

Kathleen Dahlgren: A linguistic ontology

Inhalt

1. Thema
2. Einleitung
3. Ontologie für Intelligente Textverarbeitung ITP
4. Beispiele
5. Nachteile einer einzigen Inhaltsontologie
6. Zusammenfassung

1. Thema

- Einsatz einer linguistikbasierten Ontologie für Verarbeitung von natürlichen Sprachen
- liefert auf allen Ebenen vom Parsing bis zum logischen Denken plausible Interpretationen; Klärung nicht eindeutiger Strukturen
- Probleme und Einbußen einer Methode mit nur einer Inhaltstontologie
- linguistikbasierte Ontologie repräsentiert Weltansicht, die in natürlicher Sprache enkodiert
- Weltansicht bezeichnet einfache Annahmen über das, was in der Welt existiert und wie es klassifiziert werden sollte
- Inhaltstontologie ist Teil eines naiven semantischen Lexikons

2. Einleitung

- eine linguistikbasierte Ontologie für Verarbeitung von natürlichen Sprachen genügt für plausible Interpretationen
- KI-Anwendung soll uneingeschränkten Text in uneingeschränkten Domänen parsen und interpretieren; soll plausibelste Interpretation wählen und in Kontext einbetten
- Ergebnis: kognitives Modell, welches vollständig und eindeutig ist

Definition von Ontologie

- bezeichnet das, was da ist
- Welt besteht aus unendlich vielen Objekten und ihren Beziehungen untereinander

Fortsetzung: 2. Einleitung

Definition von linguistischer Ontologie

- Ontologie, die auf bestimmter Sprache, seiner Syntax und seiner Wortbedeutung basiert
- reflektiert Kultur, wie Menschen innerhalb einer Kultur mit Objekten, Ereignissen und Relationen interagieren
- Gewichtung der Objekte in Kultur wiedergegeben in Typ und Ausmaß der Klassifizierung

Vorteile einer linguistikbasierten Ontologie:

1. verschiedene Ebenen nötig, um syntaktische und bedeutungsbasierte Unklarheiten zu beseitigen und für formale semantische Interpretationen, für zeitliches Denken und für Verfolgen von Diskursen
2. genauere Interpretation, da direkt mit Wortbedeutung verbunden
3. Transportfähigkeit
4. ein Mittel, um menschliches Denken über Wissen vorherzusagen

3. Ontologie für Intelligente Textverarbeitung ITP

Es sind zwei verschiedene Ontologien für ITP notwendig:

- Ontologie mit Typen für Wahrheitswerte (DRT-framework)
- Ontologie für linguistische Konzeptualisierungen/lexikalischen Inhalt

Aufbau der Ontologie für linguistische Konzeptualisierungen

- Dahlgren/McDowell haben für Inhaltstontologie eigenen Ansatz entwickelt (naive Semantik)
 - Klassifizierungen/Beschreibungen in Lexika spiegeln naive Theorien über Welt wieder
 - jede Wortbedeutung verbunden mit Annahmen über Objekte/ Ereignisse; in naiver Semantik nur Information, die nötig um semantische Interpretation zu ermöglichen, nicht gesamtes Wissen über Objekt

Fortsetzung: 3. Ontologie für Intelligente Textverarbeitung ITP

- semantischer Typ und syntaktische Kategorie sind orthogonal, deshalb wird folgende naive semantische Ontologie angelegt:
 - Nomen (Objekte, Ereignisse, Zustände)
 - Verben (Beziehungen, Mittel/Methoden)
 - Adjektive (Eigenschaften von Dingen)
 - Adverbien (Eigenschaften von Eigenschaften, Ereignissen oder Aussagen)

Wie soll darüber hinaus das Grundgerüst der Ontologie aussehen?

- sollte wichtige Eckpfeiler der Welt reflektieren
- zwei Hauptunterscheidungen :
 - empfindungsfähig/nicht empfindungsfähig und
 - gesellschaftlich/natürlich (bezogen auf Ursprung)

Fortsetzung: 3. Ontologie für Intelligente Textverarbeitung ITP

- Ontologieknoten empfindungsfähig am häufigsten genutzt
- in ITP Ontologie ist Unterscheidung empfindungsfähig/nicht empfindungsfähig weit oben und ein Schwesterknoten der Unterscheidung gesellschaftlich/natürlich
- alle Empfindungsfähigen („John“, „Sekretärin“, „die UN“) fallen unter einen einzigen Knoten fallen und erfüllen Einschränkungen eines Verbs, wie z.B. „sagen“
- generisches Wissen von Menschen über Menschen („hat ein Herz“, „atmet“) wird vererbt unter dem Knoten Tier, während Wissen über Rollen (beispielsweise „arbeitet für jemanden“) unter den Knoten für empfindungsfähig bzw. gesellschaftlich zu finden ist.

