



Explainable Subsymbolic Semantics

Dr. Ronald Winnemöller - Norderstedt, 02.03.2021



Lufthansa
Industry Solutions

Einleitung

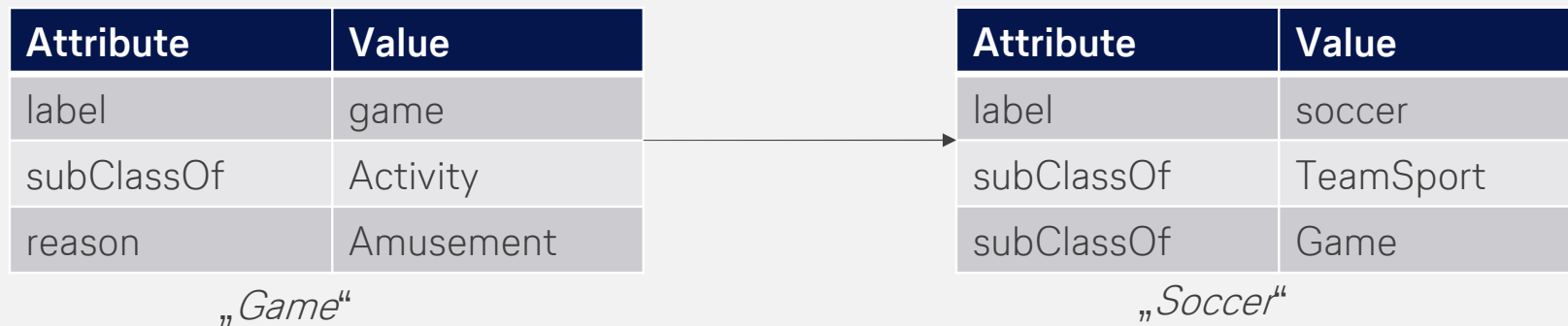
Semantik in der KI

(Oft vernachlässigt, aber
hochrelevant für KI Anwendungen)

- **Semantik:** Theorie(n) der Bedeutung der Zeichen
- **Zeichen:** Symbole, Kompositionen von Symbolen, z.B. Worte, Sätze, Diskurs, etc.
- **Meistens (implizit) eingesetzt:** Intensionale und extensionale Semantik, insbes. Merkmalssemantik
- **Abgrenzung:**
 - Semantik logischer Sprachen bzw. formaler Sprachen
 - Subsymbolische Semantik implizit durch quantitative Aspekte innerhalb von KNN und anderer ML Modelle

Symbolische (Merkmals-) Semantik

Basis aktueller Anwendungssysteme



- Bereiche: OO-Programmierung u. Modellierung, Relationale Datenbanken, etc.
- **Pro:** einfach zu verstehende einfache Modelle

