

Topic Maps zur Realisierung von Ontologien im Kontext der Vision des Semantic Web

Christine Haller
Zentrum für Graphische Daten-
verarbeitung
Fraunhoferstraße 5
D – 64 283 Darmstadt
haller@zgdv.de

Ulrik Schroeder
Institut für Mathematik & Informatik
Pädagogische Hochschule Ludwigsburg
Reuteallee 46
D-71634 Ludwigsburg
schroeder@ph-ludwigsburg.de

Abstract: Topic Maps ermöglichen, semantische Navigationsstrukturen über Dokumente im Web zu modellieren. Nach einer kurzen Vorstellung der wichtigsten Elemente des Standards untersuchen wir die Anwendbarkeit am Beispiel der Modellierung der Kompetenzen von wissenschaftlichen Mitarbeitern, um Experten für den Review von Fachartikeln auffinden und auswählen zu können. Dabei wird der Standard im Kontext der Vision des Semantic Web diskutiert und drei Modellierungsebenen als wesentliche Grundlage für die systematische Entwicklung und verständliches Design von Topic Maps herausgearbeitet. Die erste Modellierungsebene stellt die Modellierungskonstrukte zur Verfügung, die beiden folgenden ermöglichen den von Staab beschriebenen Wissensmetaprozess und Wissensprozess [St02].

1 Motivation

Der Aufstieg des World Wide Web hat innerhalb kürzester Zeit zu einer Informationsüberflutung seiner Nutzer geführt, die mit dem Problem einer enormen Menge unstrukturierter und heterogener Daten konfrontiert werden. Aktuelle Werkzeuge wie Suchmaschinen oder Kataloge offenbaren eine Lücke zwischen verwalteter Information und dem vom Benutzer gesuchten Wissen und verlangen darüber hinaus ein hohes Maß an Recherchekompetenz [Bu99]. Klassische Techniken des Information Retrieval unterstützen zwar das Finden von Informationen auf der Basis von Wortsuche, Thesauri und Indizes, setzen dabei aber nur auf Statistiken und schiere Rechenleistung. Sie ermöglichen damit nicht die intelligente Beantwortung von Anfragen, die automatische Generierung von Wissen und die maschinelle Bearbeitung aufgefundener Informationsressourcen. Dies sind die Ziele der zweiten Generation des WWW, der Vision des W3C, Daten im Web so zu beschreiben, dass sie nicht nur maschinenlesbar, sondern auch maschineninterpretierbar werden (*Semantic Web*).

Mit dem seit 1999 von der ISO standardisierten Konzept Topic Maps [ISO00] steht ein leistungsfähiger Entwurf für die Anbindung von Informationsressourcen an eine semantische Modellierung ihrer Inhalte zur Verfügung. Die breite Akzeptanz von XML (Extensible Markup Language) für Anwendungen im Internet als äußerst erfolgreicher Motor für viele aktuelle Entwicklungen [Kn02] führte zu einer XML-basierten Syntax für Topic Maps (XTM 1.0 [XTM00]). Daher kann man Topic Maps als XML-basiertes Austauschformat für semantische Netze bezeichnen. Damit adressiert der Topic Map Standard die gleiche Problematik wie die parallel verlaufende Entwicklung der W3C Recommendation RDF (Resource Description Framework) [BHL01]. Der Schwerpunkt der Topic Map Entwicklung liegt allerdings in der für den Nutzer expliziten Modellie-

rung semantischer Navigationsstrukturen und ihrer Visualisierung, während RDF eher auf das für menschliche Nutzer verborgene Wirken der Metadaten als Grundlage intelligenter Agenten abzielt.

2 Der XTM Topic Map-Standard

Die ersten Ideen zu Topic Maps reichen in die frühen 90er Jahre zurück, als die jetzige Davenport Group über Austauschformate für Computerliteratur diskutierte und schließlich die DocBook DTD [Wa99, Pe00] spezifizierte. Das Ziel, Stichwortverzeichnisse verschiedener Dokumentationen zu vereinigen, schlug damals fehl, weil es an formalen Wissensstrukturen in Form von Ontologien mangelte, mittels derer gleichbedeutende Stichworte hätten erkannt werden können.

Eine Topic Map realisiert eine thematische Landkarte, in der Themen (*topics*) in ihren Beziehungen (*associations*) und mit Bezug zu Ausprägungen im Informationspool (*occurrences*) modelliert werden. Da mehrere, z.T. verteilt entwickelte Topic Map-Modelle über einem globalen Informationsraum definiert werden können, schaffen diese eine alternative, assoziative Zugriffsstruktur auf zugrundeliegende Informationsressourcen.

