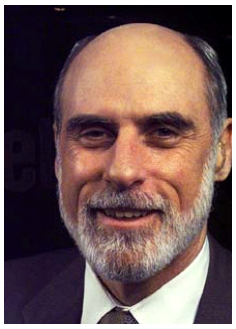
Modul 5: Semantik im WWW



Vinton Cerf (geb. 1943), einer der Väter des Internets [W1]

„The Internet will become a repository of knowledge, not only a compendium of facts."

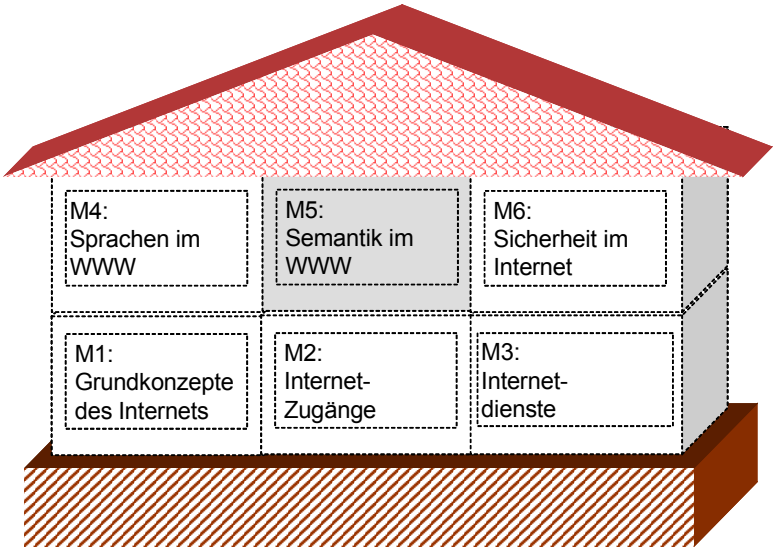
Lernziele

Interoperabilität und Metadaten werden immer wichtiger auf dem Weg zu einem **semantischen Web.** Ontologien stellen formale, semantische Modelle dar, die den Austausch von „Wissen" zwischen Maschine und Maschine und natürlich besonders zwischen Mensch und Maschine erleichtern sollen. Suchmaschinen schließlich sind heute praktisch unverzichtbare Werkzeuge in der **Erschließung des Internets.**

Metadaten

Ontologien

Suchmaschinen



Information im Internet ist heute zwar maschinenlesbar, aber sie ist nicht maschinenverarbeitbar. **Suchmaschinen** können nur auf der lexikalischen Ebene arbeiten. Somit bleibt ihnen der *semantische Zusammenhang* verborgen. Dieser Nachteil wird bei jeder Suchanfrage in das WWW immer deutlicher: meistens zigtausend Treffer, die dann die informationssuchende Person auf *relevante* Information „manuell" untersuchen muss (Bild 5.1). Außerdem ist es prinzipiell nicht möglich, Dienste auf Basis der im Web vorhandenen Informationen anzubieten, da den Maschinen der Zusammenhang, die Semantik, verborgen bleibt. Der Ansatz des **Semantic Web** versucht nun diese Schwachstelle zu beheben. Es wird versucht, semantische Information maschinenlesbar und maschinenverarbeitbar zur eigentlichen Information (Inhalt, Content) hinzuzufügen. Um diese Ziele zu erreichen, baut das Semantic Web auf bereits existierende Standards auf und adaptiert etablierte Methoden der Wissensrepräsentation. Im Gegensatz zu speziellen Systemen der Wissensrepräsentation, denen oftmals eine geschlossene und sorgfältig gewartete Wissensbasis zugrunde liegt, ist das WWW eine weltweite, heterogene, mehrsprachige und vor allem **offene „Wissensbasis".** Um in und mit einer solchen offenen Wissensbasis zu arbeiten, sind akzeptierte und verbreitete **Standards** unbedingte Voraussetzung.

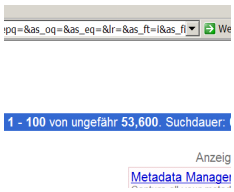


Bild 5.1 Wer soll diese 53.600 Suchergebnisse händisch durcharbeiten? Suchwort: „Metadaten" am 9. 12. 2003

Metadaten sind zur Lösung solcher Probleme ein unverzichtbarer Ansatz. Metadaten existieren prinzipiell, seit der erste Bibliothekar eine Liste über seine bestehenden Pergamentrollen anlegte (Bild 5.2). Heute stellt – neben der Fülle existierender Metadaten-Konzepte (wie z.B. Dublin Core, RDF, PICS, GEM, Warwick-Framework oder IMS) – der speziell zur Verwaltung und Auffindung von Lernmaterial („Content") vom Institute of Electrical and Electronics Engineers (IEEE), USA, vorgeschlagene Standard IEEE P1484.12/D6.1 LOM einen interessanten Ansatz dar.

Bild 5.2 Listen über vorhandene Pergamentrollen enthielten die ersten Metadaten

1 Metadaten

Metadaten („Daten über Daten“) sind Informationen über Daten und Objekte und dienen der Beschreibung und *eindeutigen* Identifizierung von Daten und Objekten zur Sicherung der Interoperabilität.

1.1 Interoperabilität

Interoperabilität ist die Möglichkeit der Austauschbarkeit von Inhalten, das heißt allgemein die Fähigkeit, Informationen über gemeinsam nutzbare Datenformate (Austauschformate) zu nutzen.

Ein Beispiel soll das deutlich machen: Die NASA archivierte sämtliche Daten der letzten 30 Jahre, und dennoch sind diese Daten völlig wertlos: Niemand hat nämlich die Millionen Magnetbänder *systematisch* **katalogisiert** und damit **suchfähig** gemacht.

Das Konzept der Metadaten existiert unabhängig vom Internet. Metadaten werden in den verschiedensten Bereichen verwendet, um die Daten und Objekte inhaltlich und formal zu beschreiben und zu strukturieren. Beispielsweise werden in Bibliotheken Metadaten in Form von Katalogkarten oder Titelaufnahmen verwendet und dienen zur Wiederfindung von Medien in Katalogen, Datenbanken und Magazinen.

Neben traditionelle Medien tritt heute das Internet als Informationsquelle. Oft existieren Inhalte sogar *nur* mehr im Internet. Damit stellt das Internet quasi eine *digitale Erweiterung* vorhandener Sammlungen dar. Bibliotheken bemühten sich dabei stets um eine Erschließung des Wissens. Dabei taucht sofort die Frage auf, *wie* diese Internetobjekte erschlossen werden sollen und *wer* diese Erschließung vornehmen soll.

Während in Bibliotheken Medien in angreifbarer Form vorliegen – sei es als Buch, CD-ROM oder Mikrofiche – existieren im Internet verschiedene Objekte (Websites, Webdokumente, Datenbanken, visuelle Objekte usw.) ausschließlich in elektronischer Form.

> Objekte im Internet werden als **elektronische Dokumente** bzw. **D**ocument **L**ike **O**bjects (DLO) bezeichnet.

Search Engines

Um das Internet zu erschließen, werden unter anderem Suchmaschinen (Kapitel 3) verwendet, die aber bei der Informationssuche viele Probleme aufwerfen (zu große Treffermengen, zu geringe Präzision). Um (einige) diese Probleme zu überwinden, sollen Metadaten zur Beschreibung dieser Objekte verwendet werden.

1.2 Grundlagen von Metadaten

Metadaten können prinzipiell definiert werden als Informationen *über* andere Daten (Dokumente, Bilder, Programme usw.).

Metadaten sollen als *Sekundärdokumente* die Recherche, das Retrieval und die Nutzung der *Primärdokumente* ermöglichen bzw. erleichtern.

Metadaten können sowohl für elektronische Dokumente als auch für dokumentenähnliche Objekte (DLOs) verwendet werden. Natürlich können Metadaten auch physisch vorhandene Objekte (Bücher, Gegenstände usw., die selbst nicht elektronisch gespeichert werden) beschreiben.

> Inhaltlich verbundene, aber heterogene Ressourcen zu einem Thema können durch Metadaten erschlossen werden.

Ziel von Forschung und Entwicklung im Bereich Metadaten ist es, eine Interoperabilität verschiedener Metadatenformate und verschiedener technischer Systeme (Datenbanktypen, Computerplattformen usw.) zu erreichen.

Die Interoperabilität verschiedener Metadatenformate ist ein wichtiger Gesichtspunkt, da es sehr viele verschiedene Metadatenformate (siehe nächste Kapitel) und maschinelle Austauschformate für Bibliotheken gibt.

Neben generellen Fragen zur Verwendung von Metadaten zur Erschließung des Internets taucht sofort die Frage auf, *wer* diese Metadaten vergeben und damit die Beschreibung von Internetobjekten vornehmen soll. Aufgrund der enormen Größe des Internets und der darin enthaltenen riesigen Menge an Objekten erscheint eine Erschließung durch Bibliotheken oder spezielle Dokumentationseinrichtungen praktisch unmöglich.

Zur Vergabe von Metadaten existieren zwei unterschiedliche Meinungen: Autoren versus Indexierer. Manche argumentieren, dass die Beschreibung des Internetdokuments durch den **Autor** vorteilhaft ist, da dieser den Inhalt und die Intention des Dokuments besser kennt und besser versteht als ein Außenstehender und diesen damit am besten beschreiben kann. Gegner meinen allerdings, dass die Autoren nicht über das notwendige Wissen über Indexierungstechniken und auch nicht die notwendige Distanz zu ihren Inhalten verfügen, um gute Metadatensätze vergeben zu können.

> Prinzipiell sind Autoren von Internetdokumenten sehr daran interessiert, ihre Objekte durch Verwendung von Metadaten *aufzuwerten* und für Suchmaschinen *besser auffindbar* zu machen.

Allerdings sind die meisten Metadatenformate so kompliziert, dass nur Experten damit umgehen können. Deshalb benötigen Autoren einfach strukturierte Metadaten oder Werkzeuge, die sie bei der Generierung von Metadaten mit standardisierten, benutzerfreundlichen Eingabemasken unterstützen (z.B. LO-Editor, siehe HOLZINGER et al. (2003)).

Im Internet unterscheiden wir zwei unterschiedliche Arten von Metadaten:

- unstandardisierte Suchmaschinen-Metadaten *(Meta-Tags)* entsprechend den verschiedenen Suchmaschinenanbietern;
- standardisierte Metadaten (z.B. Dublin Core usw.)

Prinzipiell besteht ein *Metadaten-Record* aus einem **Satz (set)** verschiedener **Attribute (attributes),** die das zu beschreibende Dokument spezifizieren.

Das einfachste Beispiel ist ein Bibliotheks-Katalog (library catalog), der aus Elementen wie z.B. Autor, Jahr, Titel, Fachgebiet, Standort usw. besteht.

Für interoperable Metadaten sind sowohl Syntax als auch Semantik wichtig:

Die **Syntax** beschreibt, in welcher Form Metadaten ausgetauscht werden können. Dafür eignet sich z.B. XML (Extensible Markup Language, siehe Modul 4) hervorragend.

Die **Semantik** hingegen beschreibt, welche Metadaten für eine Ressource eingesetzt werden können (bzw. auch dürfen). Dafür eignen sich eigene Vokabulare bzw. Schemas.

Suchmaschinen für das WWW basieren auf einer Schlüsselwortsuche in Volltexten (siehe Kapitel 3.3). Es wurde erst sehr spät und mit begrenztem Erfolg versucht, eine *systematische Auszeichnung mit Metadaten* zu bewirken und für die Informationssuche im WWW zu nutzen.

Von den Ansätzen, die bis jetzt entstanden sind, werden im Folgenden nur einige der wichtigsten angesprochen. Die Geschichte von standardisierten Metadaten begann eigentlich mit den HTML-Tags (siehe folgendes Kapitel).

Einführende Literatur zu Metadaten z.B. RUSCH-FEJA (1997).

Englische Einführungen z.B. CAPLAN (2003), HAYNES (2004).

1.2.1 HTML-Meta-Tags

Eine grundlegende Form der Erschließung von Semantik von Ressourcen ist die Auszeichnung mit **inhaltsorientierten Metadaten.**

Das vornehmlich als Dokumentenstruktursprache konzipierte HTML sah von Beginn an solche Elemente vor, die auf eine inhaltliche Beschreibung des gesamten Dokumentes abzielten. Bereits 1992 wird das TITLE-Tag als Attribut einer HTML-Seite erwähnt, das den Inhalt dieser Seite mit Hilfe einer Textzeile beschreiben soll. Ab etwa 1993 wurden so genannte **Meta-Tags** verwendet, die im Header eines HTML-Dokuments untergebracht werden. Der Inhalt wird in Form von *Attribut-Wert-Paaren* nach folgendem Beispiel charakterisiert (Bild 5.3):

```
<meta http-equiv="Content-Language" content="de">
<meta name="Abstract" content="Inhalt der Lehrbuchreihe Basiswissen Multimedia">
<meta name="Author" content="Andreas Holzinger, andreas.holzinger@uni-graz.at">
<meta name="Content-Type" content="text/html; charset=iso-8859-1">
<meta name="Copyright" content="Dr.Andreas Holzinger, 2000">
<meta name="keywords" content="Multimedia, Informationssysteme, Neue Medien"
```

Bild 5.3 Einfachstes Beispiel: HTML-Meta-Tags

Diese Meta-Tags wurden 1995 in den HTML-2.0-Standard aufgenommen, wobei ausdrücklich erwähnt wurde, dass es sich um einen „erweiterbaren Container für Meta-Informationen" handelt, der zur „Dokumentation des Inhalts, der Qualität und der Eigenschaften eines Datensatzes dient".

Erst in HTML 4.0 (Dezember 1997) ist vorgesehen, dass die Bedeutung einer Eigenschaft und die Menge der dieser Eigenschaft zugehörenden Werte in einem **Referenzlexikon (profile)** definiert werden.

Dies geschieht, indem der Universal Resource Identifier (URI) des profiles im Header des HTML-Dokuments referenziert wird. Damit erhält es Gültigkeit für das gesamte Dokument. Daneben wird mit der Einführung des Attributs „scheme" innerhalb eines Meta-Tags die Möglichkeit der Angabe eines Kontextes für Meta-Tag-Werte geschaffen, der die korrekte Interpretation dieser Werte erleichtern soll.

Beispiel:
<meta scheme="ISBN" name="identifier" content="3-8023-1899-4">

Zur Literatur über Meta-Tags siehe jedes HTML-Buch, z.B. NIEDERST (2002), BORN (2003).

1.2.2 Dublin Core

Die HTML-Meta-Tags ermöglichen zwar, HTML-Dokumente mit den Metadaten auszuzeichnen, der Nutzen wird jedoch dadurch stark relativiert, dass weder die Menge der möglichen Metadatenelemente noch deren Semantik definiert sind.

Der Ansatz der **D**ublin **C**ore **M**etadata **I**nitiative (DCMI) geht bei diesen Punkten einen entscheidenden Schritt weiter. Ziel dieser Initiative war es zunächst, eine **Kernmenge semantischer Beschreibungselemente** (Dublin Core Metadata Element Set, kurz: Dublin Core) für webbasierte Ressourcen zu schaffen.

Im März 1995 wurde zu diesem Thema in Dublin (Ohio, USA) ein erster Workshop abgehalten, wo 13 Elemente als „Kern" festgelegt wurden; daher der Name **Dublin Core (DC).**

Es zeichnete sich jedoch schon nach kurzer Zeit ab, dass diese Menge von 13 Beschreibungselementen den Anforderungen von unterschiedlichen Objekttypen und verschiedenen Communities an einen Metadatenstandard *nicht* entsprechen und somit nicht als alleiniger Standard für das Web dienen können. Daher ist es mittlerweile das Ziel von Dublin Core, lediglich einen gemeinsamen semantischen *Kern* der in den verschiedenen Disziplinen fortexistierenden *spezifischen* Metadatenstandards zu repräsentieren und eine übergreifende Suche zu ermöglichen.

Dublin Core dient mittlerweile als der kleinste gemeinsame Nenner *verschiedener* Standards, um *semantische Interoperabilität* zwischen diesen zu ermöglichen.

Zentral für den Entwurf von DC waren genau fünf Voraussetzungen:

- **Einfache Handhabung:** Die Beschreibungselemente wurden bewusst einfach gehalten, um eine größere Akzeptanz im Web zu ermöglichen. Es sollte jeder „auszeichnen" können – ohne Fachwissen oder spezielle Ausbildung zu benötigen.
- **Ausrichtung auf Resource Discovery:** Das Hauptziel war und ist es, die Suche nach Objekten im WWW zu erleichtern. Dublin-Core-Metadaten dienen in erster Linie der *Identifikation* und der *Inhaltsbeschreibung*. Andere (generelle) Aspekte von Metadaten (wie strukturelle Aspekte, Kontext usw.) werden daher nicht oder nur teilweise unterstützt.
- **Beschreibung von Document Like Objects:** Dublin Core wurde ganz primär zur Beschreibung von DLOs entwickelt; es wurde davon ausgegangen, dass dies der vorherrschende Typ im Internet sein wird.

- **Erweiterbarkeit:** Die Kernelemente stellen nur eine *partielle Beschreibung* der Objekte dar. Dublin Core wurde von Beginn an so konzipiert, dass eine Erweiterung in Anwendungsgebieten, die *mehr Attribute* benötigen, möglich sein sollte.
- **Internationaler Konsens:** Die Entwicklung erfolgte von Anfang an stets unter der Beteiligung einer breiten internationalen Community. Der Kernstandard wurde bereits in zahlreiche Sprachen übersetzt.

Die Bedeutung der Elemente ist in der ISO/IEC 11179/1-6 festgelegt: Die Syntax unterscheidet auch „Eigenschaften" und deren „Werte" (Bild 5.4):

```
<meta name="DC.Description" content="Inhalt der Lehrbuchreihe Basiswissen
Multimedia">
<meta name="DC.Creator" content="Andreas Holzinger, andreas.holzinger@uni-graz.
<meta name="DC.Type" content="text">
<meta name="DC.Rights" content="Copyright by Dr.Andreas Holzinger, 2000">
<meta name="DC.Subject" content="Multimedia, Informationssysteme, Neue Medien"
```

Bild 5.4 Beispiel für DC-Metadaten

Die 15 Elemente mit den derzeitigen Erläuterungen nach dem englischen Vorbild (in den Klammern steht der englische Begriff des Elements (Label) des Dublin Core Element Reference Set) umfassen (Bild 5.5):

1. Titel bzw. Name der Ressource (DC.TITLE)
2. Verfasser oder Urheber (DC.CREATOR)
3. Thema und Stichwörter (DC.SUBJECT); kann systematische Daten nach einer Klassifikation (SCHEME) enthalten, wie z.B. Library-of-Congress-Klassifikationsnummer oder Begriffe aus anerkannten Thesauri (wie MEdical Subject Headings (MESH) oder Art-and-Architecture- Thesaurus-(AAT)-Deskriptoren enthalten.
4. Inhaltliche Beschreibung (DC.DESCRIPTION); Abstract, Kurzbeschreibung.
5. Verleger, Herausgeber (DC.PUBLISHER)
6. Co-Autoren (DC.CONTRIBUTORS)
7. Datum (DC.DATE); empfohlener Eintrag des Datums ist eine achtstellige Zahl (JJJJMMTT), wie es in ANSI X3.30-1985 definiert ist (eine anderes Format sollte mit SCHEME spezifiziert werden).
8. Ressourcenart (DC.TYPE); Art der Ressource, z.B. Homepage, technischer Bericht, Essay usw.
9. Format (DC.FORMAT); datentechnisches Format der Ressource eingetragen, z.B. Text/HTML, ASCII, Postscript-Datei usw.
10. Ressourcen-Identifikation (DC.IDENTIFIER); eine Zeichenkette oder Zahl, die diese Ressource eindeutig identifiziert. Beispiele für vernetzte Ressourcen sind URLs, ISBN (International Standard Book Number).
11. Quelle (DC.SOURCE); Herkunft, Originaldatei usw.
12. Sprache (DC.LANGUAGE); dreistelliger Code nach ANSI/NISO Z39.53-200X, z.B. ger, eng, usw.
13. Beziehung zu anderen Ressourcen (DC.RELATION)
14. Räumliche und zeitliche Maßangaben (DC.COVERAGE); z.B. geografische Koordinaten.
15. Rechtliche Bedingungen (DC.RIGHTS); Urheberrechtsvermerk.

