

DVMD-Tagung 2011

Ontologieunterstützte Informationsaufbereitung

M. Engelhorn, Stephan Profitlich

DVMD-Tagung 2011

Man erkennt nur was man weiß

in Anlehnung an ein Zitat von J. W. von Goethe

DVMD-Tagung 2011

Fahrplan

- **Motivation**
- **Daten - Information - Wissen**
- **Informationsaufbereitung und Wissen**
- **Unterstützung durch Ontologien**
- **Auflösen von Mehrdeutigkeiten**
- **Umsetzung (Projekt) und Ausblick**

Motivation

Informationszuwachs

- **Neue Verfahren erzeugen meist automatisiert mehr Information**
- **Neue Verordnungen erzwingen mehr Information**
- **Der einfach Zugang zu Information fördert den Zuwachs an Information**
- **Die Selektion und die Bewertung der Information wird immer schwieriger**
- **Zusammenhänge können immer weniger erkannt werden**
- **Der Überblick geht verloren.**

Motivation

Informationszuwachs

- **Große Mengen an Information ist unstrukturiert**
- **Es gibt fachspezifische Ausdrücke, Abkürzungen und Synonyme**
- **Es gibt unterschiedliche Schreibweisen für Sachverhalte**
- **Die Mehrsprachigkeit ist eine weitere Hürde**
- **Eine Volltextrecherche bringt häufig nicht das erwartete Ergebnis**
- **Sehr große und sehr kleine Treffermengen sind wenig nützlich**

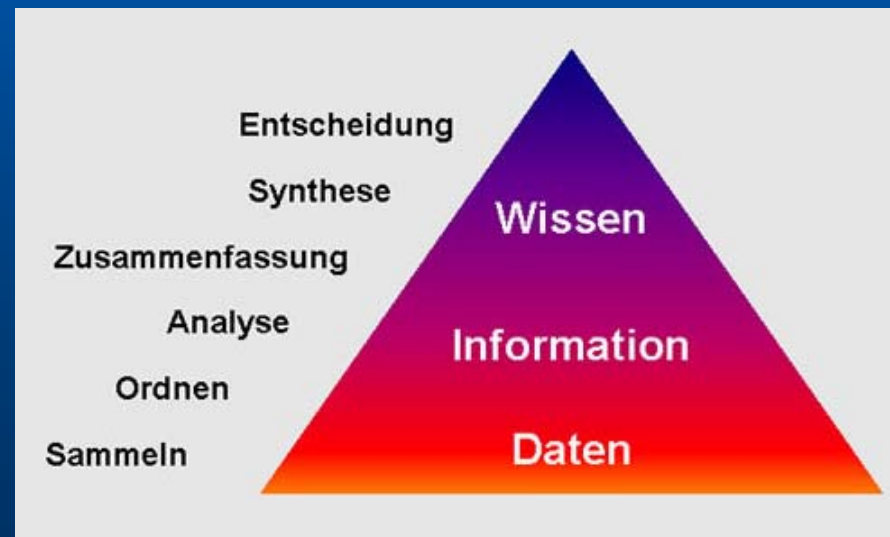
Motivation

Fazit

**Der ständige und explosionsartiger Zuwachs an Information
ist ohne maschinelle Unterstützung nicht mehr handhabbar!**

Informationsaufbereitung

Daten – Information - Wissen



Informationsaufbereitung

Informationsaufbereitung und Generierung von Wissen

- **Extraktion der Daten (63, 3.14, kg, Müller)**
- **Informationsgewinnung durch Erkennen von Zusammenhängen und generieren von „Metadaten“ (63, kg => Gewicht)**
- **Wissensgenerierung durch Anreichern semantischer Zusammenhänge und Vergleich mit vorhandenem und bekanntem Wissen (Karl Müller, aktuelles Gewicht 63kg, Gewichtsverlust in den letzten 2 Wochen)**

Informationsaufbereitung

Prozessunterstützung

- **Statistische Methoden (bei der Extraktion der Daten)**
- **Thesauri (Vokabulare)**
- **Taxonomien (Klassifikationen)**
- **Terminologien (Fachspezifische Zuordnungen)**
- **Nomenklaturen (Regelwerke)**
- **Ontologien (Wissensräume)**

Informationsaufbereitung

Automatisierung

- **Zur weitgehend automatischen Aufbereitung von unstrukturierten Informationen ist der Zugriff auf vorhandenes Wissen unerlässlich.**
- **Auch die Erweiterungen und die Pflege des Wissens muss in diesen Prozess einbezogen werden.**

Informationsaufbereitung

Nutzung von vorhandenem Wissen

- **Vorhandenes Wissen muss modelliert werden (Ontologien)**
- **Ontologien decken immer bestimmte Wissensräume ab (Geografie, Personen, Institutionen, etc.)**
- **Begrifflichkeiten sind oft mehrdeutig**

Informationsaufbereitung

Formen von Wissen

- **Explizites und implizites (verborgenes) Wissen**
- **Unvollständiges und unscharfes Wissen**
- **Strukturiertes und unstrukturiertes Wissen**
- **Falsches, widersprüchliches und veraltetes Wissen**

Informationsaufbereitung

Beispiel „Rosenthal“

- **Suche bei Google ergibt 9,31 Mio Treffer :**
 - **Bereiche: Porzellan (Rang 1),**
 - **Unterhaltung (Rang 5),**
 - **Geografie (Rang 12),**
 - **MDC (27), ...**

Informationsgewinnung

Mehrdeutigkeiten

- **Der Begriff „Rosenthal“ kann**
 - eine Person sein (Wissenschaft, Unterhaltung, ...),
 - ein geografischer Ort sein (Stadtteil von Berlin),
 - eine Institution sein (Porzellan).

