

Borst, Timo; Fingerle, Birgit; Neubert, Joachim; Seiler, Anette

Article

Wie finden Bibliotheken den Weg in das Semantic Web? Bericht von der SWIB09 in Köln

ABI-Technik

Suggested Citation: Borst, Timo; Fingerle, Birgit; Neubert, Joachim; Seiler, Anette (2010) : Wie finden Bibliotheken den Weg in das Semantic Web? Bericht von der SWIB09 in Köln, ABI-Technik, ISSN 0720-6763, de Gruyter Saur, Berlin, Vol. 30, Iss. 1, pp. 31-34

This Version is available at:

<http://hdl.handle.net/11108/13>

Kontakt/Contact

ZBW – Leibniz-Informationzentrum Wirtschaft/Leibniz Information Centre for Economics
Düsternbrooker Weg 120
24105 Kiel (Germany)
E-Mail: info@zbw.eu
<https://www.zbw.eu/de/ueber-uns/profil-der-zbw/veroeffentlichungen-zbw>

Standard-Nutzungsbedingungen:

Dieses Dokument darf zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden. Sie dürfen dieses Dokument nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Sofern für das Dokument eine Open-Content-Lizenz verwendet wurde, so gelten abweichend von diesen Nutzungsbedingungen die in der Lizenz gewährten Nutzungsrechte.

Terms of use:

This document may be saved and copied for your personal and scholarly purposes. You are not to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. If the document is made available under a Creative Commons Licence you may exercise further usage rights as specified in the licence.

►► Wie finden Bibliotheken den Weg in das Semantic Web?

Bericht von der SWIB09 in Köln

Als „Fachtagung für innovative Bibliothekare“ fand in Köln am 24. und 25. November 2009 erstmals die Veranstaltung „Semantic Web in Bibliotheken – SWIB09“ statt. Mit dem Ziel, Bibliotheken das „Semantic Web“ näher zu bringen, bot der erste Tag eine Einführung und erste Anwendungsbeispiele, die das Potenzial von Semantic Web-Anwendungen zeigten. Der zweite Tag stand im Zeichen von praxisnahen Beispielen, die die Anforderungen und Lösungsszenarien für Bibliotheken vermittelten. Die Tagung war international besetzt und wurde vom Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz) und der Deutschen Zentralbibliothek für Wirtschaftswissenschaften (ZBW) gemeinsam ausgerichtet. Aus Platzgründen war die Teilnehmerzahl auf rund 100 am ersten Tag und 60 am zweiten Tag begrenzt. Die deutlich höhere Anzahl an Anmeldungen zeigte den großen Wunsch nach Austausch zu diesem Thema.

Was ist das Semantic Web?

In seinem Eröffnungsvortrag gab Jakob Voß, Software-Entwickler bei der Verbundzentrale des GBV (VZG), eine kurze Einführung in das Semantic Web. Anhand von herkömmlichen bibliothekarischen Katalogeinträgen skizzierte er den für das Semantic Web typischen Wandel weg von einem Web ver-

netzter Dokumente hin zu einem Web vernetzter Daten (Linked Data). Verschiedene Einrichtungen aus dem akademischen Umfeld hätten ihre Daten bereits als Linked Data veröffentlicht und miteinander vernetzt. Voß wies darauf hin, dass Bibliotheken aber bislang nicht einmal alle Möglichkeiten des bisherigen Webs nutzen und beispielsweise ihre Inhalte in Datenbanken im „Deep Web“ versteckten. Mangels stabiler URLs sei es oft nicht möglich, auf herkömmliche Weise auf einen Eintrag in einem Bibliothekskatalog zu verlinken.

