

Semantic Web Paradigmen

Simon Heimler

heimlersimon@gmail.com

Master of Applied Research Computer Science

Prof. Dr. Sabine Müllenbach

Faculty of Computer Science

University of Applied Sciences Augsburg

EINLEITUNG

SEMANTIC WEB PARADIGMEN

HYPERLINKS

GRAPHENSTRUKTUR

TRENNUNG VON FAKT UND INTERPRETATION

OPEN WORLD ASSUMPTION

FAZIT

Einleitung

„Wer als Werkzeug nur einen Hammer hat, sieht in jedem Problem einen Nagel.“

Unbekannter Autor

- Semantic Web bietet einige interessante, allgemein eher unbekannte Konzepte und Ideen.
- Erweiterung des (mentalen) Werkzeug-Arsenals!

- **Semantic Web:** Offizieller, akademischer Begriff
- **Linked Data:** Selbe Technologien.
Schwerpunkt: Erstellung von (SW-kompatiblen) Daten.
- **Paradigma:** Grundsätzliche Denkweise, Weltanschauung

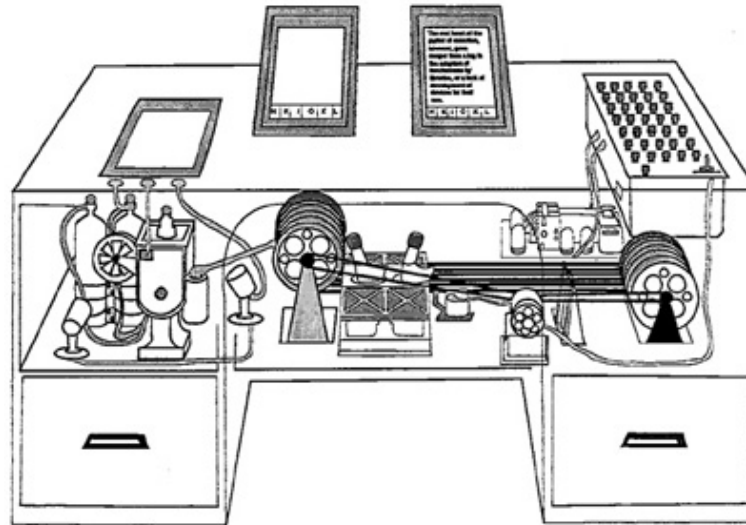
- Üblicherweise wird das Semantic Web anhand der verwendeten Technologien erklärt
- Hier: Erklärung einiger Konzepte / Paradigmen dahinter.
- Die Auswahl ist subjektiv und bei weitem nicht komplett!

Semantic Web Paradigmen

Hyperlinks

HYPERLINKS

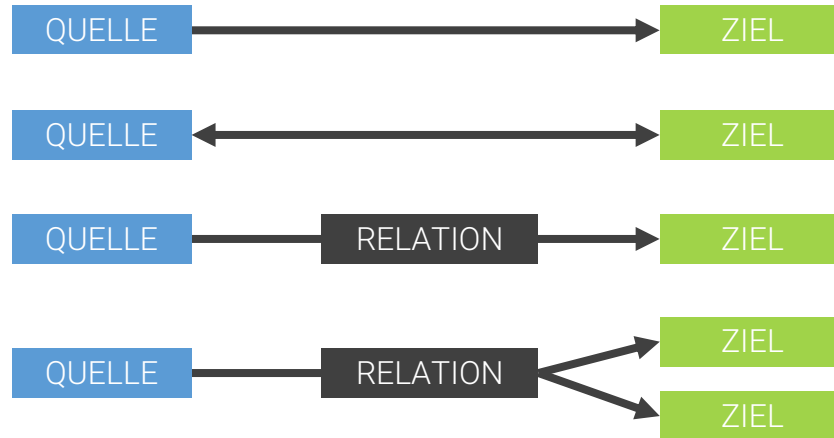
- Hyperlinks sind ein altes Konzept.
 - Geht mind. bis auf die Memex (1945) zurück.
 - Wichtiger Bestandteil der Hypertextsysteme (ab 60er Jahre).