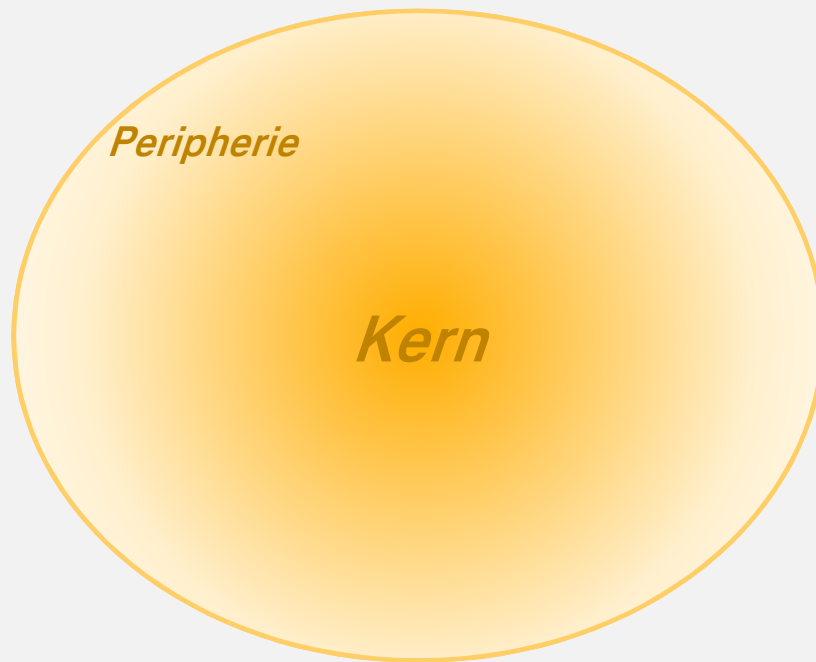
Symbolische (Merkmals-) Semantik

Schwierige Probleme

- Unvermeidbare Mehrdeutigkeit der Sprache
- Nicht-symbolische Aspekte und Kontexte gehen verloren, sind schwer abzubilden
- Unvollständigkeit ist bedeutendes Problem (z.B. keine „closed world“ Annahmen möglich)
- Keine „natürliche“ Verknüpfung zu subsymbolischen Methoden
- Eingeschränkte oder aufwändige Wartbarkeit von Implementierungen
- Nicht-trivialer Modellierungsaufwand

Prototypensemantik

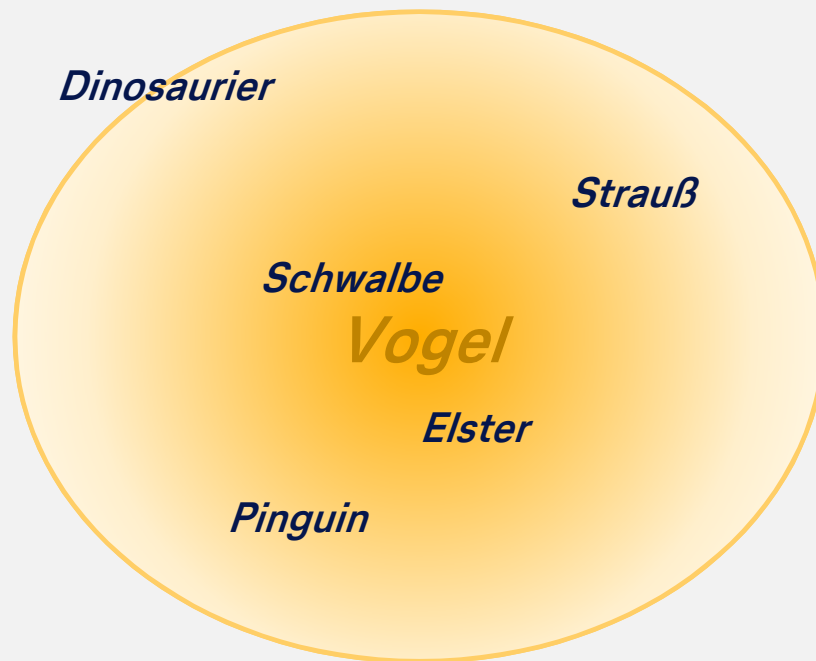
Ein alternativer Ansatz zur Merkmalssemantik



- Prototypen als zentrale normative Kategorie-Elemente
- Relation: „ist-ein“
- Quantitative Zuordnung statt qualitativer/intensionaler Zugehörigkeit
- „Typische Vertreter“ im Kernbereich, „untypische Vertreter“ in der Peripherie

