



Thema: Semantic Web

- Einführung -

Name, Vorname: Fuchs, Christian

Matrikelnummer: 288116

Kurs: WWI 07 E

Fach: Systemanalyse

Mannheim, 13.10.2008

## Inhaltsverzeichnis

|   |    |
|---|----|
| Inhaltsverzeichnis .....                        | 2  |
| Abbildungsverzeichnis .....                     | 3  |
| 1. Die Entstehung des World Wide Web .....      | 4  |
| 2. Die Grenzen des World Wide Web .....         | 5  |
| 3. Semantic Web als Lösung .....                | 9  |
| 3.1 Was ist das Semantic Web? .....             | 10 |
| 3.2 Wie funktioniert das Semantic Web? .....    | 11 |
| 3.2.1 Metadaten .....                           | 11 |
| 3.2.2 Taxonomie .....                           | 12 |
| 3.2.3 Ontologie .....                           | 14 |
| 4. Anwendungen des Semantic Web .....           | 15 |
| 4.1 Navigieren und Suchen .....                 | 15 |
| 4.2 Social Software und soziale Netzwerke ..... | 17 |
| 4.3 Tagging .....                               | 17 |
| 5. Fazit .....                                  | 18 |
| Quellen- und Literaturverzeichnis .....         | 19 |

## Abbildungsverzeichnis

|  |    |
|--|----|
| Abbildung 1: Blogosphere growth (Technorati).....    | 5  |
| Abbildung 2: Daily Posting Volume (Technorati) ..... | 6  |
| Abbildung 3: Taxonomie Bibliothek .....              | 13 |
| Abbildung 4: Falsche Taxonomie .....                 | 13 |
| Abbildung 5: Kreuzkatalog.....                       | 15 |

## 1. Die Entstehung des World Wide Web

Das World Wide Web entstand 1989, als Tim Berners-Lee die Hypertext Markup Language (HTML) erfand und somit das Web, wie wir es heute kennen, begründete. Vier Jahre später, mit der Veröffentlichung des ersten graphikfähigen Webbrowsers „Mosaik“, begann der große Aufschwung des World Wide Web. Dieser wurde kostenlos zum Download angeboten und ermöglichte nun auch Laien, das Web zu nutzen.

Dieser Meilenstein in der Geschichte machte es erst möglich, dass heute nahezu jeder Zugang zu einem internetfähigen Computer besitzt. Der Wandel der Industriegesellschaft zur Informationsgesellschaft ließ die Bedeutung des World Wide Web weiter steigen. Die Kommunikation über E-Mails wurde ermöglicht und die Wirtschaft begann im Web Fuß zu fassen. Der Online-Handel etablierte sich und das Web wurde von der Werbebranche erschlossen. Nahezu jeder kauft heute bei Online-Shops und Aktionshäusern wie Amazon oder eBay ein. Auch verdient eine Großzahl an kommerziellen Internetpräsenzen ihr Geld rein durch Werbung. Erst die massive Nutzung des Internets durch die breite Masse hat dies überhaupt ermöglicht.

Heute prägt der Begriff „Web 2.0“ das Internet. Täglich entstehen neue Weblogs, Wikis und soziale Netzwerke. Auch die Kommunikation über E-Mail und Instant-Messaging-Programme ist kaum noch wegzudenken.

Kein Medium hat solch eine universelle Informationsvielfalt wie das Internet. Es gibt uns die Möglichkeit zu jeder Zeit an jedem Ort an Informationen zu kommen, die an einem anderen Ort gespeichert wurden. Die Informationen können jederzeit auf den neusten Stand gebracht werden, was durch die vernachlässigbare Übertragungszeit zum Empfänger zu einer ständigen Aktualität der bereitgestellten Informationen führt. Diese Aktualität gekoppelt mit der ständigen Verfügbarkeit haben dazu beigetragen, das Internet zu einem weltweiten Netz für Informationstausch und Kommunikation heranwachsen zu lassen.

## 2. Die Grenzen des World Wide Web

Die zuvor beschriebene Entwicklung hat allerdings auch negative Auswirkungen. Die hierdurch entstandenen Probleme erschweren uns die Nutzung des World Wide Web.

Das Web bietet heute eine so große Menge an Informationen, dass es unüberschaubar geworden ist. Gerade durch den Erfolg des „Web 2.0“ wuchs die Informationsvielfalt des Webs sehr stark an. Früher war es für den einzelnen Menschen nicht einfach, Informationen ins Web zu bringen. Heute können nicht nur Reporter von Online-Magazinen Artikel veröffentlichen, sondern auch Privatpersonen. Jeder Interessierte kann sich sein eigenes Blog einrichten und somit sein Wissen teilen. Wer hätte gedacht, dass so viele Menschen diese Möglichkeit nutzen würden? Die Universität Leipzig hat 2007 in Kooperation mit der Suchmaschine Ask.com über 600 deutsche Internetnutzer über deren Nutzung von Blogs befragt. Fast 90% haben schon einmal ein Blog genutzt und über ein Drittel der Befragten schreiben sogar ihr eigenes Weblog. Technorati zählte ebenfalls 2007

