
Web-Engineering

Semantic Web

World Wide Web

- Netzwerk
 - Ressourcen unterschiedlicher Art
 - Hypertext: Verweise auf Ressourcen, insbesondere Textdokumente
- Implizite Bedeutung der Ressourcen und ihrer Beziehungen
 - Unterschiedliche Interpretationen möglich
 - Nur durch den Benutzer interpretierbar

Wie kommt man an Informationen ?

- Suchmaschinen:
 - Information-Retrieval: Merkmale extrahieren aus unstrukturierten Daten
 - Kontext der durchsuchten Daten kann nur schwer einbezogen werden
 - Interpretation des Suchergebnisses dadurch eingeschränkt
 - Vereinfacht gesehen: Übertragung des Prinzip der Textmustersuche in Dateisystemen auf das WWW

Metadaten: Anreicherung mit Informationen

- Verbesserung der Suche:
 - Unstrukturierte Daten mit beschreibenden Daten (= Metadaten) anreichern
 - Beispiel Datenbanken:
 - `SELECT * FROM personen WHERE name LIKE 'B%'`
 - Auswertung des DB-Schema = Metadaten möglich
- Übertragung auf das WWW:
 - WWW erweitern zu einem *semantischen* Netz

Semantisches Netz (1)

- Formales Modell für Begriffe und deren Beziehungen (Relationen)
- i.d.R. Graphen:
 - Knoten als Begriffe
 - Kanten als Beziehungen
- Einfaches Beispiel:
 - Mindmaps: zentrale Begriffe, hierarchische Beziehungen

Semantisches Netz (2)

- Regeln
 - Zur Beschreibung
 - Der Bildung von Begriffen und Relationen
 - Der Gültigkeit von Begriffen und Relationen
 - Von Einschränkungen
- Mit Begriffen, Relationen und Regeln als Metadaten kann die Bedeutung des semantischen Netzes "WWW" angegeben werden

Semantic Web (1)

- Initiative von Tim Berners-Lee
- Ziele:
 - Verbesserung der Recherche durch Einbeziehung inhaltlicher Aspekte = Bedeutung der Inhalte
 - auch für den Menschen
 - Vor allem für die automatisierte Verarbeitung durch DV-Systeme
 - Automatische Ableitung neuen Wissens
- WWW mit Metadaten anreichern in Form von *Ontologien*

Semantic Web (2)

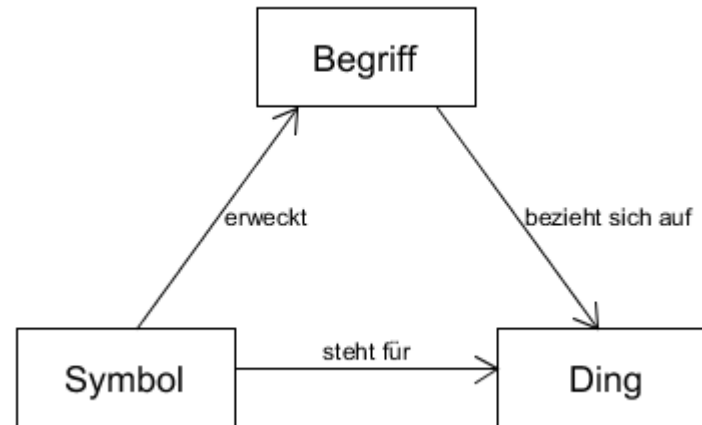
Zitat Tim Berners-Lee (2002):

"The WWW contains documents intended **for human consumption**, and those intended **for machine processing**. The Semantic Web will enhance the latter. The Semantic Web will not understand human language ... The Semantic Web is about machine languages: well-defined, mathematical, boring, but processible. **Data, not poetry.**"

Exkurs 1: Semiotik (1)

- Allgemeine Lehre von Zeichen, Zeichensystemen und Zeichenprozessen
- Bereiche (Interpretationen/Abbildungen):
 - Syntax: Zeichen \leftrightarrow Zeichen
 - Semantik: Zeichen \leftrightarrow Bedeutung
 - Pragmatik: Zeichen \leftrightarrow Benutzer und Situation

Exkurs 1: Semiotik (2)



- Semiotisches Dreieck:
 - Gegenstand (Ding) (aber auch: Sachverhalte, Ereignisse...)
 - Begriff
 - Benennung (Symbol)

Exkurs 2: Ontologie (1)

- Beschreibung von
 - Begriffen
 - Relationen
 - Regelnfür einen bestimmten Gegenstandsbereich
- Begriffe treten auf als
 - Typen
 - Instanzen
- Regeln beziehen sich auf
 - Bildung
 - Gültigkeit
 - Einschränkungen

Exkurs 2: Ontologie (2)

Abgrenzung zu anderen Konzepten:

