

Kolloquium Wissensinfrastruktur

Semantische Suche

Jochen Schirrwagen

Universitätsbibliothek Bielefeld

05. Feb. 2010

- Semantic Web - Erweiterung des gegenwärtigen Web
- Semantische Suche - Erweiterung "klassischer Suchmaschinentechnologie"
- von der Stichwortsuche zu komplexen Anfragen
- "genauere" Suchergebnisse
- vom Dokumenten zentrierten Modell zum Entitäten zentrierten Modell
- Zeitleiste zur Wissensorganisation

Was ist Semantic Search ?

Eine Anwendung des Semantic Web

zitiert aus der Wikipedia

Semantic search is a process used to improve online searching by using data from semantic networks to disambiguate queries and web text in order to generate more relevant results.

Was sind Ontologien (im Sinne der Informatik)?

Beschreibung von Konzepten einer spezifischen Domäne und ihrer Beziehungen

- ausgedrückt in formalen Sprachen: RDF und darauf aufbauend OWL
- beruhen auf einem gemeinsamen Verständnis innerhalb einer 'Community'
- im Gegensatz zu Datenbank-Schemata, UML-Diagrammen (konzeptionelle Modelle, die aber nicht ausserhalb eines Anwendungskontext "gemeinsam" genutzt werden)
- Unterscheidung zwischen Klassen, Instanzen und Eigenschaften
- Reasoning: Konsistenzcheck in der Verwendung von Termen, Klassifizierung von Konzepten, Finden von Unterklassen, Relationen, ...

in Reihenfolge mit zunehmender Komplexität

- 1 kontrollierte Vokabularien/Menge von Termen
- 2 Glossar
- 3 Semantisches Netzwerk
- 4 Thesaurus (ab hier “automated reasoning”)
- 5 Folksonomy
- 6 lightweight ontology
- 7 heavyweight ontology
- 8 Prädikatenlogik

grafisches Ontologie-Beispiel

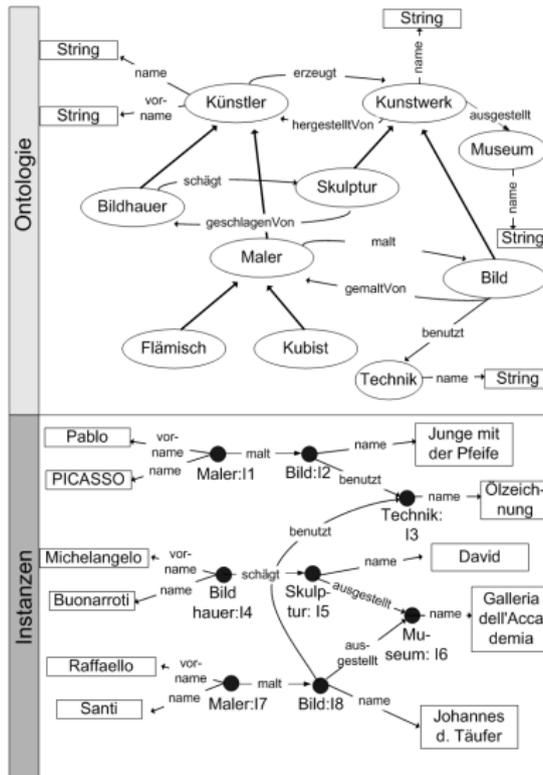


Abbildung: Quelle: Wikipedia

Eine Auswahl:

- RDF - Modellierung von Informationen über Webressourcen
- OWL - Sprachfamilie zur Erstellung von Ontologien, eine der fundamentalen Technologien des Semantic Web
- SKOS - Sprachfamilie zur Repräsentation von Thesauri, Klassifikationsschemata, Taxonomien, kontrollierte Vokabularien (Subject-Headings)), zur einfachen Publikation kontrollierter, strukturierter Vokabularien für das Semantic Web
- FOAF - Dataset zur Beschreibung von Personen, deren Eigenschaften und Beziehungen
- SPARQL - Abfragesprache über RDF im Semantic Web

Realisierung des Semantic Web:

- Annotation von existierenden Webressourcen mit Ontologie-Metadaten
- Exponieren von Datenbankinhalten durch Publikation solcher Daten und deren Schemata in einer Ontology-Sprache
- Identifizierung von Konzepten in Ontologien mit Hilfe von URIs, d.h. URIs in Daten-Ressourcen zeigen/linken auf entsprechende Konzepte und externe, frei verfügbare Ontologien
- Auf diese Weise entsteht eine fortwährende Verlinkung von Wissensressourcen im Web. (Web of Data, Web of Knowledge, Linked Open Data)

