

## **International Conference on Semantic Web and Digital Libraries**

(vom 21.-23. Februar 2007 in Bangalore / Indien)

Die Konferenz wurde ausgerichtet vom Documentation Research and Training Centre am Indian Statistical Institute, der ehemaligen Wirkungsstätte von S.R. Ranganathan. Die Digitalen Bibliotheken und das Semantic Web sind sicherlich Themen, mit denen sich mühelos drei Konferenztage mit bis zu vier Parallelveranstaltungen sowie ein umfangreicher Kongressband<sup>1</sup> füllen lassen. So wäre es sicherlich vermessen, im Rahmen eines kurzen Aufsatzes alle Facetten der Thematik und des Kongresses abbilden zu wollen. Doch vielleicht gelingt es ja, wichtige Entwicklungen und Themenstellungen aus Sicht der bibliothekarischen Welt (und gleichzeitig Sicht eines Nicht-Technikers) widerzuspiegeln?

Werfen wir zunächst einen Blick auf die Vortragenden dieser Veranstaltung: auf den Rednerlisten fanden sich sowohl Informations- und Bibliotheksspezialisten als auch IT-Spezialisten, die sich mit dem Semantic Web völlig unabhängig von bibliothekarischen Anwendungsgebieten auseinandersetzen. Auch waren die parallelen Vortragsreihen klar getrennt nach Themenkreisen zum Semantic Web und zu Digitalen Bibliotheken. Wie ist also das „und“ im Konferenznamen zu verstehen? Bleiben beide Themen letztlich unverbunden? Oder finden sich tatsächlich echte Anwendungsfälle der Semantic Web - Technologie in der Welt der Digitalen Bibliotheken?

Auffallend war zunächst, dass die bibliothekarische und die IT-Welt von unterschiedlichen Begriffswelten ausgehen, wenn in den Vorträgen von „Digital Libraries“ die Rede war. So wurde im Kontext der Semantic-Web-Vorträge oftmals jede Internetpräsenz mit umfangreichen elektronischen Daten oder Texten als Digital Library bezeichnet. Eine präzise Definition und Abgrenzung der „Digital Libraries“ von anderen Webanwendungen schien hingegen aus dieser Perspektive nicht als notwendig gesehen zu werden.

Ganz anders stellt sich hingegen die Begriffswelt in den Themenkreise zu den Digital Libraries dar. Hier zeigt eine Analyse der konkreten Vortragsthemen, dass es hier nahezu ausschließlich um fachliche und institutionelle Repositories von open-access-Publikationen (sei es born-digital Material oder auch digitalisierte Printmaterialien) ging. Eine Analyse dieser indischen Repositories hat Prof. Osswald (FH Köln) im Rahmen eines Forschungsaufenthaltes in Bangalore durchgeführt und auf der Konferenz vorgestellt.

Aus diesem Grunde fand die deutsche DINI-Initiative<sup>2</sup> zur Zertifizierung von Repositorien, die von Prof. Schirnbacher (Humboldt-Universität Berlin) vorgestellt wurde, reges Interesse. Die Entwicklung eines InDINI, also die Anwendung der deutschen Zertifizierungskriterien auch auf indische Repositorien ist in der konkreten Planung. Dieses sehr konkrete und klare Verständnis von „Digital Libraries“ bringt es mit sich, dass viele Fragestellungen, die aus der Sicht der deutschen Virtuellen Fachbibliotheken und aus Sicht von vascoda eine hohe Relevanz haben (wie etwa die Einbindung und Harmonisierung unterschiedlicher elektronischer Ressourcen, die Autorisierung, Authentifizierung und Rechteverwaltung für den unmittelbaren Zugriff auf kostenpflichtige elektronische Ressourcen oder die Einbindung von Dokumentlieferdiensten für gedruckten Materialien), in diesem Kontext kaum thematisiert wurden. Intensiv wurden hingegen Fragen nach der geeigneten Software für

---

<sup>1</sup> A.R.D.-Prasad / Devika P. Madalli (Hrsg.): International Conference on Semantic Web & Digital Libraries. Bangalore 2007

<sup>2</sup> Vgl. <http://www.dini.de>

Repositorien und Aspekte der Bewahrung des kulturellen Erbes durch Digitalisierung diskutiert.