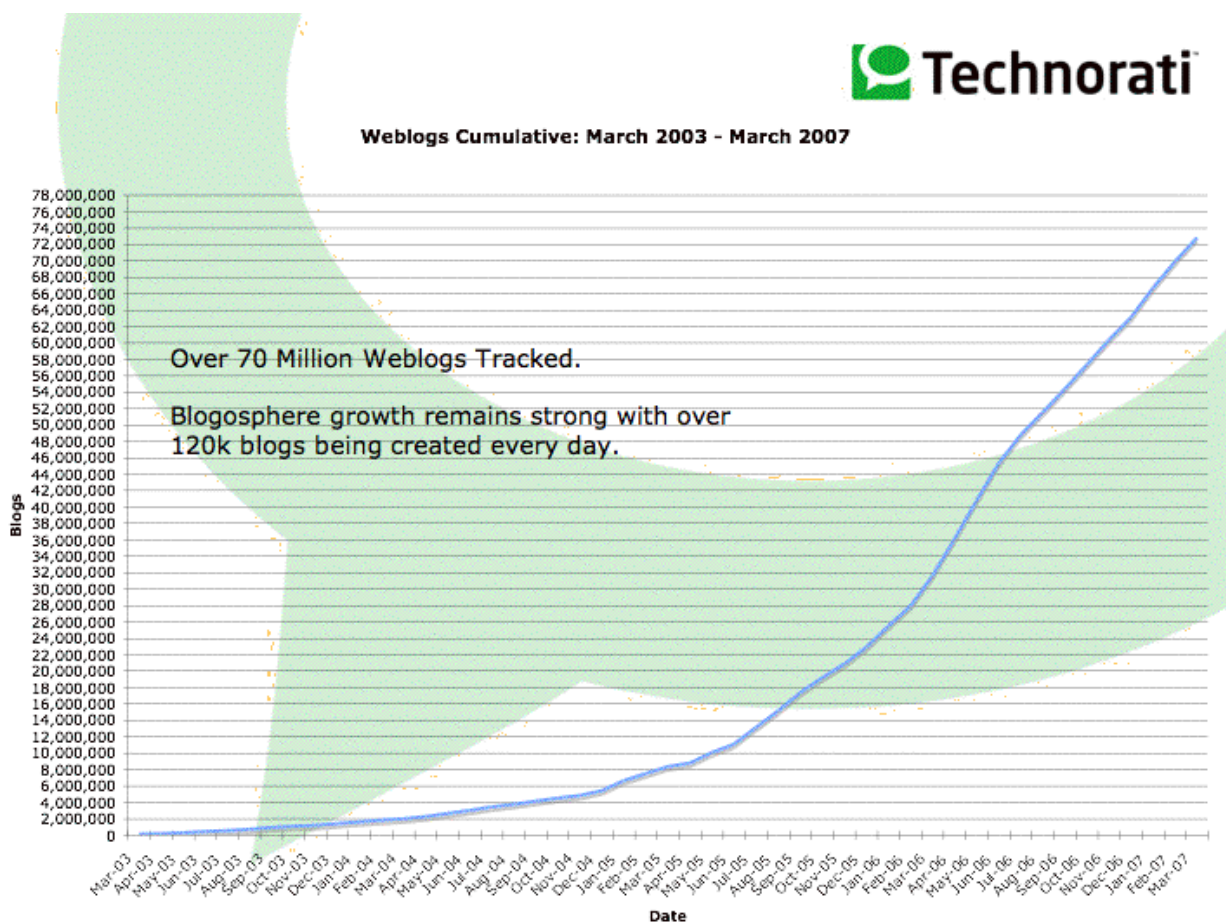


Abbildung 1: Blogosphere growth (Technorati)