<http://trevor.smith.name/memex/>

- Umgangssprachlich: Meist nur das Ziel des Links (URI) gemeint.
- Links sind eigentlich “Pfeile”.
- Links haben mehrere Komponenten.

- Können unterschiedlich “mächtig” sein:
 - Direktional oder Unidirektional (directionality)
 - Ein Ziel oder mehrere Ziele (arity)
 - Benannte Beziehung oder nicht (named, unnamed)



```
1 <a href="http://example.com#anchortag" rel="">Example.com</a>
```

- Ausschließlich direktional mit genau einem Ziel.
- Quelle ist implizit immer das Dokument selbst.
- Beziehung spielt keine nennenswerte Rolle. (rel=„“)
- Möglich auf eindeutig bezeichnete Unterelemente zu verlinken (anchor-tags)

- Relativ reduzierte, aber einfach zu verwendende Links.
- Direktionale Links ermöglichen die dezentrale Architektur des Webs, da kein Einverständnis, bzw. zentrale Datenbank nötig!
 - Dafür haben wir „tote“ Links.
- Rückblickend eine gute Entscheidung, das Web hat sich als einziges Hypertextsystem nennenswert durchgesetzt.

BEZIEHUNG ZUM SEMANTIC WEB

- Das Semantic Web baut das Konzept der Links wieder aus.
- Links werden verwendet um Inhalte und deren Beziehungen zu beschreiben.
- Dieses Konzept heißt RDF.

- Die Relation ist essentiell.
- Jedes Element ist eine URI
- URIs können sich auch auf abstrakte und reale Dinge beziehen.
- Dadurch entstehen einfache, grammatikalische Aussagen

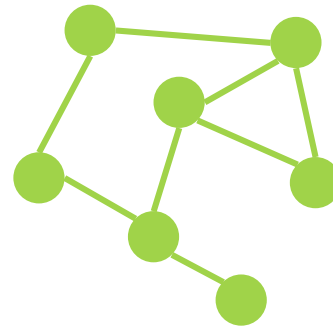
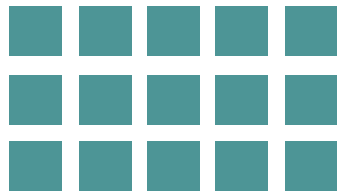
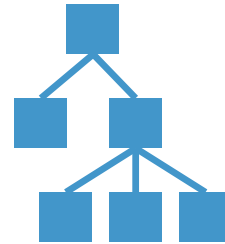


- Links können nicht nur Dokumente verlinken, sondern auch die Inhalte der Dokumente zueinander in Beziehung setzen.
- So entsteht eine einfache, minimalistische „Maschinensprache“.
- Da jedes Element eine URI ist, können auch Aussagen über verschiedene Domänen und Datenbestände hinweg getroffen und verknüpft werden.

Graphenstruktur

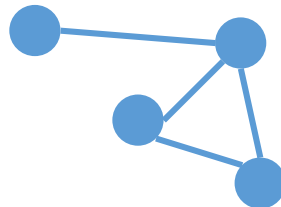
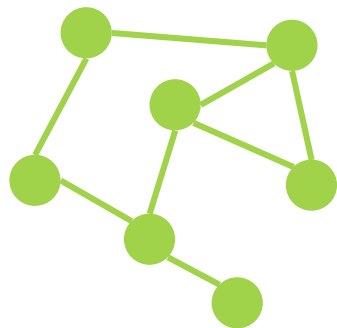
GRAPHENSTRUKTUR

- Ein Graph ist die flexibelste Datenstruktur. Sie kann alle anderen abbilden.

