

Business Maps: Topic Maps Go B2B

Natalie Jäger

Hauptstrasse 22
63110 Rodgau
natalie.jaeger@web.de

Abstract. Durch eine Vielzahl an unterschiedlichen Ontologien im Datenaustausch zwischen Unternehmen, entstand das Problem der mangelnden Interoperabilität: Wie kann man den Inhalt der bevorzugten B2B Ontologie (wie ebXML, cXML, u.ä.) auf die des Geschäftspartners abbilden? Ansätze auf dem Weg zu mehr semantischer Interoperabilität stellen Published Subject Indicators (PSIs), sowie das Hinzufügen von Metainformation dar. Einen sehr guten Lösungsansatz bieten Business Maps. Business Maps sind Topic Maps, die Informationen über Mapping Prozesse zwischen verschiedenen Ontologien speichern. In dieser Ausarbeitung wird die Bedeutsamkeit der Published Subject Indicators, sowie ebXML für mehr Interoperabilität kurz erläutert. Ausführlicher werden Business Maps, einfache Topic Map Konstrukte, als eine leicht anwendbare, portable, wiederverwendbare Lösung zur besseren Zusammenarbeit im B2B Datenaustausch vorgestellt.

1 Einführung

Aufgrund der Fülle von XML Technologien im Bereich des Wissensmanagements, wie beispielsweise ebXML (Electronic Business XML), cXML (Commerce XML) und FpML (Financial Products Markup Language) entstand ein weit verbreitetes Problem: Welche Elemente dieser verschiedenen Ontologien sind nun tatsächlich gleich und welche doch verschieden? In der heutigen Business Welt trifft man häufig auf dieses Problem, wenn es zum Austausch von Daten kommt. Denn wie vergleicht man die Inhalte und Elemente des bspw. bevorzugten cXML des eigenen Unternehmens mit dem des Geschäftspartners, der ein anderes B2B Protokoll nutzt.

Der Begriff „Ontologie“ ist in dieser Thematik von großer Wichtigkeit. Ontologien versuchen zu definieren, über welche Themen wir sprechen. Im Bereich der Informatik definieren Ontologien, welche Daten ausgetauscht werden und wie man diese referenziert.

Interoperabilität beschreibt das Problem, welches bei der Kommunikation zwischen Quellen mit verschiedenen Ontologien entsteht. Diese verschiedenen Datenquellen, aufgrund obengenannter unterschiedlicher B2B Austausch-Möglichkeiten, haben nur geringe „Zusammenarbeitsfähigkeit“, sind also wenig kompatibel, obwohl der semantische Inhalt meist derselbe ist.

„Interoperability between ontologies is a big, if not the single biggest issue in B2B data exchange.“ [1]

In Abschnitt 2 wird näher auf diese Problematik und ihre derzeitigen Lösungsansätze eingegangen. Es werden die Published Subject Indicators (PSIs) und ebXML als Ansätze vorgestellt, und schließlich Topic Maps, die das Erfassen von Mappings (Abbildungen) zwischen verschiedenen Ontologien ermöglichen. Somit stellen Topic Maps, in diesem Fall „Business Maps“ genannt, einen sehr guten Lösungsansatz dar, der in Abschnitt 3 vorgestellt wird.

2 Topic Maps Go B2B

Topic Maps:

Topic Maps spezifizieren Begriffe und deren Beziehung untereinander. Zusätzlich sind die Begriffe, in diesem Fall „Topics“ genannt, mit Informationsquellen im Internet verknüpft (ähnlich dem RDF = Resource Description Framework). Die Topics sind somit die elementaren Bestandteile einer Topic Map. Topic Maps bestehen aus vielen Topics, die separat von den Daten/Dokumenten gespeichert sind, auf welche sie sich beziehen. Zu beachten ist, dass es sich hierbei nicht um eine Art von Datenbank oder Software handelt, sondern um ein abstraktes Datenmodell. Im Gegensatz zu RDF, das eine für den Computer verstehbare Formalisierung zum Ziel hat, sind Topic Maps eher zur Strukturierung von Wissen aus Sicht der Menschen konzipiert. Topic Maps sollen die bessere Navigation und Suche in Internetressourcen und anderen Dokumenten ermöglichen und dem Austausch von Metadaten dienen. XTM ist eine XML basierte Sprache für Topic Maps. [4]