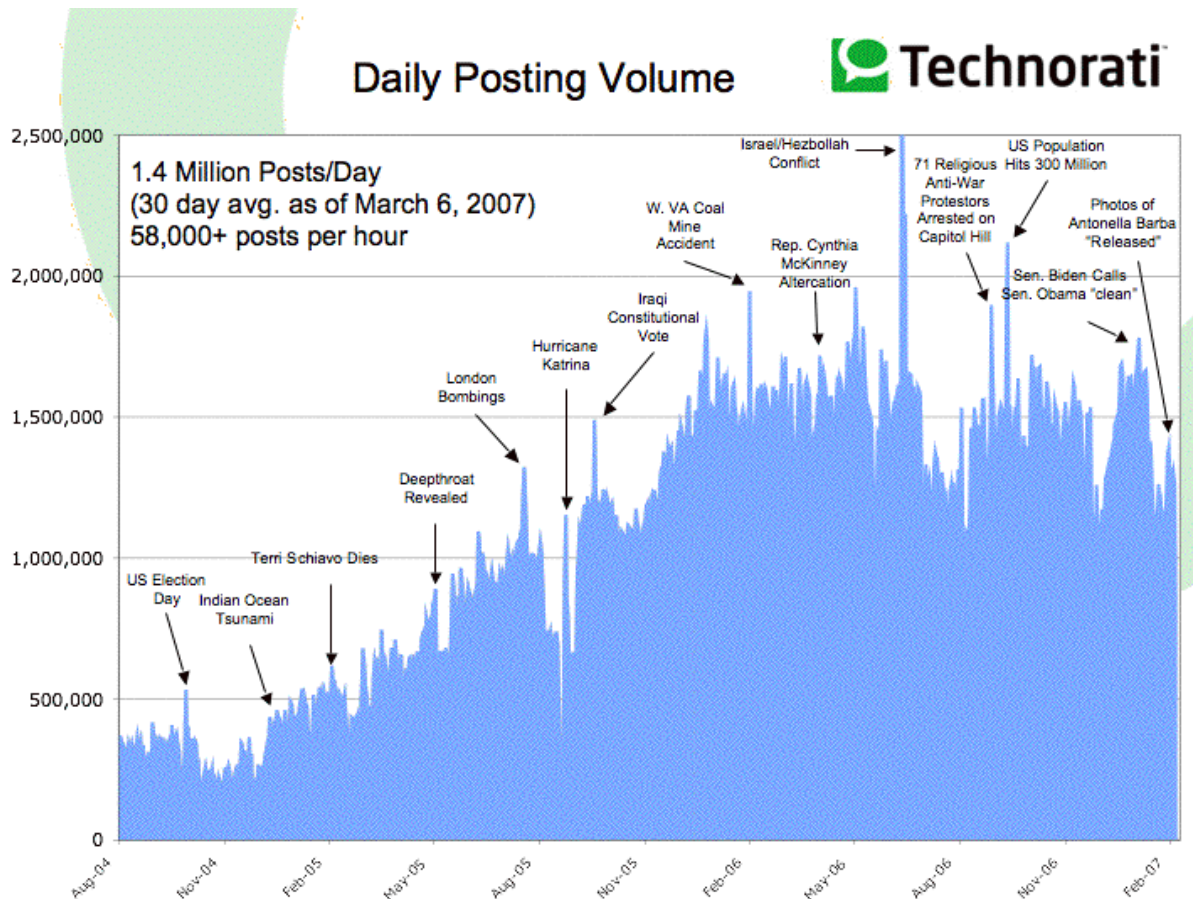


Abbildung 2: Daily Posting Volume (Technorati)

in der Summe 70 Millionen Blogs (siehe Abb. 1), wobei täglich 120.000 neue Blogs entstehen. Im Durchschnitt werden 1,5 Millionen Beiträge täglich geschrieben (siehe Abb. 2), das sind 17 pro Sekunde.

Eine weitere Möglichkeit sein individuelles Wissen im Web zu veröffentlichen stellen Wikis dar. Wikis sind Wissenssammlungen und können von Benutzern im Regelfall nicht nur gelesen, sondern auch selbst online verändert werden. Wikipedia, der bekannteste Vertreter der Wikis, hat immer mehr an Bedeutung gewonnen. Aufgrund von Verlusten in Millionenhöhe musste sich eine feste Größe wie der Brockhaus aus dem Print-Bereich zurückziehen. Im Dezember 2005 wurde in der Zeitschrift „Nature“ ein Artikel veröffentlicht, der die Qualität der Online-Ausgabe der Encyclopædia Britannica mit der der englischen Wikipedia verglich. Bei Stichproben bezüglich Korrektheit und Vollständigkeit stellte der Autor nur wenige Unterschiede fest. Mit mehr als neun Millionen Artikeln hat Wikipedia rein quantitativ längst alle bekannten Enzyklopädien überholt.

Die Frage lautet: Was bringen uns Informationen, die wir nicht finden können? Das Web wird geradezu überflutet und es wird immer schwerer die benötigten Informationen herauszufiltern. Das Navigieren und Suchen nimmt heute einen wesentlichen Teil der Zeit in Anspruch, die wir im Internet verbringen. Wir geben verschiedene Kombinationen von relevanten Begriffen in Suchmaschinen ein und folgen den Links auf Internetseiten, die uns wiederum neue Links anbieten. Wir müssen die wesentlichen Informationen aus einer Unmenge an Daten herausfiltern. Oft benötigen wir auch Informationen von verschiedenen Quellen, die wir sinnvoll zusammenfügen können, was wiederum sehr viel Zeit in Anspruch nimmt.

Warum finden die Suchmaschinen nicht genau das, was wir gerne finden möchten? Die Antwort ist einfach: Sie wissen nicht, was wir suchen. Das klassische Web richtet sich an den Menschen.

An dieser Stelle ist es notwendig, den Begriff der Semantik zu definieren. Semantik kommt aus dem griechischen und bedeutet „bezeichnen“. Sie ist die Bedeutungslehre, die Wissenschaft von der Bedeutung der Zeichen.

Der Mensch kann Informationen als solche erkennen und transformieren, er kann auf semantischer Ebene denken. Er kann eine große Menge an Informationen aufnehmen, ordnen und sie somit zu einem sinnvollen System zusammensetzen. Automatisch, ohne darüber nachzudenken, stellen wir Beziehungen zwischen einzelnen Informationen her. Wenn wir einen Text lesen, können wir den Inhalt erfassen. Eine Maschine kann das nicht. Sie kann einzelne Zeichen, Wörter oder Sätze in Texten finden, die Bedeutung bleibt ihr aber verborgen. Für den Computer ist ein Text nur ein String, bestehend aus mehreren Zeichen. Aufgrund unserer komplexen und vieldeutigen Sprache kommt es deshalb oft zu Missverständnissen zwischen Mensch und Maschine. Eine Maschine kann zum Beispiel den Unterschied zwischen einer Bank zum sitzen und dem gleichnamigen Geldinstitut nicht erkennen, wenn wir ihr nur das Wort „Bank“ liefern. Sie erkennt weder Ironie noch Sarkasmus. Auch bei Formulierungen wie „das Gegenteil von ...“ oder „der Unterschied zwischen ...“ weiß sie nicht, was gemeint ist.