Bild 5.5 Die 15 Elemente des DC Element Reference Set

Alle Eigenschaften sind optional und wiederholbar; die Reihenfolge der Angabe der Eigenschaften ist beliebig. Die Einbettung in HTML erfolgt mit dem Meta-Tag, die Einbindung in XHTML und XML erfolgt mit RDF (siehe Bild 5.5).

Literatur über Dublin Core am besten über http://dublincore.org .

1.2.3 Resource Description Framework (RDF)

Das **R**esource **D**escription **F**ramework (RDF) wurde vom W3C entwickelt und stellt eine *Infrastruktur* dar, um Codierung, Austausch und die Wiederverwendung von Metadaten zu ermöglichen. Das Datenmodell von RDF (siehe Bild 5.6) besteht aus drei Elementen:

1. Resource kann praktisch alles sein, was einen URI (Uniform Resource Identifier, das ist der technisch korrekte Begriff von URL = Uniform Ressource Locator) haben kann, d.h. z.B. alle Web-Dokumente.

2. Property ist ein Attribut (Eigenschaft) einer Resource; diese kann selbst Resource sein und einen Namen haben, z.B. Autor oder Titel; diese kann wieder eigene Properties haben.

3. Statement ist die Beziehung zwischen einer Resource, einer Property und einem Wert; ein Wert kann eine Zeichenkette oder wieder eine Resource sein. Damit ist ein Tripel definiert (Beispiel: subject, predicate und object entspricht einer Resource, einer Property und einem Wert).

Das Datenmodell orientiert sich an der Aufgabe, Informationsressourcen standardisiert beschreiben zu können. Eine Ressource, die bezeichnende Eigenschaft und der beschreibende Inhalt (Wert) dieser Eigenschaft, bilden im RDF-Kontext eine Einheit, bestehend aus Subjekt, Prädikat und Objekt. Eine RDF-Aussage kann als **gerichteter Graph** dargestellt werden: Subjekte und Objekte sind Knoten, und jedem Prädikat (Eigenschaft) entspricht eine Kante, die vom Subjekt zum Objekt weist (Bild 5.6):

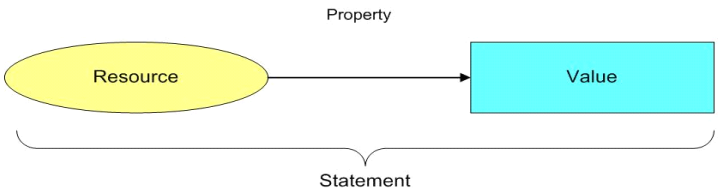


Bild 5.6 Beispiel für DC-Metadaten; solche RDF-Konstrukte werden als DLG (Directed labeled graph) bezeichnet

Als Sprache (Syntax) verwendet RDF die Extensible Markup Language (XML, siehe Modul 4).

RALPH SWICK *(W3C) und* ERIC MILLER *(OCLC) stellten im Sommer 1997 das RDF vor. Im Rahmen der gesamten Metadaten-Aktivitäten des W3C und Erfahrungen von Anwendern bei verschiedenen Digital-Library-Projekten wurde dieses Datenmodell als Verwirklichung des Warwick Framework (nach einer speziellen Konferenz in Warwick, UK benannt) entwickelt.*

RDF hat zum Ziel, die Interoperabilität zwischen *verschiedenen* Anwendungen zu sichern, die auf einen Austausch von Metadaten beruhen. Das ist z.B. nicht nur eine Informationssuche im Netz (resource discovery) oder die Katalogisierung von Webinhalten (cataloging), sondern auch e-Commerce-Anwendungen, Bewertung (content ranking), digitale Unterschriften (digital signatures) und Aspekte des Datenschutzes (privacy).

Was aber RDF nicht bietet ist, dass es selbst kein Vokabular für Metadaten definiert (wie z.B. Dublin Core); RDF ist nicht selbst durch eine XML DTD definiert, sondern direkt durch eine EBNF (Extended Backus-Naur-Form, siehe Band 2).

Bild 5.7 zeigt die Einbettung der DC-Deskriptoren (Kapitel 1.2.2) in das Resource Description Framework (RDF). PURL ist dabei die Abkürzung für Persistent Uniform Resource Locator (Bild 5.7):

```
<?xml version="1.0"?>
<rdf:RDF xmls:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:dc="http://purl.org/dc/about/index.htm">
<rdf:Description about="http://purl.org/dc/about/index.htm">
<DC.Description>Inhalt der Lehrbuchreihe Basiswissen Multimedia</DC.Description>
<DC.Creator>Andreas Holzinger, andreas.holzinger@uni-graz.at</DC.Creator>
<DC.Type>text</DC.Type>
<DC.Rights>Copyright by Dr.Andreas Holzinger</DC.Rights>
<DC.Subject>Multimedia, Informationssysteme, Neue Medien</DC.Subject>
</rdf:Description>
</rdf:RDF>
```

Bild 5.7 DC-Metadaten werden in RDF eingebunden

Einige Vorteile von RDF:

- stabiler Ansatz;
- Informationen können relativ zielgenau gefunden werden;
- auf RDF und RDF-SCHEMA aufbauend können Bayes'sche Wahrscheinlichkeitsnetze oder Fuzzy Logik neue Informationen generieren.

Einige Nachteile von RDF:

- eingeschränkte Richtung der Beziehungen zwischen den Ressourcen;
- Beziehungen müssen oftmals durch zwei Attribute neu eingegeben werden;
- die Mächtigkeit wird manchmal zu einem Nachteil.

Literatur zu RDF z.B. POWERS (2003), HJELM (2001).

1.2.4 RDF-Schema (RDFS)

RDF selbst bietet auch keine Möglichkeiten, die Beziehungen zwischen Eigenschaften und Ressourcen zu definieren und hier erforderlichenfalls Restriktionen aufzuerlegen. Um diese Beziehungen und Restriktionen zu deklarieren, werden so genannte **RDF-Schemas** (RDFS) benutzt. Das heißt, es wird z.B. definiert, welche Bedeutung „Creator“ hat und dass der Wert des Prädikates vom Typ „Person“ sein muss.

Die Schema-Vokabulare von RDFS sind einfache Ontologien (Kapitel 2) und definieren ein gültiges Prädikat in einer RDF Beschreibung und *Charakteristika* des Prädikatwertes selbst.

Das RDF-Schema wird ausgedrückt durch das RDF-Datenmodell (vgl. mit Bild 5.6). Der Aufbau des Schemas basiert auf einem Klassensystem und ist vergleichbar mit einer objektorientierten Sprache (Band 2). RDF definiert im Gegensatz zu den objektorientierten Programmiersprachen *nicht* die Eigenschaften der Instanzen einer Klasse. Ein RDF-Schema definiert *nur die Eigenschaften von Klassen einer Ressource,* zu der sie gehören (Bild 5.8):

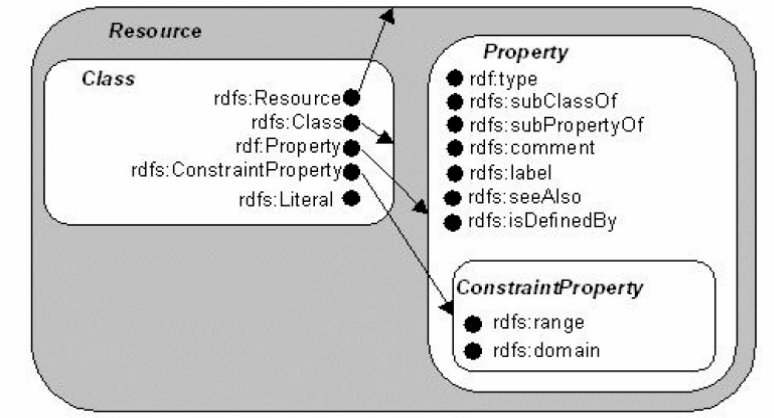


Bild 5.8 Klassen werden als abgerundete Rechtecke dargestellt

Klassen werden als abgerundete Rechtecke dargestellt. Eine Resource entspricht einem schwarzen Punkt. Die Pfeile definieren, welche Resource zu welcher Klasse gehört, bzw. welche Resource von welcher Klasse definiert wird. Subklassen werden innerhalb von Klassen dargestellt. In unserem Beispiel wäre „ConstraintProperty“ eine Subklasse von „Property“. In diesem objektorientierten Klassensystem des RDF-Schemas können unter anderem auch Einschränkungen definiert werden: Die Instanz rdfs:range der Klasse „ConstraintProperty“ wird eingesetzt, um die gültigen Werte der Eigenschaft einzugrenzen. Die Instanz rdfs:domain andererseits schränkt die Menge der Ressourcen ein, die eine Eigenschaft haben kann.

1.2.5 PICS

Beurteiler bewerten die Ausprägung bestimmter Merkmale anhand vorgegebener Kategorien (Ratingskalen)

Die Ursprünge von RDF gehen auf Arbeiten des W3C an der so genannten **P**latform for **I**nternet **C**ontent **S**election (PICS) im Jahre 1995 zurück. Bei PICS handelt es sich um ein **Rating-System.** Es ermöglicht, Inhalte von Ressourcen im WWW zu bewerten und diese Bewertungen – beispielsweise zum Schutz Minderjähriger beim Browsen – zu nutzen. Dabei sollten die Bewertungskriterien nicht allgemein und zentral fixiert, sondern durch jede Nutzergruppe individuell vergeben werden.

Es stellte sich während der Arbeit an PICS die Frage der Schaffung einer allgemeinen Architektur, mit deren Hilfe Ressourcen mit Metadaten versehen werden können. Diese Problemstellung wurde in Arbeitsgruppen des W3C unter Beteiligung von kommerziellen Anbietern und Vertretern der Metadaten-Communities aufgegriffen und resultierte u.a. in der W3C-Empfehlung zu RDF vom 22.02.1999. PICS kann somit eigentlich als ein Vorläufer von RDF gesehen werden.

Mit PICS kann eine *Bewertung (rating)* von Webseiten erfolgen – entweder durch „Self-rating“ oder „third-person-rating“.

Anhand dieser (subjektiven) Bewertung des Inhaltes erhalten die Seiten eine Kennzeichnung (label). Dadurch können z.B. Eltern ein Profil festlegen, ab wann die Übertragung einer Webseite unterbunden wird. Die inhaltliche Klassifizierung wird entweder von den Anbietern der Webseiten (Content Provider) selbst oder durch so genannte Label-Büros vorgenommen.

Entsprechende Filtersoftware (z.B. im MS Internet Explorer ab Version 4 integriert) kontrolliert das Label und lässt nur bestimmte (festgelegte) Seiten zu.

Ein PICS Label wird zwischen den Tags <head> und </head> in eine HTML-Datei eingefügt. Das vollständige Meta-Tag sieht so aus:

<META http-equiv="PICS-Label" content='labellist'>

Hier steht labellist für das Label, das das Dokument bewertet. Ein Beispiel:

<META http-equiv="PICS-Label" content='

(PICS-1.1 "http://www.rsac.org/ratingsv01.html"

l r (v 0 s 0 n 0 l 0))'>

Literatur zu PICS siehe www.w3.org/PICS

1.2.6 Gateway to Educational Materials (GEM)

Im Internet existiert sehr viel durchaus qualitativ hochwertiges Lehr- und Lernmaterial. Für die meisten Lernenden und vor allem für die Lehrenden ist aber nur mit großer Mühe relevantes Material auffindbar. Daher wurde vom U.S. Department of Education in Kooperation mit dem Educational Resources Information Center on Information Technology (ERIC/IT) ab 1995 das Projekt „**G**ateway to **E**ducational **M**aterials (GEM)“ gestartet.

Vorrangiges Ziel war es, das Verwertbarkeitsdefizit zu beheben und einen Zugang (ein „Gateway“) zu einer qualitativ hochwertigen Auswahl von Lehr- und Lernmaterial bereitzustellen. Insbesondere den Lehrenden soll das Finden von relevantem Lehrmaterial erleichtert werden. Das erfolgt durch Listen, die u.a. nach Oberbegriff, Schlüsselwörtern (keywords) und Ausbildungsstufe (Jahrgangsstufe) geordnet und direkt aus dem Gateway heraus zugänglich sind.

Es gibt ein *kontrolliertes Vokabular* zu folgenden 6 Kategorien (in Englisch):

- **Audience** (in zwei Gruppen: „Lehrseite: Wer kann es einsetzen?“ und „Lernseite: Für wen ist es am besten geeignet);
- **Format** (*alle Arten* von Filetypen, z.B.: .exe, .doc, VHS, Quicktime usw.);
- **Grade** (zwei Gruppen, einmal Klassenstufen von Kindergarten bis zur 12. Klasse (K–12) und einmal „educational level“, z.B. vocational usw.);
- **Language** (derzeit Englisch, aber geplant sind auch Französisch, Deutsch, Italienisch, Spanisch u.a.);
- **Pedagogy** (eine Fülle didaktischer Angaben, es lohnt sich der Blick auf: http://gem.syr.edu/Workbench/Metadata/Vocab_Pedagogy.html);
- **Relation** (logische Relationen);
- **Resource Type** (Art des Lernmaterials, wie z.B. Exercise, Activity, Simulation, Best practice usw.);
- **Subject** (Fach).

Die Elemente bestehen aus DublinCore "DC." + eigene "GEM.*", obligatorisch: z.B. GEM.cataloging, DC.date, DC.format, DC.title, optional: z.B. GEM.audience, DC.creator, DC.description, GEM.essentialResources.*

Der Dublin Core wurde als Basis für die GEM-Elemente gewählt. Beispiel für GEM-Meta-Tags (Bild 5.9):

Bild 5.9 GEM Meta-Tags

```
<meta name="DC.subject.levelOne.I" scheme="GEM" content="Computer Science">
<meta name="DC.subject.levelTwo.I" scheme="GEM" content="Information Systems">
<meta name="DC.subject.levelTwo.I" scheme="GEM" content="Multimedia">
<meta name="DC.subject.levelTwo.I" scheme="GEM" content="New Media">
```

1.2.7 Warwick-Framework

Der auf dem zweiten Metadata-Workshop, 1996 in England, erarbeitete Vorschlag des nach dem Veranstaltungsort Warwick benannten Frameworks stellt selbst kein weiteres Metadatenformat dar, sondern ein Modell einer Container-Architektur. Dazu wurde eine (möglicherweise rekursive) Struktur aus Containern und Packages vorgeschlagen.

Ein **Container** ist eine Datenstruktur, die eine Menge von Packages enthält. Dabei sind die **Packages** für den Container opak, d.h., der Container muss sie nicht interpretieren können. Es wird von dem Container nur verlangt, dass er die Packages (einzeln) zur Verfügung stellen kann, d.h., dass er ein Package überspringen kann. Die Packages sind getypt, d.h., sie haben einen Typ, der von einem Verarbeitungsprogramm festgestellt werden kann. Entsprechend dieses Typs kann ihr Inhalt von dem Programm verarbeitet werden. Es gibt drei Arten von Packages:

- **Metadaten:** Hier stehen die Metadaten direkt im Package. Es können z.B. Dublin-Core-Daten sein oder Daten eines anderen Beschreibungsformats, wie USMARK oder das Inspec-Format. Wie weit es sinnvoll ist, hier beliebige Beschreibungsformate zuzulassen, oder ob man sich auf eine kleine standardisierte Menge einschränken soll, ist bisher noch nicht entschieden.
- **Indirekt:** Hier handelt es sich um eine Referenz auf ein anders Objekt, in dem Metadaten gefunden werden können. Dabei kann es sich um ein Objekt handeln, zu dem wiederum eigene Metadaten existieren; es kann sich um ein Objekt handeln, auf das auch andere Objekte zugreifen, also um Metadaten, die von vielen Objekten genutzt werden, und schließlich kann das Objekt auf einem anderen Server liegen.
- **Container:** Schließlich kann ein Package wieder ein Container sein.

Das Metadatenkonzept, das dem Warwick-Framework zugrunde liegt, geht über intrinsische Daten, wie sie durch den Dublin Core beschrieben werden, hinaus. Hier können auch spezielle Metadaten eingebunden werden, die z.B. Informationen über jeweilige Zugriffsrechte bzw. Abrechnungsmodalitäten oder Zugriffsstatistiken enthalten.

Der Vorteil des Warwick-Frameworks ist die Flexibilität in der Darstellung und Übertragung komplexer Metadaten unter Verwendung der einfachen Beschreibungselemente des Dublin Core. Dadurch wird Interoperabilität mit anderen Metadatenformaten gewährleistet.

Literatur: LAGOZE (1996).

1.2.8 Instructional Managing System (IMS)

Bereits 1994 wurde in USA die **N**ational **L**earning **I**nfrastructure **I**nitiative (NLII) von einem Konsortium (heute etwa 200 Mitglieder) verschiedener Institutionen und Organisationen aus dem Bildungsbereich zusammen mit Industriepartnern gegründet. Daraus entstand das Projekt **I**nstructional **M**anagement **S**ystem (IMS), das zunächst die Entwicklung von Spezifizierungen und Prototypen für ein verteiltes Informationssystem zum Ziel hatte.

Qualität von Content im Internet

IMS-Metadaten bauen auf dem Dublin Core auf und dienen nach IMS-Spezifikation einem effizienteren Suchverfahren mit dem vordergründigen Ziel einer Ermittlung der Qualität von Lehr- und Lernmaterial im Internet. Aber auch Fragen des Managements der Materialien und des Schutzes intellektuellen Eigentums stehen im Fokus.

Das IMS-Metadaten-Dictionary wurde für das Bildungswesen entwickelt. Mit entsprechenden Metadaten (Identifyers) werden z.B. Lernziele, Art der Interaktivität, Lerndauer und pädagogisch-didaktische Methoden beschrieben. Ein typisches Beispiel für so genannte Identifiers (Bild 5.10):

Bild 5.10 Die Baumstruktur des LOM-Ansatzes

Elementname	*Beschreibung*
identifyer	Global eindeutiges Label für ein Lernobjekt
title	Titel des Lernobjektes
catalog	Katalog, worin das Objekt zu finden ist (z.B. ISBN)
language	Sprache, in der das Objekt zur Verfügung steht
description	Beschreibung des Lernobjektes
keyword	Schlüsselwörter
version	Version des Lernobjektes
typcialagerange	Typisches Alter der angepeilten Zielgruppe
difficulty	Schwierigkeit für einen typischen Lernenden aus der angepeilten Zielgruppe
typicallearnertime	Durchschnittliche Lernzeit, die zur Durcharbeit des Lernobjektes notwendig ist

Die IMS **Q**uestion & **T**est **I**nteroperability Specification (QTI) gibt eine Struktur für Prüfungsfragen und Tests an und erläutert, wie die Antworten der Lernenden gehandhabt werden können. Dies hat wiederum zum Ziel, solche On-linetests zwischen verschiedenen Lern- und Autorensystemen austauschen zu können und damit Interoperabilität zu erreichen. Die durch die Spezifikation zur Verfügung gestellten XML-Elemente definieren nicht nur ein großes Set an standardisierten Fragetypen, sondern lassen gezielt auch proprietäre Typen zu, die dann aber außerhalb der Spezifikation genauer definiert werden müssen und natürlich dann nicht mehr allgemein austauschbar sind.