Doch nun zur Verbindung von Semantic Web und Digital Libraries. Hier finden sich durchaus Konzepte zur Anwendung der Semantic-Web-Technologie, hingegen bislang nur wenige Beispiele einer praxistauglichen Realisierung. Beginnen wir zunächst mit einer begrifflichen Klärung: „Das semantische Web (engl. Semantic Web) ist eine Erweiterung des World Wide Web (WWW) um maschinenlesbare Daten, welche die Semantik der Inhalte formal festlegen. ... Informationen sollen zusätzlich zu der für Menschen lesbaren Form auch formal, in einer für Maschinen verarbeitbaren Form repräsentiert werden, damit Programme darauf operieren können, so dass Anfragen aufgrund ihres Bedeutungsinhalts anstelle ihrer Schreibweise bearbeitet werden können.“<sup>3</sup> Diese Definition macht schon klar, dass in vielen bibliothekarischen Anwendungsfeldern, in denen Metadaten (bibliographische Daten) im Zentrum stehen, das Problem der Bedeutungsinhalte durch die Datenkategorien zumindest teilweise gelöst ist. Während auf einer beliebigen Homepage keineswegs für die Maschine erkennbar ist, ob ein Begriff zum Beispiel ein Personennamen ist, sind bibliographische Daten klar den unterschiedlichen Kategorien zuzuordnen.

Es bleibt jedoch in der Behandlung der bibliographischen Daten das Problem der semantischen Heterogenität, also insbesondere die Verwendung von unterschiedlichen Thesauri. Ein Lösungsansatz dafür bilden die Aktivitäten am Informationszentrum Sozialwissenschaften in Bonn zur Verknüpfung von Thesauri über Crosskonkordanzen bzw. Konkordanzen, die von Prof. Krause vorgestellt wurden.<sup>4</sup>

Verlässt man jedoch die Welt der bibliographischen Daten und betrachtet Webseiten, in denen Informationen enthalten sind, denen keine Begriffskategorien zugeordnet werden, so stellen sich die Probleme des Semantic Web noch ungleich komplexer dar. Hier lassen sich zwei unterschiedliche Herangehensweisen unterscheiden: Im ersten Falle werden Ontologien erstellt, mit deren Hilfe dann den Inhalten „fremder“ Webseiten Bedeutungsinhalte hinterlegt werden. "Unter einer Ontologie versteht man in der Informatik im Bereich der Wissensrepräsentation ein formal definiertes System von Begriffen und Relationen. Zusätzlich enthalten Ontologien Inferenz- und Integritätsregeln, das heißt, Regeln der Schlussfolgerung und zur Gewährleistung ihrer Gültigkeit. ...Ontologien haben mit der Idee des „Semantic Web“ in den letzten Jahren einen Aufschwung erfahren. Eine allgemeine Definition des Begriffes „Ontologie“ ist schwierig, da je nach Autor verschiedene Systeme darunter subsumiert werden. So lassen sich mehrere bereits länger existierende Formate und Ansätze wie Frames und Semantische Netze aus der Künstlichen Intelligenz oder Klassifikationen und Thesauri aus der Dokumentationswissenschaft wahlweise als Vorläufer, Alternativen oder als spezielle Formen von Ontologien auffassen. ...Ontologien dienen in verschiedenen Bereichen als Mittel der Strukturierung und zum Datenaustausch, um bereits bestehende Wissensbestände zusammenzufügen ...“<sup>5</sup>

Dieses Verfahren findet zumeist fachbezogen und für ausgewählte Webseiten Anwendung, da eine interdisziplinäre Erstellung von leistungsfähigen Ontologien, die dann auf eine sehr große Anzahl an Webseiten angewendet werden, sicherlich ein zu ehrgeiziges Unterfangen wäre. Allerdings sind durchaus Anwendungsszenarien realisierbar, in denen mehrere fachliche Ontologien miteinander verknüpft werden.