Informationsaufbereitung

Nutzung von Wissen im „Expertennetz“

- Nutzung des vorhandenen Wissens in den Ontologien
- Berechnen einer wahrscheinlichen Zuordnung zu einer Ontologie
- Erweitern der Ontologie um neue Einträge

Informationsaufbereitung

Beispiel „Rosenthal“

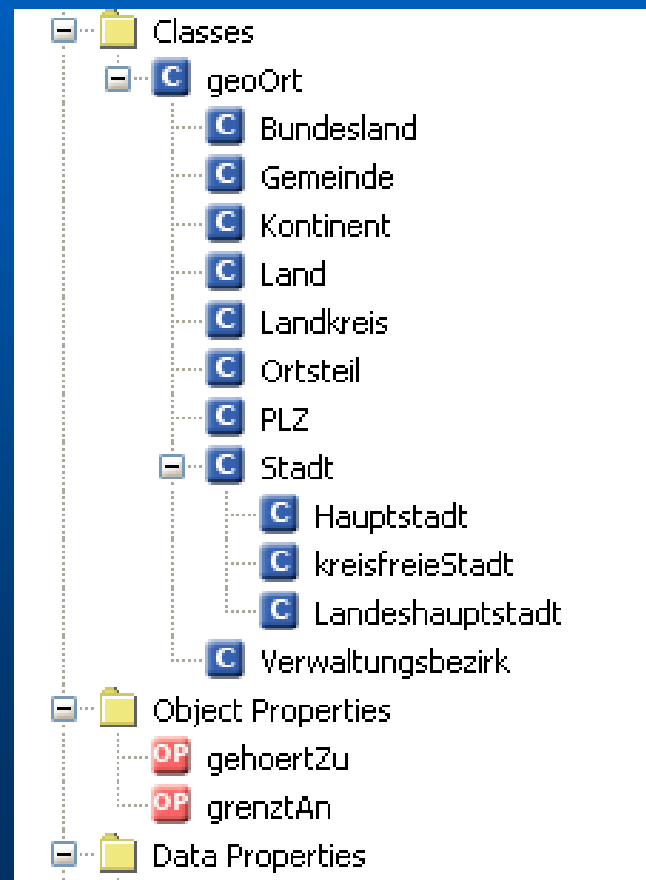
- **Durch Betrachtung des Kontexts:**
 - „Prof.“ deutet auf eine Person hin
 - „Wiss. Vorstand des MDC“ deutet ebenfalls auf eine Person hin
 - „MDC“ als Institution ist in Berlin-Buch angesiedelt
- **Zuordnung:**
 - Personenontologie
 - **Nicht** Geografieontologie
- **Ggf. Ontologien (Personen und Institutionen) erweitern**

Informationsaufbereitung

Nutzung von implizitem Wissen

- Nutzung des impliziten Wissen in den Ontologien durch „reasoning“
- Berechnen einer „Distanzfunktion“
- Nutzung der Distanzfunktion im Expertennetz

Geografieontologie



Informationsaufbereitung

Beispiel „Rosenthal“

- Implizites Wissen über die „gehörtZu-Relation“ explizit machen
- **Durch Betrachtung des Kontexts:**
 - (1) Rosenthal gehört im geografischen Kontext „näher“ zu „Berlin“, falls Berlin in Deutschland liegt
 - (2) Rosenthal gehört im geografischen Kontext „**nicht** näher“ zu „Berlin“, falls Berlin in den USA liegt
- **Zuordnung:**
 - 1: Geografieontologie
 - 2: **Nicht** Geografieontologie

Informationsaufbereitung

Ablauf

- **Aufbau von Ontologien**
- **Erschließen des impliziten Wissens durch Inferenzmaschinen (Reasoner)**
- **Erkennen und markieren von widersprüchlichem Wissen**
- **Entfernen von falschem und überflüssigem Wissen**

Informationsaufbereitung

Erweiterte Prozesskette

- **Extraktion der Daten**
- **Anreicherung mit Metainformation**
- **Auflösen von Mehrdeutigkeiten**
- **Generierung von Wissen**
- **Einbringen von neuem Wissen in das System**

Projekt/Ausblick

Pilotprojekt „Wissensportal“

- **Implementiert**
 - **Statistische Extraktion der Daten**
 - **Aufbau von mehreren Ontologien (Personen, Institutionen, Geografie, Chemie, Biologie, Medizin)**
 - **Expertennetz**
- **Work in Progress**
 - **Integration der verschiedenen Pakete**

Projekt/Ausblick

Offene Fragen

- **Umgang mit unvollständigem und unscharfem Wissen**
- **Umgang mit Zeitabhängigkeiten (Zeiträume)**
- **Datenqualität und Glaubwürdigkeit**
- **Performance bei großen Datenbeständen und Ontologien**

Danke

**Ich bedanke mich für Ihre
Aufmerksamkeit**