Nächste Schritte:

- Sammeln aus / Abfragen von Datenquellen
- Aggregation
- logisches Verknüpfen (reasoning) auf Basis einer Ontologie mit den Ergebnissen

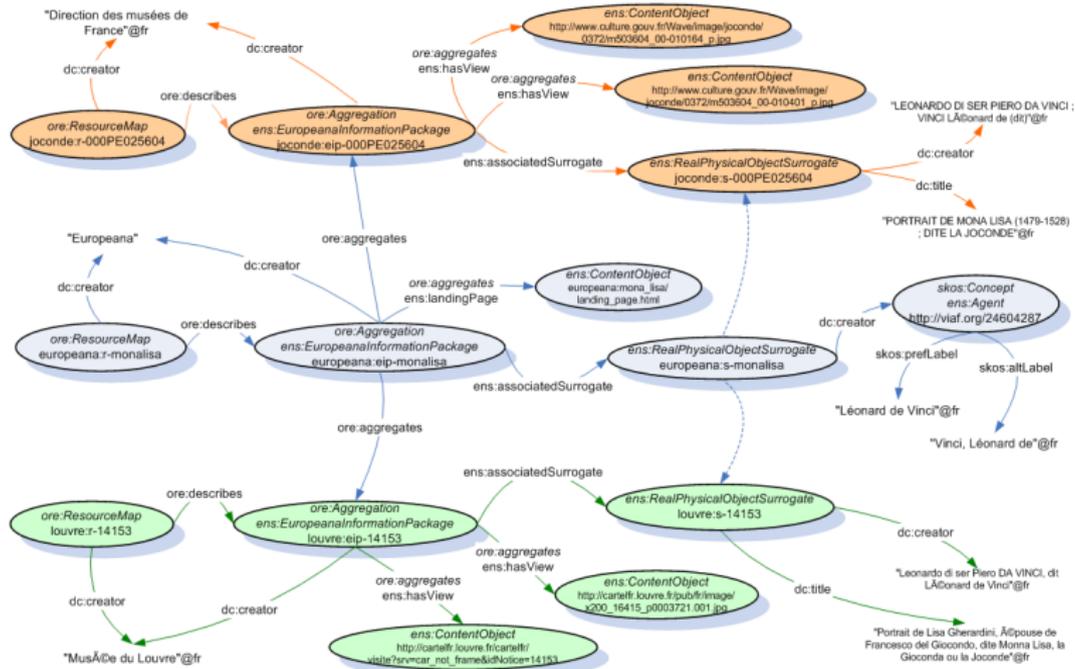
Die Information hat eine maschinell auswertbare Semantik erhalten. Es gibt nun zwei Möglichkeiten mit den semantisch angereicherten Daten umzugehen:

- Exponieren dieser Daten oder
- Abfragen über SPARQL

Europeana Data Model wird modelliert mit

- OAI-ORE - zur Definition von Aggregationen
- DC - zur Zuweisung von Eigenschaften an Objekte
- SKOS - Kontextualisierung von Objekten

Europeana Data Model



- Notes:
- many (HTTP) URIs were abbreviated (e.g. with "europeana:")
 - some values for dc:creator may be modelled as instances of ens:Agent in the "real" semantic layer
 - color code: orange (resp. green, blue) denotes information contributed by the Joconde portal (resp. the Louvre museum, Europeana)
 - in the resource nodes (resp. arrow labels), *italics* refer to the classes instantiated by the resource (resp. a superproperty of the property)

Abbildung: Quelle: Europeana

Architektur Semantisches Suchsystem

Am Beispiel von Watson "Semantic Web Gateway"

