

Austausch von Lernressourcen durch Peer-to-Peer Netzwerken mit Hilfe von Lokalen Ontologie-Basierten Repositorien

Luka Divac-Krnica, Cornelia Seeberg und Ralf Steinmetz

KOM - Multimedia Kommunikation
Fachbereich Elektrotechnik und Informationstechnologie
Technische Universität Darmstadt
Merckstr. 25, 64283 Darmstadt

Abstract: In Abwesenheit des "Semantic Web" sucht man nach alternativen Lösungen, um lokale semantische Netze zu erstellen. Peer-to-Peer-Netzwerke können verwendet werden, um auf ontologie-strukturierten Repositorien Suchabfragen durchführen zu können. Es kann behauptet werden, dass Peer-to-Peer-Netze schneller performant und effektiv werden können, als das allumfassende "Semantic Web" entstehen kann. Sichere Nutznießer von solchen semantisch angereicherten Peer-to-Peer-Netzen wären Universitäten, die sich durch Austausch der Lerninhalte gegenseitig bereichern würden.

1. Einleitung

Die Erweiterung der heutigen Peer-to-Peer-Anwendungen (P2P) um eine semantische Repräsentation der Objekte in den Repositorien würde einen Auftrieb zur Präzision der Suchresultate geben. Verschiedene Bildungsinstitutionen wie Universitäten würden durch Verbreitung der Lerninhalte, insbesondere derjenigen mit multimedialen Elementen, durch solche P2P-Netzwerke profitieren.

Ein Verbund von Lehrstühlen mit verwandten Themen würde den wissenschaftlichen Austausch fördern und die Lerninhalte, die auf Festplatten der teilnehmenden Peers liegen, in einem verteilten System verfügbar machen.

Teil 2 dieser Arbeit zeigt Implikationen der fehlenden Semantik in P2P-Netzwerken und das Potential des künftigen Semantic Web. Teil 3 zeigt die verwandten Arbeiten hierzu auf. Im Abschnitt 4 stellen wir unseren Vorschlag, Domain Specific Ontologies zu verwenden, vor.

2. Die fehlende Semantik in Peer-to-Peer Netzwerken

Eine rechnergestützte semantische Suche nach Objekten im heutigen Web ist nicht möglich, da Rechner nicht fähig sind, die Semantik eines Objektes zu interpretieren. Die Vision des "Semantic Web" (siehe unten) setzt hierbei Maßstäbe für Zukunftsentwicklungen. Jedoch ist die Vollendung dieser Idee aufgrund des Umfangs des heutigen Webs und der darin gehaltenen Daten, die um eine weitere semantische Schicht erweitert werden müssten, nicht abzusehen. Bevor das Semantic Web mit all seinen Möglichkeiten tatsächlich in Betrieb geht, kann auf Technologien zurückgegriffen werden, die schon länger etabliert sind und schnell weiterentwickelt werden können, wie z.B. das P2P-Kommunikationsmodell. Diese Technologie lässt dem Benutzer völlige Kontrolle über seine Ressourcen und zielt in dem oben skizzierten Anwendungsszenario auf diejenigen Benutzerkreise, die bereit sind, den Aufwand, Metadaten zu erstellen, auf sich zu nehmen.

2.1 Semantic Web

Das "Semantic Web", eine Vision, die von Tim Berners-Lee entworfen wurde, wird ein mächtiges Tool sein, mit dessen Hilfe es "für jeden möglich sein wird, irgendetwas über irgendetwas zu sagen" [Übersetzung aus THL01]. Die Verwendung des "Semantic Web" wird nicht nur auf die verbesserte Suche beschränkt sein, sondern auch das Aussuchen und Abrufen verschiedener Dienste auf dem Web möglich machen. Ebenso können diese Dienste aufgrund plötzlicher Ereignisse und Benachrichtigen von Benutzern über diese Änderungen angepasst werden. Autonome Software-Agenten sollen durch das Web wandern, aus Dokumenten Informationen extrahieren, nach anderen Agenten suchen und mit ihnen Daten austauschen und zwar ohne menschliches Eingreifen.