Prototypensemantik

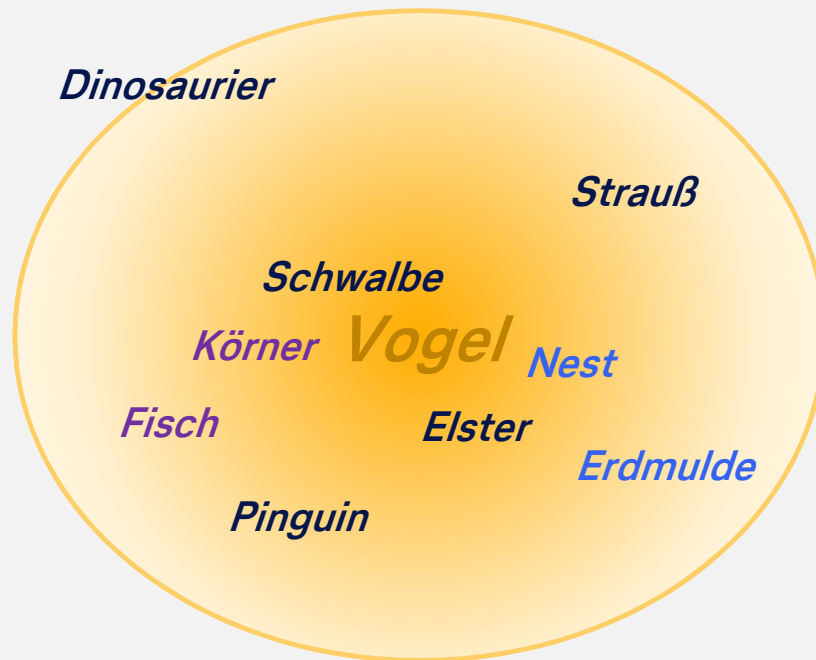
Beispiel Begriff „Vogel“



- Eine Schwalbe ist „typischer Vogel“ als ein Pinguin

Erweiterungen der Prototypensemantik I

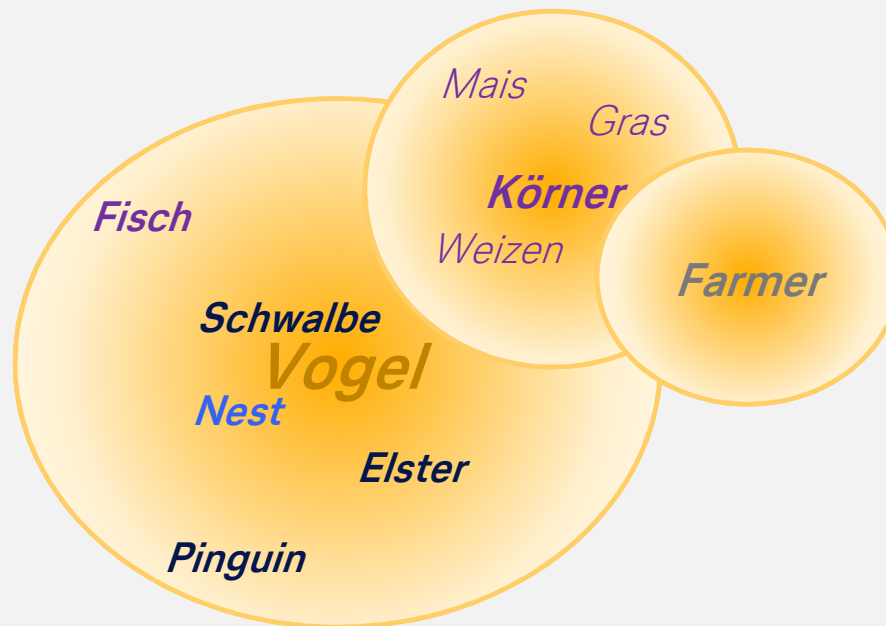
Typen von Relationen



- **Klassisch:** Relation „ist-ein“
- **Neu:** Relationen „lebt-in“, „frisst“
- Relation kann auch vage sein:
z.B. „steht-in-Beziehung-zu“
- Taxonomie der Relationen
- Merkmalssemantik wird Teilmenge der erweiterten Prototypensemantik