- Thesaurus:
 - Systematisch geordnete Sammlung von Begriffen, die in thematischer Ordnung zueinander stehen
 - Deskriptoren: beschreibende Attribute
 - Synonyme: gleichbedeutende Worte
 - Homonyme: Worte, die verschiedene Begriffe bezeichnen
 - Ober-/Unterbegriffe

Exkurs 2: Ontologie (3)

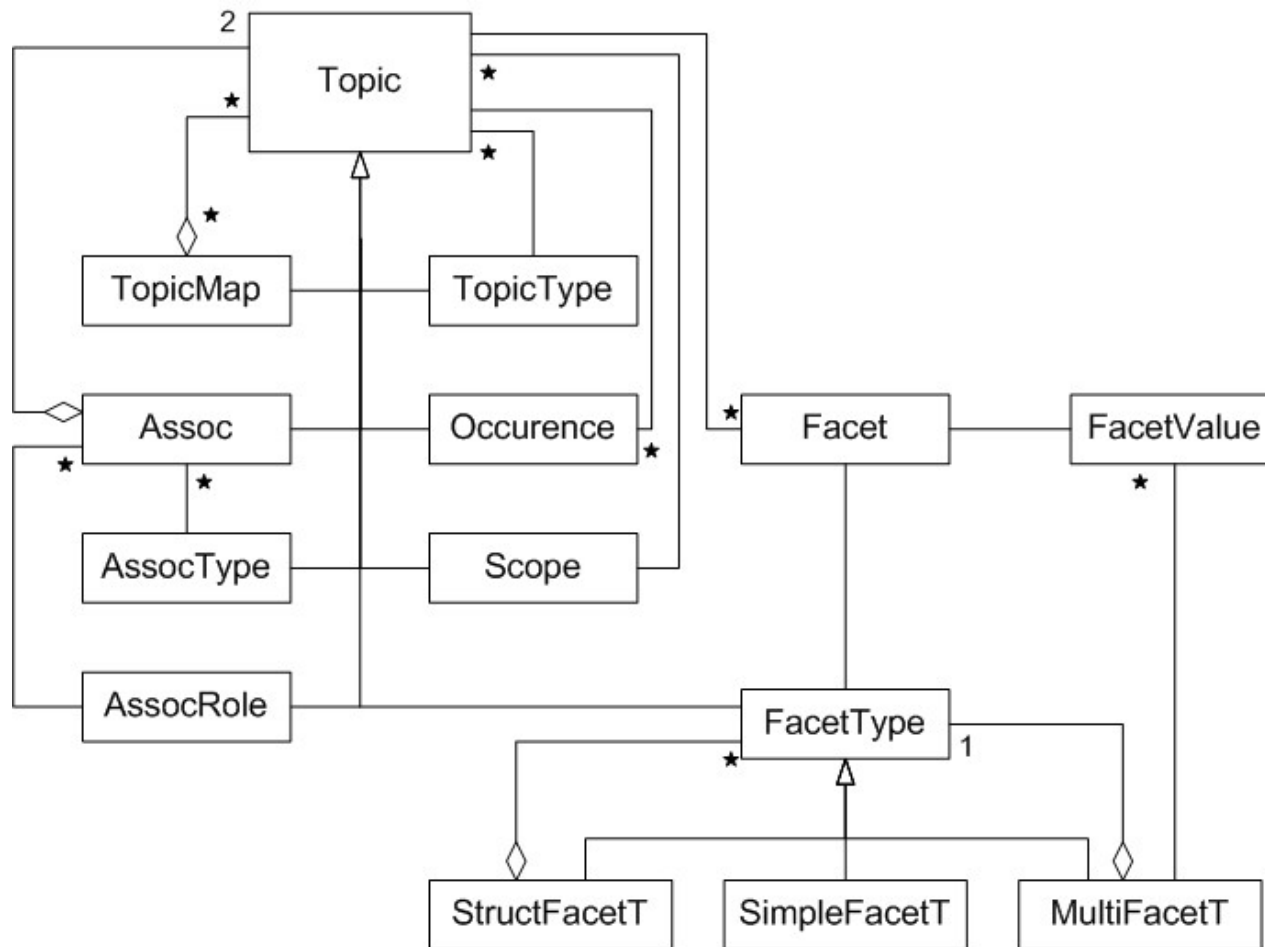
(Forts.) Abgrenzung zu anderen Konzepten:

- Taxonomie:
 - Klassifizierungen, systematische Verschlagwortung
- Folksonomy: (neues Kunstwort aus folk und taxonomy)
 - Taxonomie durch Laien, i.d.R. mit "Sozialer Software"
- Glossar:
 - Zusammenstellung von Begriffen