Ein Topic kann ein Gegenstand, Person, Ort, Eigenname usw. sein, also Dinge oder Konzepte im allgemeinen Sinn. Beispielsweise sind der Film „Lord of the Rings“, die Schauspielerin „Nicole Kidmann“, der Ort „Hollywood“ und die Regisseure „Quentin Tarantino“ und „Sofia Coppola“ gleichberechtigte Topics. Jedes Topic ist Instanz von mindestens einer oder beliebig vielen *Topic Types*, Klassen denen das jeweilige Topic zugeordnet werden kann. Im obigen Beispiel gibt es vier Topic Types: Film, Ort, Regisseur und Schauspieler. Jedes Topic besitzt außerdem beliebig viele *Occurrences (Vorkommen)*, das heißt, dass es mit einem URI (Uniform Resource Identifier) auf Dokumente im Internet verweisen. *Occurrence Roles* geben schließlich an, um welche Art von Informationsquelle es sich handelt, wie bspw. eine Illustration, einen Artikel, ein Kommentar usw.

Topics werden zudem durch *Associations (Beziehungen)* verknüpft, wodurch Begriffe in einen inhaltlichen Zusammenhang gestellt werden. Beispiel für eine Assoziation: Sofia Coppola *arbeitet in* Hollywood. Hierbei stellt *arbeitet in* die Association dar.

B2B:

B2B ist die Abkürzung für Business-To-Business und steht für (elektronische) Kommunikationsbeziehungen zwischen Unternehmen, im Gegensatz zur Kommunikation mit Privatpersonen, Mitarbeitern, usw.

2.1 Probleme im B2B Datenaustausch

Wie erreicht man nun mehr Kompatibilität – Zusammenarbeitsfähigkeit – im Wissensmanagement?

Der naive Ansatz besteht darin, ein komplett neues Konzept zur Strukturierung von Daten zu entwickeln, das alles abdeckt und welches dann ausschließlich genutzt wird. Diese Vereinheitlichung ist praktisch aber nicht umsetzbar, da es einerseits viel zu lange dauern würde bis diese einheitliche *Sprache* entwickelt wäre und andererseits müsste jedes Unternehmen die bereits bestehenden Datenstrukturen der neuen Sprache anpassen.

Das Hauptproblem sollte vielmehr darin bestehen, den reibungslosen Datenaustausch zwischen den verschiedenen Technologien im Wissensmanagement zu ermöglichen.

Ein weiterer Lösungsansatz zu mehr Interoperabilität ist der Gebrauch von *Published Subject Indicators (PSIs)*, welche in Topic Maps genutzt werden. Ein *Published Subject* ist eine Sache für die mindestens ein PSI existiert. Ein *Subject Indicator* ist eine Quelle, auf die eine Anwendung verweist, um die Identität (ID) eines Subjects einem Menschen eindeutig anzuzeigen, während der *Subject Identifier* die adressierbare Bedeutung eines Subjects für den Computer darstellt. Der *Published Subject Indicator (PSI)* wird an einer angegebenen Adresse veröffentlicht, der sogenannten *Public Library*. Der *Published Subject Identifier* ist die Quelle eines PSIs, die von seinem Veröffentlichler als die Quelle gewählt wurde, die von Anwendungen genutzt werden soll, um das *Published Subject* zu identifizieren.

Diese Lösung beruht also darauf, jeder Sache eine eindeutige ID zu geben, und die verschiedenen IDs werden in Bibliotheken (auf die jeder Zugriff hat) gespeichert. Ein Beispiel:

| | |
|-----------------------------|-----------------------|
| <u>Topic:</u> „Zicari“ | <u>PSI:</u> Professor |
| <u>Topic:</u> „Wotschke“ | <u>PSI:</u> Professor |
| <u>Topic:</u> „Waldschmidt“ | <u>PSI:</u> Professor |

Durch die PSIs lässt sich darauf schließen, dass alle drei Themen gleiches beinhalten. Allerdings können PSIs nur dort erfolgreich eingesetzt werden, wo sich die Dinge leicht standardisieren lassen (wie geographische Benennungen, chemische Verbindungen, usw.). PSIs setzen zudem echte semantische Gleichheit der Topics voraus, die aber oft nicht existiert. Denn beispielsweise kann man dem Topic „Apple“ als PSI *Frucht* zuordnen oder *Unternehmen*. Neben der semantische Gleichheit ist auch die Gleichheit von Datenformaten wichtig, die bspw. schon bei der Angabe des Datums meist nicht gegeben ist, denn das Datum 23.07.2004 ist nicht gleich (für eine Maschine) dem 07/23/2004.