---

<sup>3</sup> Vgl. [http://de.wikipedia.org/wiki/Semantic\\_Web](http://de.wikipedia.org/wiki/Semantic_Web) (Wikipedia-Definition Stand 21.3.07)

<sup>4</sup> Vgl. z.B. Walter u.a.: Crosskonkordanzen als Mittel zur Heterogenitätsbehandlung in Informationssystemen (2006) <http://eprints.rclis.org/archive/00007256/>

<sup>5</sup> Vgl. [http://de.wikipedia.org/wiki/Ontologie\\_%28Informatik%29](http://de.wikipedia.org/wiki/Ontologie_%28Informatik%29) (Wikipedia-Definition Stand 21.3.07)

Im zweiten Falle wird eine Internetpräsenz durch die Einbindung semantischer Strukturen aufbereitet. Ein Beispiel für eine solche semantische Aufbereitung ist das „Semantic WIKI“. Dargestellt wurde dieses am Beispiel der online-Enzyklopädie WIKIPEDIA, wobei allerdings – entgegen der gewählten Begrifflichkeit - die Organisation als WIKI, an dem zahlreiche Personen kollaborativ mitarbeiten, keineswegs das entscheidende Kriterium für diesen Anwendungsfall ist. Die semantische Aufbereitung wäre natürlich auch bei Webseiten, die von Einzelpersonen oder kleineren Gruppen erstellt werden, möglich.

Betrachtet man einmal exemplarisch die Einträge einer Online-Enzyklopädie zu Städten, so lässt sich feststellen, dass sich in nahezu allen Einträgen bestimmte Aspekte und Verlinkungen wiederholen. So ist in der Regel die Einwohnerzahl genannt, die Region (das Bundesland) und der Staat, zu dem diese Stadt gehört usw. Bereitet man diese Informationen mit „typisierten Links“ (z.B. einer Link-Struktur „gehört zu“) semantisch auf, so ist es auch maschinell möglich, die Beziehungen zwischen den jeweiligen Informationen erfassen. Dies würde insbesondere die Recherchemöglichkeiten innerhalb der Enzyklopädie entscheidend verbessern. So könnte in unserem Beispiel der Städteinträge gezielt nach allen Städten in Deutschland mit mehr als 500.000 Einwohnern gesucht werden. Oder aus dem Artikel zu Nordrhein-Westfalen heraus könnten alle zugehörigen Städte (die mit einer entsprechenden „gehört zu“ Link-Struktur versehen wurden) recherchiert werden, ohne dass diese im Artikel selbst hinterlegt wurden. Doch sind durchaus auch noch wesentlich komplexere Recherchemöglichkeiten realisierbar, in denen noch mehr unterschiedliche semantische Beziehungen miteinander verknüpft werden.

Zusammenfassend lässt sich feststellen, dass die meisten Anwendungsfälle des Semantic Web in der Welt der Digitalen Bibliotheken bisher eher prototypisch und noch nicht praxistauglich zu sein scheinen. Aus Sicht eines Bibliothekars und Nicht-Technikers ergibt sich das Bild eines guten Entwicklungsstandes der zugrunde liegenden Technologien. Nunmehr gilt es jedoch, alltagstaugliche Ontologien zu erschaffen sowie vorhandene Digitale Bibliotheken semantisch aufzubereiten. Hier ist von Bibliotheks- und Informationsspezialisten noch viel inhaltlich-konzeptionelle Arbeit zu leisten.

Ralf Depping  
Universitäts- und Stadtbibliothek Köln