Damit eine präzise Suche stattfinden kann, müssen die Inhalte der Dokumente von Rechnern interpretiert werden können. Das macht die Annotation mit Metadaten notwendig, die die semantische Bedeutung der Dokumente ausdrückt, solange die Inhalte der Texte, Bilder, Audio- und Videoströme selbst nicht interpretiert werden können. Relationale Informationen zwischen Dokumenten müssen spezifiziert und Dokumente müssen mit diesen markiert werden. Hierzu sollen Ontologien für ein spezifisches Bereich der realen Welt (Domain Specific Ontology) erstellt werden. Die Objekte sind hierbei in Ober- und Unterbegriffe angeordnet, die zueinander durch hierarchische und andere semantische Relationen verbunden sind.

Beim Aufbau des Semantic Web werden RDF (Resource Description Framework) und RDFS (Resource Description Framework Schema) eine entscheidende Rolle spielen [RDF99]. RDF stellt Aussagen mittels "triples" bereit, mit denen Relationen zwischen den Objekten durch gerichtete Korrelationen dargestellt werden können. Zum Aufbau komplexer Ontologien existieren einige Beschreibungssprachen wie Topic Maps XTM [XTM00], DAML+OIL [DAO01], OWL [OWL02].

Semantic Web mit seiner globalen Reichweite bleibt aber vorerst eine Vision. Es fehlen u.a. entsprechende Tools für eine effektive Anreicherung der Objekte mit Metadaten.

2.2 Peer-to-Peer Netzwerke

P2P stellt ein Kommunikationsmodell dar, in dem jeder Teilnehmer gleiche Leistungsfähigkeiten hat und gleiche Funktionen bietet, in Gegensatz zur Client/Server Architektur. Die P2P-Technologie kann Dienste von jedem vernetzten Rechner zu jedem anderen vernetzten Rechner bereitstellen. Die Motivation von P2P-Anwendungen ist heute meistens unterhaltungsorientiert, weniger von pädagogischer Natur. Die Suche ist schlüsselwort-basiert und zum Beispiel bei Musik-Tausch-Börsen nur auf den Titel und den Namen des Künstlers beschränkt, die im Dateinamen integriert sind.

Aber wenn zum Beispiel nach "Fingerhut" gesucht werden sollte, könnte das Objekt, das bei einem Peer gefunden wird, völlig verschiedene Bedeutungen haben, weil Fingerhut zum einen eine Pflanze (latein: "Digitalis") ist und zum anderen eine Kapsel als Schutz für einen Finger beim Nähen darstellt. Wenn also Benutzer an Botanik interessiert sind, erwarten sie nicht die Adresse eines Ladens, der Kurzwaren verkauft. Man braucht mehr Ausdruckskraft in Suchanfragen (wie: "Botanische Beschreibung des Fingerhuts"), damit die Suche eingeschränkt wird und genaue Treffer erzielt werden können. Der Grund hierfür liegt in der größeren semantischen Komplexität eines un spezifizierten Lernmaterials. Ein Benutzer soll auch in der Lage sein, Inhalte zu finden, deren Begriff er nicht kennt, sondern nur Relationen und andere Begriffe, die mit seinem Suchobjekt in irgendeiner Beziehung stehen (Beispiel: „Pflanzen, die als Medikament genutzt werden“, Ergebnis hiervon könnte Fingerhut sein).