Das Manko der *Published Subject Indicators* ist aber, dass sie das Problem der vielen verschiedenen Ontologien eigentlich nur in Bibliotheken verschieben. [1]

Ein andere Ansatz besteht darin, Meta-Information hinzuzufügen, wie in ebXML verwirklicht: „electronic Business XML ermöglicht jedem, irgendwo mit

irgendjemand Geschäfte zu betreiben“, so das anspruchsvolle Ziel der Initiative. ebXML basiert auf einer Menge von Kern-Komponenten, die als primitive Datentypen im B2B Datenaustausch benutzt werden. Diese Kernkomponenten beinhalten keinerlei Semantik, sondern sind nur syntaktische Konstrukte. Benutzt werden die Kern-Komponenten, um Dokumente zusammen zu stellen, die dem B2B Datenaustausch dienen. Der Prozess des Zusammenstellens eines Dokuments basiert auf *Kontext Regeln (Context Rules)*, welche von Unternehmen oder Industrien definiert werden. Die Kontext Regeln bestimmen wie die verschiedenen Kern-Komponenten schließlich in Dokumente verwandelt werden. Dem Dokument wird während dieses Prozesses der semantische Inhalt hinzugefügt. Abb. 1 zeigt diesen Vorgang zum besseren Verständnis:

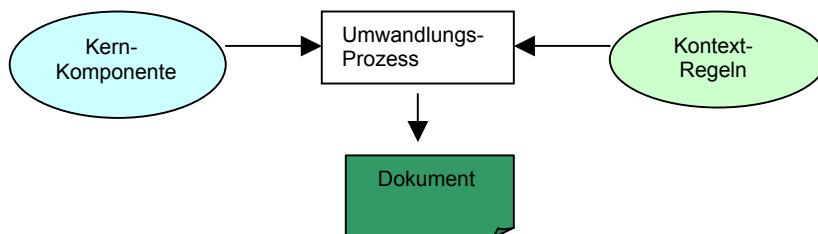


Abb. 1. ebXML: Prozess des Erstellens eines fertigen Dokumentes aus den für ebXML typischen Kernkomponenten

Die Kontext Regeln nutzen *Context Driver (Kontext Treiber)*, um Informationen über den Kontext, also die Metainformation, hinzu zu fügen. Ein solcher *Context Driver* verfügt über Informationen der folgenden Bereiche: Region, Industriezweig, Zeitrahmen, Business Prozess, Produkt, Offizielle Auflagen, Funktion, Service Ebene, unternehmerische Absichten, virtueller Handel, und vieles mehr. [3] Durch den Gebrauch von Context Drivers können bspw. Dokumente landesspezifisch erstellt werden durch einen Region-Context Driver. Das heißt, ein bestimmtes Dokument wird für die USA und für Deutschland aus den Kern-Komponenten erstellt, aber bestimmte Inhalte, wie beispielsweise das bereits erwähnte Datum-Problem, werden dem Bestimmungsort angepasst.

Darin liegt natürlich der Vorteil dieses Lösungsansatzes, da durch den Gebrauch der immer gleichen Kern-Komponenten die Interoperabilität klar erleichtert wird. Ist das nun die optimale Lösung für einen weltweiten, branchenübergreifenden, offenen und interoperablen Standard für den Handel zwischen Geschäftspartnern („Business-to-Business“) sowie zwischen einem Geschäftspartner und dem Kunden („Business-to-Consumer“)?

Einen Nachteil gibt es leider doch, denn alle bereits existierenden Dokumente müssen erst in ebXML umgeschrieben werden.