Erweiterungen der Prototypensemantik II

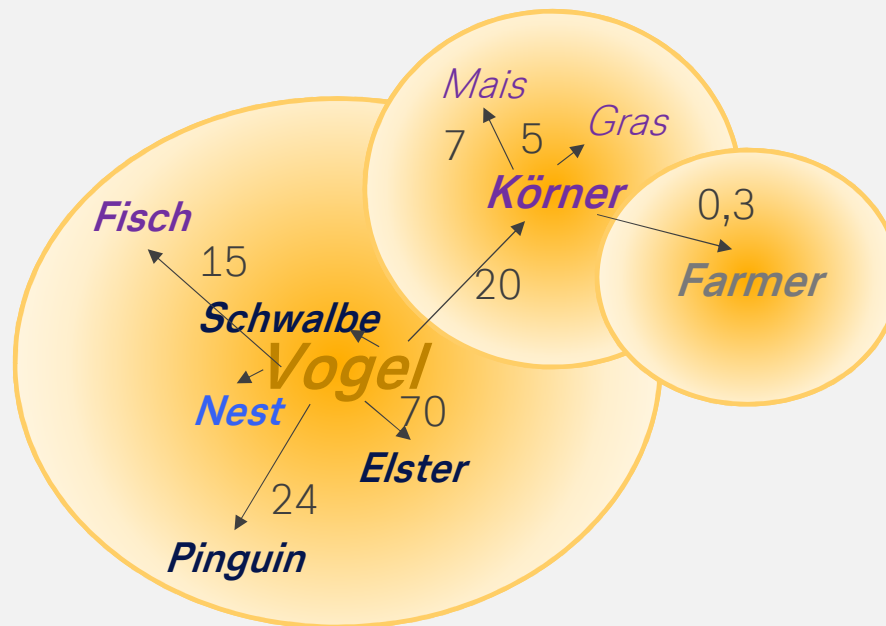
Semantischer Raum



- **Klassisch:** Betrachtung eines Begriffes
- **Neu:** Beziehung der Begriffe untereinander
- Ketten von Relationen
- Transitivität
- Gesamtheit der Begriffe und ihrer Beziehungen:
Der semantische Raum

Erweiterungen der Prototypensemantik III

Quantifikation des semantischen Raumes



- **Klassisch:** vage Begriffe von „mehr“ vs. „weniger“
- **Neu:** Zuordnung numerischer Werte
- Wirkung Transitivität:
Immer in Bezug auf den Zentralbegriff (hier „Vogel“)
- Gesamtheit der Begriffe und ihrer *quantifizierten* Beziehungen:
Der semantische Raum
- Semantik wird operationalisierbar!

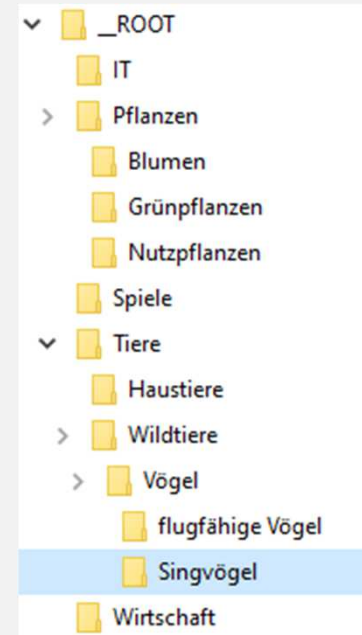
Operationalisierung Prototypensemantik

Grundlagen und Datenbasis

- Benötigt wird ein (sehr großes) „Dokumenten“-Verzeichnis
 - Verzeichnis-Strukturparadigma bestimmt Relationentypen
 - Begriffliche Zuordnung ergibt sich aus Inhalten der „Dokumente“

Weiterführend:

- Verknüpfte Verzeichnisse:
Unterschiedliche Relationentaxonomien, Strukturen oder Inhalte kein Problem
- Natürliche oder künstliche Verzeichnisse