Gnutella-Peers [GNU00] suchen per Flooding blind auf allen bekannten Nachbarn. Napster [NAP00] hatte ein hybrides System, das aus einem zentralen Server besteht, der ein Verzeichnis mit allen Dateien aller Peers enthält, wobei Peers untereinander die Dateien austauschen. Im Fall von Freenet [FRE01] findet eine gemeinsame Verwendung von Speicherplatz statt, wobei die Peers nicht wissen, was sich auf ihren eigenen Festplatten befindet, weil Dateien verschlüsselt gehalten werden. Hierbei sind Dateien als einzelne Einheiten in verschiedenen, willkürlich vom Benutzer genannten Verzeichnissen und Unterverzeichnissen gespeichert, die in keiner semantischen Relation zueinander stehen.

Die redundante Natur eines P2P-Netzwerks hat auch große Nachteile. Anfragen können oft in gar keiner Antwort resultieren, weil Peers plötzlich das Netz verlassen haben oder das ganze Netz überlastet ist. Das macht sie weniger zuverlässig als das traditionelle Internet, in dem Ressourcen mit einer viel größeren Zuverlässigkeit zugänglich sind, bzw. nicht vom Verhalten einzelner Benutzer abhängig sind. Diese Probleme können mit anderen Eigenschaften der P2P-Netze gemildert werden, wie z.B. durch Replikation der Ressourcen zwischen Peers. Dennoch wird es auf lange Zeit nicht möglich sein, Performance-Garantien in P2P-Netzen zu gewährleisten.

3. Verwandte Arbeiten

In diesem Abschnitt werden Arbeiten und Projekte vorgestellt, die P2P-Systeme mit Metadaten für Dokumente semantisch zu erweitern suchen.

3.1 Edutella

Edutella [EDU02] bietet eine RDF-basierte Metadaten Infrastruktur für P2P-Anwendungen. Die Hauptfunktionen sind Query (Suche), Replikation (Verfügbarkeit und Workload Balancing), Mapping (Übersetzung zwischen verschiedenen Metadaten), Mediation (Abgleich zwischen überlappenden oder widersprüchlichen Informationen) und ein Annotations-Dienst. Edutella verwendet ein eigenes Datenmodell ECDM und eine Abfragesprache RDF-QEL-i. Edutella arbeitet auf JXTA [JXT01], ein von Sun Microsystems entwickelter Satz von XML-basierten Protokollen zum Auffinden von Peers, Bilden von Peer-Gruppen, Bilden von Pipes zwischen Peers zur direkten Kommunikation und Peer-Überwachung.

Edutella konzentriert sich erstens auf Metadatenaustausch zwischen Peers und soll auch für einen Austausch von Lernressourcen dienen. Hierbei wird auch ein Einsatz von Ontologien beabsichtigt, die ein hierarchisch definiertes Vokabular der Begriffe eines Lernbereichs darstellen. Die Begriffe dienen als Schlüsselwörter, auf deren Basis die Suche nach Lernressourcen erfolgen kann [OML03]. Beim Hinzufügen einer neuen Lernressource in das Repositorium eines Peers muss diese mit Begriffen aus der Ontologie annotiert werden [DLR02]. Edutella basiert darüberhinaus auf dem Einsatz von sogenannten Hubs, die zentral die Anfragen der Peers steuern und unter den Peers vermitteln sollen [EDU02]. Diese zentrale Strukturinstanz hat den Vorteil, dass sie den Peers viel Arbeit spart, stellt aber häufig ein "single point of failure" dar, der beim Ausfall die Funktionsfähigkeit weiterer Teile des P2P-Netztes beeinträchtigt.

3.2 SON- Semantic Overlay Networks

Arturo Crespo und Hector Garcia Molina schlagen in [SON02] ein Clustern der Knoten vor, die einen semantisch ähnlichen Inhalt besitzen, in sogenannte Semantic Overlay Networks (SONs). So sollte die Sucheeffizienz gesteigert werden, indem die SONs bestimmt werden, die am besten geeignet sind die Anfragen zu beantworten, während andere SONs nicht angesprochen werden sollen. Hiermit werden Verbindungen zwischen Knoten strukturiert. Empirische Untersuchungen sind an Musik-Tauschbörsen vorgenommen worden wegen ihres großen Ausmaßes und weil die Semantiken der Musik-Dateien "reich" genug sind, um verschiedene Klassifikationen zu erlauben [SON02]. Bei Lernmaterialien, die komplexere semantische Beschreibungen enthalten, ließen sich sehr schwierig kohärente SONs bestimmen, sodass nicht klar wäre, nach welchen Kriterien Knoten in welche SONs eintreten sollen.