Zentraler Bestandteil des Topic Map-Konzeptes ist der benannte Topic, der einen beliebigen Inhalt (Person, Ort, Ereignis, Konzept) repräsentiert. Eine Klassifikation, Assoziationen oder Ausprägungen werden einem Topic in Form von Attributen zugewiesen, die wiederum selbst als *topics* modelliert werden. Da zum einen mehrdimensionale, ungerichtete Assoziationen und unterschiedliche Rollen der daran beteiligten *topics* definiert werden können und zum anderen alle Topic Map-Elemente Objekte erster Ordnung sind, lassen sich beliebige Beziehungen und Strukturen ausdrücken. Weiterführende Konzepte sind *scopes* (Gültigkeitsbereiche), *public subject* (allgemein bekanntes Thema) und *mergeMaps* (Anpassung von Definitionen beim Vereinigen verschiedener Maps). Diese Konzepte unterstützen die Modularisierung von Topic Map-Modellen und damit die verteilte Entwicklung und Wieder- bzw. Mehrfachverwendung von (Teil-)Modellen.

3 Die Kompetenz-Topic Map (KTM)

Für die Verwaltung von Kompetenzen wissenschaftlicher Mitarbeiter in einem internationalen Forschungsverbund wurde eine Topic Map-Anwendung für die Unterstützung des internen Reviewprozesses entwickelt. Die Ausgangsdaten lagen dabei in unterschiedlichen Formaten in den verschiedenen organisatorischen Einheiten vor. Die zu realisierende semantische Navigationsstruktur sollte zwei vornehmliche Ziele verfolgen: eine präzise Auswahl von Experten für den Review von Fachartikeln zu ermöglichen und die Zusammenstellung standortübergreifender Projektteams adäquat zu unterstützen. Dabei war insbesondere der Aspekt, verschiedene Sichten auf die im Netz vorhandenen Kompetenzen zu haben, von Interesse. Eine detaillierte Beschreibung der Anwendung findet sich in [Ha01].

Der Kern der Arbeit bestand in der Entwicklung von zwei Ontologien: die wissenschaftlichen Arbeitsgebiete und die Organisationsstrukturen des Forschungsnetzwerks. Diese wurden in XTM realisiert und mittels XSLT-Skripten explizite, semantische Navigationsstrukturen auf die referenzierten Webdokumente in Form von Webseiten

(HTML) bzw. SVG Graphiken [Di01] generiert. Bei der ersten Ontologie handelt es sich um eine relativ einfache Entwicklung, die Modellierung von Organisationsstrukturen eines internationalen Forschungsnetzwerks, das Verwaltungseinheiten und Zuordnungen von Mitarbeitern in verschiedenen Rollen zu diesen Einheiten ermöglicht. Inhaltlich schwieriger war die Modellierung thematischer Fachthemen. Ausgangspunkt hierfür bildeten Schlüsselwort-Taxonomien der ACM, die an die Forschungsthemen des Netzwerks angepasst wurden. In zahlreichen Interviews mit Experten der Fachgebiete zeigte sich, dass eine disjunkte Einteilung der Themen und die Herausbildung baumartiger Strukturen, wie sie klassisch unternommen werden, nicht hilfreich sind. Daher wurden zusätzlich zu den vorgegebenen Hierarchien, Querverbindungen der Themen modelliert und zusätzliche Themengebiete aufgenommen.

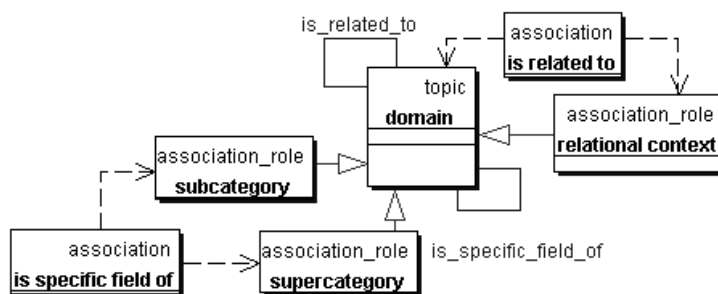


Abbildung 1: Relationen von Forschungsthemen als XTM Assoziationsmodell

Hier zeigt sich eine Stärke des Topic-Map-Ansatzes: die beliebige Modellierbarkeit von Assoziationen, die nicht auf binäre Beziehungen beschränkt sind. Die in Abbildung 1 als UML Modell dargestellte Relation *is_specific_field_of* mit den Rollen *Super-* bzw. *Subcategory* teilt die Themen in eine normale Taxonomie ein. Orthogonal wurde die Relation *is_related_to* modelliert, die Domänen einem Beziehungskontext zuordnet. In diesem Beziehungskontext können beliebige, über verschiedene Kategorien verteilte Domänen in Relation stehen. Eine entsprechende Visualisierung kann dann z.B. Themen des Semantic Web zusammenfassen, auch wenn Technologien wie RDF, Topic Maps oder XML und XSLT oder Information Retrieval in verschiedenen Teilbäumen der Domänentaxonomie angeordnet sind.