Entität → (abstrakt vs konkret) & (Individuum vs Kollektiv)

Konkret → (physisch vs temporal vs empfindungsfähig) & (natürlich vs gesellschaftlich)

Fortsetzung: 3. Ontologie für Intelligente Textverarbeitung ITP

	natürlich	gesellschaftlich
physisch	Stein	Messer
empfindungsfähig	Mann	Programmierer
temporal	Erdbeben	Party

- nicht einzig mögliche Ontologie für Verarbeitung von natürlicher Sprache
- von Vorteil mehrere unterschiedliche Ontologien parallel laufen lassen

4. Beispiele

- ITP-Ontologie um Tausende Knoten erweitert; Anwendung auf verschiedenen Ebenen der Verarbeitung getestet

4.1. Disambiguierung von Wortbedeutung

Jede Wortbedeutung wird aufgelistet und mit ontologischem Anhang versehen.
Beispiel Bedeutungen für Symphonie:

- (1) soziales Ereignis – das Spielen der Symphonie
Kopie – die Papierversion
Verlauf – die Struktur der musikalischen Noten

ITP-Disambiguierungsmodul wählt korrekte Bedeutung in folgenden Sätzen:

- (2) Fred sah Hanna bei der Symphonie. (soziales Ereignis)
Fred hob die Symphonie auf. (Kopie)
Die Symphonie dauerte eine Stunde. (soziales Ereignis)
Fred hat die Symphonie im letzten Jahr geschrieben. (Verlauf)

Die Worte disambiguieren einander häufig, da sie einander ausschließen.

Fortsetzung: 4. Beispiele

4.2. Anaphora

- (1) Die Sekretärin kaufte einen brandneuen Maserati.
Die Frau mochte schnelle Autos.
- (2) Die Sekretärin kaufte einen brandneuen Maserati.
Das Fahrzeug war rot.
- (3) Die Sekretärin kaufte einen brandneuen Maserati.
Die Entscheidung fiel ihr nicht schwer.

5. Nachteile einer einzigen Inhaltstontologie

- liegt in Vielfalt von Weltansichten der Menschen und in Komplexität sprachlicher/konzeptueller Interaktion
- besser mehrere unterschiedliche Ontologien parallel laufen lassen, da Objekte kognitiv so gespeichert werden
 - z.B. Pflanzen anhand Form (Bäume, Gebüsch, Grass, etc)
oder biologischen Gesichtspunkten (laubabwerfend/immergrün)
- wenn beide Gliederungsarten wichtig genug sind, ist Kombination sinnvoll

6. Zusammenfassung

- für Interpretation von natürlicher Sprache zwei Arten von Ontologien erforderlich:
 - formale semantische Ontologie
 - konzeptionelle Ontologie
- für konzeptionelle Ontologie genügt eine linguistischbasierte Ontologie für eindeutige Interpretation
- linguistischbasierte Ontologie konstruiert zur Darstellung psychologischer Kategorisierungen und linguistischer Einschränkungen
- in Praxis: englisches Lexikon liefert korrekte Interpretation eines Satzes

James Pustejovsky: Lexical Semantics and Formal Ontologies

Inhalt

1. Thema
2. Linguistikbasierte Ontologien
 - 2.1. *Principle of Orthogonal Typing*
(Generatives Lexikon, *qualia structures*,
qualia roles, komplexe Typen)
3. Zusammenfassung

1. Thema

- Methode für Aufbau von ontologischen Typen; sie berücksichtigt linguistische Generalisierungen und ontologische Gesichtspunkte
- Bedarf für multiple Vererbungsstrategie
- Technik namens *Principle of Orthogonal Typing*; bei Erstellung von Konzepten wird einfache Semantik berücksichtigt, die rekursiv und ausdrucksstark genug ist, um komplexe Konstruktionen auszudrücken

2. Linguistikbasierte Ontologien

- wissensbasierte Systeme folgen Standardstrategie; Modell geschaffen mit Unterkategorien; Unterscheidung zwischen Dingen, Handlungen und Eigenschaften
- in Domäne von Entitäten/Dingen wäre multiple Vererbung ein Mittel, um Beziehung z.B. zwischen den Konzepten *Nahrungsmittel* und *Apfel* oder zwischen *Tier* und *Haustier* zu formen
- Ziel des Autors ist ein Konzept, in dem von multiplen Basistypen vererbt wird
- Unterscheidung zwischen natürlichen/künstlichen Objekten; Funktion spielt wichtige Rolle

Fortsetzung: 2. Linguistikbasierte Ontologien

Wie wird in einer Ontologie kategorisiert?