Solche Ausdrucksweisen oder Formulierungen, die der Mensch benutzt, um die eigentliche Information näher zu beschreiben, bzw. um die Bedeutung für den Empfänger eindeutig festzulegen, erkennt das klassische Web nicht als solche. Im

Web ist ein String ein String, er hat keine Bedeutung. Somit ist das klassische Web nicht semantisch.

Aufgrund dieser Tatsache, dass eine Maschine, und damit auch das Web, die Bedeutung hinter Informationen nicht versteht, kann es Informationen auch nicht verarbeiten und somit auf Neues schließen. Der Mensch kann Informationen, die nicht explizit vorliegen, implizieren. Wenn wir beispielsweise wissen, dass der Zeitunterschied von Mannheim nach Sydney neun Stunden beträgt, können wir daraus schließen, dass wir um 19Uhr nicht in Sydney anrufen sollten, wenn wir niemanden wecken wollen. Wenn wir jetzt eben diese Information im Web suchen, werden wir sie wahrscheinlich nicht finden. Wir finden nur den Zeitunterschied und müssen uns den Rest selbst denken.

Das klassische Web stellt uns aufgrund des steigenden Informationsgehaltes und dem fehlenden Verständnis der Semantik also vor zwei Probleme:

1. Es fällt uns schwer, die gesuchte Information zu finden, vorausgesetzt sie ist explizit vorhanden.
2. Die Denkleistung beim Implizieren von nicht vorhandenen Informationen müssen wir selbst erbringen.

### 3. Semantic Web als Lösung

Stellen Sie sich folgendes Szenario vor:

*„Toms Sehstärke hat sich geändert, deshalb ist er auf der Suche nach einem Augenarzt in seiner Heimatstadt. Er gibt in eine Suchmaschine die Begriffe "Augenarzt" und "Mannheim" ein. Als Ergebnis bekommt er eine lange Liste an Links. Zuerst muss er alle Links herausfiltern, die nicht direkt auf Internetseiten von Ärzten zeigen. Natürlich hat die Suchmaschine nicht nur Augenärzte gefunden, die anderen Ärzte muss er auch herausfiltern. Da er kein Auto besitzt, sucht sich Tom alle Augenärzte zusammen, die mit öffentlichen Verkehrsmitteln zu erreichen sind. Dazu muss er die genaue Lage der Adressen in einem Stadtplan herausfinden. Tom vergleicht die Sprechstunden der verbliebenen Ärzte mit seinem Kalender. Übrig bleiben drei Augenärzte. Er ruft bei allen an und bekommt bei nur einem einen passenden Termin innerhalb der nächsten Woche, diesen aber auch nur montags morgens. Jetzt muss Tom nur noch seinem Chef Bescheid geben, den Termin in seinen Kalender eintragen, die optimale Verkehrsverbindung zur Praxis herausfinden und eine Fahrkarte kaufen.“*

Ein solches Szenario kennen wir alle. Eine simple Angelegenheit, die aber einige Zeit in Anspruch nimmt und einiges an Aufwand mit sich bringt, da sie sich in viele Teilaufgaben spaltet. Das klassische Web kann uns dabei helfen, beispielsweise durch die Online-Suche oder die Kommunikation über E-Mails. Es kann jedoch keine Beziehung zwischen den einzelnen Teilaufgaben herstellen, um uns Zeit und Aufwand zu ersparen. Selbst wenn Toms Terminkalender, der Stadtplan, die Kontaktdaten zu seinem Chef und die restlichen benötigten Informationen vorliegen, weiß das Web nicht, was das eine mit dem anderen zu tun hat.

Hier stellt man sich doch die Frage, ob das ganze nicht auch einfacher und schneller funktionieren könnte?

An dieser Stelle tritt wieder der Urheber des World Wide Web in den Mittelpunkt des Geschehens. Seine Antwort auf diese Frage lautet eindeutig: „Ja“. 1999 schrieb er:

*“I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A ‘Semantic Web’, which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The ‘intelligent agents’ people have touted for ages will finally materialize.”*

Grundsätzlich gibt es zwei Möglichkeiten, die sich mit der Lösung dieses Problems beschäftigen. Die Idee besteht darin, die Informationen der Maschine auf eine Art zur Verfügung zu stellen, die es ihr erlaubt auf semantischer Ebene damit zu arbeiten. Das Web muss die Bedeutung der vorhandenen Informationen verstehen und sie automatisch verarbeiten können.

Eine Herangehensweise ist die künstliche Intelligenz, welche die kognitiven Aufgaben, die ein Mensch bei der Verarbeitung von Informationen aus dem Web bewältigt, in maschinenlesbare Daten umwandelt. Der Computer soll also die Leistungen des menschlichen Gehirns nachbilden. Bisher ist es aber noch nicht gelungen, eine künstliche Intelligenz für diesen Zweck effektiv und zuverlässig einzusetzen.

Die andere, und weitaus vielversprechendere, Herangehensweise ist das Semantic Web.

### 3.1 Was ist das Semantic Web?

Das Semantic Web unterscheidet sich insofern von der künstlichen Intelligenz, dass es die Informationen im Vorfeld auf eine Art und Weise zur Verfügung stellt, die deren Verarbeitung durch Maschinen ermöglicht. Es stellt also Methoden zur Verfügung, mit denen das Web Informationen verarbeiten und verknüpfen kann. Durch Semantic Web können komplexe Systeme maschinell erfasst und abgebildet werden. Die Bedeutung der einzelnen Informationen ist mitgespeichert und kann ausgelesen werden, was auch das Implizieren von neuen Informationen ermöglicht.