4 XTM Entwicklungsstrategie

Bei einer systematischen Entwicklung wird jede Topic Map auf drei Abstraktionsebenen spezifiziert, die sinnvollerweise in verschiedenen Dokumenten erstellt werden, um eine modulare, verteilte Entwicklung mit unterschiedlichen Zuständigkeiten zu ermöglichen.

Die abstrakteste Modellebene (M0) beschreibt die XTM-Basiskonstrukte, die verwendet werden, um das konzeptuelle Modell (M1) für das zugrundeliegende Informationsarchiv zu definieren. Die Modellebene M0 wurde in der Arbeit [Ha01] ursprünglich selbst als *topics* definiert, inzwischen sind sie aber größtenteils als XTM-Kernelemente¹ verfügbar. Das konzeptuelle Modell M1 wiederum definiert, vergleichbar mit einem Daten-

¹ <http://www.topicmaps.org/xtm/1.0/core.xtm>

bankschema, die Beziehungen der Abstraktionen der Individuen (Klassen, Typen, Kategorien oder Begriffe eines allgemeinen Kompetenzstrukturmodells) als Typen, die für die Spezifikation der Anwendung (M2) benötigt werden. Diese Modellebene lässt sich inzwischen in Form von Topic Map Templates realisieren. Abbildung 2 verdeutlicht die Architektur der Topic Map-Anwendung für die Verwaltung von Kompetenzen in einem F&E-Institut für den internen Review-Prozess [Ha01].

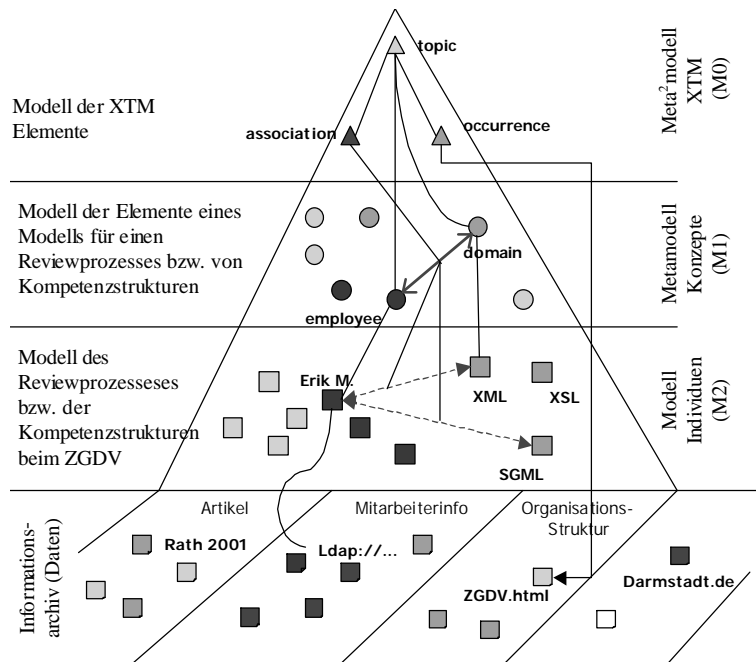


Abbildung 2: Modellierungsebenen von Topic Maps

Bei der Entwicklung der Ontologien für die wissenschaftlichen Kompetenzen der Mitarbeiter und deren Einbindung und Beziehungen innerhalb der Organisationsstrukturen erwies es sich als absolut notwendig, graphische Hilfsmittel zur Kommunikation der Konzepte hinzuzuziehen. Hierfür eigneten sich UML-Klassendiagramme als Werkzeug, da sie alle Metamodellierungskonzepte von Topic Maps abbilden und den Ansprechpartnern aus der Informatik geläufig sind.