2.2 Das Wissen in Mappings (Abbildungen)

Die zuerst vorgestellte Lösung, in der alle Daten-Definitionen, alle Dokumente, einer einzigen einheitlichen Ontologie (*Single Ontology*) entsprechen müssen, versagte sofort. ebXML und die *Published Subject Indicators (PSIs)* sind gute Ansätze auf dem langen Weg zu mehr semantischer Interoperabilität; in beiden Fällen können die lokal verschiedene Ontologien erhalten werden, aber ihr Vokabular entstammt einem weltweit gemeinsam benutzten Wortschatz. Sie sind als Mischform (*Hybrid Approach*) des *Single Ontology*- Ansatzes und einer Lösung, die viele Ontologien zulässt (*Multiple Ontologies*), zu sehen. [3] Der Nachteil von *Single Ontologies* und den Mischformen ist in beiden Fällen, dass bestehende Ontologien erst umgeschrieben werden müssen, was zeitaufwendig und kostenspielig ist.

Das soll nicht bedeuten, dass diese neuen vereinheitlichenden Ontologien unnützlich wären. Im Gegenteil, weltweit akzeptiert, sind sie zukunftsweisend für mehr Interoperabilität im B2B Datentransfer.

Trotzdem ist es wichtig zu realisieren, dass eine Vielzahl an verschiedenen Ontologien bestehen bleiben werden, und es müssen Wege gefunden werden auch Austauschmöglichkeiten zwischen diesen verschiedenen Ontologien zu erreichen.

Die oben bereits erwähnten *Multiple Ontologies* haben den großen Vorteil, dass bereits bestehende Ontologien erhalten werden können, solange *Mappings* zwischen verschiedenen Ontologien zur Verfügung gestellt werden. *Mappings* sind Übersetzungen zwischen zwei oder mehreren verschiedenen Ontologien, mathematischer ausgedrückt: Abbildungen, nicht unbedingt bijektiver Art. So ermöglichen *Mappings* bspw. die Übersetzung von cXML in eine andere Ontologie. Es würde sehr viel Arbeit bedeuten, wenn man immer zwischen jeweils zwei Ontologien eine Abbildung vornehmen müsste, aber das ist nicht der Fall, da Abbildungen wiederverwendbar sind.

Es gilt, das Wissen über den *Mapping Prozess* an sich zu erfassen!

Ein Beispiel: Amazon möchte seine Kundenrechnungen durch ein kleineres Unternehmen bearbeiten lassen. Die entsprechenden Daten über Kunden und zu zahlende Beträge gelangen zum kleineren Unternehmen. Dieses verwendet nun die „Lieferadresse“ von Amazon als die Rechnungsadresse in ihren Rechnungsformularen. Allerdings ist die „Lieferadresse“ nicht zwangsläufig auch die „Rechnungsadresse“, da es bei Amazon auch möglich ist, Waren als Geschenk zu versenden.

Es wäre also von Vorteil diese Information, ob nun „Lieferadresse“ wirklich immer als „Rechnungsadresse“ benutzt werden kann oder nicht, „irgendwo“ zu speichern. Auch wenn das für Amazon klar ist, so doch nicht für Außenstehende. Würde man also auch diese Art der Information speichern, so wäre es viel einfacher, Abbildungen zwischen verschiedenen Ontologien zu erstellen.

Die Antwort auf das „irgendwo“ speichern, liegt in den *Business Maps*.

Im folgenden Modell (Abb. 2) ist dargestellt, welches Wissen man aus dem Prozess des Abbildens zwischen Ontologien ziehen kann. In der linken Seite der Graphik handelt es sich um Dokumente gespeichert nach Wichtigkeit der Geschäftspartner. Zu beachten ist, dass die Daten nicht vollständig gespeichert werden, sondern es genügt,

den Dokumenten Variablen zu zuweisen, welche die Daten dann sicher identifizieren. Die rechte Seite der Graphik zeigt das eigentliche „Mapping“, also die Abbildung.