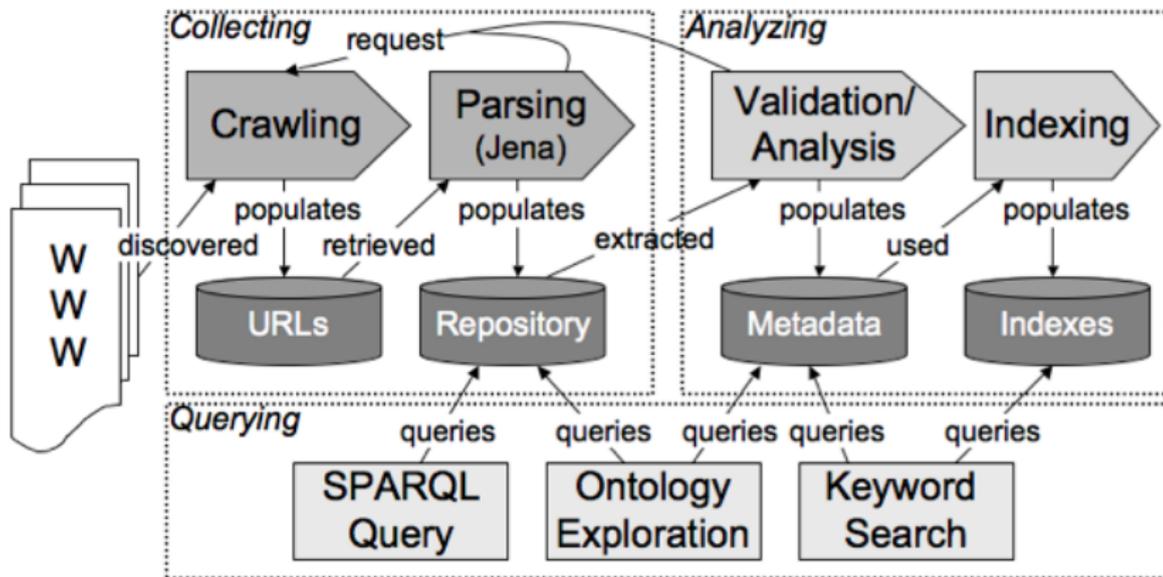


Abbildung: Quelle: <http://watson.kmi.open.ac.uk/Overview.html>

- **DBpedia** - extracted data from Wikipedia: 2,18 Mio Konzepte
→ beschrieben durch 218 Mio Triple in 11 verschiedenen Sprachen - <http://wiki.dbpedia.org/OnlineAccess>
- **Europeana** - Prototyp einer semantischen Suchmaschine - <http://eculture.cs.vu.nl/europeana/session/search>
- **Hakia** - Phrasenanalyse auf Basis natürlicher Sprache - <http://www.hakia.com/>
- **Semantic MediaWiki** - freie Erweiterung des MediaWiki - http://semantic-mediawiki.org/wiki/Semantic_MediaWiki
- **Watson** - Semantic Web Gateway (sammelt, analysiert, indexiert semantische Dokumente) - <http://watson.kmi.open.ac.uk/WatsonWUI/>
- **Wolfram Alpha** - computational knowledge engine - <http://www.wolframalpha.com/>

- **OpenLink Data Explorer** (Firefox AddOn)
<http://esw.w3.org/topic/OpenLinkDataExplorer>
- **Protégé** - Semantic Editor zur Bearbeitung von Ontologien -
<http://protege.stanford.edu/>
- **SIMILE** - Semantic Interoperability of Metadata and Information in unLike Environments <http://simile.mit.edu/>
- weitere Informationen : *Semantic Web Development Tools*
<http://www.w3.org/2001/sw/wiki/Tools>

- Kommt das Ende der klassischen Suchmaschine oder der Suche überhaupt?
- Nova Spivack über “ Eliminating the Need for Search – Help Engines”

Zitat

...For example, choosing a location for an event, or planning a trip itinerary, or choosing what medicine to take, deciding what product to buy, who to hire, what company to work for, what stock to invest in, what website to read about some topic. These kinds of activities require a lot of research, evaluations of choices, comparisons, testing, and thinking. A lot of clicking. And they also happen to be some of the most monetizable activities for search engines. Existing search engines like Google that make money from getting you to click on their pages as much as possible have no financial incentive to solve this problem — if they actually worked so well that consumers clicked less they would make less money.

- Hildebrand et al. (2007) *Semantic Search Survey*
http://swuiwiki.webscience.org/index.php/Semantic_Search_Survey
- Kruk, S.R.; McDaniel, B. (2009) *Semantic Digital Libraries*, Springer
- Mika, P. (2007) *Social Networks and the Semantic Web*, Springer
- Spivack, N. *Eliminating the Need for Search – Help Engines*
<http://www.novaspivack.com/uncategorized/eliminating-the-need-to-search>
- Sutter, J.D. (2009) *New search engines aspire to supplement Google*
<http://edition.cnn.com/2009/TECH/05/12/future.search.engine/>

Vielen Dank für Eure Aufmerksamkeit!