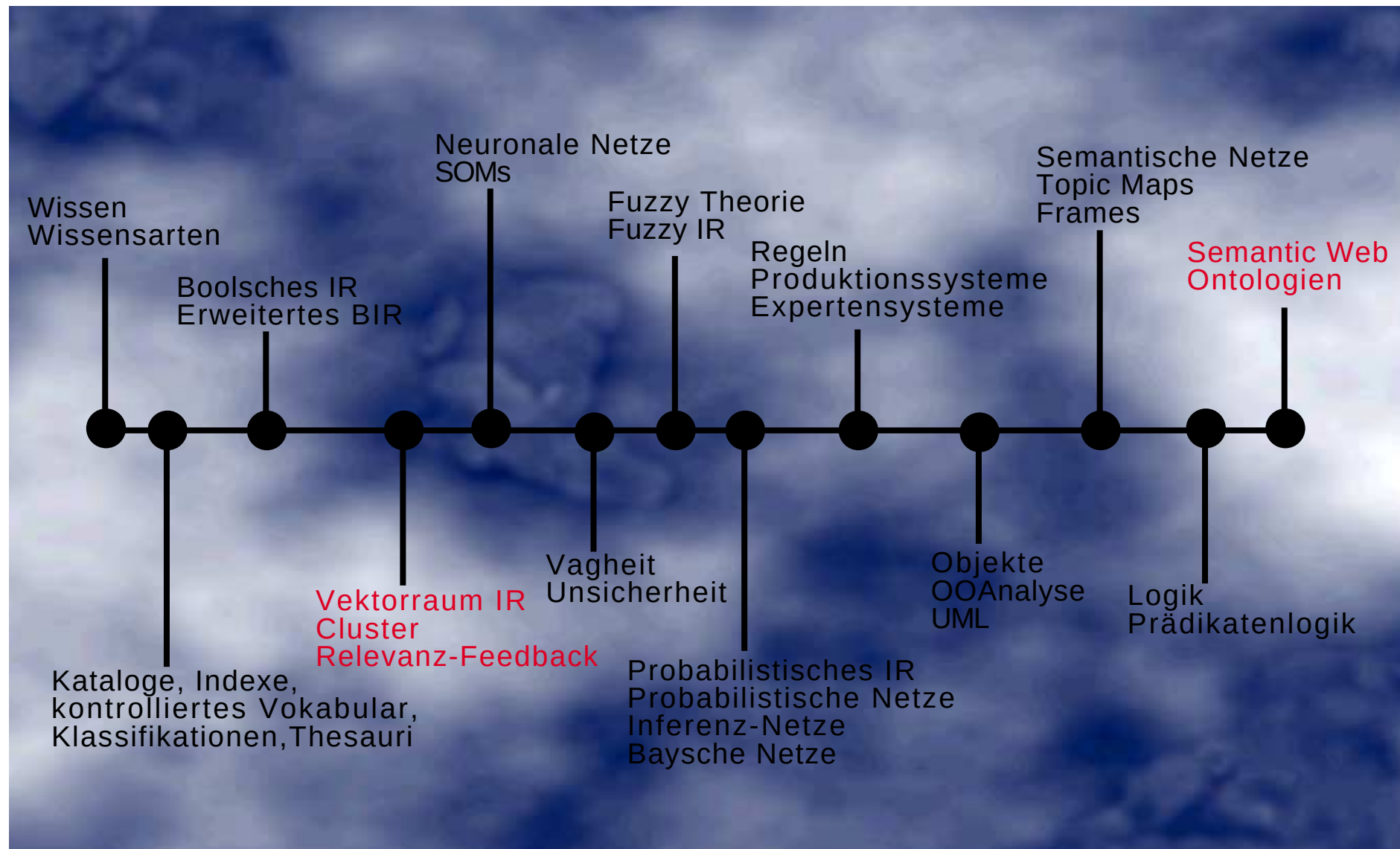
Ontologien als Basis für Information Retrieval und Knowledge Retrieval



Frage:

Welche Formen der
Wissensrepräsentation
und des Information Retrievals
werden behandelt?

Themenüberblick



Referenzierte Literatur

Bachelier, Günter: Polyrepräsentation, Relevanz-Approximation und aktives Lernen im Vektorraummodell des Information-Retrievals. 2002.

Ingwersen, Peter: Information Retrieval Interaction. London, 1992.

Ingwersen, Peter: Polyrepresentation of Information Needs and Semantic Entities. In: Croft & Rijsbergen (1994), SIGIR 1994, S. 101 - 110.

Panyr, J.: Automatische Klassifikation und Information Retrieval. Tübingen, 1986.

Wolpert, David H.; Macready, William G.: No Free Lunch Theorems for Search. The Santa Fe Institute Technical Report SFI-TR-95-02-010, 1996.

Zobel, Justin; Moffat, Alistair: Exploring the Similarity Space. In: IGIR Forum, Vol. 32, 1998, Nr. 1, S. 18 - 34.