Mit Linked Data, so Voß, würden sich hingegen neue Möglichkeiten zur Arbeitsteilung erschließen. So sei es beispielsweise möglich, bibliografische Daten, die ein Verlag über eine Publikation veröffentlicht, von Bi-

Die SWIB09 – erste internationale Fachtagung zum Thema Semantic Web in Bibliotheken – zieht Gäste aus Europa und den USA



SWIB09 – Ein Forum für den fachlichen Austausch auf internationaler Ebene



► Wie finden Bibliotheken den Weg in das Semantic Web?

blibliotheksseite mit weiteren Daten zu verknüpfen, etwa mit Personennormdaten und dem genauen Standort in einer Bibliothek. Dabei könnten auch Daten, die andere Bibliotheken oder Personen zu dieser Publikation veröffentlicht haben, nachgenutzt werden. Den eigenen Nutzerinnen und Nutzern könnten ausgewählte Informationen in einer auf ihre Bedürfnisse zugeschnittenen Webseite präsentiert werden. Die Vorteile, die sich aus solchen Linked Data-Nutzungsszenarien ergeben, lägen auf der Hand. Weil das Web dabei selbst die Datenbank darstellt, müsste weniger Spezialsoftware eingesetzt werden und es wären weniger Datenkonvertierungen notwendig. Aufwändiges Datensammeln in der eigenen Datenbank entfielen ebenso wie Doppelarbeit. Denn die relevanten Informationen würden aus verschiedenen Quellen zusammengesucht und es würden lediglich die benötigten Daten selbst veröffentlicht werden, die nicht schon an anderer Stelle vorliegen.

Warum sollten sich Bibliotheken mit dem Semantic Web beschäftigen?

Bibliotheken sollten sich laut Prof. Dr. Stefan Gradmann, Inhaber des Lehrstuhls für Wissensmanagement an der Humboldt-Universität zu Berlin (HU Berlin) mit dem Semantic Web beschäftigen, um einerseits in einer zukünftigen Wissensgesellschaft überhaupt noch wahrgenommen zu werden. Andererseits könnten Bibliotheken auf der Basis von Semantic Web-Technologien zukünftig attraktive und innovative Angebote für ihre Nutzerinnen und Nutzer entwickeln, wie z. B. bessere Suchfunktionalitäten. Zudem könnten gerade Bibliotheken einen spezifischen Beitrag zum Semantic Web leisten. Gradmann erinnerte in diesem Zusammenhang daran, dass Tim Berners-Lee, der Erfinder des Web und Semantic Web-Pionier, bereits im Jahr 2000 „Catalogs on the web“ als eine Killer Application des Semantic Web bezeichnet hatte. Bibliotheken besäßen im

Gegensatz zu vielen anderen Datenanbietern zuverlässige Metadaten und Kontextualisierungsdaten (wie PND- oder GKD-Normdaten, Klassifikationen und Thesauri). Diese seien für eine Verknüpfung mit anderen Daten höchst attraktiv, und für ihre Überführung ins Semantic Web liegt mit SKOS (Simple Knowledge Organization System) ein geeigneter Standard vor, der speziell für Thesauri und Klassifikationen entwickelt wurde. Bei bislang ungenügend beachteten Aspekten, wie dem Umgang mit Bedeutungswandel und der Langzeitverfügbarkeit von Informationen und Wissen im Semantic Web, könnten Bibliotheken ihre Erfahrungen einbringen.

Ebenso wie Gradmann sieht Dr. Bernhard Haslhofer, Assistent am Lehrgebiet für Verteilte Multimediasysteme an der Universität Wien, ein großes Potenzial darin, klassische Mechanismen der Informationsorganisation durch Bibliotheken (Identifizier, Metadaten, Kontrollierte Vokabulare) als „Linked Open Data“ in das Semantic Web einzubringen. Mit Linked Data werde versucht, diese Mechanismen im Web weiterzuführen. Bibliotheken könnten damit möglicherweise sogar einen Baustein zu einer freien und Community-getriebenen Alternative zu kommerziellen Anwendungen wie Google Books beitragen. Voraussetzung sei allerdings, dass Bibliotheken nicht an ihren Datenbeständen in Form geschlossener Datensilos festhalten, sondern ihre Daten unter einer entsprechenden Lizenz frei im Internet zur Verfügung stellen. Jürgen Kett von der Deutschen Nationalbibliothek (DNB) betrachtete Universitäten, Bibliotheken, Museen und Archive in seinem Vortrag als integralen Bestandteil des Semantic Web, da sie ihm die nötige Verlässlichkeit und Stabilität verliehen.