5 Bewertung & Fazit

Das KTM-Beispiel demonstriert, wie Metainformationen verteilt erstellt und gepflegt werden, um eine assoziative Zugriffsstruktur auf die im Intranet verfügbaren Informationen zu realisieren: Die organisationellen Daten werden in Form von Corporate Webseiten vom Institut vorgehalten und verteilt in den Abteilungen gepflegt. Diese enthalten Metadaten zu den Mitarbeitern, deren Rollen und Beziehungen, ihren Projekten und Arbeitsgebieten, die als Topic Maps von den Abteilungen spezifiziert werden. Jeder Mitarbeiter pflegt seine eigene Homepage und die Metadaten zu den eigenen Publikationen, den belegten Seminaren, den Arbeitsgebieten den betreuten Diplomarbeiten etc. Die

Bibliothek wiederum verwaltet die Publikationen mit den Schlüsselworten aus der Kompetenz-Ontologie. Die Verknüpfung der verteilten Informationen erfolgt über die im Papier skizzierte Kompetenz-Topic Map.

Die Ausdrucksmächtigkeit von Topic Maps besteht vor allem darin, dass alle Topic Map-Elemente Objekte erster Ordnung sind und sich beliebige auch mehrdimensionale Beziehungen ausdrücken lassen. Somit sind die Konzepte beliebig erweiterbar. Zudem stellen Topic Maps mit Templates und dem *mergeMap*-Konstrukt ein Modularisierungskonzept zur Verfügung, die die verteilte Entwicklung und Pflege von Ontologien erlauben und damit das Organisationsmodell des WWW auf die Metadaten übertragen.

Diese Mächtigkeit hat allerdings ihren Preis. Sie macht den Einsatz kompliziert, insbesondere für den in der Metamodellierung Unerfahrenen. Nur ein wirklich gut vorbereitetes Design macht es möglich, die Übersicht zu behalten. Der Einsatz von UML-Klassendiagrammen (Unified Modeling Language) als Design- und Visualisierungsunterstützung hat sich bei der Realisierung der Kompetenz-Topic Map bewährt, vor allem um die innere Strukturen von Topic Maps (M0-M2) herauszuarbeiten.

Die größte Herausforderung für die Weiterentwicklung des Web liegt allerdings weniger in der Entwicklung und Bereitstellung der Technologien, sondern vielmehr in der Entwicklung der Ontologien selbst, die eine intellektuelle Analyse und Aufbereitung solcher Wissensstrukturen erfordern. Es ist absehbar nicht zu erwarten, dass eine automatische Indexierung Ähnliches leisten kann.

6 Literaturverzeichnis

- [BHL01] Berners-Lee, T. Hendler, J.; Lassila, O.: The Semantic Web, Scientific American, Mai 2001.
- [Bu97] Burneleit, H.-D.: Neue Formen der Suche. In: Buder, M., Rehfeld, W., Seeger, T. und Strauch, D. (Hrsg.): Grundlagen der Praktischen Information und Dokumentation, 4. Ausgabe, KG Saur, München, New Providence, London, Paris: 1997, S. 1016.
- [Di01] Diegelmann, A.: Visualisierung einer Topic Map mittels skalierbarer Vektorgrafiken. Diplomarbeit, FH Darmstadt, Informations- und Wissensmanagement, 2001.
- [Ha01] Haller, C.: XML-Topic Maps für Verwaltung und Retrieval von Kompetenzen in einem F&E-Institut, Diplomarbeit, FH Darmstadt, Fachbereich Informations- und Wissensmanagement, Juni 2001.
- [ISO00] International Organization for Standardization (2000): Information Technology - Document Description and Processing Languages, ISO 13250:2000 - Topic Maps, ISO, Geneva. URL: <http://www.y12.doe.gov/sgml/sc34/document/129.pdf>.
- [Kn02] Knorz, G.: Visualisierung von Zusammenhängen – Von der Wissenskarte zur interaktiven graphischen Topic Map, in: Menne-Haritz, A. (Hrsg.): Online-Findbücher, Suchmaschinen und Portale, Archivschule Marburg, Bd. 36, Marburg 2002.
- [Pe00] Pepper, S.: The TAO of Topic Maps: finding the way in the age of infoglut, Infostream, 2000, <http://www.gca.org/papers/xml europe2000/papers/s11-01.html>.
- [St02] Staab, St.: Wissensmanagement mit Ontologien und Metadaten. In GI Informatik Spektrum, Band 25, Heft 3, Springer Verlag, Heidelberg, Juni 2002; S. 194 – 209.
- [TM01] TopicMapsOrg: Overview of UML Syntax Used for the Models. URL: http://www.doctypes.org/xm/docs/UML_Syntax.pdf.
- [Wa99] Walsh, N; Muellner, L.: DocBook: The Definitive Guide, O'Reilly & Associates, Inc., 1999, <http://www.docbook.org/tdg>.
- [XTM00] XML Topic Maps (XTM) 1.0. <http://www.topicmaps.org>.