1.2.9 IEEE Learning Objects Metadata (LOM)

Das **L**earning **T**echnology **S**tandards **C**ommittee (LTSC) der IEEE stellte bereits 1999 als Working Draft den Standard P1484.12 mit der Bezeichnung **L**earning **O**bjects **M**etadata (LOM) vor. Dieser steht in Verbindung mit der **L**earning **T**echnology **S**ystem **A**rchitecture (LTSA, Standard P1484.1). In der LTSA wird eine Definition notwendiger Prozesse zur Datenhaltung und zum Datenfluss gemacht, aber keine Implementierungsrichtlinien festgelegt.

Lernobjekte werden bei LOM als beliebige digitale und auch nicht digitale Einheiten definiert, wie z.B. Video-Objekte, Bilder, Audio-Objekte, Texte, aber auch Software-Tools, Simulationen, Bücher usw.

LOM definiert nicht – wie IMS – ein generelles Schema für Lernobjekte, sondern stellt neun unterschiedliche **Kategorien** zur Verfügung:

1. General: allgemeine Informationen über ein Lernobjekt.

2. Lifecycle: Informationen über den Lebenszyklus und aktuellen Stand (current state) des Lernobjekts.

3. Meta-metadata: Informationen über die verwendeten Metadaten über das Lernobjekt.

4. Technical: Informationen über technische Details (requirements) des Lernobjekts.

5. Educational: Informationen über pädagogische und speziell didaktische Eigenschaften des Lernobjekts.

6. Rights: Informationen über die Nutzungsbedingungen (property rights) des Lernobjekts.

7. Relation: Informationen über Beziehungen (relationships) zu anderen Lernobjekten.

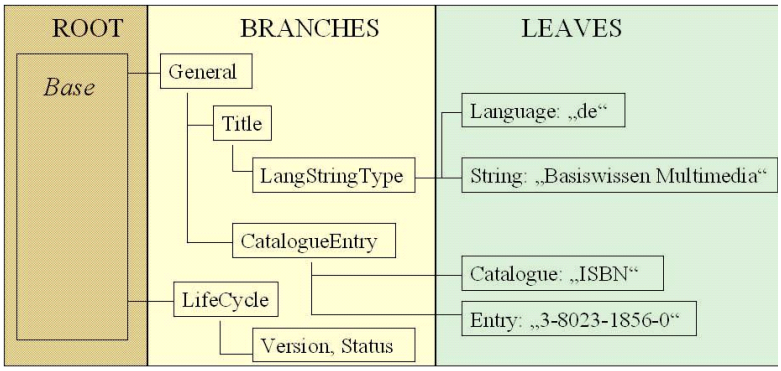
8. Annotation: Kommentare (comments) sowohl zum pädagogischen als auch didaktischen Einsatz (educational use) des Lernobjekts und zusätzliche Informationen.

9. Classification: Information über die Zuordnung des Lernobjekts in eine bestimmte Lernobjekt-Klasse.

Die Elemente in LOM sind also hierarchisch in Kategorien angeordnet.

Sie besitzen verschiedene festgelegte und standardisierte Datentypen, die jeweils obligatorisch (compulsory) oder fakultativ (optional) anzugeben sind. Es besteht eine gewisse Kompatibilität zwischen LOM und Dublin Core: LOM ist zwar komplexer als der Dublin Core, doch lassen sich DC-Elemente komplett auf die LOM-Unterelemente abbilden. Die Kategorien bilden eine Art **Baumstruktur,** bestehend aus Stamm (Root) und Zweigen (Branches), wobei die einzelnen Werte nur in den Blättern (Leaves) stehen dürfen (Bild 5.11):

Bild 5.11 Die Baumstruktur des LOM-Ansatzes



Besonders interessant ist die **Kategorie 5 Educational,** die durch folgende **Typen** beschrieben wird:

5.1 Interaktivität (Interactivity Type) zwischen den Lernenden und den Lernobjekten (Bild 5.11):

Bild 5.12 Die Baumstruktur des LOM-Ansatzes

Größe (Size)	Datentyp (Data Type)	Wert (Value)
1 (single value)	Vocabulary	Active Expositive Mixed Undefined

„Active“ z.B. ist ein Lernobjekt nur dann, wenn Interaktion vorhanden ist, das ist z.B. bei Simulationen, Fragenquiz usw. der Fall, nicht aber bei reiner Navigations-Interaktion, wie bei Hypertexten. Das didaktische Ziel ist dabei **„learning by doing“.**

„Expositive“ ist ein Lernobjekt dann, wenn sich die „Interaktion“ auf eine reine Präsentation beschränkt, wie z.B. typische Power-Point-Folien, Hypertexte usw. Das Ziel ist dabei **„learning by reading“.**

Siehe dazu z.B. einen Artikel von SCHULMEISTER (2002) und Basiswissen Multimedia, Band 1.

5.2 Art der Resource (Learning Resource Type), entspricht Dublin Core „ResourceType“ (Bild 5.13):

Größe (Size)	Datentyp (Data Type)	Wert (Value)
kleinstes erlaubtes Maximum: 10 Einträge (smallest permitted maximum: 10 items)	Vocabulary	Exercise Simulation Questionnaire Diagram Figure Graph Index Slide Table Narrative Text Exam Experiment Problem Statement Self Assessment

Bild 5.13 ResourceType in LOM

5.3 Grad der Interaktivität *zwischen* den Lernenden und den Lernobjekten (Interactivity Level, Bild 5.14):

Größe (Size)	Datentyp (Data Type)	Wert (Value)
1 (single value)	Vocabulary	very low low medium high very high

Bild 5.14 Interactivity Level

5.4 Subjektive Nützlichkeit im Verhältnis von Aufwand/Lernzeit (Semantic Density, Bild 5.15):

Größe (Size)	Datentyp (Data Type)	Wert (Value)
1 (single value)	Vocabulary	very low low medium high very high

Bild 5.15 Semantik Density

5.5 Zielpersonen (Intended End User Role, Bild 5.16):

Größe (Size)	Datentyp (Data Type)	Wert (Value)
kleinstes erlaubtes Maximum: 10 Einträge	Vocabulary	Teacher Author Learner Manager

Bild 5.16 Intended End User Role

5.6 Zielgruppe (Context, Bild 5.17):

Bild 5.17 Context

Größe (Size)	Datentyp (Data Type)	Wert (Value)
kleinstes erlaubtes Maximum: 10 Einträge	Vocabulary	Primary Education Secondary Education Higher Education University First Cycle University Second Cycle University Postgraduate Technical School First Cycle Technical School Second Cycle Professional Formation (Berufsausbildung) Continuous Formation (Berufsfortbildung) Vocational Training (Berufsschule) other

5.7 Altersgruppe (Typical Age Range, Bild 5.18):

Bild 5.18 Typical Age Range

Größe (Size)	Datentyp (Data Type)	Wert (Value)
1 (single value)	LangString* (smallest permitted maximum: 1000 characters)	-

*z.B. 0-5, 7-9, 15, 18-, suitable for children over 7, adults only usw.

5.8 Schwierigkeit für die typischen Lernenden (Difficulty, Bild 5.19):

Bild 5.19 Difficulty

Größe (Size)	Datentyp (Data Type)	Wert (Value)
1 (single value)	Vocabulary	very easy easy medium difficult very difficult

5.9 Bearbeitungsdauer (Typical Learning Time, Bild 5.20):

Bild 5.20 Typical Learning Time

Größe (Size)	Datentyp (Data Type)	Wert (Value)
1 (single value)	Date*	-

* PT1H30M, PT1M45S usw.

5.10 Einsatz des Lernobjekts (Description, Bild 5.21):

Bild 5.21 Description

Größe (Size)	Datentyp (Data Type)	Wert (Value)
1 (single value)	LangString* (smallest permitted maximum: 1000 characters)	-

* Kommentare, Empfehlungen, Tipps usw., *wie* das Lernobjekt eingesetzt werden kann.

5.11 Sprache (Language, Bild 5.22):

Größe (Size)	Datentyp (Data Type)	Wert (Value)
1 (single value)	LanguageID = Langcode*	-

Bild 5.22 Language

* „en“ = Englisch, „de“ = Deutsch usw.

Wichtig ist, dass sich **Relationen** zwischen Lernobjekten in LOM herstellen lassen. Diese lassen sich in drei Kategorien unterteilen:

- strukturelle Beziehungen, die bei der Modularisierung von Dokumenten die vormals existierende Dokumentstruktur widerspiegeln,
- inhaltliche Beziehungen, die aus semantischen Abhängigkeiten zwischen Lernobjekten abgeleitet werden können,
- ordinale Beziehungen, die sich aus der Einordnung von Lernobjekten auf einer Skala, z.B. einer Zeitleiste, ergeben.

Strukturelle Beziehungen: Über Relationen, die mit Objekten der *Kategorie „LOM.Relation“* der LOM-Datensätze formuliert werden können, werden die Lernobjekte zu einem größerem Lernobjekt, beispielsweise einem Kurs, einer Lerneinheit oder einem Modul, miteinander verbunden. Dieser Ansatz eignet sich insbesondere für modularisierte Lehrbücher, Hochschul-Skripte und Foliensätze. Diese Materialien sind über Kapitel und Unterkapitel per se strukturiert. Eine solche bereits bestehende Strukturierung sollte in einem Metadatenschema des Projektes abbildbar sein. Eine derartige Vorgehensweise besitzt folgende Vorteile:

- Bereits bestehende strukturelle Informationen können in LOM abgebildet werden und stehen für Information-Retrieval-Dienste bereit;
- zusammengesetzte Lernobjekte werden implizit mit abgebildet;
- feingranulare Lernobjekte lassen sich neu zu grobgranularen Lernobjekten zusammenfügen.

Inhaltliche Beziehungen: Sollen Lernobjekte abhängig vom Wissen und den Wünschen des Lernenden zusammengestellt werden, dann müssen inhaltliche Beziehungen zwischen den Lernobjekten ausgewertet werden. *Inhaltliche Verbindungen* zwischen Medienobjekten lassen sich in semantische, rhetorische und pragmatische Links unterteilen:

- **Semantische Links** stellen vor allem Verbindungen zwischen ähnlichen, gegensätzlichen oder in „ist-ein“- bzw. „ist-Teil-von“-Beziehung zueinander stehenden Wörtern und Themen her. Semantische Links sind hauptsächlich auf der Medienobjekt-Ebene, vereinzelt aber auch auf der Seiten- oder Kapitelebene anzutreffen.

- **Rhetorische Links** dienen dazu, die Lernenden durch eine Folge von Informationen zu führen, um ein ganz bestimmtes Lernziel zu erreichen. Rhetorische Links sind somit vor allem ein Werkzeug des Autors, mit dem Definitionen, Erklärungen, Illustrationen, Exkurse und ähnliche Elemente in einen Hypertext integriert werden können.
- **Pragmatische Links** stellen im Gegensatz zu rhetorischen Links keine Verbindungen zwischen Medienobjekten, sondern vielmehr zwischen dem behandelten Thema und dem vorgegebenen Lernziel bzw. auch der aktuellen Lernsituation der Lernenden her. Beispiele für pragmatische Verbindungen sind Erfahrungen, Warnungen, Hinweise auf benutzerbezogene Beispiele und Benutzungshinweise selbst.

Im folgenden Beispiel werden inhaltliche Beziehungen zwischen Lernobjekten durch rhetorisch didaktische Relationen definiert (Bild 5.23):

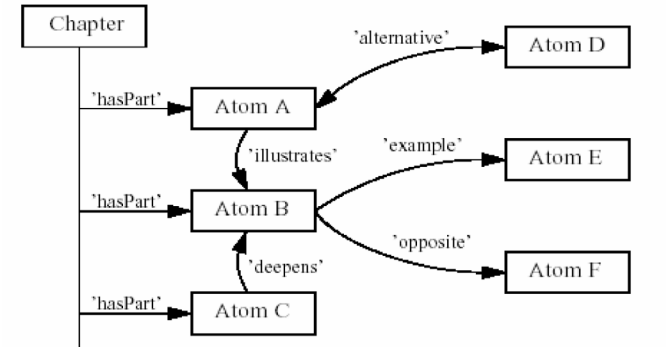


Bild 5.23 Beispiel für Beziehungen zwischen Lernobjekten, aus: Laatz et al. (2001)

- Beispiel: Lernobjekt E enthält ein Beispiel zu Lernobjekt B.
- Illustration: Lernobjekt A enthält eine Illustration zu Lernobjekt B.
- Instanz: Lernobjekt B ist inhaltlich Lernobjekt A untergeordnet.
- Restriktion: Lernobjekt B schränkt das in Lernobjekt A Gesagte ein.
- Erweiterung: Lernobjekt B enthält weiterführende Zusatzinformationen zum Inhalt von Lernobjekt A.
- Fortsetzung: Lernobjekt B führt den in Lernobjekt A aufgegriffenen Gedanken fort.
- Vertiefung: Lernobjekt C vertieft den Inhalt von Lernobjekt B.
- Gegenteil: Der Inhalt von Lernobjekt B steht im Gegensatz zum Inhalt von Lernobjekt F.
- Alternative: Lernobjekt A und D sind inhaltlich identisch, unterscheiden sich jedoch in Form oder Codierung.

Literatur z.B. STEINACKER et al. (1999), HOLZINGER et al. (2000), LAATZ et al. (2001) und generell unter http://ltsc.ieee.org/wg12

1.2.10 SCORM

SCORM, Referenzmodell für wiederverwendbare Inhalte

Das **S**harable **C**ontent **O**bject **R**eference **M**odel (SCORM) ist ein Referenzmodell zur Integration *verschiedener* Standards. Das Ziel ist es, system- und plattformunabhängige **Lernobjekte** (Learning Objects, LO) verwenden, verarbeiten und austauschen zu können. Die vier Hauptziele des SCORM-Modells sind zusammengefasst unter dem Acronym RAID:

- **R**eusability (Wiederverwendbarkeit),
- **A**ccessibility (Zugreifbarkeit, Zugänglichkeit),
- **I**nteroperability (Austauschbarkeit, Unabhängigkeit) und
- **D**urability (Dauerhaftigkeit, Beständigkeit).

Das **Referenzmodell SCORM** soll Wiederverwendbarkeit (reusability) und die Austauschbarkeit (interoperability) von Lernobjekten (Learning Objects, LO) gewährleisten.

SCORM wurde von der Initiative ADL (Advanced Distributed Learning) gegründet. ADL wurde u.a. vom US-Verteidigungsministerium gegründet. Auch die US-Luft- und Raumfahrt steht dahinter. SCORM integriert mehrere Lerntechnologiestandards durch die Beteiligung namhafter Initiativen (LTSC, IMS, ARIADNE, AICC), was es zu einem Erfolg versprechenden Standard gemacht hat (PAWLOWSKI, *2002). Ausgangspunkt für die Entwicklung von SCORM war die Tatsache, dass es eine riesige Auswahl von Learning-Management-Systemen (LMS) gibt. Einige populäre sind z.B. Hyperwave e-Learning Suite (eLS), WebCT, Clix, Blackboard, Topclass usw. Der schnell wachsende Markt hat für die Anbieter von Lerninhalten ein ernstes Problem geschaffen: Obwohl alle LMS Standard-Protokolle benutzen, haben alle ein Problem, wenn ein Kurs von einem LMS auf ein anderes LMS übertragen werden soll. Ohne eine gemeinsame Spezifikation zum Aufbau von e-Learning-Kursen strukturieren alle Anbieter ihre Inhalte jeweils in einer Form, die diese für die richtige halten. In der Praxis muss daher oft der komplette Kurs „rekonstruiert" werden – natürlich unter dem entsprechenden Aufwand.*

SCORM bezeichnet jede instruktionale Einheit als **S**harable **C**ontent **O**bject (SCO). Das SCO wiederum besteht aus wieder verwendbaren (reusable) **Assets.** Mit den jeweils zugehörenden Metadaten (z.B. IEEE LOM) können die Assets und SCOs von einem SCORM-konformen **Content Repository** verwaltet werden. SCORM besteht im Prinzip aus zwei Teilen:

- Content Aggregation Model (das auf dem IMS Learning Resource Metadata Information Model aufbaut, das seinerseits wiederum auf der IEEE-LOM-Spezifikation beruht, und
- Run-Time Environment (Spezifikation einer Lernumgebung, der Kommunikation und dem Tracking von Inhalten).

Content-Aggregationsmodell

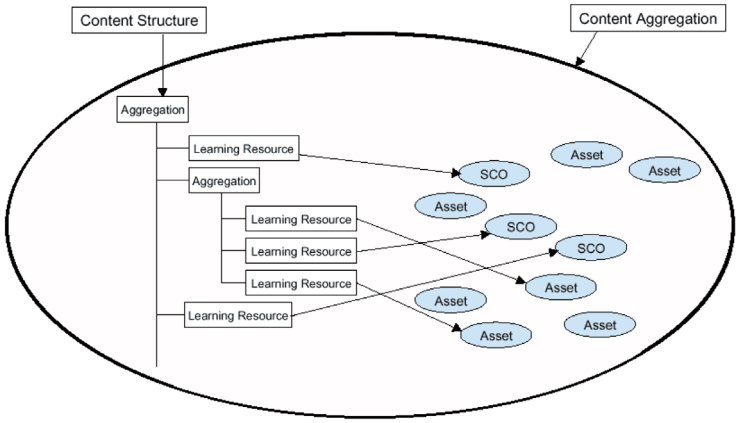
Das SCORM **C**ontent **A**ggregation **M**odel (CAM) unterstützt das Erzeugen, Zusammentragen und Aggregieren (Anhäufen) einfacher Lern-Ressourcen zu komplexen Lerneinheiten und besteht aus folgenden drei Teilen:

- **1) Content Model,**
- **2) Metadata Dictionary** und
- **3) Content Packaging.**

Content Model

1) Content Model. Dieses enthält eine Nomenklatur zur Beschreibung der benötigten Komponenten für den Zusammenbau von Lernerfahrungen (Learning Experiences) aus wieder verwendbaren Lern-Ressourcen. Das Modell legt fest, wie verteile, wieder verwendbare Low-Level-Lern-Ressourcen zu High-Level-Lerneinheiten *aggregiert* werden können. Das Content Model definiert dafür folgende Komponenten: Assets, Shareable Content Objects (SCO) und Content Aggregation (Bild 5.24):

Bild 5.24 SCORM Content Model



Die folgenden vier Elemente erlauben es, (relativ kleine) *wieder verwendbare* Komponenten zu größeren Einheiten zusammenzufassen:

- Asset (A),
- Sharable Content Object (SCO),
- Block (B) und
- Content Structure Format (CSF).

Drei Metadatentypen (Raw media metadata, Content metadata und Course metadata) beschreiben diese Elemente. Die Block-Elemente und Content-Structure-Format-Elemente enthalten keine eigentlichen Daten, sondern nur Verweise auf die Daten, die sie aggregieren.