3.3 SWAP - Semantic Web and Peer-to-Peer

Im Projekt SWAP [SWP02] wird der Aufbau eines Knowledge Management Systems angestrebt, das vor allem in einem Intranet schnell und effektiv Informationsaustausch und Aktualisierung der verteilten Dateien in einem P2P-Netzwerk ermöglichen soll. Das Ablegen der Dateien der Benutzer in verschiedene Verzeichnisse, Email-Folder und Datenbanken wird auf eine Ontologie abgebildet, sodass das Verhalten des Benutzers und seine Art der Anordnung von Dateien als eine hierarchische Wissenstrukturierung interpretiert wird. Diese "Sicht" des einzelnen Benutzers (Peers) soll so umgestaltet werden können, dass andere Peers es auch interpretieren und nutzen können. Da die Methode der automatischen Erstellung der lokalen Ontologien bei Peers nicht vollständig entwickelt ist, stehen Evaluationsergebnisse noch aus. Die Betonung liegt hierbei auf schnelle Wissensverbreitung und nicht auf Austausch und Suche von Dateien.

4. Ansatz mit Domain Specific Ontologies

Hier wird ein Methode vorgestellt, die Repositorien der Peers mittels Ontologien so zu strukturieren, dass sich relationale Zusammenhänge aus der realen Welt hierin widerspiegeln. Im Vergleich zu verwandten Arbeiten wird hier eine lokale Strukturierung der Lernmaterialien zwecks deren leichterem Auffinden angestrebt.

Durch eine Domain Specific Ontology [COA02] wird eine hierarchische Repräsentation eines Anwendungsgebietes der realen Welt abgebildet, beispielsweise eines Fachgebietes einer Universität. Zu Begriffen dieser Ontologie können Metadaten durch Relationen zugeordnet werden, bei denen jeder Metadatensatz eine spezielle Lernressource kennzeichnet. An diesem Metadatensatz wird schließlich durch einen relativen oder absoluten Link eine Ressource selbst angebracht (Abbildung 1).

Eine Anreicherung der Inhalte mit Metadaten kann unter Verwendung eines spezifizierten Vokabulars erfolgen, wie beispielsweise LOM [LOM02]. Wir haben an unserem Institut einen LOM-Editor [Ste02, SHR03] erstellt, der als Java-Applet mit einer Xindice [XIN02] Datenbank arbeitet. Beim Erstellen eines neuen Metadatensatzes für einen speziellen Inhalt können mehrere Felder automatisch, abhängig von Business-Model, mit Werten ausgefüllt werden, während weitere Felder manuell ausgefüllt werden müssen. Die automatisch gefüllten Felder enthalten Informationen, die leicht maschinell, also ohne menschliches Zutun, erstellt werden können, wie Größe, Titel und Datum der Erstellung.

Die Bereitstellung einer Ontologie, die die Inhalte des Repositoriums eines Peers repräsentiert, kann deutlich die Suche nach und Administration von Inhalten verbessern. Verwandte Lehrstühle, die ihre Lerninhalte miteinander austauschen würden, müssten hierzu eine einheitliche Ontologie-Repräsentation ihrer Wissensdomäne verwenden.