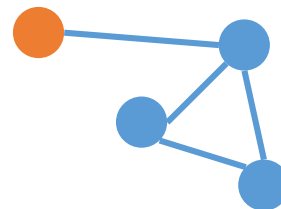
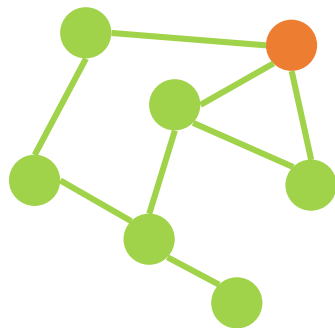


GRAPHENSTRUKTUR

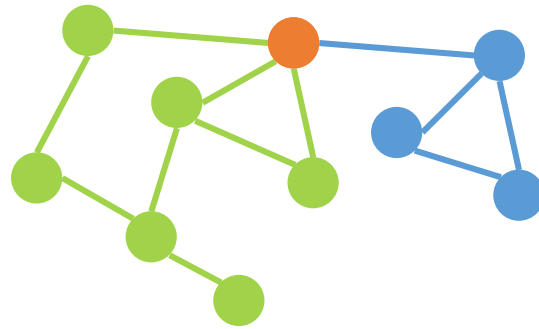
- Graphen können problemlos miteinander “gemergt” werden.
- Bei Tabellen und Bäumen gibt es Probleme. Oft sind zusätzliche Anweisungen nötig. (Schema)



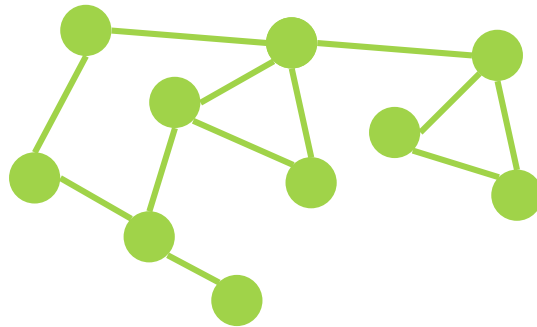
- Sobald sich zwei Knoten auf dasselbe beziehen (selbe ID), können sie zusammengelegt werden:



GRAPHENSTRUKTUR



GRAPHENSTRUKTUR



- Viele Strukturen sind eigentlich Graphen und Netzwerke.
- Problem: Graphenstrukturen sind für Menschen oft schwer zu interpretieren und zu verstehen.
- Deswegen abstrahieren / vereinfachen wir gerne auf einfachere Strukturen, wie Bäume.
- Bsp. Unternehmenshierarchie.
Soziale Beziehungen sind allerdings trotzdem Netzwerke!

- Eine Datenstruktur die näher an der “realen” Struktur ist, hat Vorteile:
 - Kann performanter sein (Bsp. Freunde 3. Grades)
 - Evtl. einfacher zu modellieren, da nicht so viel Abstraktion nötig ist.
 - Abstraktion ist oft (immer?) subjektiv.

- **Schema-los:** Nicht notwendig vorher ein Schema festzulegen. (Im Gegensatz zu relationalen DB)
 - Graphen erlauben es Daten aufzunehmen, ohne sie vorher abstrahieren zu müssen.
 - Die Struktur entsteht organisch mit den eingetragenen Daten.
- Graphendatenbanken fallen damit in die Kategorie NoSQL.

- RDF / Hyperlinks erzeugen einen Graphen, wenn sie zusammengelegt werden.
- Sinnvolle Struktur für das verteilte Wissen im Internet, da sie flexibel ist und sich leicht kombinieren lässt.

Trennung von Fakt und Interpretation

ONTOLOGIE

- Graphen sind schema-los.
- Es ist dennoch möglich ein Schema (optional) zu verwenden.
- Im Semantic Web werden diese Ontologien genannt.

ONTOLOGIE

- Durch Ontologie / Schemas können weitere Beziehungen zwischen den Daten hergestellt werden.
- Regeln wie:
 - Wenn A die Tochter von B ist, dann ist B der Vater von A.
 - Wenn Person A Person B kennt, kennt auch B Person A. (invers)
 - Personen sind Lebewesen und erben ihre Eigenschaften (inheritance)
 - Komplexere Regeln, bis hin zur Unentscheidbarkeit. (first order logic)
- Zu komplexes Thema um dies weiter zu vertiefen.