Was machen andere Bibliotheken?

Mehrere Beiträge zeigten, dass mit Linked Data mittlerweile eine geeignete Technologie

Ed Summers. Library of Congress (USA)



vorliegt, um das Semantic Web für Bibliotheken Wirklichkeit werden zu lassen. Ed Summers, Software-Entwickler bei der Library of Congress (LoC) und, mit den LoC Subject Headings zuständig für einen der bislang größten Linked Data-Bestände aus Bibliotheken, verwies in seinem Vortrag auf die vielfältigen Verknüpfungsmöglichkeiten und Anwendungen von Linked Data. So könnten Schlagworte aus verschiedenen Vokabularen in Beziehung gesetzt werden und darauf basierend bessere Suchmöglichkeiten angeboten werden. Oder es könnten Presseartikel wie beim Projekt „Chronicling America: Historic American Newspapers“ mit „Flickr“-Bildern und Geokoordinaten verknüpft werden, die das Ereignis bebildern bzw. räumlich verorten.

Wie ein klassischer Bibliothekskatalog in Linked Open Data und – darauf aufbauend – in eine moderne Webanwendung überführt werden kann und welche Rolle dabei Semantic Web-Technologien spielen können, berichtete Anders Söderbäck von der Schwedischen Nationalbibliothek. Unter dem Thema „Linked Applications“ entwarf Felix Ostrowski, IT-Entwickler beim hbz, das Modell einer tiefen Integration von Ontologien direkt in die verwendete Programmiersprache am Beispiel von Repository Software. Andre Hagenbruch, Mitarbeiter der Universitätsbibliothek Bochum, stellte ein Projekt für ein Linked Data basiertes Bibliotheksportal vor, das neben bibliographischen Daten auch administrative Informationen wie Personaldaten oder Öffnungszeiten integrieren soll.

Interessante Einblicke zur Linked Data-Strategie der DNB bot Jürgen Kett. Dazu gehört die Veröffentlichung der Personennamendatei (PND) zwecks kostenfreier Weiterverwendung. In einem weiteren Vortrag gingen Kett und sein Kollege Dr. Lars G. Svensson auf Details ihres aktuellen Linked Data-Projekts ein, das in einer ersten Stufe URI- und SPARQL-Zugriff (SPARQL Protocol and RDF Query Language) auf die SWD- und PND-Datensets sowie ihre Verknüpfungen zu internationalen Normdatenbeständen ermöglichen soll. Ein Beta-Dienst soll bis Mitte 2010 online gehen. Dr. Timo Borst, Leiter der IT-Entwicklung der ZBW, griff dies auf und demonstrierte den Nutzen des geplanten DNB-Linked Data-Services anhand der Integration von Personennormdaten in den ZBW-Dokumentenserver EconStor. Als technischer Zugangspunkt für Linked Data der PND wurde dabei ein experimenteller SPARQL-Endpoint der DNB genutzt, dessen Abfrage eine automatische Vorschlagsfunktion für PND-Einträge ermöglicht und somit eine kontrolliertere Erfassung von Autoren Daten. Die Einbindung des Standard-Thesaurus Wirtschaft (STW) als Linked Data in EconStor soll analog dazu die kontrollierte Verschlagwortung

unterstützen und zudem per Suchterweiterung die Suchfunktionalitäten verbessern.

Was können Bibliotheken aus anderen Bereichen lernen?