Die Erstellung und Pflege einer Ontologie erfordert explizites Modellieren des Wissens, das in eigenen Ressourcen liegt, sodass das Wissen in Unter- und Oberbegriffe

klassifiziert werden kann. Dieses Vorgehen verursacht einen nicht unerheblichen Aufwand, sodass die primäre Schwierigkeit sein wird, die Autoren zu dieser Maßnahme zu bewegen. Autoren müssen gegebenenfalls jedes Mal beim Aktualisieren ihres Repositoriums einen Link zu ihrer Ontologie hinzufügen (oder entfernen). Wenn diese Aufgabe nicht sorgfältig ausgeführt wird, können fehlerhafte Semantiken zu einer ungenauen Suche bei anderen Peers führen. Dieses Problem kann durch Einführung eines Belohnung/Bestrafungs-Schemas für Peers angesprochen werden. Ein Peer, der falsche oder ungenaue Semantiken erzeugt, wird basierend auf dem Feedback anderer Peers, als unzuverlässig eingestuft. Die Bewertung durch andere Peers erfolgt, nachdem sie den gefundenen Inhalt geprüft und bewertet haben, ob es tatsächlich das ist, was sie gesucht haben. Hierdurch können die Peers einerseits basierend auf Zuverlässigkeits-einstufung entscheiden, ob sie von einem bestimmten Peer die Inhalte anfordern wollen oder nicht. Andererseits können als Bestrafung/Belohnung die Wartelisten der unzuverlässigen/zuverlässigen Peers langsamer oder schneller abgearbeitet werden. Hierbei soll betont werden, dass mit dieser Methode das Recall (Verhältnis zwischen gefundenen und existenten Begriffen) einer Ontologie nicht erfassbar ist. Das bedeutet, dass Peers, die bestimmte Lernressourcen besitzen, die aber wegen einer fehlerhaften Verlinkung überhaupt nicht gefunden werden können, trotzdem als „zuverlässige“ Peers gelten werden.

Das P2P-Kommunikationsmodell bietet eine leichte Administration der Daten bei gleichzeitiger Kontrolle des Autors über alle seine Ressourcen. (Der Gegensatz hierzu wäre eine zentrale Ressourcenhaltung auf einem Webserver, der gepflegt werden muss und der ein "Single point of failure" darstellt.) Ontologien stellen da ein zusätzliches Tool für Knowledge Management dar, das auf Präzision und leichte Auffindbarkeit der Ressourcen ansetzt. Durch dessen Hilfe ist es möglich, auch Begriffe zu finden, die man nicht kennt, aber Lernressourcen über diese sucht. Beispiel: "Pflanzen, aus denen Herzmedikamente gewonnen werden". Es ist hierbei nicht notwendig, den Begriff "Fingerhut" oder "Digitalis" zu kennen. Es genügt mindestens einen Begriff und eine Relation zu kennen, die den gesuchten Begriff beschreiben.

Die vorgeschlagene Lösung kann als ein Tool für eine relativ geschlossene Gesellschaft angesehen werden, das nicht für eine breite Öffentlichkeit geeignet ist. Da eine anfängliche Konstellation des P2P-Netzwerks für einen Austausch der Lernressourcen zwischen interessierten Teilnehmern wahrscheinlich wenige Knoten besitzt, spielt die Skalierbarkeit eher eine geringe Rolle. In diesem Netzverbund würde eine Suche nach den Ressourcen durch Flooding keine große Datenflut auslösen, sodass sie durchaus für Suchanfragen brauchbar sein wird. Von einem solchen kleinerem Netzverbund ausgehend meinen wir, dass ein dezentrales P2P-Netzwerk in der Lage sein wird, eine effektive Suche und den Austausch der Lerninhalte zu unterstützen. Ein größeres Hindernis hierbei stellen Flaschenhälse im Internet dar, die durch dediziertes Routing überwunden werden können. Das Routing in solchen Netzen ist nicht das Thema dieses Papiers.