## 3.2 Wie funktioniert das Semantic Web?

Die Fragen, die man sich an dieser Stelle stellt, lauten: Wie funktioniert das Ganze? Wie kann die Bedeutung von Informationen maschinenlesbar erfasst werden? Wie können Relationen abgebildet werden? Wie können nicht vorhandene Informationen impliziert werden?

Hierzu werde ich auf drei Begriffe näher eingehen:

1. Metadaten
2. Taxonomie
3. Ontologie

### 3.2.1 Metadaten

Metadaten sind Daten, die Informationen über andere Daten enthalten. Das Prinzip der Metadaten ist, im Vergleich zum Begriff selbst, keineswegs neu. Sie liefern Informationen über ein Dokument, wie etwa Autor, Titel oder Zeitpunkt der Veröffentlichung Informationen über ein Buch sind. Hier wäre das Buch das eigentliche Dokument. Autor, Titel und Zeitpunkt der Veröffentlichung wären die Metadaten.

Eine allgemeingültige Unterscheidung von Daten und Metadaten ist nicht immer gewährleistet. Oft ist es eine Frage des Standpunkts. Autor und ISBN-Nummer sind für den Leser eines Buches Metadaten, während der Inhalt die eigentliche Information darstellt. Für einen Herausgeber eines Bücherkatalogs wären Autor und ISBN-Nummer allerdings wichtige Informationen.

Metadaten können auf zwei verschiedene Arten gespeichert werden.

1. Im Dokument selbst: Titel und Autor eines Buches stehen im Normalfall direkt auf dem Buch, werden also im Dokument selbst gespeichert.
2. In zugeordneten Nachschlagewerken: Die Metadaten eines Buches können zum Beispiel auch in einem Bücherkatalog der Bibliothek hinterlegt werden.

Wichtig für die Realisierung des Semantic Web ist der Begriff der Interoperabilität. „Inter“ heißt übersetzt „zwischen“ und „operabel“ bedeutet soviel wie „es kann damit gearbeitet werden“. Interoperable Metadaten sind also Metadaten aus verschiedenen Datenquellen, mit denen aber gemeinsam gearbeitet werden kann. Dies ist wichtig, weil im Web viele Informationen aus vielen verschiedenen Quellen zur Verfügung stehen. Nur wenn die Interoperabilität gewährleistet ist, können Relationen zwischen Informationen hergestellt werden, was eine Grundvoraussetzung des Semantic Web darstellt.

Durch die einheitliche, genormte Speicherung von Metadaten, sogenannten Standards, kann die Interoperabilität von Metadaten sichergestellt werden. Im Semantic Web werden Metadaten mit Hilfe des „Resource Description Framework“ (RDF) organisiert und strukturiert. RDF basiert auf sogenannten „triples“, welche eine Erweiterung der klassischen Schlüssel-Werte-Paare, wie sie beispielsweise von HTML Meta-Tags verwendet werden, darstellen. Ein Schlüssel-Werte-Paar wäre zum Beispiel „Tom, Mannheim“. Hier wären „Tom“ der Schlüssel und „Mannheim“ der Wohnort. Ein „tripel“, wie RDF es verwendet, ist in Subjekt, Prädikat und Objekt aufgeteilt. Es kann, im Gegensatz zu Schlüssel-Werte-Paaren, semantische Beziehungen zwischen zwei Objekten herstellen. Ein Beispiel dafür wäre: „Mannheim ist Heimatstadt von Tom“. Hier wäre „Mannheim“ das Subjekt, „ist Heimatstadt von“ das Prädikat und „Tom“ das Objekt.

### 3.2.2 Taxonomie

Eine Taxonomie ist eine Klassifikation mit monohierarchischer Struktur. Jeder Klasse wird genau eine Oberklasse zugeordnet. Bildlich kann man sich eine Taxonomie wie eine Baumstruktur vorstellen. Jede Klasse erbt allgemeine Eigenschaften ihrer Oberklasse.

Taxonomien werden in der Informatik schon seit langer Zeit benutzt. Zum Beispiel kann in einer Bibliothekssoftware die Wurzel (= oberste Klasse) das Gebäude sein. Unterklassen des Gebäudes sind die einzelnen Stockwerke, welche wiederum

einzelne Regale als Unterklassen haben. Die unterste Schicht sind die einzelnen Bücher in den Regalen (siehe Abb. 3).

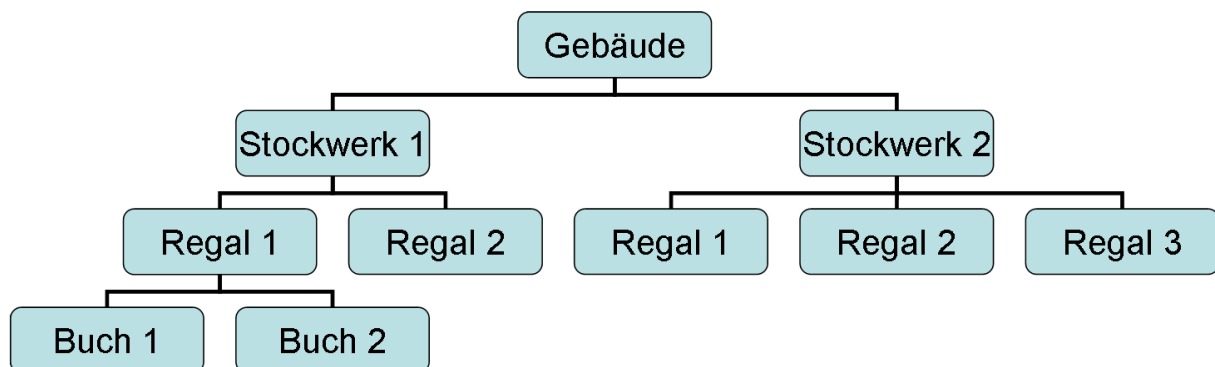


Abbildung 3: Taxonomie Bibliothek

Es gibt jedoch auch vernetzte Hierarchien, welche nicht durch Taxonomien abgebildet werden können.