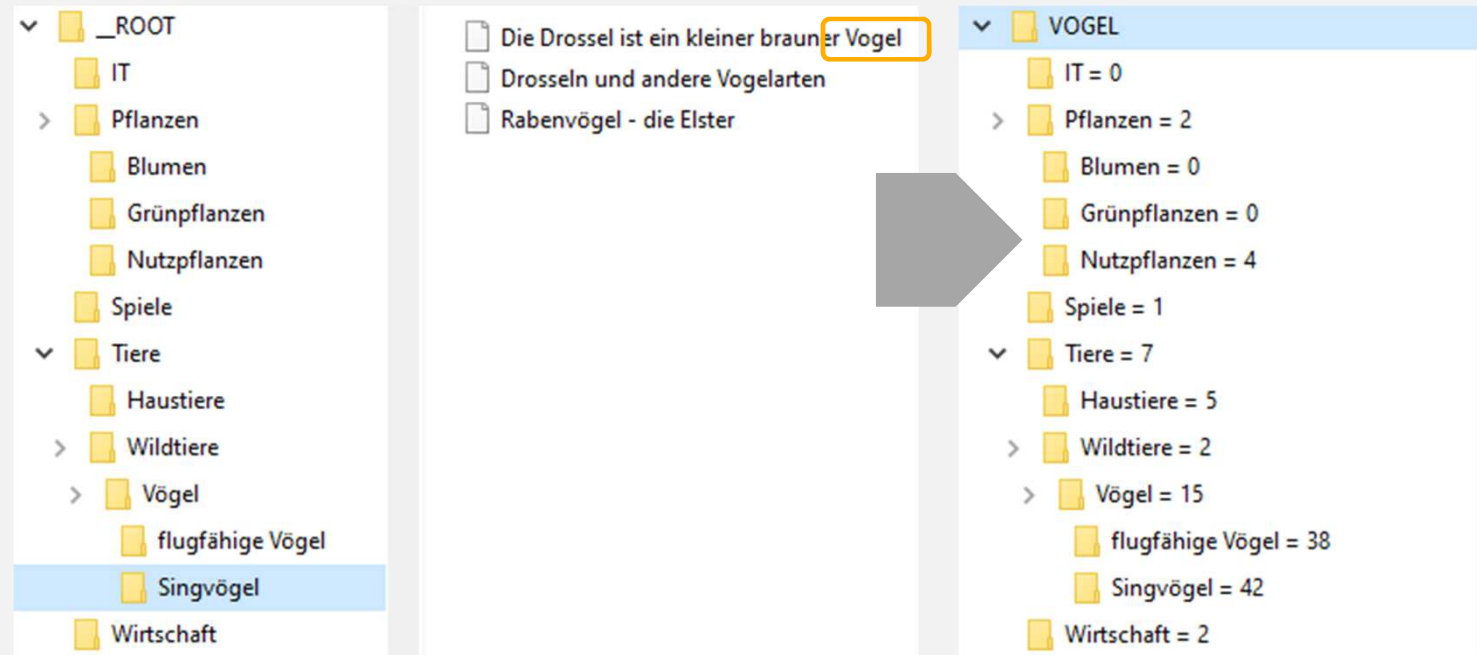


- Die Drossel ist ein kleiner brauner Vogel
- Drosseln und andere Vogelarten
- Rabenvögel - die Elster

Operationalisierung Prototypensemantik

Verfahren

- Einfach : Occurrence
- (Besser: TFXIDF)