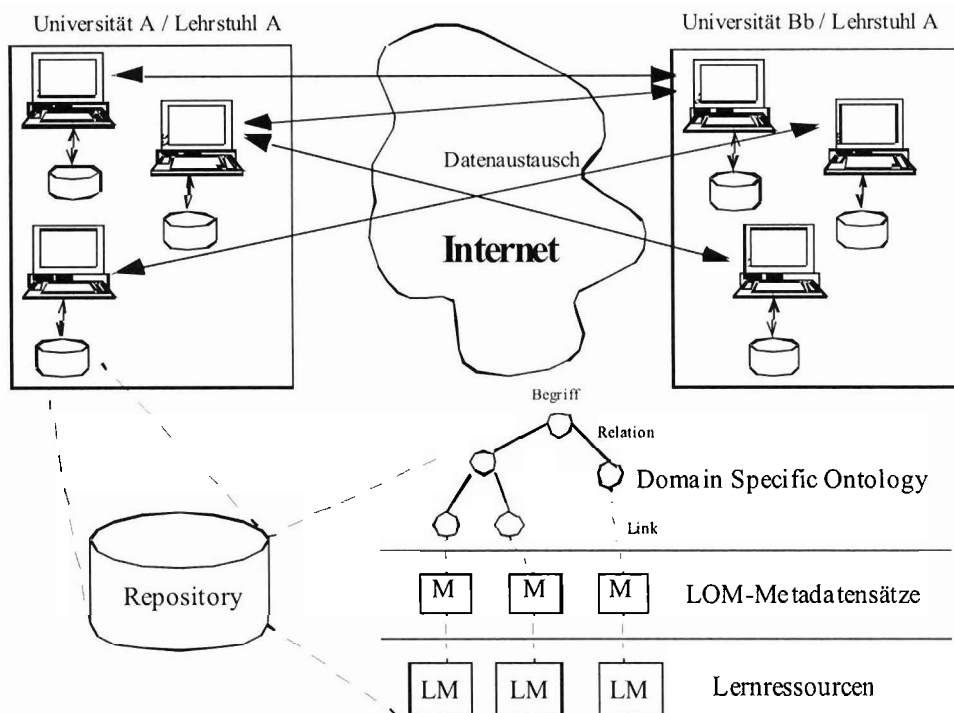


Abbildung 1: Dezentrales P2P-Netzwerk mit einer ontologie-basierten Organisation und Administration der Lernressourcen

5. Zusammenfassung

Bei der Konzipierung des obigen Modells sollen keine proprietäre Lösungen angestrebt werden, sondern neu entstehende Standards verwendet werden, die später das Rückgrat der Wiederverwendung von Dokumenten und Objekten bilden werden. Je mehr multimediale Ressourcen aufwändig erstellt werden, desto höher wird das Interesse der Universitäten sein, Lerninhalte auszutauschen. Hierbei wird vom Interesse der Universitäten an einem gleichberechtigten, reziproken, zwischen ähnlichen Lehrstühlen ausgeglichenen Austausch der Lernressourcen ausgegangen, von dem alle Autoren profitieren können.

Diese Arbeit zeigt eine Lösung zum Aufbau eines eher homogenen geschlossenen Benutzerkreises für einen Austausch von Lernressourcen über einen P2P- Netzwerk mit semantisch angereicherten Repositorien. Die Motivation für einen solchen Aufbau stellt die schwer abschätzbare Inbetriebnahme des "Semantic Web". Es wird ein weniger performantes, dezentrales Netzwerk vorgeschlagen, das mit einer ontologie-basierten Organisation von Ressourcen bei einzelnen Benutzern eine verbesserte Suche für darauf

zugreifende Interessenten bietet. Eine solche Lösung würde natürlich nur einen Fragment des Internets umfassen und somit zu einer Improvisierung eines niedriger performanten Verbundes mehrerer „semantischer Netze“ führen. Es existiert ein großes Potential im Ansatz des P2P-Paradigma, wenn man die steigende Anzahl der vernetzten Rechner betrachtet. Diese Technologie fördert Entfaltung dynamischer und anpassungsfähiger Netzstrukturen und wird sehr wahrscheinlich noch weitere Verwendung erfahren.