Stellen wir uns folgende Taxonomie vor (siehe Abb. 4):

Die Erde hat die Kontinente als Unterklassen, welche die Oberklassen der einzelnen Länder darstellen. Jetzt werden aber in manchen Ländern mehrere verschiedene Sprachen gesprochen. Ebenso werden aber manche Sprachen in mehreren verschiedenen Ländern gesprochen.

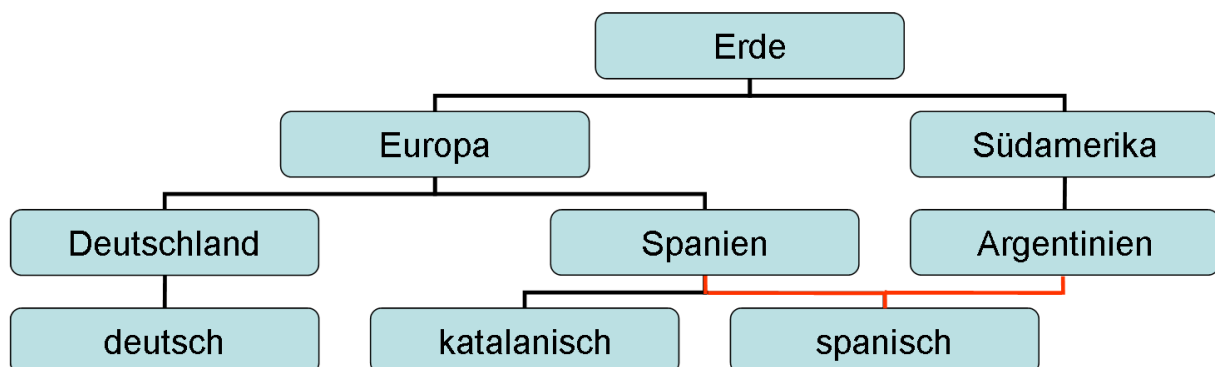


Abbildung 4: Falsche Taxonomie

Hier kommen wir mit einer Taxonomie nicht weiter. Wir brauchen ein komplettes Netz, um diesen Tatbestand abzubilden. Die Lösung hierfür ist die Ontologie.

### 3.2.3 Ontologie

Eine Ontologie kann, im Gegensatz zur Taxonomie, ein komplettes Netzwerk von Informationen abbilden. In diesem Netz können Informationen über logische Beziehungen verknüpft werden.

Eine Ontologie setzt sich aus folgenden Bestandteilen zusammen:

1. **Begriffe:** Begriffe sind Beschreibungen gemeinsamer Eigenschaften. Sie können auch als Klassen bezeichnet und in einer Klassenstruktur angeordnet werden.  
(Bsp.: Stadt, Land (wobei Land Oberklasse von Stadt wäre))
2. **Instanzen:** Instanzen sind Objekte in der Ontologie. Sie werden anhand von Begriffen erzeugt.  
(Bsp.: Mannheim ist Instanz von Stadt, Deutschland ist Instanz von Land)
3. **Relationen:** Relationen beschreiben Beziehungen zwischen verschiedenen Instanzen.  
(Bsp.: Mannheim liegt in Deutschland)
4. **Vererbung:** Eigenschaften von Begriffen und Relationen können vererbt werden, wobei Mehrfachvererbung möglich ist.  
(Bsp.: Mannheim erbt die Eigenschaft „liegt in Europa“ von Deutschland)
5. **Axiome:** Axiome sind Aussagen, welche innerhalb der Ontologie zwingend wahr sind. Diese werden meistens dazu verwendet, um Informationen zu repräsentieren, welche nicht aus den Begriffen der Ontologie abgeleitet werden können.  
(Bsp.: „zwischen Amerika und Europa besteht keine Landverbindung“.)

In einer Ontologie sind Informationen also formal beschrieben und es gibt feste Regeln zu deren Zusammenhang. Deshalb kann durch Ontologien auch auf nicht eingegebene Informationen geschlossen, d.h. neue Informationen impliziert werden.

Wenn beispielsweise die Relationen „Deutschland liegt in Europa“, „Italien liegt in Europa“ und das Axiom „zwischen Orten innerhalb eines Kontinents besteht immer eine Landverbindung“ haben können wir auf die neue Eigenschaft „Zwischen Deutschland und Italien besteht eine Landverbindung“ schließen.

Im Semantic Web werden Ontologien durch die „Web Ontology Language“ (OWL) umgesetzt. OWL basiert auf der RDF-Syntax, ihre Ausdrucksmöglichkeiten gehen aber weiter über die des RDF hinaus.