1. rock (individual, with no declared purpose)
2. sandwich (individual, with purpose, artifactual)
3. George (agentive individual)
4. my father (relational noun of individual)

- Mechanismen für Konstruktion von Konzepten
- Annahme: Mechanismus hat generative Basis
- von Menschen erstellte Konzepte besitzen drei grundlegende Eigenschaften:
 - sie sind begründet
 - sie sind rekursiv, d.h. Konzepte können wiederverwendet werden und miteinander kombiniert werden
 - sie sind kompositional, d.h. die Bedeutung der Konzepte erschließt sich aus ihren Einzelteilen

Fortsetzung: 2. Linguistikbasierte Ontologien

2.1. *Principle of Orthogonal Typing*

- Theorie des Autors: *Principle of Orthogonal Typing* ist allgemeine Strategie, um spezifische Konzepte (Typen) mit Eigenschaften zu erstellen; Grundlage ist ein generatives Lexikon

Generatives Lexikon

- generatives Lexikon: ein System mit vier *basic levels* linguistischer Repräsentation:
 - *Argument Structure*: Zahl u. Typ der logischen Argumente
 - *Event Structure*: Definition des Ereignistyps (Prozess, Zustand, Verwandlung/Übergang); Zeitfaktor
 - *Qualia Structure*: strukturelle Unterscheidung der Worte; Wortbedeutung
 - *Lexical Inheritance Structure*: stellt fest, in welcher Beziehung eine lexikalische Struktur zu einer anderen steht (Generalisierung)

Fortsetzung: 2.1. *Principle of Orthogonal Typing*

Qualia Structure

- Wortbedeutung der *qualia structure* strukturiert in vier generative Faktoren (*qualia roles*); reflektieren, wie Menschen Objekte und Beziehungen verstehen; liefern Erklärung für linguistisches Verhalten v. lexikalischen Items
- teilt Worten funktionale Tags zu; diese verbinden Worte mit Netzwerk aus Konzepten
 - ***Formal***: die Basiskategorie (z.B. konkret/abstrakt)
 - ***Constitutive***: die Beziehung zwischen einem Objekt und den Teilen aus denen es besteht (z.B. unzählbare Entitäten)
 - ***Telic***: die Funktion/der Zweck eines Objekts
 - ***Agentive***: Faktoren, die sich auf den Ursprung des Objekts beziehen bzw. auf ihre Entstehung (z.B. natürliche/künstliche Objekte)

Fortsetzung: 2.1. *Principle of Orthogonal Typing*

- *qualia structure* ist Kern des Lexikons; liefert allgemeine Strategie für Erstellung von spezifischen Konzepten
- Beispiel: Haustier Katze namens Eno
 - *Formal*: Katze
 - *Telic*: Haustier
- Kombination der *qualia roles* ergibt orthogonales Set mit Eigenschaften, welches Typen (*unified types*) definiert
- daher: *Principle of Orthogonal Typing*
- Typen (*unified types*) sind wie orthogonale Dimensionen strukturiert und weniger wie multipel vererbte Konzepte
- Konzepte, die Eigenschaften erben, sind entweder
 - typisierte orthogonale Vererbungsstrukturen
 - oder komplexe Typen

- typisierte orthogonale Vererbungsstrukturen:
qualia structures liefern typisierte orthogonale Informationen für orthogonale Vererbungsstrukturen

Komplexe Typen

- komplexe Typen: Konzept nur wohlgeformt, wenn es von einem einzigen Elternknoten erbt; Darstellung komplexerer Worte oder Ausdrücke durch Unifizierung der *qualia structures*
 - Buch: physisches Objekt / Information, die es enthält
 - Mittag: Ereignis / Essen
 - Verabredung: Ereignis / Person

3. Zusammenfassung

- *Principle of Orthogonal Typing* ist allgemeine Strategie, um spezifische Konzepte (Typen) mit Eigenschaften zu erstellen; Grundlage ist ein generatives Lexikon