Operationalisierung Prototypensemantik

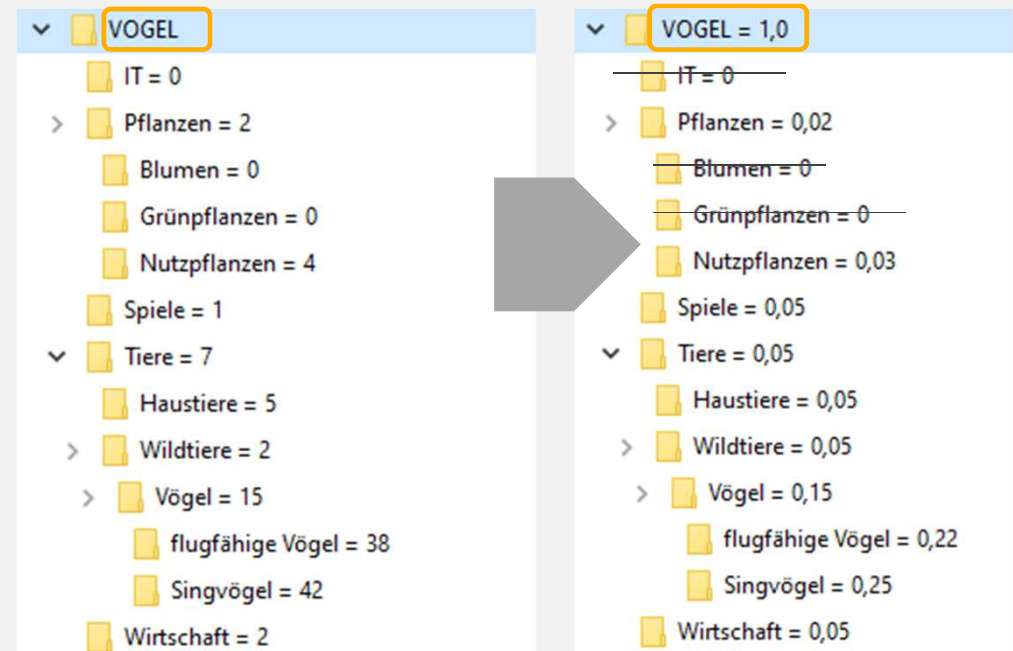
Verfahren

■ Normalisierung

- Summe aller Werte = 1,0
- „Hochreichen“ von Werten in der Struktur
- Eliminierung von Leer-Knoten
- Ggf. Schwellwert

- **Anwendung** auf *alle* Worte innerhalb des Verzeichnisses

- **Semantischer Raum:**
Verzeichnisstruktur + Wort-Bäume



Operationalisierung Prototypensemantik

Operationen

- **ADD** : knotenweise Addition und folgende Baum-Normalisierung
= Abbildung auf Textfragmente – Ausdrücke, Sätze, Abschnitte, Dokumente, Sammlungen, ...
- **FILTER** : knotenweise Filterung durch einen Referenzbaum
= Thematische Fokussierung
- **SIM** : Ermittlung des „Similarity-Wertes“ zweier Bäume
= Semantische Ähnlichkeit (in Bezug auf Baumstruktur)
- Vielfalt weiterer Operationen (teilw. umgesetzt)

Anwendung Prototypensemantik

Applikationen

Beispielanwendung Auflösung von Wortmehrdeutigkeiten

- **Szenario:** Ein Wort vs. Mehrere Bedeutungen (SENSEVAL Competition/Datasets)
- **Aufgabe:** Feststellung der Bedeutungsvariante in Beispieltexten
- **Hilfestellung:** Definition der einzelnen Bedeutungen (als WORDNET Sense Def.)
- **Anwendung:**
 1. Wort-Bäume gefiltert durch jeweilige „Sense“-Bäume
 2. Similarity Beispieltext-Baum vs. gefilterter Wort-Baum
 3. Höchster Wert gewinnt
- **Weitere Anwendungen:** Language Identification, Discourse Analytics, Semantic Search, Recommender Systems, etc.

Fazit

Erkenntnisse zusammengefasst

- Erweiterte Prototypensemantik als mächtiges theoretisches Fundament
- Verknüpfung von symbolischer und subsymbolischer Semantik
- Direkte Implementierung durch verfügbare Datenmengen
- Nachvollziehbare („explainable“) Darstellung und Operatoranwendung
- Kein aufwändiges Training notwendig

Stattdessen: einfache Mechanik der Selbstoptimierung

- Vielseitige Einsetzbarkeit



Vielen Dank für die Aufmerksamkeit

Gerne Fragen, Diskussionen, Anregungen



Lufthansa
Industry Solutions