Aus dem Museumsbereich stellte Karin Teichmann, Leiterin der Grafischen Sammlung im Deutschen Buch- und Schriftmuseum der DNB, „CIDOC Conceptual Reference Model“ vor. Es liegt als ISO-Standard vor und kann unterschiedlich strukturierte und verteilte Informationen durch ein übergreifendes Meta-Modell zusammenführen und so Integrationsprobleme lösen – was auch im Bibliotheksbereich eine verbreitete Anforderung ist.

Prof. Dr. Stefan Gradmann und Marlies Olensky, Forscherin an der HU Berlin, präsentierten die semantische Datenschicht der Europeana, die den W3C-Standard SKOS und Linked Data nutzt, um Benutzeranfragen, Browsing und Ergebnispräsentation zu unterstützen.

Über die umfassenden Linked Data-Aktivitäten innerhalb der DBPedia, einer Semantic Web-Version der Wikipedia, berichtete Anja Jentzsch, Forscherin an der Freien Universität Berlin. Neben der Infrastruktur für Linked Data veranschaulichte sie die Integration von DBPedia-Daten in Browser-, Such- und Portaloberflächen. DBPedia stelle einen wichtigen Knotenpunkt für Linked Data dar aufgrund der Vielzahl an Daten und Verknüpfungen zu anderen Linked Data-Sets.

Einen Einblick in Ontologien, die eine weitere Säule des Semantic Web bilden, in ihren Zusammenhang mit natürlicher Sprache und in ihre Nutzungsmöglichkeiten gab Elena Semanova, Forscherin und selbstständige Ontologie-Expertin.

Wie könnten erste Schritte ins Semantic Web aussehen?

Anders Söderbäck berichtete, dass es für Linked Data-Projekte hilfreich sein kann, einen anderen Blickwinkel einzunehmen und den Bibliothekskatalog als ein Netzwerk zu betrachten. Ein fundamentaler Schritt in Richtung Semantic Web bestehe für Bibliotheken dann in der rechtlich abgesicherten Veröffentlichung ihrer Daten als Linked Data. Joachim Neubert, IT-Entwickler bei der ZBW, fasste praktische Erfahrungen aus den Linked Data-Projekten der ZBW zusammen und gab Hinweise darauf, wie ganz konkret mit dem Semantic Web begonnen werden könnte. So sollte zunächst ein überschaubares Gebiet und Datenset ausgesucht werden.

► Wie finden Bibliotheken den Weg in das Semantic Web?

Anschließend müssten Lizenzierungs- und Attributierungsfragen für die Daten geklärt werden. Die Datenspeicherung und -pflege sowie bibliothekarische Geschäftsgänge sollten möglichst nicht verändert werden.

Die Verwendung stabiler HTTP URIs (URLs) zum Beispiel für Katalogisate und Normdaten sei eine Grundregel, auf die Bibliotheken bei der Publikation von Linked Data achten sollten, wie Jakob Voß in seinem Einführungsvortrag schilderte. Gemäß der Regeln für Linked Data, die von Tim Berners-Lee formuliert wurden, würde die Formulierung eines Angebots an relevanten Informationen zu diesen URIs in den Semantic Web Standards RDF (Resource Description Framework) oder SPARQL einen weiteren Schritt darstellen. Dafür könnten Ontologien eingesetzt werden, die beschreiben, welche Arten von Objekten es gibt, welche Eigenschaften sie besitzen und in welcher Beziehung diese Objekte zueinander stehen. Ein letzter Schritt wäre die Verknüpfung mit externen URIs, um auf diese Weise weitere Ressourcen zu erschließen.