Metadata Dictionary

2) Das **Metadata Dictionary** ist ein Mechanismus zur Beschreibung von Instanzen des Content Model (Assets, Shareable Content Objects und Content Aggregation) und enthält ein **Set von Schlüsselwörtern** zur Beschreibung einer Lernressource. Das SCORM-Metadaten-Elemente-Set wurde von IMS übernommen und entspricht somit dem IMS Learning Resource Metadata Information Model, das auf IEEE LOM basiert (Bild 5.25):

Element Name	C	S	A	Anzahl
general	M	M	M	1
identifier	—	—	—	—
title	M	M	M	1
catalogentry	M	M	O	0 ... 10
catalogue	M	M	O	0 ... 1
entry	M	M	O	0 ... 1
language	O	O	O	0 ... 10
description	M	M	M	0 ... 10
keyword	M	M	O	0 ... 10
coverage	O	O	O	0 ... 10
structure	O	O	O	0 ... 1
aggregationlevel	O	O	O	0 ... 1
lifecycle	M	M	O	0 ... 1
version	M	M	O	0 ... 1
status	M	M	O	0 ... 1
contribute	O	O	O	0 ... 30
role	O	O	O	0 ... 1
centity	O	O	O	0 ... 40
date	O	O	O	0 ... 1
metametadata	M	M	M	1
identifier	—	—	—	—
catalogentry	O	O	O	0 ... 10
catalogue	O	O	O	0 ... 1
entry	O	O	O	0 ... 1
contribute	O	O	O	0 ... 10
role	O	O	O	0 ... 1
centity	O	O	O	0 ... 10
date	O	O	O	0 ... 1
metadatascheme	M	M	M	0 ... 10
language	O	O	O	0 ... 1
technical	M	M	M	1
format	M	M	M	0 ... 40
size	O	O	O	0 ... 1
location	M	M	M	0 ... 10
requirement	O	O	O	0 ... 40
type	O	O	O	0 ... 1
name	O	O	O	0 ... 1
minimumversion	O	O	O	0 ... 1
maximumversion	O	O	O	0 ... 1
installationremarks	O	O	O	0 ... 1
otherplatformrequirements	O	O	O	0 ... 1
duration	O	O	O	0 ... 1

Element Name	C	S	A	Anzahl
educational	O	O	O	0 ... 1
interactivitytype	O	O	O	0 ... 1
learningresourcetype	O	O	O	0 ... 10
interactivitylevel	O	O	O	0 ... 1
semanticdensity	O	O	O	0 ... 1
intendededenduserrole	O	O	O	0 ... 10
context	O	O	O	0 ... 10
typicalagerange	O	O	O	0 ... 5
difficulty	O	O	O	0 ... 1
typicallearningtime	O	O	O	0 ... 1
description	O	O	O	0 ... 1
language	O	O	O	0 ... 10
rights	M	M	M	1
cost	M	M	M	1
copyrightandotherrestrictions	M	M	M	1
description	O	O	O	0 ... 1
relation	O	O	O	0 ... 100
kind	O	O	O	0 ... 1
resource	O	O	O	0 ... 1
identifier	—	—	—	—
description	O	O	O	0 ... 1
catalogentry	O	O	O	0 ... 10
catalog	O	O	O	0 ... 1
entry	O	O	O	0 ... 1
annotation	O	O	O	0 ... 30
person	O	O	O	0 ... 1
date	O	O	O	0 ... 1
description	O	O	O	0 ... 1
classification	M	M	O	0 ... 40
purpose	M	M	O	0 ... 1
taxonpath	O	O	O	0 ... 15
source	O	O	O	0 ... 1
taxon	O	O	O	0 ... 15
id	O	O	O	0 ... 1
entry	O	O	O	0 ... 1
description	M	M	O	0 ... 1
keyword	M	M	O	0 ... 40

Legende
C = Content Aggregation, S = SCO, A = Asset.
M = Mandatory, O = Optional.

Bild 5.25 SCORM Metadata Dictionary basiert auf IEEE LOM

Bild 5.25 zeigt alle Elemente, die im Elemente-Set definiert sind. Dabei wird unterschieden zwischen obligatorischen (M – mandatory) und optionalen (O – optional) Elementen. Welche Elemente obligatorisch sind und welche nicht, hängt von der jeweiligen Metadaten-Komponente ab.

Im Gegensatz zu LOM wird bei SCORM die Verwendung der Metadaten-Elemente in Verbindung mit XML festgelegt.

Wie die Metadaten-Elemente mittels XML verwendet werden, beschreibt SCORM mit der **Metadata XML Binding.** Das ist eine Referenz auf die *IMS Learning Resource XML Binding Specification.* Das Binding-Schema gibt für *jedes* Metadaten-Element eine Richtlinie zur Benutzung an. Einige Elemente werden einem bestimmten Elementtyp zugeordnet. Das trifft in der Regel für die Elemente zu, die in irgendeiner Form *nicht eindeutig* sind, z.B. kann ein System nicht erkennen, in welcher Sprache der folgende Titel verfasst ist: <title>Basiswissen IT</title>. Die Elemente, die mehrsprachig angegeben werden können, enthalten einen speziellen Elementtyp: langstring. Das Beispiel würde dann so aussehen: <title><langstring xml:lang="de"> Basiswissen IT </langstring></title>. Nun kann das System erkennen, dass der Titel in Deutsch (de) angegeben ist (weiterhin gibt es Datums- und Uhrzeitangaben, elektronische Visitenkarten usw.). Das folgende Beispiel zeigt die Minimalvariante einer Metadatendatei für ein Asset. Es werden nur die obligatorischen Elemente verwendet. Durch Angabe der Document Type Definition (DTD) kann das XML-Dokument durch ein Lernmanagementsystem oder das Erstellungsprogramm validiert werden (Bild 5.26).

Bild 5.26 Beispiel für eine XML-Bindung

```
<?xml version="1.0" encoding="UTF-8"?>
<!doctype lom system "http://www.imsproject.org/xml/IMS_METADATAv1p1.dtd">
<lom>
   <general>
      <title>
         <langstring xml:lang="de">Beispiel Asset Titel</langstring>
      </title>
      <description>
         <langstring xml:lang="de">Beispiel Asset Beschreibung</langstring>
      </description>
   </general>
   <metametadata>
      <metadatascheme>LOM-1.0</metadatascheme>
   </metametadata>
   <technical>
      <format>text/html</format>
      <location type="URI">http://www.beispiel.de/scorm/asset/</location>
   </technical>
   <rights>
      <cost>no</cost>
      <copyrightandotherrestrictions>yes</copyrightandotherrestrictions>
   </rights>
</lom>
```

Content Packaging

3) Content Packaging definiert, wie das beabsichtigte Verhalten und die Struktur einer Menge von Lern-Ressourcen beschrieben (Content Structure) und für den Austausch zwischen Systemen oder Werkzeugen gepackt wird (Content Packaging). Es ist ein Überbegriff für einen *standardisierten Weg,* digitale Lernressourcen zwischen verschiedenen Systemen auszutauschen.

Literatur zu SCORM am besten über: www.adlnet.org

Einfacher ausgedrückt ist ein Package eine Archivdatei, die alles enthält, was zur Darstellung eines Kurses notwendig ist. Dazu gehören 3 Dinge: eine im Package referenzierte Liste der Ressourcen, eine (optionale) Organisation des Kurses, die Kursstruktur und Verhalten festlegt, sowie Metadaten über das Package selbst. Diese Bestandteile befinden sich in einer Datei, die Manifest genannt wird.

2 Ontologie und semantisches Web

2.1 Ziel: Semantisches Web

Das Web ist für *menschliche Benutzer* konzipiert. Suchmaschinen (Kapitel 3) helfen zwar immer besser, aber es muss dann immer noch aus zigtausend Treffern „händisch" die passende Information gesucht und eingeordnet werden. Das ist nicht nur zeitraubend, sondern auch sehr frustrierend. Hauptgrund dafür ist, dass Wörter verschiedene Bedeutungen haben und **nur im Kontext** klar sind.

Bank

Beispiel: Bank. Dabei kann es sich um eine Parkbank oder um eine Geldbank handeln, unter Umständen wird Blutbank, Datenbank oder Werkbank darunter verstanden.

Die Idee des semantischen Webs ist es, die Inhalte von Webseiten so zu gestalten, dass diese durch Computerprogramme gelesen und maschinell verarbeitet werden können. Es wird also vor allem der Inhalt auf **logische Zusammenhänge** der Daten untersucht.

> Ziel des semantischen Webs ist, Menschen von der *Überflutung der Information* aus dem Web zu befreien und nur die wirklich relevanten und gewünschten Inhalte anzubieten.

Das „Semantik-Web" setzt voraus, dass Daten im Web richtig **definiert** werden und **Regeln** (Ontologien) angewandt werden, um die **Beziehungen** zwischen den Daten zu ermitteln.

DAVENPORT *formulierte das Hauptproblem so: „People can´t share knowledge if they don´t speak a common language". Das ist kein technisches Problem, denn obwohl Engländer und Deutsche das gleiche Alphabet, die gleichen Zeichen verwenden, können sie kein Wissen untereinander austauschen (man muss sich auf eine Sprache einigen).*

> Für ein semantisches Web ist eine **gemeinsame Sprache** notwendig, die aus einem **wohldefinierten Vokabular** besteht.

Nach TIM BERNERS-LEE muss ein semantisches Web auf den folgenden (technischen) Grundlagen beruhen (Bild 5.27):

1) Uniform Resource Identifier (URI) als Strings, mit denen Objekte im Web referenziert werden. Syntax ist detailliert im RFC 2396 beschrieben. Jedes Objekt im WWW wird durch mindestens eine URI beschrieben. Allerdings kann nicht entschieden werden, ob zwei unterschiedliche URIs das gleiche Objekt referenzieren.

2) Unicode-Standard (siehe Band 1) ist ein plattformunabhängiges, weit verbreitetes Codierungssystem, das eine Zuordnung von Zeichen zu Zahlen vornimmt. Buchstaben und Interpunktionszeichen verschiedener Sprachen werden *eineindeutig* (das ist *kein* Schreibfehler, sondern ein mathematisch strenger Begriff) auf Zahlen abgebildet. So können diese von Maschinen verarbeitet werden. Damit liefert der Unicode-Standard das „Alphabet", also die *zur Verfügung stehenden atomaren Zeichen,* für ein semantisches Web.

3) XML (XML-Schema und Namespace NS) soll das *syntaktische Grundgerüst* des semantischen Webs bilden. Ein Vorteil ist die Möglichkeit, aufbauend auf ein einheitliches Metamodell, relativ frei Inhalte formulieren zu können.

4) Resource Description Framework (RDF) sichert die Interoperabilität. Ein RDF-Modell kann mit XML als Syntax serialisiert werden. Das RDF-Modell (siehe Kapitel 1.2.3) kennt drei Datentypen: Resources, Properties und Statements. Eine Resource kann alles beschreiben, was durch ein URI identifiziert werden kann: Dokumente, Personen, Bücher usw. Properties sind Charakteristiken, Attribute, Aspekte oder Relationen, die eine Resource beschreiben. Statements sind Aussagen über eine Resource in der Form {Subjekt, Prädikat, Objekt}, wobei Subjekt die beschriebene Resource ist, Prädikat die beschreibende Property und Objekt der Wert des Property. Ein Objekt im Sinne des Statements kann ebenfalls eine Resource sein oder ein einfacher String (Literal). Des Weiteren bietet RDF die Unterstützung von *Containerstrukturen* (Bag, Sequence und Alternative). **Bag** ist eine *ungeordnete Liste* von Resources oder Literalen (z.B. eine Liste von Studierenden). Eine **Sequence** ist eine *geordnete Liste* von Resources oder Literalen (z.B. die Aufzählung von Arbeitsschritten in fester Reihenfolge). Eine **Alternative** ist eine Liste von Resources oder Literalen, die alternative Werte einer Property enthält (z.B. Aufzählung von gespiegelten Web-Seiten).

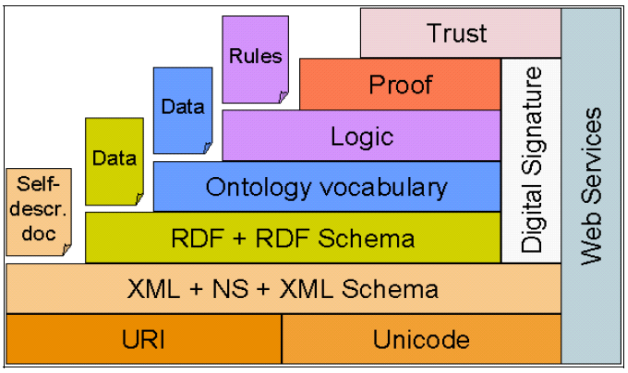


Bild 5.27 Der technische Aufbau eines semantischen Webs nach Tim Berners-Lee (2001)

Literatur zum semantischen Web z.B.: BERNERS-LEE et al. (2001), HYVÖNEN (2001), KAPPEL et al. (2003), FENSEL (2003).

2.2 Ontologie als Basis

Der Begriff Ontologie bezeichnet in der Philosophie die Lehre vom Sein bzw. vom Seienden. In der Informatik verstehen wir darunter etwas anderes:

> Eine **Ontologie** stellt eine *formale Beschreibung* von Objekten und Beziehungen dar, die dann für eine Gruppe von Personen begriffsbildend sind (GRUBER, 1993).

Eingesetzt werden Ontologien u.a. in der Wissensverarbeitung zur Formalisierung für Repräsentation und Austausch von Wissen. Die Festlegung von Begriffshierarchien, von Relationen und Attributen erlaubt die Verwendung spezieller Softwarewerkzeuge für Inferenz, Extraktion von Information und Erzeugung suchbarer Dokumentenindexe.

> Eine Ontologie ist ein **semantisches Modell,** das den Austausch von „Wissen" zwischen *Mensch und Maschine* erleichtert.

Dies wird erreicht durch ein wohldefiniertes Vokabular an Symbolen und einem einheitlichen Verständnis, welche Begriffe und Beziehungen jeweils die Symbole beschreiben. Die Verwendung einer Ontologie erlaubt es, das Kontextwissen, wie z.B. die Beziehung zwischen Ober- und Unterbegriffen, Verwandtschaft von Themengebieten usw., zu spezifizieren.

Symbole können Wörter sein. Ein solches Symbol erweckt in einem Akteur einen Begriff, der eine gedankliche Projektion auf die Wirklichkeit darstellt. Der Begriff bezieht sich also auf einen Gegenstand der Wirklichkeit. Sender und Empfänger sollten dabei dieselbe Vorstellung von dem gleichen realen Objekt haben. Es gibt jedoch manchmal – geprägt von sprachlichen, kulturellen und erfahrungsbedingten Unterschieden – mehrere mögliche Interpretationen eines Symbols (Bild 5.28).

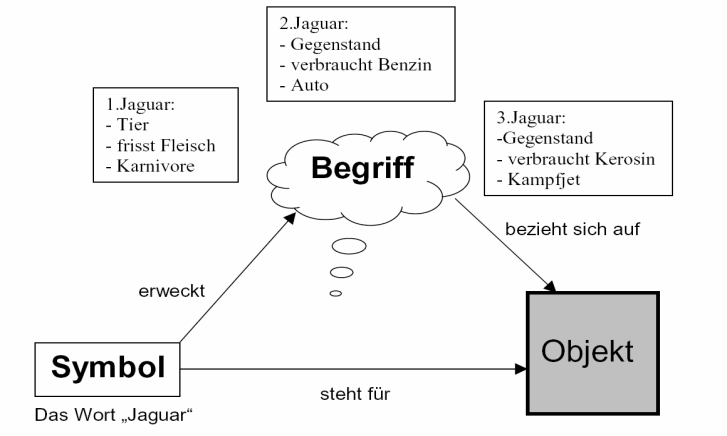


Bild 5.28 Die Mehrdeutigkeit eines Begriffs

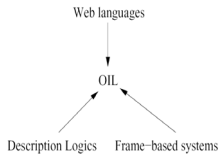


Bild 5.29 Die drei Basisaspekte von OIL Fensel et al. (2000)

2.3 OIL

Die **O**ntology **I**nterchange **L**anguage (OIL, zuerst als Ontology Inference Layer bezeichnet) wurde für die Repräsentation von Ontologien im Rahmen eines EU-Projektes (On-To-Knowledge) entwickelt. In OIL sind drei bedeutende Basisaspekte in einem Ansatz vereint (Bild 5.29).

Es werden die standardisierten und allgemein akzeptierten Datenaustausch- und Beschreibungsformate des Internets als Grundlage verwendet. Somit kann die Syntax der Beschreibungssprache mittels dieser Formate festgelegt werden. Die dokumentenunabhängige Beschreibung und der Austausch von Metadaten werden mittels RDF gemacht. Die Sprache OIL selbst setzt auf den Modellierungsprimitiven von RDFS auf und erweitert diese.

OIL erbt seine formale Semantik und die daraus folgende Unterstützung effizienter Inferenzdienste von den Beschreibungslogiken.

Beschreibungslogiken stellen eine Teilsprache der Prädikatenlogik erster Stufe dar, die eine hohe Ausdrucksmächtigkeit in Bezug auf die Darstellung von strukturiertem Wissen besitzt, ohne dabei auf entscheidbare und effiziente Inferenzprozeduren zu verzichten.

Analog zu HTML (Modul 4) wurde bei OIL mit einem kompakten, aber daher wohldefinierten Kern begonnen und sukzessive Erweiterungen, so genannte *Extensions,* gebildet.

Für diese Kernsprache existiert eine wohldefinierte Semantik, worauf die Erweiterungen aufbauen können und daher mit der Zeit mächtige Ontologiesprachen für verschiedene Anwendungsbereiche entstehen können.

OIL lässt sich auf drei unterschiedlichen Ebenen betrachten:

- Konkrete Instanzen einer Ontologie werden auf der Objekt-Ebene beschrieben.
- Auf der Meta-Ebene erfolgt die eigentliche Definition der Ontologien. Diese Ebene wird ontology definition genannt.
- Die Meta-Meta-Ebene befasst sich mit beschreibenden Anteilen einer Ontologie. Dafür wird der Standard des Dublin Core Paketes angewandt. Sie wird als ontology container bezeichnet.

Literatur z.B.: FENSEL (2003).

2.4 DAML+OIL

Bild 5.30 Das Logo von DAML+OIL

DAML+OIL (Logo in Bild 5.30) wurde vom W3C im Jahr 2000 veröffentlicht und ist eine Symbiose aus zwei Standards: der amerikanischen **D**ARPA (Defense Advanced Research Projects Agency), **A**gent **M**arkup **L**anguage (DAML) und der europäischen **O**ntology **I**nference **L**anguage (OIL).

Mit beiden Sprachen lassen sich Ontologien beschreiben, aber DAML+OIL als Kombination bietet wesentliche Erweiterungen gegenüber XML und RDF, um Objekte und deren Beziehung untereinander zu definieren (also eine Ontologie zu erstellen).

DAML+OIL baut auf den Prinzipien von XML und RDF auf (Bild 5.31).

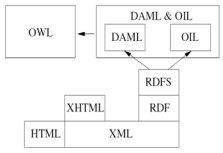


Bild 5.31 DAML+OIL baut auf XML und RDF auf und bildet die Basis für OWL (Kapitel 2.5)

Die DAML+OIL-Vision besteht darin, aus dem World Wide Web ein Semantic Web zu machen, in dem nicht länger mehrdeutige Daten dargestellt werden, sondern durch Beschreibung der Semantik eindeutige Informationen gewonnen werden, die nicht von jedem Nutzer individuell herausgearbeitet werden müssen.

DAML+OIL ist wie jede Ontologiesprache dahingehend ausgerichtet, die relevanten **Strukturen** einer Domäne zu beschreiben. Hierzu dienen die Konzepte Klasse und Property. Ontologien bestehen in DAML+OIL aus einer Reihe von Axiomen, die die charakteristischen Merkmale der Klassen und Properties ausdrücken. Da DAML+OIL eine Erweiterung von RDFS darstellt, bleiben natürlich Konzepte der Subsprache erhalten und werden nicht neu definiert. Ausdrücke, die Ressourcen als Instanzen von Klassen ausweisen, können gut in RDFS formuliert werden.