Topic Maps (1): Konzept

- Semantisches Netz zur Repräsentation von Wissen
- Geeignet zur Repräsentation von Ontologien
- TAO:
 - Topics: Person, log. Einheit, Konzept, ...
 - Associations: Beziehungen zwischen Topics
 - Occurrences: Erscheinungsformen von Topics
- IFS:
 - Identity: eindeutige Adressierbarkeit eines Topic
 - Facets: Eigenschaften
 - Scopes: Kontext, in dem T/A/O interpretiert werden

Topic Maps (2): Metamodell



Standards zur Repräsentation (1)

- Topic-Maps:
 - ISO 13250
 - XTM: XML Topic Maps, Redefinition ISO 13250 mit XML
- RDF (W3C): Resource Description Facility
 - XML basiert, d.h.
 - Strukturierung mit einer einheitlichen Syntax
 - Eindeutige Definition von XML-Vokabularen mit XML-Namespaces
 - Syntax von XML-Vokabularen mit XML-Schema detailliert beschreibbar

Standards zur Repräsentation (2)

- (Forts.) RDF:
 - Beschreibung einzelner Ressourcen:
 - Sammlung von Aussagen in der Form *Subjekt Prädikat Objekt*
 - Objekte können wiederum Ressourcen sein, die weiter beschrieben werden
 - Es ergibt sich ein gerichteter Graph aus Ressourcen und ihren Eigenschaften
- RDF-Schema (W3C):
 - In RDF Klassen von Objekten bilden
 - Eigenschaften in Werte- / Gültigkeitsbereiche zu ordnen

Standards zur Repräsentation (3)

- Web Ontology Language (OWL) (W3C):
 - Erweitert RDF-Schema um Beschreibungsmöglichkeiten
 - Zur Einschränkung von Eigenschaften (z.B. Kardinalitäten)
 - Angabe gemeinsamer Eigenschaften unterschiedlicher RDF-Klassen (Schnittmengen)
- Werkzeuge, Anwendungen zu RDF, RDF-Schema und OWL:
 - z.B. *dbpedia.org* : Wikipedia aufbereitet
 - *librdf.org*: Redland RDF Libraries (C), mit Bindungen zu Perl, PHP, Python, Ruby

In der Praxis: Dublin-Core (1)

(Quelle: dublincore.org, wiki.dublincore.org)

- "Dublin Core Metadata Initiative"
 - Dublin: Konferenzort in Ohio (!) im Jahr 1995
 - Dokumente mit recherchierbaren Metadaten anreichern
- "set of fifteen generic elements for describing resources"
 - Creator, Contributor, Publisher, Title, Date, Language, Format, Subject, Description, Identifier, Relation, Source, Type, Coverage, and Rights
- Weitere Elemente zur optionalen Verfeinerung der Beschreibung

In der Praxis: Dublin-Core (2)

(Quelle: de.wikipedia.org)

```
<head profile="http://dublincore.org/documents/dcq-html/">
  <title>Dublin Core</title>
  <link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
  <link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
  <meta name="DC.format"          scheme="DCTERMS.IMT" content="text/html" />
  <meta name="DC.type"            scheme="DCTERMS.DCMIType" content="Text" />
  <meta name="DC.publisher"       content="Jimmy Wales" />
  <meta name="DC.subject"        content="Dublin Core Metadaten-Elemente, Anwendungen" />
  <meta name="DC.creator"        content="Björn G. Kulms" />
  <meta name="DCTERMS.license"    scheme="DCTERMS.URI" content="http://www.gnu.org/copyleft/fdl.html" />
  <meta name="DCTERMS.rightsHolder" content="Wikimedia Foundation Inc." />
  <meta name="DCTERMS.modified"   scheme="DCTERMS.W3CDTF" content="2006-03-08" />
</head>
```

In der Praxis: LinkedData (1)

(Quellen: linkeddata.org und linkeddatabook.com)

Prinzipien nach T. Berners-Lee:

- Use URIs as names for things.
- Use HTTP URIs, so that people can look up those names.
- When someone looks up a URI, provide useful information, using the standards.
- Include links to other URIs, so that they can discover more things.

In der Praxis: LinkedData (2)

- Identifizierung per URI:
 - für alle Zusammenhänge, auch Begriffe, Erscheinungsformen, Relationen
 - Nicht nur auf Web-Dokumente anwenden
- Zugriff per HTTP
- Bereitstellung der Informationen in standardisierter Form
 - HTML
 - RDF
- Beziehungen mit Hilfe von Hyperlinks herstellen
 - Nicht nur Web-Dokumente (s.o.)
 - Mit Bedeutung anreichern (welche Beziehung?) statt untypisierter Hyperlinks wie bei Web-Dokumenten

In der Praxis: LinkedData (3)

Bereitstellung der Informationen:

- Lesbar für Benutzer
- Maschinell verarbeitbar
- Unterschiedliche Auslieferungen
 - "content negotiation" ("Verhandlung über den Inhalt"): im http-Header wird angegeben, welches Format erwartet wird
 - Für Benutzer z.B. HTML
 - Für Software z.B. RDF