

## Überlegungen zur Sprache als Basis der Klassifikation

Informationssuche ist in der Regel Suche nach/in Text/Sprache.  
(Zu berücksichtigen ist die Verschiedenheit von Dokument- und Suchsprache.)

Einige eher sprachliche Phänomene:

Morphologische Reduktion/Grundformerzeugung (Lemmatisierung)	Sprache	Sprachen
Dekomposition	Verbrechen	Verb/Rechen

Komposita: Goldfisch - Kalbsleberwurst - Transcontainerfließfertigung

Bedeutungsdifferenz / Trennung von sprachlicher Form und Bedeutung

Die Bank ist geschlossen

Die Bank ist schmutzig

Die Bank betreibt schmutzige Geschäfte

Die Homonymie erfordert somit Differenzierung durch Kontext.

Semantisches Ranking  
(und/oder/nicht)

„Sport im Alter“  
Senioren-sport

aber: Leistungssport im Entwicklungsalter

Hilfreich ist Einbindung in Semantisches Netz

Synonyme:

Wort für Wort:  
„intellektuell“  
Kontext:

Mundart - Dialekt  
Norbert Wiener - Kybernetik  
The Australian Stage (1984) .  
Von Grubenhunden und aufgebundenen Bären  
Im Blätterwald (1981)

Allgemein sind die Grenzen automatischer Sprachverarbeitung dort erreicht, wo die Intelligenz beginnt.

Konsequenzen u.a. für Computerlinguistik in der Informationserschließung:

Ausreichende Textbasis für Inhaltsanalyse nötig.  
Text muß ausreichende Aussagen über Dokumenteninhalt ermöglichen.  
Datenbasis muß hinreichend homogen sein.

Die Einbeziehung von Notationen ermöglicht sprachunabhängige Erschließung.  
(zumindest teilweise)