Der Unterschied von DAML+OIL zu XML und RDF (Bild 5.32):

Functionality	XML DTD	XML Schema	RDF(S)	DAML+OIL
bounded lists				X
cardinality constrains	X	X		X
class expressions				X
data types		X		X
defined classes				X
enumerations	X	X		X
equivalence				X
extensibility			X	X
formal semantics				X
inheritance			X	X
local restrictions				X
qualified constrains				X
reification			X	X

Bild 5.32 In diesem Bild wird rasch klar, wie sich DAML+OIL von XML und RDF unterscheidet

DAML+OIL hat eine hohe Verwandtschaft mit **Beschreibungslogiken.** Es kann als Erweiterung der Beschreibungslogik SHIQ gesehen werden, mit

den Zusatzoptionen, Klassen existenziell zu beschreiben und Datentypen zu verwenden. Klassen können in DAML+OIL sowohl durch Namen (d.h. URIs) als auch durch Klassenausdrücke formuliert sein. Für die Bildung von Klassenausdrücken stehen Konstruktoren zur Verfügung (Bild 5.33):

Bild 5.33 Konstruktoren
C_i = Konzeptnamen;
P = Property
x_i = Individuen
n = ganze Zahl
(Horrocks et al. (2001))

Konstruktor	DL Syntax	Beispiel
intersectionOf	$C_1 \sqcap \ldots \sqcap C_n$	$Human \sqcap Male$
unionOf	$C_1 \sqcup \ldots \sqcup C_n$	$Doctor \sqcup Lawyer$
complementOf	$\neg C$	$\neg Male$
oneOf	$\{x_1 \ldots x_n\}$	$\{john, mary\}$
toClass	$\forall P.C$	$\forall hasChild.Doctor$
hasClass	$\exists P.C$	$\exists hasChild.Lawyer$
hasValue	$\exists P.\{x\}$	$\exists citizenOf.\{USA\}$
minCardinalityQ	$\geqslant n\ P.C$	$\geqslant 2\ hasChild.Lawyer$
maxCardinalityQ	$\leqslant n\ P.C$	$\leqslant 1\ hasChild.Male$
cardinalityQ	$= n\ P.C$	$= 1\ hasChild.Male$

Konstruktoren können beliebig verschachtelt werden. Axiome erlauben in DAML+OIL, Beziehungen und Merkmale von Klassen und Eigenschaften auszudrücken. Des Weiteren lassen sich für Properties Eigenschaften wie beispielsweise Transitivität festlegen (Bild 5.34):

Bild 5.34 Axiome
(Horrocks et al. (2001))

Axiom	DL Syntax	Beispiel
subClassOf	$C_1 \sqsubseteq C_2$	$Human \sqsubseteq Animal \sqcap Biped$
sameClassAs	$C_1 \equiv C_2$	$Man \equiv Human \sqcap Male$
subPropertyOf	$P_1 \sqsubseteq P_2$	$hasDaughter \sqsubseteq hasChild$
samePropertyAs	$P_1 \equiv P_2$	$cost \equiv price$
disjointWith	$C_1 \sqsubseteq \neg C_2$	$Male \sqsubseteq \neg Female$
sameIndividualAs	$\{x_1\} \equiv \{x_2\}$	$\{President\_Bush\} \equiv \{G\_W\_Bush\}$
differentIndividualFrom	$\{x_1\} \sqsubseteq \neg\{x_2\}$	$\{john\} \sqsubseteq \neg\{peter\}$
inverseOf	$P_1 \equiv P_2^-$	$hasChild \equiv hasParent^-$
transitiveProperty	$P^+ \sqsubseteq P$	$ancestor^+ \sqsubseteq ancestor$
uniqueProperty	$\top \sqsubseteq \leqslant\ 1\ P$	$\top \sqsubseteq \leqslant\ 1\ hasMother$
unambiguousProperty	$\top \sqsubseteq \leqslant\ 1\ P^-$	$\top \sqsubseteq \leqslant\ 1\ isMotherOf^-$

In DAML+OIL sind Ontologien durch Axiome und Definitionen gebildet, wodurch insbesondere die Nutzung von Tools, die Frames als elementare Modellierungsprimitive ansehen, zu Schwierigkeiten führen kann.

Weitere Literatur zu DAML+OIL: McGuiness et al. (2002).

2.5 Web Ontology Language (OWL)

Im Herbst 2001 wurde vom W3C die Web Ontology Working Group gegründet. Ziel dieser Arbeitsgruppe ist eine Spezifikation einer „echten" Web-Ontologie-Sprache für das semantische Web.

OWL baut auf DAML+OIL auf (Bild 5.28) und basiert ebenfalls auf XML, RDF und dem RDF-Schema.

3 Suchen im WWW

Heute bietet sich einer Suchmaschine im WWW eine dramatische Situation: Es existieren viele Milliarden Webpages (Verdoppelung alle 3 bis 6 Monate), mehrere 100 Millionen Suchterme müssen pro Tag indiziert und hunderte Millionen Suchanfragen pro Tag beantwortet werden.

Trotz dieser Probleme bietet das WWW auch einen großen Vorteil, der es eben zum Web macht: Die strukturelle Organisation als Hypertext. Aus der Struktur des Hypertext und der Benennung von Links lassen sich aussagekräftige Metainformationen ableiten.

3.1 Suchmöglichkeiten

Es können 2 gegensätzliche Sucharten unterschieden werden: Matching und Browsing. **Matching** entspricht der Vorgehensweise einer klassischen Suchmaschine. Dabei werden eingegebene Suchbegriffe mit Indexbegriffen aus erfassten (indexierten) Dokumenten aus dem Web verglichen.

Matching: klassischer Vergleich von Indexbegriffen

- **Vorteile:** zielgerichtetes Vorgehen, das den Suchenden „zwingt", sein Informationsproblem zu durchdenken, um zu geeigneten Begriffen zu gelangen. Es besteht auch die Möglichkeit einer automatisierten Form der Relevanzbeurteilung für die Ergebnismenge in der Suchmaschine.
- **Nachteile:** liegen auf Seiten des Suchenden in der Formulierung und Konzeptionalisierung seines Informationsproblems, bedingt durch die Erfordernis, mit dem Vokabular des Problembereiches vertraut sein zu müssen, um eine erfolgreiche Anfrage erstellen zu können.

Browsing ist durch mehr oder weniger zielgerichtete Navigation von einem Hyperlink zu einem anderen gekennzeichnet, so dass sich ein zurückgelegter *Pfad von besuchten Seiten* ergibt (siehe Basiswissen Multimedia, Band 2). Es wird unterschieden zwischen *gerichtetem Browsing,* bei dem nach gezielter Information gesucht wird, und *ungerichtetem Browsing,* wo kein besonderes Problem im Hintergrund steht. Beim *assoziativen Browsing* wird Hyperlinks auf entsprechenden Seiten so lange nachgegangen, bis sich die erreichten Dokumente als irrelevant erweisen oder man das Interesse daran verliert.

Browsing: Der Suchende lässt sich von interessanten Dingen leiten (Serendipity-Effekt)

- **Vorteile:** Browsing ist eine Möglichkeit zur sukzessiven Eingrenzung des Problembereiches. Es ist nicht erforderlich, die gesuchte Information mit terminologisch genauen Begriffen zu bezeichnen. Hypertext-Wissensstrukturen sollen auch eher mit der assoziativen Denkweise des Menschen harmonieren und der Kreativität entgegenkommen.
- **Nachteile:** Browsing ist wenig zielorientiert und deshalb aufwendiger, das außerdem das Risiko der *Ablenkung* von der eigentlich beabsichtigten Problemlösung (Serendipity) enthält. Frustration kann entstehen.

Bei jeder Suche können allgemein drei Möglichkeiten unterschieden werden:

- 1) Suche auf lokalen WWW-Servern,
- 2) katalog- und verzeichnisbasierte Suche und
- 3) roboterbasierte Suche mit Suchmaschinen.

3.1.1 Suche in lokalen WWW-Servern

Lokale Suche ist eine **Stichwortsuche,** die auf das *Dokumentverzeichnis* des lokalen Web-Servers zugreift. Dieses einfache Verfahren wurde schon von den Web-Entwicklern am CERN durch HTML und HTTP ermöglicht. Über <ISINDEX> wird die Eingabe von Suchwörtern innerhalb einer HTML-Seite definiert. Der Browser bietet innerhalb dieser Web-Seite ein Eingabefeld für Suchbegriffe an. Dort eingegebene Begriffe werden mit einem vorangehenden „?“ und durch ein „+“ voneinander getrennt an die Dokumentadresse geschickt und die Anfrage an den Server gesendet.

Soll die Suche nicht nur auf Volltextsuche und einfache Retrieval-Methoden beschränkt sein, sondern auch Dokumentstruktur, bestimmte Felder (URL, HTML-Elemente usw.) oder Relevanzgrade in die Suche miteinbezogen sowie der Gebrauch von Operatoren ermöglicht werden, so müssen zusätzliche Komponenten in Verbindung mit Datenbanken auf der Server-Seite die Anfragebearbeitung erledigen. Auf dem Web-Browser können Daten über Eingabefelder bzw. Formulare eingegeben werden, die dann auf Server-Seite an Hintergrundprogramme z.B. über eine CGI-Schnittstelle weitergeleitet werden.

Web-Server bieten lokale Indizes an, um den Benutzern eine professionelle Suche in den **lokalen Dokumentenbeständen** zu ermöglichen.

Ein Vorteil dieser Gateway-Lösung ist, dass die Suche nicht nur auf Web-Dokumente beschränkt sein muss. Viele Datenbankanbieter und auch Produzenten des klassischen Informationsmarktes nutzen diese Technik, um die Bestände ihrer Datenbanken über komfortable Web-Schnittstellen anzubieten.

Das befreit die Benutzer natürlich nicht automatisch von der Benutzung proprietärer Retrievalsprachen, die bei kommerziellen Datenbankanbietern für die Recherche in den Datenbeständen erforderlich sind. Beispiele für derartige Retrievalsprachen sind **D**ata **S**tar **O**nline (DSO) des Anbieters DIALOGTM (DataStar), Messenger von STN/Fachinformationszentrum (FIZ) Karlsruhe und SPIRS des Anbieters Silverplatter. Die Suche in den Datenbanken kommerzieller Anbieter ist meistens kostenpflichtig. Nicht immer enthalten diese Datenbanken die Volltexte der gefundenen Artikel, häufig werden nur die bibliografischen Angaben und eine Inhaltszusammenfassung (Abstracts) ausgewiesen. Allerdings kann mit diesen Angaben das Dokument über Dokumentenlieferdienste (z.B. Subito) bestellt werden.

3.1.2 Katalog- und verzeichnisbasierte Suche

Ausgehend von der Navigation in hierarchisch aufgebauten Sachgebieten ist Browsing das klassische Suchverfahren in Katalogen. Die Katalogrubriken werden als *Hub Pages* bezeichnet und sind bei der Informationssuche ebenso relevant wie die eigentlichen *Authoritative Pages,* d.h. jene Dokumente, die die relevanten Informationen zur Problemlösung enthalten. Der Aufbau der Katalogrubriken erfolgt manuell. Als hochwertig *empfundene* Seiten werden lokalisiert und nach bestimmten Kriterien (thematisch) in Kategorien eingeordnet. Bei dieser Vorgehensweise ist häufig auch das Anmelden eigener Web-Seiten durch die Benutzer für die Aufnahme in den Katalog möglich bzw. erwünscht. Das klassische Beispiel dafür ist Yahoo.com, einer der ersten intellektuell erzeugten Kataloge. Typisch für solche Kataloge ist eine Begutachtung (Review, Rating) der vorgeschlagenen Seiten, um ein Mindestmaß an Qualitätssicherung zu erreichen.

Eine Möglichkeit ist die Erstellung eines Katalogs durch eine Benutzergruppe. Dabei können Benutzer freiwillig für den Inhalt einer Katalog-Rubrik verantwortlich sein und Links, Content usw. dafür sammeln und über den jeweiligen Katalog der Allgemeinheit zur Verfügung stellen. Beispielhaft dafür sind Community-Sites und Expertenforen. Die für jeweils eine oder mehrere Katalog-Rubriken verantwortlichen Editoren übernehmen die Qualitätssicherung der Beiträge aus der Öffentlichkeit in Form vorgeschlagener bzw. eingetragener Links. Des Weiteren werden bei diesen Katalog-Ausprägungen mitunter auch Vorschläge für neue Rubriken und Unterrubriken des Katalogs von der Allgemeinheit entgegengenommen. Open Directory (DMOZ) ist z.B. ein Dienst, der die Suchtechnologie vieler Suchmaschinenbetreiber um einen hierarchischen Katalog ergänzt.

Ein **Portal** ist ein Einstiegspunkt für den Zugang zu einer großen Menge an Informationen und Angeboten. Portale verwenden Informationsquellen und Suchwerkzeuge, die den wirtschaftlichen Interessen der Portal-Anbieter entsprechen (z.B. durch Verträge mit Fluglinien, Banken, Kaufhäusern usw.). Meistens ist zusätzlich ein eigener, redaktionell aufgearbeiteter Inhalt (Content) zu bestimmten Themengebieten über ein Portal zugänglich. Ein Merkmal von Portalen ist auch die Möglichkeit der Personalisierung zur Anpassung des Portals an individuelle Interessen und Präferenzen.

3.1.3 Roboterbasierte Suche: Suchmaschinen

Suchmaschinen sind die meistbenutzten Suchdienste des Internets. Um diese richtig einzusetzen und damit einen hohen Nutzen daraus ziehen zu können, bedarf es neben der Kenntnis und Anwendung von Suchmethoden und Suchoperatoren auch eines Einblicks in deren Funktionsweisen. Es werden ja nicht alle Dokumente oder Dokumentteile gleich gut erschlossen. Die nächsten Kapitel widmen sich daher ausschließlich den Suchmaschinen.

3.2 Definitionen bei Suchmaschinen

Die folgenden Definitionen, die in anderen Fachgebieten anders verwendet werden, gelten im Bereich der Suchmaschinen, um diese zu bewerten und Suchergebnisse und Dokumente möglichst objektiv zu charakterisieren:

- **Precision** ist der Anteil an *relevanten* Suchergebnissen (z.B. das Ergebnis befindet sich unter den Top 10). *Je weniger Präzision, desto mehr irrelevante Dokumente („falsche Alarme") liefert die Suchmaschine.*
- **Recall** ist der Anteil der gefundenen *relevanten* Dokumente an der Gesamtheit aller relevanten Dokumente. Dieser Wert sagt noch nichts über die Güte der Treffer. *Im Web hat Recall meist eine geringe Bedeutung, weil es eine „praktisch unendliche" Zahl relevanter Dokumente zu einer Suchanfrage gibt.*
- **Relevance** ist die Bedeutsamkeit eines Dokuments *in einem bestimmten Kontext.* Die Relevanz eines bestimmten Dokuments wird relativiert durch die Relevanz anderer Dokumente zur selben Anfrage. Es sollen nur die besten Dokumente in den Top 10 Suchergebnissen erscheinen, da es viele Tausend Dokumente geben kann, die irgendwie relevant sind.
- **Ranking** ist ein (oft komplexer) Algorithmus zur Relevanzbewertung. Suchergebnisse werden oft in absteigender Reihenfolge ihres Ranking-Wertes sortiert dargestellt. *Wie* dieser Wert berechnet wird, ist oftmals entscheidend für den Erfolg einer Suchmaschine – und natürlich auch für den Erfolg einer Web-Seite. Der Ranking-Algorithmus ist bei jeder Suchmaschine anders optimiert und oft ein Betriebsgeheimnis.
- **Authority** (auch: authoritativity) heißt, dass eine Webseite kompetente, aktuelle und verlässliche Informationen zu einem bestimmten Thema enthält. Manche Experten gehen davon aus, dass man gute Authorities an der großen Anzahl eingehender Kanten (Backlinks) von guten Hubs erkennen kann, wobei diese nur den durch die jeweiligen Suchergebnisse gegebenen Teilgraph des Webs betrachten.
- **Hub** ist praktisch komplementär zur Authority, d.h. z.B., eine Webseite enthält viele gute Links zu einem Thema. Gute Hubs zeichnen sich durch eine große Anzahl ausgehender Kanten (Links) zu Authorities aus, wobei aber wieder nur der durch die jeweiligen Suchergebnisse gegebene Teilgraph des Webs betrachtet wird.

3.3 Funktionsweise von Suchmaschinen

Die Anfragenbearbeitung läuft meistens als ein einfacher Zugriff auf eine **Index-Datenbank** ab. Die Benutzer geben auf einem Web-Formular die Suchbegriffe ein, diese werden dann von der Retrievalsoftware mit der Datenbank abgeglichen und die Ergebnismenge nach Relevanz sortiert dem Benutzer wieder zurückgeschickt.

Viel interessanter ist natürlich der Aufbau der Datenbank: Ganz am Anfang, also beim Aufbau einer Suchmaschine, steht eine Start-URL-Liste, die zunächst vom Betreiber aufgestellt wird und sich aus bekannten Web-Seiten, mitunter auch aus Katalogrubriken bzw. Hub-Pages anderer Anbieter zusammensetzen kann. Diese Liste wird dann vom Roboter Adresse für Adresse abgearbeitet. Die so erreichten Seiten werden inhaltlich erschlossen und die gefundenen Verweise an die URL-Liste angehängt. Dann werden die noch nicht verarbeiteten Adressen der URL-Liste nach dem gleichen Schema verarbeitet.

Die meisten Suchmaschinen „sehen" das Web als (riesigen) **gerichteten Graphen.** Die *Knoten* stellen die Dokumente und die *gerichteten Kanten* die Verweise dar, die von einem Dokument ausgehen.

Von einem bestimmten Knoten aus wird der Graph entlang den Kanten abgearbeitet. Bei jedem so erreichten Dokument wird von der Suchmaschine eine lexikalische Analyse durchgeführt, bei der inhaltsrelevante Terme aus dem Dokument extrahiert und in der Datenbank (DB) abgelegt werden. Das Abrufen einzelner Web-Dokumente erledigen parallel ablaufende „Agenten"-Prozesse. Diese geben der Suchmaschine das gewünschte HTML-Dokument oder eine Fehlermeldung – warum nicht zugegriffen werden konnte.

Die Aufgaben einer Suchmaschine sind in vier Teilaufgaben zerlegbar:

- 1) Dokumentenbeschaffung (Akquisition),
- 2) Indexierung,
- 3) Aktualisierung und die
- 4) Anfragenbearbeitung.

3.3.1 Dokumentbeschaffung (Akquisition)

Als erstes stellt sich die Frage, wie Suchmaschinen an Startseiten für eine weitere rekursive Erkundung herankommen. Hierfür werden zwei URL-Quellen unterschieden: **Roboter-Treffer** oder **URL-Submission.** Wenn auch roboterbasierte Suchdienste automatisch das Web erkundschaften, so ist der Nachweis meistens nur von bekannten und schon in Katalogen verzeichneten Dokumenten gesichert. Suchmaschinen benutzen oft Kataloge, inzwischen auch andere Suchmaschinen, um Startseiten für die automatische Suche zu bekommen. Ansonsten können bei den Suchdiensten URL-Vorschläge manuell in einer dafür eingerichteten WWW-Seite eingetragen werden. Dabei können oft auch zusätzliche Informationen über die Seite (Autor, Kommentare, e-Mail usw.) angegeben werden. Über diese Quellen werden Roboter auf Dokumente „aufmerksam" gemacht. Der Rest wird über rekursives Folgen der Hypertextstrukturen automatisch weiterverfolgt. Die so erreichten Seiten werden verarbeitet, ihre URLs in einem Register gespeichert und in regelmäßigen Abständen wieder besucht und aktualisiert. Die Tiefe der rekursiven Verfolgung der Links ist von Suchmaschine zu Suchmaschine unterschiedlich.