TRENNUNG VON FAKT UND INTERPRETATION

- Da Schemas optional sind, können sie von den Daten unabhängig sein - und umgekehrt.
- Es ist möglich mehrere Ontologien mit der selben Faktenbasis zu verwenden.
- Ontologien können weiterentwickelt werden, ohne dass der Datenbestand dafür migriert werden muss.
- Unterschiedliche Interpretationen und Weltanschauungen sind also möglich!

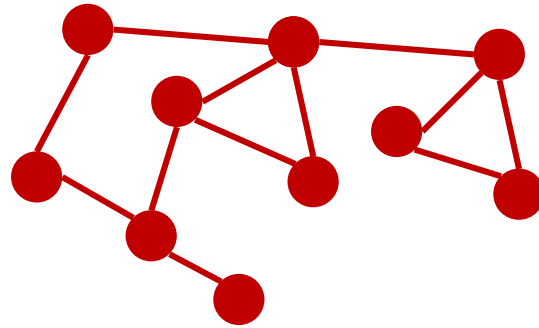
TRENNUNG VON FAKT UND INTERPRETATION



- **Fakten**, **Interpretation** und damit auch die **Schlussfolgerungen** sind klar getrennt.

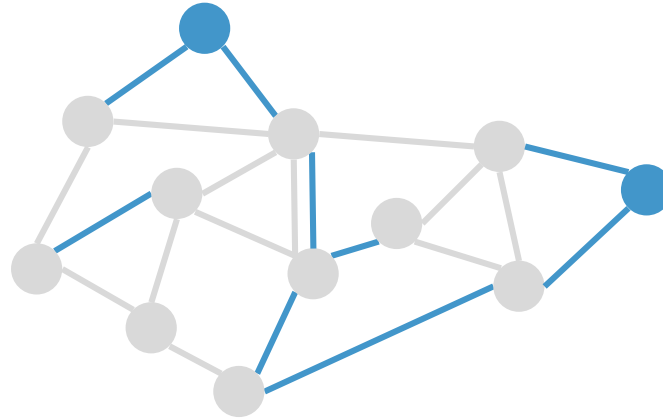
TRENNUNG VON FAKT UND INTERPRETATION

FAKTENBASIS



TRENNUNG VON FAKT UND INTERPRETATION

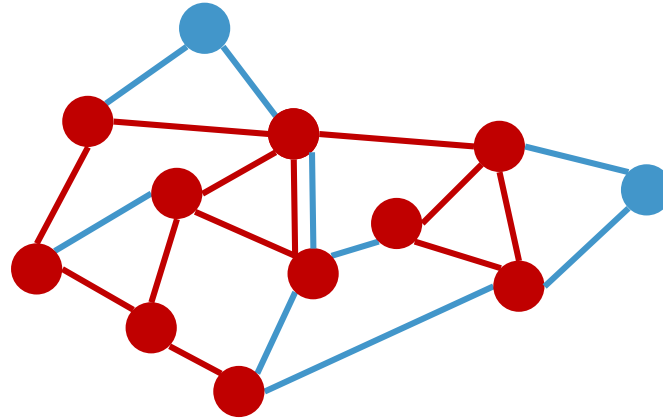
SCHLUSSFOLGERUNGEN



TRENNUNG VON FAKT UND INTERPRETATION

WISSENSBASIS

QUERIES



FAZIT

- Da die Daten erfasst werden können, ohne vorher ein Schema (und damit Interpretation) festzulegen, sind diese neutraler.
- Stellt sich die Interpretation als falsch heraus, erstellt man eine neue und kommt mit den alten Daten zu anderen Schlüssen!
- Sowohl die Daten als auch das Schema kann dynamisch mit den Anforderungen mitwachsen.

Open World Assumption

OPEN WORLD ASSUMPTION

- Die Open World Assumption (OWA) geht davon aus, dass eine Wissensbasis immer unvollständig ist.
- Das Gegenkonzept ist die Closed World Assumption.
- Die meisten Systeme (z.B. relationale DB) sind Closed World.