Bei der Veröffentlichung von Linked Data sollten Bibliotheken so weit wie möglich auf bereits vorhandene Standards anstatt auf Eigenentwicklungen setzen und sich lieber an der Weiterentwicklung dieser Standards beteiligen. Dieser Standpunkt wurde auch in mehreren anderen Vorträgen geäußert. Gradmann berichtete, dass bei Europeana von der zunächst geplanten Eigenentwicklung auf die Nachnutzung bereits vorhandener Standards umgeschwenkt wurde. Bei den Ontologien kann auf einige Quasi-Standards zurückgegriffen werden, wie SKOS, Dublin Core, FRBR, FOAF (friend of a friend, eine Ontologie zur Beschreibung von Personen und sozialen Netzwerken) oder OAI-ORE (Open Archives Initiative – Object Reuse and Exchange). Zu einem Standard könnte sich auch die „Bibliographic Ontology“ (bibo) entwickeln, die Jakob Voß in einem weiteren Vortrag vorstellte. Die inzwischen in Version 1.3 vorliegende Ontologie basiert auf existierenden Vokabularen wie Dublin Core oder FOAF und wird in einem offenen Community-Prozess von derzeit rund 150 Personen weiterentwickelt. Mit ihr könnten Typen von Dokumenten, Sammelwerken, und Ereignissen, ferner der Publikationsstatus beschrieben werden.

Da es derzeit keine andere so umfangreiche, gut dokumentierte und offen diskutierte Ontologie für bibliographische Daten in RDF gibt, existiert nach Meinung von Voß keine Alternative zu bibo, wenn bibliographische Daten ins Semantic Web gebracht werden sollen. In diesem Zusammenhang warf er die provozierende Frage auf, ob die bibo die Nachfolger bibliographischer Da-

tenformate antreten werde. Svensson dagegen sah bibo in seinem Vortrag zu stark auf das Zitationsmanagement ausgerichtet und stellte daher RDA (Resource Description and Access) ins Zentrum der DNB-Planungen. Sein Fazit lautete allerdings, dass keine optimale Ontologie für Bibliotheksdaten existiere.

Eine Reihe praktischer Tipps, Tutorials und Tools zur Erstellung von Linked Data benannte Neubert. Er gab Empfehlungen, wie die Persistenz von URIs garantiert werden könne und wie die URIs praktisch gestaltet werden sollten. Zudem berichtete er von seinen positiven Erfahrungen mit Mailinglisten. Dort könne die Linked Data Community um Feedback zu eigenen Anwendungen gebeten werden.

Schließlich sollten Bibliotheken auf eine lizenzrechtlich möglichst unbeschränkte Veröffentlichung ihrer Daten achten, um deren weitestgehende Nachnutzbarkeit zu gewährleisten. Dies verdeutlichte vor allem Patrick Danowski, Emerging Technologies Librarian am CERN, in seinem leidenschaftlichen Plädoyer für „Free Data“: Die ganze Mächtigkeit seiner Netzwerkeffekte könne Linked Data nur entfalten, wenn die Daten unter einer freien Lizenz veröffentlicht würden, wobei Danowski für eine Bereitstellung als „public domain“ plädierte.

Wie geht es weiter?

Im Verlauf der SWIB09 wurde deutlich, dass es im Hinblick auf das Semantic Web auch noch einige offene Fragen und Probleme gibt. Einige davon fasste Söderbäck zusammen: Das gegenwärtige Bibliothekssystem und das verbreitete bibliografische Denken seien ebenso wenig wie das vorhandene Rechtssystem für Linked Data geeignet. Söderbäck verband dies allerdings mit der Hoffnung „All this will change...“.

Eine von Jürgen Kett moderierte Abschlussrunde sammelte Ideen zur Frage „Wie geht es weiter?“. Um den Austausch über Linked Data-Projekte fortzuführen, wurde unter anderem vorgeschlagen, eine Kommunikationsplattform einzurichten, die im Umfeld der Semantic Web Aktivitäten des W3C angesiedelt ist. Vielfach wurde zudem der Wunsch nach einer Neuauflage der Tagung in 2010 geäußert.

Das Programm der Tagung, Abstracts und Vorträge sind online unter <http://www.swib09.de>.

Dr. Timo Borst, Birgit Fingerle, Joachim Neubert (ZBW), Anette Seiler (hbz)