3.3.2 Indexierung

Das Angebot von Suchmethoden und Suchoperatoren ist in erster Linie von der Indizierung und der daraus resultierenden Datenbank abhängig. Dabei sind die Analysemethoden und der Umfang der Indizierung der einzelnen Web-Seiten von großer Bedeutung. Wie nun die vom Roboter laufend zusammengetragenen HTML-Seiten tatsächlich indiziert werden, lassen die einzelnen Suchmaschinenanbieter nur zum Teil erkennen.

Indizierung gehört zu den Kernkompetenzen der Suchmaschinenbetreiber, denn abhängig von dieser können in der Recherchekomponente mehr oder weniger fortschrittliche Suchoperatoren angeboten werden. Multimediale Elemente und die Mischung verschiedener Informationstypen bereitet der Indizierung aber noch einige Schwierigkeiten.

Grundsätzlich werden klassische Retrieval-Methoden verwendet, angepasst an die Gegebenheiten von HTML und WWW:

- Wortextraktion mit mehrsprachigen Stoppwortlisten;
- exakte Wortschreibweisen (Groß-, Kleinschreibung, Bindestriche usw.);
- Position der Wörter;
- Berechnung von Dokumentähnlichkeiten;
- Funktion der Wörter (URL, Titel, Überschrift, Link usw.);
- HTML-Elemente (Filenamen von Bildern, Java-Applets, Kommentare, unbekannte Elemente, die nicht vom Browser angezeigt werden, usw.);
- Verweisstrukturen (aus- und eingehende Links zu Dokumenten usw.).

Im Umfang der Indizierung werden unterschiedliche Strategien verfolgt:

- **Volltext:** Bei den meisten Suchmaschinen werden inhaltsbedeutende Begriffe oder Elemente aus der gesamten HTML-Seite (mehrsprachige Stoppwortlisten) indiziert;
- **Teilindex:** Suchmaschinen mit einem Teilindex indizieren meistens URL, Titel (TITLE-Element) und Überschriften („Hx-Elemente“) oder auch die ersten paar Zeilen der Web-Seite;
- **inhaltsbeschreibende Bereiche:** das META-Tag ist ein spezielles HTML-Element, über das der Autor eines Dokuments selbstständig Deskriptoren (z.B. Author, Copyright, Resource Type, Keywords usw.) und Zusatzinformationen über seine Web-Seiten strukturiert hinterlegen kann. Suchmaschinen, die solche META-Elemente unterstützen, extrahieren aus diesen die Metainformationen, so dass keine eigene Analyse bzw. Indizierung der Seite gemacht wird. Dieses Verfahren wird gerne bei Frame-Dokumenten angewandt, da viele Suchmaschinen diese nicht verarbeiten können.

3.3.3 Aktualisierung

Für die Aktualisierung bei der Übertragung einer Web-Seite ist durch ein „IF-MODIFIED-SINCE"-Feld im HTTP-Protokoll ein sehr wichtiger Mechanismus vorhanden. Über die Angabe dieses Feldes kann beim Laden eines Dokuments die Übertragung *von der letzten Änderung* (Datum/Uhrzeit) abhängig gemacht werden, d.h., falls das Dokument seit dieser Zeitangabe geändert wurde, wird das Dokument übertragen, sonst nicht.

Grundsätzlich gibt es in der Aktualisierungsfrequenz bei verschiedenen Suchmaschinen große Unterschiede in Art und Zeit. Meist wird mit einer zeitabhängigen Frequentierung gearbeitet. Die Angaben für die zeitliche Aktualisierung einzelner Web-Seiten bei den Suchmaschinen schwanken zwischen einem Tag und sechs Wochen. Oft wird dies von der Zugriffshäufigkeit auf ein Dokument abhängig gemacht.

Ein Problem, das bei Suchmaschinen aufgrund der hohen Anzahl von Dokumenten im Web und deren Streben nach einer möglichst umfassenden Abdeckung (Coverage) auftritt, wird durch eine nicht unerhebliche Menge von **Dead-Links** (Dangled Links) deutlich. Diese ergeben sich dann, wenn in der Index-Datenbank der Suchmaschine noch Einträge für Seiten des Webs in Form von Links enthalten sind, die sich nicht mehr an der zum Zeitpunkt der Indexierung gültigen URL befinden. Solche Dokumente, die nach mehrmaligen Zugriffsversuchen zu unterschiedlichen Zeiten durch die Suchmaschine nicht zugreifbar werden, werden aus der Datenbank entfernt. Dieses Problem z.B. tritt bei Hyperwave-Servern nicht auf.

Schwieriger wird der Fall, dass eine als Suchergebnis nachgewiesene Seite zwar an der zum Zeitpunkt der Indexierung aktuellen URL noch vorhanden ist, zwischenzeitlich aber eine inhaltliche Aktualisierung erfahren hat, nun die Suchbegriffe nicht mehr vorhanden sind und damit kein Bezug mehr zur Suchanfrage gegeben ist. Eine solche Variante kann erst nach Laden und erneutem Erfassen des Dokumenteninhalts aufgedeckt werden.

3.3.4 Anfragebearbeitung

Die Funktionalität bei der Anfragebearbeitung ist abhängig von der Inhaltserschließung der Dokumente. Je besser Analyse und Indizierung der HTML-Seiten sind, desto umfangreicher ist das Angebot an Suchmethoden und Operatoren. Die Benutzerschnittstelle (User Interface) muss daher nach Funktionalität ausgerichtet werden:

- verschiedene Suchmodi (Einfache/Erweiterte/Profi-Suche);
- formularbasierte Suchmasken mit diversen Einstellmöglichkeiten;
- Voreinstellungen über Buttons, Menüs, Listen usw.;
- Ergebnislisten mit Ranking, Sortierung, Vorschaufunktionen usw.

Die Treffermenge wird dem Benutzer sortiert nach einer internen Relevanzberechnung (Ranking) präsentiert. Meistens werden rein statistische Methoden verwendet:

- Anzahl aller gefundenen Suchbegriffe aus der Anfrage; falls nicht durch Boole'sche Suchoperatoren eingeschränkt, werden Dokumente, die alle Suchbegriffe enthalten, besser positioniert;
- Position (Funktion) der gefundenen Begriffe (Begriffe aus URL und dem TITLE-Element oder den Überschriften immer stärker gewichtet);
- Häufigkeit eines Suchwortes in einem Dokument (Term Frequency);
- Nähe der Suchbegriffe untereinander innerhalb des Textes (Proximity);
- Gesamtzahl eines Suchbegriffs in der Datenbank (je größer die Gesamthäufigkeit eines Begriffes innerhalb der Datenbank ist, desto niedriger ist auch der *inhaltswiedergebende* Wert dieses Begriffes (Inverse Document Frequency). Zusammen mit der term frequency bildet dieser Wert die Grundlage für ein verbreitetes, statistisches Verfahren namens TFIDF (Term Frequency times Inverse Document Frequency) zur Ermittlung von Relevanz und dem Ranking von Suchergebnissen;
- Popularität eines Dokuments: Je häufiger ein Dokument hohe Rankingwerte erhält, desto stärker wird es bei einem erneuten Rankingverfahren gewichtet. Solche Dokumente dürften gegenüber anderen Web-Seiten tatsächlich auch eine höhere Relevanz haben;
- Anzahl und Qualität von Hyperlinks, die auf ein Dokument verweisen und von einem Dokument ausgehen. Es werden dabei die indexierten Dokumente der Suchmaschine als ein *gerichteter Graph* betrachtet und für jedes verknüpfte Dokument werden Informationen aus den eingehenden Links als Anzahl und einem berechneten Wert für die Relevanzbewertung dieser Links und der Anzahl der abgehenden Links benutzt, um ein zusätzliches Ranking-Kriterium zu erhalten.

Durch die Verknüpfung von roboterbasierten Verfahren und Web-Katalogen ergeben sich Synergieeffekte bezüglich der Relevanzbeurteilung: Durch die Datenbank nachgewiesene Dokumente, die auch im Katalog verzeichnet sind, erhalten einen höheren Relevanzgrad.

Manche Suchmaschinen erlauben aber auch die Sortierung der Trefferliste nach anderen Kriterien wie Größe, Alter des Dokuments oder nach Servern. Die Sortierung nach Servern (z.B. bei Excite und Lycos) bietet dazu eine Erleichterung der Relevanzbeurteilung für die Benutzer, da sich auf einem Server meist gleichartige Dokumente befinden. Es genügt oft, nur wenige Dokumente pro Server zu betrachten, um die Relevanz der zugehörigen Web-Seite zu beurteilen.

Fortgeschrittene Verfahren der Relevanzbeurteilung und der Positionierung im Ranking werden möglich durch die Nutzung von Informationen, die sich aus der Hyperlink-Struktur vernetzter Dokumente gewinnen lassen.

Außer der Zahl eingehender Verweise (Backlinks) und abgehender Verweise (Forward Links) kann auch die Qualität der Web-Seite, von der der Link ausgeht, in die Berechnung einer Hyperlink-basierten Relevanz einbezogen werden.

So können Dokumente mit nur wenigen Verweisen von einer qualitativ hochwertigen und mit entsprechend hohem Relevanzurteil bedachten Seite (z.B. aus einem manuell erstellten Katalog) auch für die Seite, auf die verwiesen wurde, zu einem höheren Relevanzwert führen als eine große Anzahl von Verweisen, die von Seiten mit geringerer Qualität ausgehen.

Dieses Verfahren findet Anwendung für die Ermittlung einer Hyperlink-basierten Relevanzkomponente namens PageRank (siehe nächstes Kapitel).

Ein anderes Ranking-Verfahren basiert auf der *Popularität von Dokumenten,* die durch Beobachtung der Benutzer der Suchmaschine im Umgang mit den erzielten Suchergebnissen gewonnen wird. Es sollen Hinweise für ein Ranking der Dokumente erlangt werden, indem festgestellt wird, ob und welche Seiten des Suchergebnisses auch wirklich aufgesucht werden, wie lange diese einzelnen Dokumente dann betrachtet werden usw. Umgesetzt wurde ein solches Verfahren bei der inzwischen eingestellten Suchmaschine DirectHit.

Da Suchmaschinen Werkzeuge für ein breites Publikum darstellen, werden von diesen nicht nur hochwertige Suchergebnisse erwartet, sondern auch eine nach Aspekten bester Usability entwickelte Benutzerschnittstelle (Basiswissen Multimedia, Band 3). Dazu gehört beispielsweise die unter dem Begriff **Effizienz** betrachtete Zeitspanne zur Bearbeitung einer Suchanfrage usw.

In der Praxis der Suchmaschinen muss immer ein Kompromiss zwischen Effektivität und Effizienz eingegangen werden.

Eine dafür charakteristische Situation ergibt sich aus der Beschränkung der Kapazität des Hauptspeichers, in dem die Index-Datenbank(en) für einen zeitnahen Abgleich mit den Suchbegriffen und damit auch einer kurzen Antwortzeit gegenüber dem Nutzer abgelegt ist. Da diese Datenbanken mittlerweile eine Größe aufweisen, die vertretbare Hauptspeicherdimensionen überschreitet, wird meistens nicht mehr der gesamte Index für einen Abgleich mit den Suchbegriffen bereitgestellt, sondern nur noch ein Teil davon, der sich z.B. durch Caching der letzten oder zumindest der am häufigsten angeforderten Dokumente ergibt. Somit wird darauf verzichtet, mögliche Suchergebnisse auszuweisen – zugunsten kurzer Antwortzeiten.

Kurze Antwortzeiten und hohe Qualität der Suchergebnisse charakterisieren eine gute Suchmaschine.

Bild 5.35 Google wurde benannt nach dem Wort Googol für die Zahl 10 hoch 100

3.4 Beispielsuchmaschine: Google

Google (Bild 5.35) ist wahrscheinlich die bis jetzt *erfolgreichste* Suchmaschine. Die Betreiberfirma Google Inc. wurde 1998 von LARRY PAGE (geb. 1972) und SERGEY BRING (geb. 1970) gegründet.

Googles Erfolg wurde maßgeblich durch die von PAGE initiierte so genannte **PageRank-Technologie** (Seitenbewertung, Name wurde aber vom Autor abgeleitet) ermöglicht. Diese bewertet eine Web-Seite danach, wie viele Links aus dem Web auf sie verweisen. Dabei werden die einkommenden Verweise von Seiten, die selbst viele eingehende Links enthalten, stärker berücksichtigt. Die Web-Robots (Crawler, Roboter) von Google erfassen also die Stärke und Struktur des Link-Netzes im Web. Seiten, bei denen die eingegebenen Suchbegriffe nahe beieinander im Dokument auftreten oder bei denen der Suchbegriff im Link-Text auftaucht, werden höher bewertet. Auf dieser Grundlage findet man (fast immer) die aussagekräftigsten Seiten am Anfang der Suchergebnisliste. Hierbei ist besonders wichtig, dass für jede Suche ein eigenes Ranking durchgeführt wird – also jeweils die für die angegebenen Suchbegriffe *relevantesten Web-Seiten* angezeigt werden. Dieses automatische Ranking wird (bis jetzt) *nicht manuell* (und bis jetzt nicht gegen Bezahlung, wie bei manchen anderen Suchmaschinen) beeinflusst.

> Google sieht die Struktur des Webs als (gerichteten) Graphen, der Informationen über die Bedeutung einzelner Web-Seiten enthält. Die PageRank-Idee ist eng an die Zitat-Analysen von wissenschaftlicher Literatur angelehnt.

Verwandte Algorithmen:

Das **Webquery-System** setzte diese Idee bereits 1995 für die Bewertung von Webseiten ein. Allerdings wurde bei Webquery nur eine simple Zählung der „Backlinks“ vorgenommen – was relativ leicht manipulierbar ist.

Der **Companion-Algorithmus** und der **Cocitation-Algorithmus** zur Bestimmung ähnlicher Seiten gehen analog vor. Auch hier wird die Linkstruktur zumindest über zwei Ebenen von Knoten analysiert, was die Manipulation durch einen Webseiten-Betreiber schon erheblich erschwert.

Der **Kleinberg-Algorithmus** zur Bewertung von Hubs und Authorities geht noch einen Schritt weiter. Gewichte werden rekursiv berechnet, aber nur in der Treffermenge einer ganz konkreten Suchanfrage. Die Berechnung der Hub- und Authority-Gewichte muss daher online für jede einzelne Suchanfrage erfolgen, was aber die Antwortgeschwindigkeit herabsetzt.

PageRank setzt diese Ideen konsequent um. Für jede Seite in der Datenbasis, die eine möglichst große „Stichprobe" des gesamten Webs repräsentieren sollte, wird aus der Struktur ihrer Referenzen (Backlinks) iterativ ein globaler „Bedeutsamkeitswert" berechnet.

Ein weiterer, nicht zu unterschätzender Vorteil von Google ist die (bis jetzt) weit gehende Werbefreiheit. Die Geschwindigkeit vieler Suchmaschinen wird durch das Laden von Bannerwerbung und Interstitials begrenzt. Dazu kommt, dass Google *kein Portal* ist. Es gibt auf der Google-Homepage nur die Suchfunktion. Beiwerk wie Börsenkurse, Tagesnachrichten usw. fehlen (allerdings können aktuelle Kurse eingeblendet werden). Außerdem bietet Google eine Archivfunktion (Cache). Alle angezeigten Seiten liegen in gespeicherter Form vor, so dass diese auch bei Server- oder Verbindungsausfällen noch eingesehen werden können.

Die Ausstattung von Google ist eindrucksvoll: 2002 arbeitete ein Cluster aus über 10.000 Linux-Computern. 150 Millionen Anfragen pro Tag wurden in Sekundenschnelle bearbeitet. Die ständig aktualisierte Indexdatenbank umfasste zwei Milliarden Webdokumente. Darunter befindet sich ein ständig wachsender Anteil von Nicht-HTML-Quellen (z.B. pdf, MS Word, Powerpoint). Das Usenet-Archiv enthält über 700 Millionen Diskussionsbeiträge aus 20 Jahren.

Google bietet auch eine wahrscheinlichkeitstheoretisch fundierte Korrekturhilfe („did you mean …?", in Deutsch: „meinten Sie vielleicht …?". Dafür ist aber Wildcard-Suche nicht möglich (z.B. Vogel-Buch*), es muss immer der ganze Suchbegriff eingeben werden (z.B. Vogel-Buchverlag).

Anders als bei vielen Suchmaschinen benutzt Google als Standard die UND-Verknüpfung. Dies reduziert nämlich die Zahl aussageschwacher Suchergebnisse erheblich, führt aber auch meistens dazu, dass mehrere Suchanfragen nacheinander gestellt werden müssen.

Literatur über Google z.B. BRIN & PAGE (1998).

3.5 Positionierung eigener Web-Seiten

Mit der zunehmenden Konkurrenz ist eine Suchmaschinen optimierte Erstellung von Web-Seiten immer wichtiger geworden. Inzwischen gibt es eigene Branchen, die sich mit Search Engine Optimization (SEO) befassen.

Um eine Web-Seite in der Trefferliste einer Suchmaschine höher einzustufen, lassen sich die nachfolgenden Empfehlungen geben. Literatur z.B.: MAZE, MOXLEY & SMITH (1997), TUNENDER & ERVIN (1998), SONNENREICH & MACINTA (1998), LAURSEN (1998), PRINGLE, ALLISON & DOWE (1998).

Search Engine Persuasion umfasst **unlautere Mittel,** um eine Webseite bekannter zu machen. Ein Beispiel dafür ist das so genannte **Spamming.** Das ist die Wiederholung von Schlüsselbegriffen, um eine Suchmaschine dazu zu bewegen, eine Webseite in der Rangfolge höher zu stufen. Dazu z.B. werden dieselben Schlüsselbegriffe in Kommentarzeilen (dem HTML-Tag <!-- Kommentar -->) oder Feldern für Indexierungsbegriffe (META-Tag keywords) mehrfach wiederholt oder in der gleichen Schriftfarbe wie die Bildschirmhintergrundfarbe in das Webdokument eingefügt. Ein unlauterer Trick ist die Verwendung von Begriffen, die nichts mit dem eigentlichen Inhalt der Seite zu tun haben, die aber häufig gesucht werden. Zur Abwehr solcher unlauteren Methoden haben die meisten Suchmaschinenanbieter Strategien gegen Spamming entwickelt. Gegen widerrechtliche Verwendung von geschützten Begriffen (Firmen- und Produktnamen, eingetragenen Warenzeichen) können die Inhaber der Namensrechte juristisch vorgehen (vgl. die Seite Meta Tag Lawsuits in Search Engine Watch).

Eine vielfache Wiederholung von Begriffen im META-Element usw. bringt dabei gar nichts. Viele Suchmaschinen nehmen solche Seiten nicht mehr in ihre Datenbanken auf.

Viele Ranking-Algorithmen berücksichtigen Auftreten und Häufigkeit von Schlüsselwörtern (keywords) in Titel und URL. Es ist sinnvoll, **relevante Schlüsselwörter** in genau diesen Bereichen unterzubringen. Insbesondere Schlüsselbegriffe im TITLE-Element und in der Überschrift werden stärker gewichtet. Zudem wird der Titel von vielen Suchmaschinen in der Trefferliste angezeigt sowie die ersten Zeichen des Seitentextes (Bild 5.36):

Bild 5.36 Entscheidend ist, dass das Suchergebnis bereits auf der 1. Bildschirmseite sichtbar ist; mehr als die „TOP10" werden selten beachtet

Bekommen die eigenen Seiten bei gezielten Suchbegriffen nur einen schlechten Ranking-Platz, so sollte man die benutzten Begriffe im Dokument überprüfen. Ein Blick in die besser positionierten Webseiten hilft in diesem Hinblick oft weiter.