OPEN WORLD ASSUMPTION

CLOSED WORLD

Alles was nicht explizit wahr oder bekannt ist, ist **falsch**.

Das System weiß alles.

Widersprüche führen zu Fehlern.

Einfache Validation.

OPEN WORLD

Alles was nicht explizit als wahr oder falsch bekannt ist, ist **unbekannt**.

Falsch ist nur, was zu unauflösbaren Widersprüchen führt.

Das System ist immer unvollständig.

Widersprüche werden, falls möglich, aufgelöst und neues Wissen erzeugt.

Validation ist aufwendig und erfordert Reasoner.

GESCHLOSSENE WELTEN

- Open World ist nicht „besser“ als Closed World, sondern anders.
- Viele Systeme müssen geschlossen sein, da Unsicherheit keine Option ist.
- „Kleine Welten“ mit überschaubarer Komplexität sind oft einfacher als geschlossene Welt umzusetzen.
- Closed World ist technisch weniger aufwendig.

OFFENE WELTEN

- Sinnvoll bei Kombination und Schöpfung von neuem Wissen aus vielen, unterschiedlichen Quellen.
- Kann mit inkompletten und widersprüchlichen Informationen „nativ“ umgehen.
- Weitere Themen:
 - **Fuzzy logic:** Unschärfelogik, Wahrscheinlichkeiten
 - **Provenance:** Herkunft
 - **Trust:** Vertrauenswürdigkeit

Fazit

FAZIT

- Nimmt man die vorgestellten Konzepte zusammen, erhält man einen Teil des Semantic Webs
- Die Paradigmen sind nicht auf das Semantic Web begrenzt.
- Bsp. Graphendatenbanken (Neo4J).

- Relativ komplexe Technologien im Hintergrund
- Tools sind meist noch unausgereift oder Forschungsprojekte
- Performance kann zum Problem werden
- Sollte gut überlegt sein, auf eher „experimentelle“ Technologien und Konzepte zu setzen.
- Wenn man mit den alten Werkzeugen nicht mehr weiterkommt, ist es vielleicht Zeit ein neues zu versuchen!

Fragen?

Danke für die
Aufmerksamkeit!

www.fannon.de/url/swp-ppt

Anhang

Mensch-Maschine Kooperation

- Es gibt viele Gebiete in denen Maschinen bessere Leistungen bringen als Menschen.
- Dasselbe gilt auch umgekehrt: Trotz KI Forschung gibt es viele Bereiche in denen Menschen nicht ersetzt werden können – oder sollten.
- Wenn Mensch und Computer produktiv zusammenarbeiten, entstehen oft die besten Ergebnisse.

- Das Semantic Web kann als „Mensch-Maschine Kooperation“ Initiative für das Web verstanden werden.
- Das aktuelle Web ist fast ausschließlich für Menschen optimiert.
- Maschinen können die Inhalte und deren Bedeutung oft nur erraten, dabei kommt es zu Fehlern und Ungenauigkeiten.

- Wenn Maschinen die Information im Web klar und eindeutig verstehen und interpretieren können, eröffnet dies viele neue Möglichkeiten.
- Maschinen können automatisiert Informationen abrufen, zusammenstellen und daraus neues Wissen schließen.
- Dadurch können sie Menschen bei Informationssuche und repetitiven Aufgaben besser unterstützen.
- Es wird ein intelligenteres und besser verknüpftes Web möglich.

- Dies wird durch Semantische Annotation möglich:
- Webseiten enthalten nicht nur eine menschen-lesbare Version, sondern zusätzlich eine maschinen-lesbare.
- Nicht nur einfache Meta-Daten, die nur das Dokument an sich beschreiben – sondern eine Beschreibung der Inhalte!

- Google, Yahoo, Microsoft und Yandex riefen **schema.org** ins Leben, ein gemeinsames Vokabular zur Beschreibung von Inhalten im Web.
- Facebooks OpenGraph Protokoll verwendet ebenfalls Semantic Web Technologien um Inhalte maschinenlesbar zu definieren.
- Könnte für SEO zukünftig eine große Rolle spielen!