Literaturverzeichnis

- [COA00] Jilianne van Zyl, Dan Corbett, *Population for Understanding and Classifying Ontology Applications*, Proc. 14th European Conference on Artificial Intelligence ECAI'00, Berlin, Germany August 20-25, 2000
- [DAO01] DAML-OIL Reference Description <URL:<http://www.w3.org/TR/daml+oil-reference>>
- [DLR02] Wolfgang Nejdl, Mario Schlosser, Wolf Siberski, Martin Wolpers, Bernd Simon, Stefan Decker, Michael Sintek, *RDF-based Peer-to-Peer-Networks for Distributed (Learning) Repositories*, Technical Report, 2002
- [EDU02] Wolfgang Nejdl, Boris Wolf, Changtao Qu, Stefan Decker, Michael Sintek, Ambjörn Næve, Mikael Nilsson, Matthias Palmer *Edutella: A P2P Networking Infrastructure Based on RDF*, 11th International World Wide Web Conference (WWW2002), Hawaii, USA, May 2002
- [FRE01] Freenet, <URL: <http://freenet.sourceforge.net>>
- [GNU00] Gnutella, <URL: <http://gnutella.wego.com>>
- [JXT01] Sun Microsystems, JXTA v1.0 Protocols Specification v1.2, <URL: <http://spec.jxta.org/v1.0/docbook/JXTAProtocols.html>>
- [LOM02] IEEE 1484.12.1 LOM draft standard, <URL: <http://ltsc.ieee.org>>
- [MES02] Stefan Saroui, P. Krishna Gummadi, Steven D. Gribble, Dept. of Computer and Engineering, Univ. of Washington, Seattle, WA, 98195-2350, *A Measurement Study of Peer-to-Peer File Sharing Systems*, Proceedings of the Multimedia Computing and Networking (MMCN), San Jose, January, 2002
- [OML03] "Ontologies and Metadata for eLearning", aus "Handbook on Ontologies", im Druck bei Springer Verlag, Herausgeber: Staab, Studer, 2003
- [OWL02] OWL Web Ontology Language, <URL: <http://www.w3.org/TR/owl-semantics/>>
- [RDF99] Resource Description Framework, <URL: <http://www.w3.org/RDF/>>
- [SHR03] Stefan Hoermann, Stefan Schneider, Ulrich Glowalla, Ralf Steinmetz, *Erstellung von SCORM-kompatiblen Kursen in k-Med*, In "Rechnergestützte Lehr- und Lernsysteme in der Medizin", pages 103-113. Shaker Verlag, Aachen, April 2003.
- [SON02] Arturo Crespo and Hector Garcia-Molina, Computer Science Department, Stanford University, CA 94305-2140, USA, *Semantic Overlay Networks for P2P Systems*, submitted for publication 2002
- [Ste02] Achim Steinacker, *Medienbausteine für web-basierte Lernsysteme*, Dissertationsschrift, Technische Universität Darmstadt, 2002
- [SWP02] Merc Ehrig, Cristoph Tempich, Jeen Broeckstra, Frank van Harmelen, Marta Sabou, Ronny Siebes, Steffen Staab, Heiner Stuckenschmidt, *SWAP: Ontology based Knowledge Management with Peer-to-Peer Technology*, Workshop Ontologie-basiertes Wissensmanagement (WOW2003)
- [THL01] Tim Berners-Lee, James Hendler, Ora Lassila, *Semantic Web*, Scientific American, May 2001
- [XIN02] Apache Xindice, <URL: <http://xml.apache.org/xindice/>>
- [XTM01] XML Topic Maps(XTM) 1.0, <URL: <http://www.topicmaps.org/xtm/1.0/>>