Gute Webseiten müssen fortlaufend überprüft, aktualisiert und verbessert werden. Der Inhalt muss **relevant** sein!

Einige META-Tags <meta name = robots> <content = siehe Liste> sind trotzdem sehr hilfreich:

- **noindex** – Dokument soll nicht indexiert werden.
- **index** – Dokument soll indexiert werden.
- **nofollow** – es sollen keine abgehenden Links verfolgt werden. Die Indexierung des aktuellen Dokuments ist allerdings erlaubt.
- **follow** – das Dokument soll indexiert werden und abgehenden Links kann durch Crawling nachgegangen werden.
- **all** – entspricht index und follow.
- **revisit-after** (content=„Anzahl Tage“) – soll den Crawler veranlassen, in der angegebenen Anzahl Tagen diese Seite erneut aufzusuchen.
- page-topic (content= „Stichworte“) – Angaben zum Themenbereich.
- page-type (content= „Stichworte“) – hier kann die Ressourcenart angegeben werden, z.B. Grafik, Linkliste, Eingabemaske usw.

Eine weitere Möglichkeit besteht darin, Metaangaben nicht direkt in die HTML-Datei einzufügen, sondern durch Verlinken mit einer externen Metadaten-Datei zu realisieren. Dafür kommt das Link-Tag zum Einsatz, das gleichfalls im Header der HTML-Datei festgelegt wird.

Da der jeweilige Verwendungszweck der Metatags leider durch das Fehlen eines Standards nicht eindeutig geregelt ist, kann daher auch nicht von einer einheitlichen semantischen Nutzung eines gleich bezeichneten Attributs unter mehreren Autoren bzw. Webseiten ausgegangen werden. Daher ist es günstig, auf den Dublin Core zurückzugreifen (siehe Kapitel 1.2.2).

Auch für Metatags nach der Dublin-Core-Spezifikationen gibt es im Web Werkzeuge zur Erzeugung von HTML-Quellcode, der direkt in die Datei eingefügt werden kann. Zu benennen wären der DC-Meta-Maker aus Baden-Württemberg und das Dublin Core Metadata Template des Nordic Metadata Projekts.

Es existieren auch Tools, die HTML-Code analysieren, ob er von Suchdiensten leicht durch das Crawling erfasst und indexiert werden kann und bei welchen Suchdiensten die betreffende Seite nach der Anmeldung (Submit) bereits im Index vorhanden ist. Auch eine Optimierung der Keywords, z.B. durch einen Vergleich mit häufig gesuchten Begriffen bestimmter Suchmaschinen, ist möglich. Beispiele: Webmasterplan.com und Makemetop.

Bei Grafiken erkennen viele Suchmaschinen nur den Dateinamen und die Bildbeschreibung im Alt-Text des HTML-Tags (<img src= „Dateiname“ alt= „Bildbeschreibung“>). Der Alt-Text sollte folglich eine aussagekräftige Beschreibung der Grafik enthalten. Außerdem ist es möglich, im File selbst relevante Schlüsselwörter unterzubringen.

Scripting: Bisher können die meisten Suchmaschinen mit Javascript und anderen Skriptsprachen nichts anfangen. Auch Javascript-Links werden meist nicht berücksichtigt. Ist beispielsweise die Navigation einer Seite mit Javascript erstellt, werden die tiefer liegenden Seiten nicht indiziert. Zudem haben viele Benutzer Javascript aus Sicherheitsgründen deaktiviert.

Frames: Die Spider der meisten Suchmaschinen bearbeiten keine Frame-Dokumente. Soll trotzdem nicht auf Frames verzichtet werden, so sollten folgende Aspekte bei der Gestaltung berücksichtigt werden: die Verwendung von Metatags auf der Masterseite des Framesets. Suchmaschinen lesen zumindest den Noframe-Bereich. Suchdienste, die keine Metatags beachten, z.B. Google, stellen sogar den Text im Noframes-Bereich in der Ergebnisliste dar. Meist angezeigte Sätze wie „Ihr Browser unterstützt keine Frames“ sind da wenig hilfreich. Der Text sollte vielmehr möglichst aussagekräftig sein und relevante Hinweise auf die anderen Seiten enthalten. Fehlen Links, haben Suchmaschinen keine Möglichkeit, zur Indizierung auf die anderen Seiten des Webauftritts zuzugreifen.

Verlinkung: Die Anzahl der Links, die auf eine Seite verweisen, haben einen Einfluss auf die Positionierung in der Trefferliste. Jeder Link wird als eine Empfehlung betrachtet, und je mehr solche Empfehlungen eine Seite hat, desto höher steigt sie im Ranking. Allerdings ist Link nicht gleich Link. Ein Eintrag im Webkatalog von Yahoo! ist beispielsweise mehr wert als ein Link von irgendeiner privaten Homepage. Generell wirken sich Links von Seiten, auf die selbst viele Links zeigen, günstiger auf das Ranking aus als Links von Seiten, auf die nur wenige Links zeigen. Vorteilhaft ist auch ein Verweis von einer Seite mit gleichem oder ähnlichem Thema; diese Anbieter werden als Experten für ihr Themengebiet betrachtet. Eine Empfehlung von ihnen in Form eines Links wird deshalb höher bewertet. Dabei zählen natürlich nur Links von externen Homepages auf die eigene Webseite. Die *Linkpopularität* einer Webseite lässt sich bei einigen Suchmaschinen über den Befehl link:URL feststellen (z.B. AltaVista, AllTheWeb).

Daneben gibt es Tools zur Überprüfung der Linkpopularität, die gleichzeitig mehrere Suchmaschinen überprüfen z.B. LinkPopularity.com, MarketLeap. Suchdienste, die dafür bekannt sind, dass sie Linkpopularität als Ranking-Kriterum verwenden, sind u.a. Google, AltaVista und Teoma. Durch häufige und umfangreiche Aktualisierung der Startseite eines Webauftritts kann man eine häufigere Indizierung erreichen.

Cloaking (aus engl. verhüllen) heißt, dass zwei oder mehrere verschiedene Versionen einer Homepage existieren: eine oder mehrere für die Roboter der Suchmaschinen optimierte und eine „normale" für die Benutzer. So kann die Positionierung im Ranking der Suchmaschinen verbessert werden, ohne dass in der Version für den Benutzer Abstriche bei der Gestaltung gemacht werden müssen. Zum anderen wird verhindert, dass der Besucher über den Seitenquelltext Zugriff auf Informationen über Metadaten und andere auf der Homepage verwendete Optimierungstechniken erhält.

Allerdings besteht die Gefahr des Missbrauchs. Die Suchbegriffe auf der für die Suchmaschine bestimmten Seite müssen nichts mit dem eigentlichen Inhalt der Homepage zu tun haben. Besucher werden unter Vorspiegelung falscher Tatsachen auf eine Seite gelockt. Bei Suchmaschinen können Cloaking-Seiten dauerhaft aus dem Index entfernt werden.

Spamming ist die zu häufige Wiederholung desselben Schlüsselwortes im Seitentext. Eine Rate von 1 bis 8 % wird im Allgemeinen akzeptiert. Nicht als Wiederholung zählen verschiedene Variationen eines Schlüsselwortes, z.B. Singular/Plural, Substantiv/Adjektiv oder zusammengesetzte Begriffe.

Linkfarmen die aus einem Netzwerk miteinander stark verlinkter Seiten bestehen. Einziges Ziel der Links ist es, die Linkpopularität zu erhöhen.

Mehrfachanmeldung einer Web-Seite unter verschiedenen URLs. Bei der redaktionellen Begutachtung fällt es im Allgemeinen auf, wenn eine Seite mehrmals auftaucht. Konsequenz ist dann die Streichung aller Seiten.

Evaluierung ist die Bewertung von Suchmaschinen, wo **Retrievaltests** durchgeführt werden: An unterschiedliche Suchdienste werden die gleichen Anfragen gestellt und die Trefferlisten miteinander verglichen. Tests solcher Art sind aber oft subjektiv, zu wenig methodisch und oft wird die Versuchsanordnung nicht bekanntgegeben.

NOTESS (2002) hat die Anzahl der Ergebnisse von 25 Einwort-Suchanfragen an 10 Suchmaschinen (AllTheWeb, AltaVista, DirectHit, Google, HotBot, iWon, MSN Search, Northern Light, Teoma, WiseNut) untereinander verglichen: Mit Abstand die meisten Treffer fand Google (8.371) vor WiseNut (über 5.009) und AllTheWeb (über 4.388). Weit abgeschlagen waren Teoma (1.839), iWon (778) und DirectHit (259). Auch bei den einzelnen Anfragen belegte Google in fast allen Fällen Platz 1.

Nicht die Anzahl, sondern die **Relevanz** der Ergebnisse ist entscheidend: Alle Treffer sollen relevant sein, aber es sollten alle relevanten Webseiten in der Trefferliste enthalten sein.

Mit der Relevanz der Treffer befasste sich eine Untersuchung der Stiftung Warentest (2001). An verschiedene deutschsprachige Suchmaschinen (darunter Google AltaVista, Lycos, Metager, Yahoo, Fireball, Dino, Web.de, Dino, MSN, Abacho, Acoon) wurden 10 Anfragen aus verschiedenen Bereichen gestellt und die Relevanz der ersten 10 Treffer nach verschiedenen Kriterien beurteilt. Lediglich Google lieferte auf alle Anfragen mindestens eine völlig zutreffende Antwort. Alta Vista und Metager konnten neun, Lycos acht Fragen beantworten. MSN dagegen fand nur für eine, Acoon sogar für gar keine Anfrage eine relevante Webseite.

Zu ähnlichen Ergebnissen führte ein Experiment der Zeitschrift Tomorrow (2002). Verglichen wurden Google, AltaVista, AllTheWeb, Hotbot, WiseNut, Teoma, Lycos, Abacho, Fireball und Acoon. Wiederum nur Google fand zu fast allen Anfragen die wichtigen Webseiten. Vor allem bei der Suche mit Boole'schen Operatoren oder Phrasen zeigten einige Suchdienste wie z.B. Lycos enorme Schwächen.

Ähnliche Ergebnisse zeigte bereits eine ältere Untersuchung der Zeitschrift Chip (2000), die unter anderem Google, Fireball, AltaVista, Lycos und Acoon miteinander verglich. Google fand mehr relevante Seiten und führte seltener zu toten Links als alle anderen.

Von GRIESBAUM, RITTBERGER & BEKAVAC (2002) wurden die *deutschsprachigen* Suchmaschinen AltaVista.de, Fireball.de, Google.de und Lycos.de bezüglich der Retrievalleistung untersucht. Anders als in den vorher beschriebenen Untersuchungen wurde eine wesentlich größere Anzahl von Anfragen (56) eingesetzt und von 28 Testpersonen beurteilt. Testsituation und Beurteilung der Treffer waren kontrolliert und standardisiert, z.B. konnten die Beurteiler nicht erkennen, von welchen Suchdiensten die Ergebnisse stammten. Berücksichtigt wurden nur die ersten 20 Ergebnisse einer Trefferliste. Bei 56 Suchanfragen waren somit maximal 1120 Treffer möglich. Als *relevant* galten Dokumente, die sich entweder selbst auf das Thema bezogen oder auf andere, relevante Seiten verwiesen. Kriterium für die Effektivität einer Suchmaschine war die Anzahl relevanter Treffer im Verhältnis zur Gesamtzahl der Treffer (Precision). Dabei wurde der Anteil relevanter Treffer für jede einzelne Suchanfrage bestimmt und über alle Suchanfragen hinweg. Es erzielte Google.de auch bei dieser Untersuchung die besten Ergebnisse.

Die Relevanz der Treffermenge hängt auch mit der Aktualität des Index einer Suchmaschine zusammen. Nur wenn der aktuelle Inhalt einer Seite erfasst ist, kann man entscheiden, ob sie relevante Information enthält. Außerdem dürfte auch der Anteil toter Links umso geringer sein, je häufiger Webseiten von den Spidern der Suchmaschinen besucht werden.

4 Modulkurzzusammenfassung

Durch **Metadaten** kann die **Interoperabilität** (Austauschbarkeit) von Inhalten (contents) gewährleistet werden. Ein Metadaten-Record besteht aus einem Satz (set) verschiedener Attribute, die das zu beschreibende Dokument spezifizieren. Die Syntax beschreibt, in welcher Form Metadaten ausgetauscht werden (z.B. XML), die Semantik beschreibt, *welche* Metadaten für eine Ressource eingesetzt werden (Vokabular, Schema). Der **Dublin Core (DC)** dient als *kleinster gemeinsamer Nenner* verschiedener Standards. **RDF** stellt eine Infrastruktur dar, um *Codierung, Austausch und Wiederverewendung* von Metadaten zu ermöglichen. RDF besteht aus Ressource, Property und Statement. Ein erster Ansatz, Lernmaterial zugänglich zu machen, ist das Projekt **GEM.** Eine *Erweiterung von DC* ist das **Warwick-Framework. IMS**-Metadaten bauen auf DC auf und dienen zur Ermittlung der Qualität von Lernmaterial im Web. Der weitverbreitetste Standard ist **IEEE LOM,** bei dem beliebige digitale „atomare" Objekte (Learning Objects) definiert werden. Das *Referenzmodell* **SCORM** gewährleistet die Wiederverwendbarkeit (reusability) und die Austauschbarkeit von LOM-basierenden Objekten. Das SCORM-Content-Aggregationsmodell besteht aus einem Content Model, aus einem Metadata Dictionary und aus einem Content Packaging. Letzteres definiert, wie das beabsichtigte Verhalten und die Struktur einer Menge von Lern-Ressourcen beschrieben und für den Austausch zwischen verschiedenen Systemen (z.B. Lernplattformen) gepackt wird.

Das Ziel eines semantischen Webs ist es, Menschen vor der Überflutung von Information aus dem Web zu befreien und wirklich relevante und gewünschte Inhalte anzubieten. Dafür ist eine gemeinsame Sprache wichtig. Nach BERNERS-LEE (2001) soll das **semantic web** auf vier Grundsäulen basieren: URI, Unicode-Standard, XML und RDF.

Die **Ontologie** stellt eine *formale Beschreibung von Objekten und Beziehungen* dar, die dann für eine Gruppe von Personen begriffsbildend sind. Damit hat man ein semantisches Modell, das den Austausch von „Wissen" (wir wissen aus Basiswissen Multimedia aber, dass wir nur *Information* austauschen können) zwischen Mensch und Maschine erleichtert. Dazu existieren Ontologie-Sprachen wie z.B. OIL, DAML+OIL und OWL.

Bei der **Suche im Web** unterscheidet man zunächst zwischen der Suche auf lokalen WWW-Servern und einer katalog- bzw. verzeichnisbasierten Suche und eine roboterbasierten Suche. Bei **Suchmaschinen** sind die folgenden Definitionen sehr nützlich: Precision, Recall, Relevance, Ranking, Authority, Hub. Die Funktionsweise einer Suchmaschine wird durch die vier Grundfunktionen beschrieben: Dokumentbeschaffung (Akquisition), Indexierung, Aktualisierung und Anfragebearbeitung.

5 Modulanhang

5.1 Literatur

5.1.1 Bücher

ALBRECHT, F. (1993): *Strategisches Management der Unternehmensressource Wissen. Europäische Hochschulschriften Bd. 1367*. Frankfurt: Lang.

POTEMPA, T.; FRANKE, P.; OSOWKSI, W.; SCHMIDT, M. (1998): *Informationen finden im Internet. Leitfaden für die gezielte Online-Recherche.* München: Hanser.

HENNE, H. (1972): *Semantik und Lexikographie. Untersuchungen zur lexikalischen Kodifikation der deutschen Sprache.* Berlin: de Gruyter.

HESS, K.; BRUSTKERN, J.; LENDERS, W. (1983): *Maschinenlesbare deutsche Wörterbücher. Dokumentation, Vergleich, Integration.* Tübingen: Niemeyer.

HOFMANN, M.; SIMON, L. (1995): *Problemlösung Hypertext. Grundlagen – Entwicklung – Anwendung.* München, Wien: Hanser.

KAPPEL, G.; PRÖLL, B.; REICH, S.; RETSCHITZEGGER, W., HRSG. (2003): *Web Engineering: Systematische Entwicklung von Web Anwendungen.* Hannover: dpunkt.

KOLKE, ERNST-GERD VON (1996): *Online-Datenbanken: systematische Einführung in die Nutzung elektronischer Fachinformation.* 2. Auflage. München: Oldenbourg.

KÖNIGER, PAUL; REITHMAYER, WALTER (1998): *Management unstrukturierter Informationen.* Frankfurt/Main: Campus.

KUHLEN, R. (1991): *Hypertext. Ein nicht-lineares Medium zwischen Buch und Wissensbank.* Berlin: Springer.

MEISS, BRIGITTE (1998): *Information Retrieval und Dokumentmanagement im Multimediazeitalter.* Frankfurt/Main: DGD.

OTTO, MICHAEL (1998): *Suchstrategien im Internet.* Bonn: Thomson.

RIEHM, U.; BÖHLE, K.; WINGERT, B.; GABEL-BECKER, I. (1992): *Elektronisches Publizieren: eine kritische Bestandsaufnahme.* Berlin et al.: Springer.

ZIMMER, DIETER (2000): *Die Bibliothek der Zukunft: Text und Schrift in den Zeiten des Internet.* Köln: Hoffmann und Campe.

5.1.2 Artikel

BEKAVAC, BERNARD (1996): Suchverfahren und Suchdienste des World Wide Web. *Nachrichten für Dokumentation,* 47, 195–213.

BLUMENTHAL, A.; LEMNITZER, L.; STORRER, A. (1988): Was ist eigentlich ein Verweis? Konzeptuelle Datenmodellierung als Voraussetzung computergestützter Verweisbehandlung. In: HARRAS, GISELA (Hrsg.): *Das Wörterbuch: Artikel und Verweisstrukturen.* Düsseldorf: Schwann, 351–373.

CARNEVALI, MICHAEL (1996): Lost in Cyberspace? Informationssuche mit Search Engines im World Wide Web. *Cogito,* 4/96, 4–8.

GRIESBAUM, J.; RITTBERGER, M.; BEKAVAC, B. (2002): Deutsche Suchmaschinen im Vergleich: AltaVista.de, Fireball.de, Google.de und Lycos.de. In: HAMMWÖHNER, R., WOLFF, C., UND WOMSER-HACKER, C. (Eds.): *Proceedings des 8. Internationalen Symposiums für Informationswissenschaft*, 201–223.

LENNARTZ, S. (1999): Ich bin wichtig! Promotion-Maßnahmen für suchdienstgerechte Webseiten. *Magazin für Computertechnik,* 23, 180–186.

SCHWEIBENZ, W. (1999): Proactive Web Design: Verbesserung der Auffindbarkeit von Webseiten durch Suchmaschinen. *Nachrichten für Dokumentation,* 50 (7), 389–396.

SCHULMEISTER, R. (2002): Taxonomie der Interaktivität von Multimedia – Ein Beitrag zur aktuellen Metadaten-Diskussion. *it+ti,* 44, 4, 193–199.

PANYR, J. (1987): Information-Retrieval-Systeme: State of the Art. *HMD,* 133, 15–16.

RUSCH-FEJA, DIANN (1997): Metadaten und Strukturierung elektronischer Information. *Nachrichten für Dokumentation,* 48, 295–302.

STEINACKER, A.; SEEBERG, C.; FISCHER, S.; STEINMETZ, R. (1999): MultiBook: Meta-data for Webbased Learning Systems. In: *Proceedings of the 2nd International Conference on New Learning Technologies.*

5.1.3 Books in English

BAEZA-YATES, RICARDO; RIBEIRO-NETO, BERTHIER (1999): *Modern Information Retrieval.* New York: ACM.