## 4. Anwendungen des Semantic Web

Das Semantic Web kann in nahezu allen Bereichen des Internets eingesetzt werden. In diesem Kapitel gehe ich auf folgende Anwendungsmöglichkeiten näher ein:

1. Navigieren und Suchen
2. Soziale Netzwerke
3. Tagging

### 4.1 Navigieren und Suchen

In den Anfangszeiten des World Wide Webs gab es keine andere Navigationsmöglichkeit als das Browsen, das man umgangssprachlich auch als „surfen“ bezeichnet. Man klickt sich über Links durch die einzelnen Webseiten, was bei der heute vorliegenden Menge an Informationen natürlich undenkbar ist.

Schon bald gab es Suchmaschinen, die Linklisten nach Begriffen durchsuchten. Später entwickelte Yahoo! einen Verweiskatalog. Ein bedeutender Fortschritt gelang Google mit der Einbeziehung von Verweisen auf die Qualität von Links.

Zum durchsuchen bestimmter Teilgebiete des Webs gibt es heute sogenannte

The image shows a search interface for a forum. It is divided into several sections:

- Suche mit Schlüsselwörtern:** A text input field for keywords and a dropdown menu set to "Gesamten Beitrag durchsuchen".
- Suche nach Benutzernamen:** A text input field for usernames and a dropdown menu set to "Beiträge vom Benutzer suchen", with a checked checkbox for "Exakter Name".
- Suchoptionen:** A section containing:
  - Suche Themen mit Präfix:** A list box with "Gaming" selected, and other options like "Einheitenvorstellung", "News", "Poker", and "BLOG".
  - Suche Themen mit:** A dropdown set to "mindestens" and a text input "0" followed by "Antworten".
  - Suche nach Beiträgen:** A dropdown set to "jeden Datums" and a dropdown set to "und neuer".
  - Sortiere Treffer nach:** A dropdown set to "letztem Beitrag" and another set to "in absteigender Sortierung".
  - Treffer anzeigen als:** Radio buttons for "Themen" (selected) and "Beiträge".
- Suche im Forum:** A list box with various forum categories like "Witzecke", "Reallife", "Nightlife", "Multimedia Forum", "Anime und Manga", "Sport Forum", "Bundesliga Tippspiel", "Technik Ecke", "Kaufberatung", "Informationen", "Häufig gestellte Fragen", "Feedback", and "Archiv". A checked checkbox "Unterforen ebenfalls durchsuchen" is at the bottom.

At the bottom of the interface are two buttons: "Suchen" and "Felder zurücksetzen".

Abbildung 5: Kreuzkatalog

Kreuzkataloge (siehe Abb. 5), mithilfe derer man genauer angeben kann wo, und was man suchen will. Diese werden oft in Diskussionsforen verwendet. Man kann die Suche zum Beispiel nach Erstellungsdatum, Autor und Themengebiet eingrenzen. Damit dies funktionieren kann, müssen zu allen Themen und Beiträgen mit deren Erstellung automatisch Metadaten, wie eben Autor und Erstellungsdatum, hinterlegt werden.

Auch eine semantische Suchmaschine basiert auf der Nutzung von Metadaten. Im Gegensatz zu Suchmaschinen wie etwa Google lässt sie Eingaben in natürlicher Sprache zu. Google ignoriert Eingabewörter wie „wann“ oder „wie“ nahezu komplett, während eine semantische Suchmaschine gerade auf diese Eingaben Wert legen muss, da sie der Frage einen komplett anderen Sinn geben können. Beispielsweise sollte eine semantische Suchmaschine auf die Fragen „Wann/Wie oft war Deutschland Fußballweltmeister?“ komplett verschiedene Ergebnisse liefern, während eine gewöhnliche Suchmaschine keine der beiden Fragen richtig beantworten würde. Um dies zu realisieren, muss sie Beziehungen zwischen den Fragewörtern und der gesuchten Information herstellen.

Die Aufgabe des Semantic Web liegt hier also darin, die Summe aller verfügbaren Informationen als System zu verstehen. Auf Anfrage liefert es die gesuchte Information oder impliziert diese Information, wenn möglich, falls sie noch nicht vorhanden ist. Bei oben genanntem Beispiel könnte man etwa aus der Information, wann Deutschland Weltmeister war auf das wie oft schließen, indem man die Anzahl der unterschiedlichen Ergebnisse zusammenzählt. Um auf diese Lösung zu kommen, muss die Suchmaschine allerdings wissen, dass es sich bei 1954, 1974 und 1990 um Jahreszahlen, also Datumsangaben handelt.

Es gibt inzwischen schon einige semantische Suchmaschinen. Allerdings sind diese noch nicht soweit ausgereift, dass sie mit den führenden Suchmaschinen mithalten können.

Beispiele für semantische Suchmaschinen sind „powerset“ ([www.powerset.com](http://www.powerset.com)) oder „Start“ ([start.csail.mit.edu](http://start.csail.mit.edu)). Beide versuchen die menschliche Sprache sinnvoll zu interpretieren und auf Fragen direkt zu Antworten. Sie geben zum Beispiel korrekte Antworten auf die Frage: „How many people live in Germany?“.

## 4.2 Social Software und soziale Netzwerke

Social Software behandelt Menschen und Personengruppen als primäre Objekte innerhalb eines Systems und ermöglicht Kommunikation in sozialen Netzwerken. Formen von Social Software sind zum Beispiel Instant-Messaging-Programme, Diskussionsforen, Blogs, Wikis oder virtuelle Welten wie Second-Life.