CAPLAN, PRISCILLA (2003): *Metadata Fundamentals for All Librarians.* London: Library Association Publications.

CLEVELAND, DONALD B.; CLEVELAND, ANA D. (1990): *Introduction to Indexing and Abstracting.* Second Edition. Englewood (CO): Libraries Unlimited.

FENICHEL, CAROL H.; HOGAN, THOMAS H. (1984): *Online Searching: A Primer.* 2nd Edition. Medford (NJ): Learned Information.

FENSEL, DIETER (2003): *Spinning the Semantic Web.* Boston (MA): MIT.

FENSEL, DIETER (2003): *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce.* Berlin et al.: Springer.

FOSKETT, A. C. (1982): *The Subject Approach to Information.* 4th Ed. London: Bingley.

HARTER, STEPHAN (1986): *Online Information Retrieval. Concepts, Principles, and Techniques.* Orlando (FL): Academic Press.

HAYNES, D. (2004): *Metadata for Information Management and Retrieval.* London: Library Association Publications.

HJELM, JOHAN (2001): *Creating the Semantic Web with RDF.* New York: Wiley.

MAZE, SUSAN; MOXLEY, DAVID; SMITH DONNA J. (1997): *Neal-Schuman Authoritative Guide to Web Search Engines.* New York/London: Neal-Schuman Publishers.

POWERS, SHELLEY (2003): *Practical RDF.* London: O'Reilly.

SONNENREICH, WES; MACINTA, TIM (1998): *Web Developer.com Guide to Search Engines.* New York, NY: John Wiley & Sons.

WILEY, DAVID (2000): *Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy.* Online Book „The Instructional Use of Learning Objects", University of Utah (http://reusability.org/read).

5.1.4 Articles in English

BHARAT, KRISHNA; BRODER, ANDREI (1998): A Technique For Measuring The Relative Size and Overlap of Public Web Search Engines. *Computer Networks and ISDN Systems,* 30, 379–388.

BERNERS-LEE, TIM; HENDLER, JAMES; LASSILA, ORA (2001): The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American,* 284 (5), May 2001, 34–43.

BRIN, SERGEY; PAGE, LAWRENCE (1998): The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems,* 30 (1–7), 107–117, Online at: http://www-db.stanford.edu/pub/papers/google.pdf.

BHARAT, KRISHNA; CHANG, BAY-WEI (2001): Web Search Engines: Algorithms and User Interfaces. *Tutorial at CHI 2001,* Seattle, April 2001.

DONG, Y., FENSEL, D., STORK, H.-G. (2003). The Semantic Web: from Concept to Percept. *Journal of the Austrian Society for Artificial Intelligence,* 22 (1), 5–18.

FENSEL, D.; HORROCKS, I.; VAN HARMELEN, F.; DECKER, S.; ERDMANN, M.; KLEIN, M. (2000): OIL in a Nutshell. In: DIENG, R.; CORBY, O. (Eds.). *Lecture Notes in Artificial Intelligence,* 1935, 1–16, Berlin et al.: Springer.

GRUBER, T. R. (1993): A translation approach to portable ontologies. *Knowledge Acquisition,* 5 (2), 199–220.

HEFLIN, JEFF (2003): Web Ontology Language (OWL) Use Cases and Requirements, Online at: www.w3.org/TR/2003/WD.webont-req-20030331

HORROCKS, I.; PATEL-SCHNEIDER, P.; VAN HARMELEN, F. (2001): Reviewing the Design of DAML+OIL: An Ontology Language for the Semantic Web. In: *Eighteenth National Conference on Artificial Intelligence*, Edmonton, Alberta, AAAI Press.

HOLZINGER, A.; WASCHER I.; STEINMANN, C. (2003): Design and Development of a LO-Editor for the Virtual Medical Campus Graz. In: BODE, A.; DESEL, J.; RATHMAYER, S.; WESSNER, M. (Ed.) *Lecture Notes in Informatics* (LNI), P-37 DeLFI 2003, Bonn: Köllen Verlag, 440–449.

HOLZINGER, A.; KLEINBERGER, T; MÜLLER, P. (2001): Multimedia Learning Systems based on IEEE Learning Objects Metadata (LOM). *Educational Multimedia, Hypermedia and Telecommunication,* Association for the Advancement of Computing in Education, Charlottesville, VA, 772–777.

HYVÖNEN, EERO (2001): The Semantic Web: The Internet of Meanings. In: *Proceedings of Semantic Web Kick-Off,* Finland (May, 2001), 3–25.

LAWRENCE, STEVE.; GILES, C. LEE (1998): Searching the World Wide Web. *Science,* April 3, Vol. 280. 98–100.

LYNCH, CLIFFORD (1997): Searching the Internet. *Scientific American,* 276, 3, 44–48.

SHERMAN, CHRIS; PRICE, GARY (2001): The Invisible Web: Uncovering Information Sources Search Engines Can`t See. Medford (NJ): CyberAge.

TUNENDER, HEATHER; ERVIN, JANE (1998): How to Succeed in Promoting Your Web Site: The Impact of Search Engine Registration on Retrieval of a World Wide Web Site. *Information Technology and Libraries,* 17 (3) September 1998, 173–179.

PRINGLE, G.; ALLISON, L.; DOWE, D.L. (1998): What is a tall poppy among Web pages? *Computer Networks and ISDN Systems,* 30, (1998) 369–377.

MCGUINESS, D.; FIKES, R.; HENDLER, J.; STEIN, L. (2002): DAML+OIL: An Ontology Language for the Semantic Web. *IEEE Intelligent Systems,* Sep./Oct. 2002.

MINTZ, ANNE P. (Hrsg.) (2002): Web of Deception: Misinformation on the Internet. Medord (NJ): CyberAge.

CARRIERE, J.; KATZMAN, R. (1997): WebQuery: Searching and visualizing the Web through connectivity. In: *Proceedings of the 6th International WWW Conference,* 1997. Online: http://www.cgl. uwaterloo.ca/Projects/Vanish/webquery-1.html

DEAN, J.; RAUCH-HENZINGER, M. (1999): Finding related pages in the world wide web. *Computer Networks,* 31 (11–16), 1467–1479.

KLEINBERG, JON M. (1999): Authoritative sources in a hyperlinked environment. *Journal of the ACM,* 46 (5), 604–632. *http://www.cs.cornell.edu/home/kleinber/auth.pdf*

MCBRYAN, OLIVER A. (1994): GENVL and WWW: Tools for Taming the Web. In O. NIERSTARSZ, ed., Proceedings of the first International World Wide Web Conference, 15, CERN. http://www.cs.colorado.edu/~mcbryan/mypapers/www94.ps.

PAGE, LAWRENCE (1999): Pagerank: Bringing order to the web. Online Powerpoint presentation at *http://hci.stanford.edu/~page/papers/pagerank/index.htm*

PAGE, LAWRENCE; BRIN, SERGEY; MOTWANI, RAJEEV; WINOGRAD, TERRY (1998): The pagerank citation ranking: Bringing order to the web. Technical report, 1998. http://stanford.edu/~backrub/ pageranksub.ps

5.2 Internet-Links

Aktualisierte Internet-Links zu diesem Modul sind auf der Buchhomepage www.basiswissen-it.at unter IN – Modul 5: Semantik verfügbar!

5.3 Prüfungsfragen

Fragen-Typ 1: Dichotome Ja/Nein-Entscheidungen:

01	Im RDF bezeichnet Property eine Beziehung zwischen einem Wert und einer Ressource.	❒ Ja ❒ Nein	
02	Im RDF-Schema werden Klassen als abgerundete Rechtecke und Ressourcen durch einen schwarzen Punkt gekennzeichnet.	❒ Ja ❒ Nein	
03	Ein PICS-Label wird zwischen den Tags <body> </body> in eine HTML-Datei eingefügt.	❒ Ja ❒ Nein	
04	Das Warwick-Framework ist kein Metadatenformat, sondern ein Modell einer nicht rekursiven Container-Architektur.	❒ Ja ❒ Nein	
05	In IEEE LOM bezeichnet die Kategorie „Relation" Informationen über die verwendeten Metadaten über das jeweilige Lernobjekt.	❒ Ja ❒ Nein	
06	Das SCORM CAM besteht aus dem Content Model, einem Metadata Dictionary und dem Content Packaging.	❒ Ja ❒ Nein	
07	Recall ist der Anteil an relevanten Suchergebnissen: Je weniger Recall, desto mehr irrelevante Dokumente liefert eine Suchmaschine.	❒ Ja ❒ Nein	
08	Die Suchmaschine Google bewertet eine Webseite auch danach, wie viele Links aus dem Web auf sie verweisen.	❒ Ja ❒ Nein	
09	Das META-Tag <meta name = robots> <content = nofollow> veranlasst Suchmaschinen, nur die „Startseite" zu indexieren.	❒ Ja ❒ Nein	
10	Cloaking ist die häufige Wiederholung eines Schlüsselwortes in einem Webseitentext.	❒ Ja ❒ Nein	

Fragen-Typ 2: Mehrfachauswahlantworten (Multiple Choice):

01	SCORM ... ❒ a) ... ist ein Referenzmodell zur Integration verschiedener Standards. ❒ b) ... bezeichnet jede instruktionale Einheit als SCO. ❒ c) ... erlaubt die Interoperabilität von Lernmaterial. ❒ d) ... besteht aus *Content Aggregation Model* und *Run-Time Environment*.	
02	DAML-OIL ... ❒ a) ... baut auf den Prinzipien von XML und RDF auf. ❒ b) ... erlaubt es, Objekte und deren Beziehungen untereinander zu definieren. ❒ c) ... verwendet die Konzepte *Klasse und Property,* um Strukturen zu beschreiben. ❒ d) ... bildet eine Erweiterung von RDFS.	
03	Zentral für den Entwurf von DC waren ... ❒ a) ... die Möglichkeit zu einem „third-person-rating". ❒ b) ... das Verwertbarkeitsdefizit von Lernmaterial zu beheben. ❒ c) ... die Sicherung eines internationalen Konsens. ❒ d) ... die Beschreibung von DLOs.	
04	Die Syntax ... ❒ a) ... beschreibt, *welche* Metadaten für eine Ressource eingesetzt werden können. ❒ b) ... benötigt eigene Vokabulare bzw. Schemas. ❒ c) ... kann sehr gut mit XML dargestellt werden. ❒ d) ... stellt dar, in welcher Form Metadaten ausgetauscht werden können.	
05	Zur Kategorie „Educational" in IEEE LOM zählt ... ❒ a) ... Interaktivität. ❒ b) ... Zielgruppe. ❒ c) ... Schwierigkeit für die typische Zielgruppe. ❒ d) ... typische Bearbeitungsdauer.	
06	Ranking bezeichnet ... ❒ a) ... den Anteil an relevanten Suchergebnissen. ❒ b) ... die Reihenfolge der Auflistung nach Relevanzbewertung. ❒ c) ... Anzahl der gefundenen relevanten Dokumente aus der Grundgesamtheit. ❒ d) ... den Grad der Zuverlässigkeit der gefundenen Information.	
07	Die Aufgaben einer Suchmaschine umfassen ... ❒ a) ... Indexierung. ❒ b) ... Dokumentbeschaffung. ❒ c) ... Aktualisierung. ❒ d) ... Anfragebearbeitung.	
08	Zu den besten Mitteln, um eigene Web-Seiten besser zu positionieren, zählen ... ❒ a) ... die vielfache Wiederholung von Keywords im META-Element. ❒ b) ... die *fortlaufende* Aktualisierung, Überprüfung und Ergänzung des Contents. ❒ c) ... die Verwendung mehrerer Web-Seiten in verschiedenen Versionen. ❒ d) ... die Verwendung eines aussagekräftigen Titels und Kurzbeschreibung.	

5.4 Übungen

- Wählen Sie Begriffe aus Ihrem Umfeld. Verwenden Sie jetzt semantische Relationen (d.h. Synonymie, Antonymie, Hyperonymie, Hyponomie, Meronomie, Troponomie), um die Begriffe in Beziehung zu setzen!
- Testen Sie drei verschiedene Suchmaschinen mit einem Testwort und vergleichen Sie die Ergebnisse! Was fällt Ihnen auf? Welche Vorteile und Nachteile ergeben sich jeweils für einen bestimmten Suchzweck?
- Bewerten Sie Metadaten-Ansätze aus Ihrer Sicht (Lernender). Machen Sie Verbesserungsvorschläge. Wiederholen Sie das Ganze aus der Sicht eines Lehrenden. Erstellen Sie ein SCORM-kompatibles Lernobjekt.

5.5 Diskussionsfragen

- Diskutieren Sie die Begriffe Daten, Information und Wissen (vgl. Band 1). Wie hängen diese Begriffe nach diesen Quellen zusammen? Sind Sie mit den Definitionen einverstanden?
- Spekulieren Sie über die Vorteile eines zukünftigen „Semantic Web". Welche Probleme könnten gelöst werden? Diskutieren Sie mit Kollegen über den Ansatz und stellen Sie Ihren Standpunkt dar.
- Bilden Sie verschiedene Gruppen. Jede Gruppe soll nun eine Suchmaschine vorstellen und möglichst deren Vorteile den anderen Gruppen „on-line" demonstrieren.

5.6 Timeline: Semantik

100 v. Chr. Bibliothekare der Antike verwenden Metadaten zur Archivierung von Pergamentrollen.

1520 Buchtitel bei gedruckten Büchern werden erstmals zur Suche eingesetzt, also als Metadaten.

1945 VANNEVAR BUSH sieht in seinem MEMEX (siehe BW MM Band 2) das rasche Wiederfinden von Information als essenziell an.

1991 Die ersten Webinhalte (Content) werden manuell in einer zentralisierten „WWW Virtual library" katalogisiert.

1991 April, das Programm WAIS (publisher-fed full text search engine) wird von BREWSTER KAHLE vorgestellt.

1992 Veronica wird als Gopher crawler search engine vorgestellt.

1993 Content im Web wird erstmals außerhalb des CERN manuell über das „WWW Virtual library system" indexiert.

1994 Der Lycos WWW Crawler wird von MAULDIN (Carnegie Mellon) vorgestellt.

1995 Infoseek WWW Crawler search engine wird vorgestellt.

1995 März, STUART WEIBEL und ERIC MILLER leiten den 1. Workshop der Dublin Core Metadata Initiative (DCMI, dublincore.org) in Dublin, Ohio.

1996 Der zweite DC Workshop in Warwick (UK) findet statt.

1998 Google wird an der Stanford University vorgestellt.

2002 Dezember, die Suchmaschine Google erfasst exakt 3,083,324,652 Webseiten und ist damit der Favorit unter den Suchmaschinen, insbesondere im wissenschaftlichen Bereich (SCIRCUS).

5.7 Glossar

ANSI American National Standards Institute. Arbeitet mit Industrie, ISO und IEC zusammen.

ASN.1 Abstract Syntax Notation One ist eine Norm für die computerunabhängige Darstellung von Daten sowie deren Umwandlung.

Bots Auch Web Bots, Spider oder Crawler genannt. Das sind von Suchservern geschickte Programme, die im WWW Informationen sammeln.

Cloaking ist das IP-abhängige Ausliefern unterschiedlicher Versionen ein und derselben HTML-Seite. Das kann natürlich für Spamming missbraucht werden.

DC Dublin Core. Nach Dublin (Ohio, USA) benannter Metadaten-Standard.

DTD Document Type Definition. Definiert aus einer Grammatik (z.B. SGML) einen bestimmten Dokumenttyp (z.B. HTML) und legt fest, was und in welcher Struktur dort auftauchen kann.

EBNF Extended Backus-Naur-Form. Formales Beschreibungsmittel der Syntax einer Programmiersprache.

GEM Gateway to Educational Material. Projekt in USA, um den Zugang zu hochwertigen Lern- und Lehrmaterialien zu ermöglichen.

IEEE Institute of Electrical and Electronics Engineers. Gesprochen: „ai-trippl-i". Arbeitet an technischen Standards, die oft weltweite Anerkennung finden.

IMS Instructional Managing System. Spezifikation, die auf DC aufbaut und speziell für das Bildungswesen entwickelt wurde.

ISO International Organization for Standardization. Wird in den USA durch die ANSI und in Deutschland durch das Deutsche Institut für Normung (DIN) vertreten.

LOM Learning Objects Metadata. IEEE Standard P1484.12, definiert und beschreibt Lernobjekte in Kategorien.

Metadaten sind Information über Information. Ein Mittel, kontextuelle semantische Informationen zu liefern, die dem Empfänger hinreichende Auskunft über ihre Quelle geben.

MPEG Motion Picture Experts Group. Komitee zur Entwicklung von Multimedia-Standards.

NISO National Information Standards Organization. Eng mit ANSI kooperierendes Standardisierungsinstitut, liefert z.B. NISO Z39.50.

PICS Platform for Internet Content Selection. Einfacher Metadaten-Mechanismus, um Webinhalte zu bewerten (content rating).

Ressource ist die beschriebene Informationsquelle. Ursprünglich im Zusammenhang mit Dublin Core nur digitaler Natur: HTML-Seiten, JPEG-Dateien usw.

RDF Resource Description Framework. Direkt durch EBNF definiertes Datenmodell als Verwirklichung des Warwick-Frameworks.

Fuzzy Suche ist eine „unscharfe Suche", bei der durch komplizierte Berechnungen basierend auf dem eigentlichen Suchwort ähnliche Begriffe generiert und dann zusätzlich nach diesen gesucht wird.

Imagemaps sind Grafiken, die innerhalb eines Bildes zu verschiedenen anderen Seiten verweisen. Das kann eine Reihe Schalter sein, die zu einem Bild zusammengefasst sind, oder eine Landkarte, die bei einem Klick auf ein Land auf die jeweils zugehörende Seite verweist.

Link Relevancy Je mehr Links auf eine Seite verweisen, desto höher das „Gewicht" und damit der Platz in der Ergebnisliste. Die Realisierung gestaltet sich enorm aufwendig. Ein komplexes Muster von Wertigkeiten muss erfasst werden, denn neben der Zahl der Links wird auch festgestellt, wie oft die Seite, von der der Verweis stammt, ihrerseits verlinkt ist. Ein Link von einer Seite, auf die häufig verwiesen wird, wiegt „schwerer" als der von einer weniger oft verlinkten Seite.

Meta-Tags stehen im Header eines HTML-Dokuments und werden vom Browser nicht angezeigt. Meta-Tags erlauben z.B. Stichwörter oder eine Zusammenfassung des Seiteninhalts, die von Suchmaschinen ausgewertet wird.

Robots sind automatische Programme, die neben dem Sammeln von Informationen eine Reihe anderer Aufgaben erfüllen. Robots sind ein wichtiger Bestandteil von Suchmaschinen. Siehe auch Bots.

Spamdexing ist ein Kunstwort aus Spamming und Indexing und bezeichnet das „Zumüllen" von Suchmaschinen mit Webseiten, um bei möglichst vielen oder zumindest bestimmten Suchanfragen ganz oben in den Trefferlisten zu stehen.

Wildcard ist ein Platzhalter (z.B. *), der für verschiedene, beliebige Zeichen stehen kann und bei der Sucheingabe für die Trunkierung verwendet wird.

5.8 Lösungen

Lösungen zu Fragen-Typ 1: 01 Nein; 02 Ja; 03 Nein; 04 Nein; 05 Nein; 06 Ja; 07 Nein; 08 Ja; 09 Ja; 10 Nein

Lösungen zu Fragen-Typ 2: Richtig sind: 01 a) b) c) d); 02 a) b) c) d); 03 c) d); 04 c) d); 05 a) b) c) d); 06 b); 07 a) b) c) d); 08 b) d)