Eine besondere Form von Social Software stellen Portale zur Kontaktverwaltung in sozialen Netzwerken dar, da sie mehrere dieser Formen zusammenführen. Bekannte Vertreter sind MySpace, Facebook und StudiVZ.

Da bei der Social Software der Mensch im Mittelpunkt steht, lebt diese natürlich von den Daten, mit denen sie von den Usern gefüttert werden. Die Aufgabe des Semantic Web liegt hier darin, die Informationen zu erfassen, aufzubereiten und zu präsentieren.

StudiVZ zum Beispiel beschreibt alle Kontakte einer Person als „Freund“. Hier kann man nicht herauslesen, ob der „Freund“ ein sehr guter Freund oder nur ein flüchtiger Bekannter ist. Könnte man hier genauere Angaben machen, könnte StudiVZ etwa die Verflechtungen der Kontakte analysieren und Textzeilen wie „Morgen hat ihr alter Schulfreund Tom Geburtstag“ ausgeben, obwohl man Tom vielleicht gar nicht mehr kennt und nicht zur eigenen Freundesliste hinzugefügt hat.

Ebenso könnte man falsche Angaben filtern, indem man etwa die Angaben zu früher besuchten Schulen mit der Liste der alten Schulfreunde vergleicht.

## 4.3 Tagging

„Tag“ ist englisch und bedeutet „Anhänger oder „Etikett“. Tags sind Schlagwörter, die dazu dienen, Informationen wie Texte oder Bilder zu kategorisieren. Sie können auch mit den Informationen assoziiert werden. Das Bild eines Fernseherers könnte zum Beispiel mit „Fernseher“, „TV“, aber auch mit „Film“ getaggt werden. Ebenso könnten „Gitarre“ und „Musik“ tags eines Textes über den Gitarristen Carlos Santana sein. Wichtig hierbei ist, dass jede Information eine Vielzahl an tags besitzen kann.

Bekannte Internetseiten, die das tagging vorangetrieben haben sind zum Beispiel „flickr“ (flickr.com), eine Online-Bilderverwaltung und „Delicious“ (del.icio.us), eine Webanwendung für Social Bookmarking (Linksammlungen zu einzelnen Schlagwörtern).

Heute werden tags häufig in der Blog-Szene verwendet. Man fügt zu den einzelnen Artikeln der Blogs tags hinzu. Diese tags dienen dann zum einen dem Auffinden verwandter oder ähnlicher Artikel innerhalb des eigenen Blogs. Zum anderen kann aber auch von externen Websites darauf zugegriffen werden. Das Paradebeispiel hierfür ist Technorati ([www.technorati.com](http://www.technorati.com)), eine Suchmaschine speziell für Weblogs.

## 5. Fazit

Obwohl seit 1999, als Tim Berners-Lee den Begriff des Semantic Web einführte, beachtliche Fortschritte erzielt wurden, steht das Semantic Web noch am Anfang. Die technischen Grundlagen sind bereits gelegt, die Realisierung der Vision hat aber gerade erst begonnen. Die Zukunft des Semantic Web ist noch gestaltbar.

Wie weitreichend die Auswirkungen letztendlich sein werden steht noch in den Sternen. Die Möglichkeiten jedoch, soviel ist bereits klar, sind atemberaubend. Durch das Semantic Web wird es in naher Zukunft vielleicht zum ersten Mal möglich sein, das geballte Wissen der Menschheit auf produktive Weise zu filtern und zu kombinieren.

## Quellen- und Literaturverzeichnis

### Literatur:

Pellegrini, T (Hrsg.) + Andreas Blumauer (Hrsg.):

Semantic Web. Wege zur vernetzten Wissensgesellschaft, Springer Verlag (2006)

Berners-Lee, T + Fischetti, M:

Weaving the Web. Origins and Future of the World Wide Web (1999)

Jiles, J:

Special Report Internet encyclopaedias go head to head, Nature 438, 900-901 (2005)

### Internet:

Universität Leipzig: Blogstudie 2007. Informationssuche im Internet – Blogs als neues Recherchetool (Ergebnisbericht) (09.02.2007)

[http://www.blogstudie2007.de/inc/blogstudie2007\\_ergebnisbericht.pdf](http://www.blogstudie2007.de/inc/blogstudie2007_ergebnisbericht.pdf), 13.10.2008

Sifry, D: The State of the Live Web (05.04.2007)

<http://technorati.com/weblog/2007/04/328.html>, 13.10.2008

Schwarz, O: Ein Volkssport namens tagging

<http://hasematzel.de/blog/2006/05/06/ein-volkssport-namens-tagging>, 13.10.2008

Deutsche Wikipedia: Vergleich einiger Enzyklopädien

<http://de.wikipedia.org/wiki/Wikipedia:Größenvergleich>, 13.10.2008

Deutsche Wikipedia: Semantik

<http://de.wikipedia.org/wiki/Semantik>, 13.10.2008

Deutsche Wikipedia: Semantisches Web

[http://de.wikipedia.org/wiki/Semantisches\\_Web](http://de.wikipedia.org/wiki/Semantisches_Web), 13.10.2008

Deutsche Wikipedia: Ontologie (Informatik)

[http://de.wikipedia.org/wiki/Ontologie\\_\(Informatik\)](http://de.wikipedia.org/wiki/Ontologie_(Informatik)) , 13.10.2008