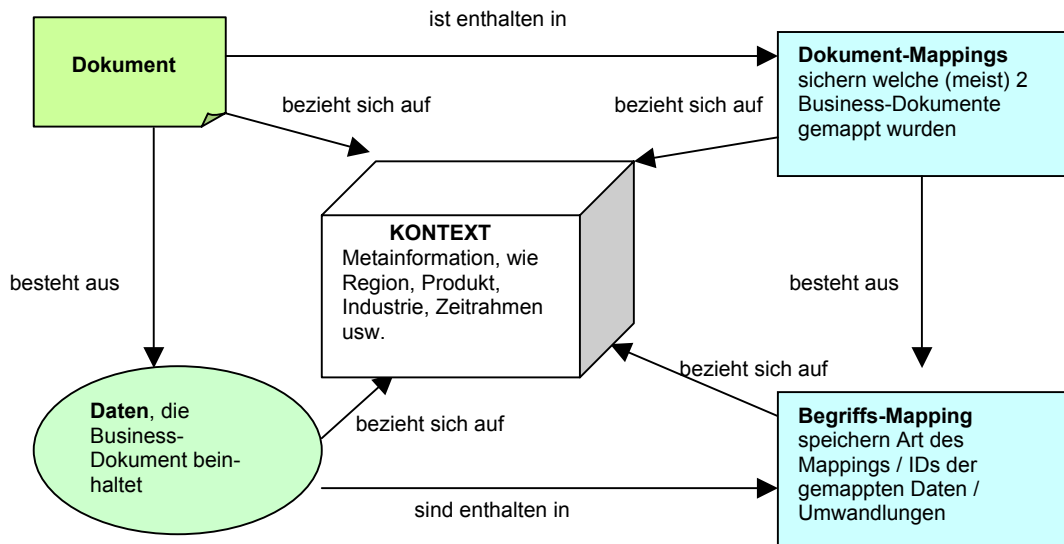


Abb. 2. Das Wissen in *Mappings* -- Linke Seite: Dokumente gespeichert nach Wichtigkeit der Geschäftspartner. Rechte Seite: stellt das eigentliche Mapping (Abbilden) dar.

2.3 Business Maps

In einer Business Map wird gespeichert, was abgebildet wird (Daten des Quell-Dokuments und die des Ziel-Dokuments), welche Art der Beziehung vorliegt zwischen den Daten, entweder Gleichheit, Teilmenge oder Obermenge, und welcher Kontext, also Metainformation vorliegt (gleich dem Context Driver bei ebXML).

Topic Maps eignen sich bestens um die Informationen beim Abbilden zu speichern.

“Topic Maps are an excellent vehicle to store such information, thus yielding Business Maps.” [1]

Die in Business Maps gespeicherten Abbildungen sind zum großen Teil wiederverwendbar, was den Prozess des *Mappings* zwischen verschiedenen Ontologien erleichtert. Durch den Gebrauch von Business Maps lässt sich somit das Abbilden einer Ontologie in ein andere, an sich speichern, außerdem die Information über die Erstellung der Map (wer hat sie erstellt), und auch Widersprüche werden angezeigt (wie im oben genannten Beispiel: Die „Lieferadresse“ in der Ontologie von Amazon entspricht nicht der „Rechnungsadresse“ in der Ontologie des anderen Unternehmens.) und gespeichert, um bei der nächsten Abbildung vermieden zu werden.

Das Auftreten solcher Ungenauigkeiten und Widersprüche ist nicht zu umgehen, da Abbildungen von Menschen erstellt werden, nicht von Maschinen. Das Feld der möglichen Einsatzmöglichkeiten für Topic Maps ist deshalb so vielfältig, weil der Standard an sich so abstrakt, offen und flexibel ausgelegt ist. Die Struktur mit Topics und Associations ist nahe am menschlichen Denken angelehnt und daher bestens geeignet zum Speichern und Austauschen von Wissen über Abbildungen. Zudem sind Business Maps (Topic Maps) leicht für Menschen lesbar, und das bezieht sich nicht nur auf Computerspezialisten, denn besonders hilfreich sind solche Business Maps auch für Geschäftsleute, insbesondere für Business Analysts.

Ein weiteres Plus von Topic Maps ist die Möglichkeit zwei unterschiedliche Topic Maps zu vereinigen, durch *Merging*. Für das Mischen von Topic Maps gibt der XTM-Standard explizite Regeln an. Eine dieser Regeln bewirkt, dass zwei Topics, die auf dasselbe öffentliche Topic referieren, miteinander vereinigt werden. Das so entstandene neue Topic enthält alle Eigenschaften der beiden ursprünglichen Topics. Durch diese Möglichkeit lassen sich zwei verschiedene Abbildungen sehr gut vergleichen. Angenommen es soll eine Business Map für „Gewinn Europa“ erstellt werden, und es liegen bereits die Business Maps „Gewinn Deutschland“ und „Gewinn Spanien“ vor. Dann kann durch einfaches Merging schneller Zugang zu allen relevanten *Mapping* Informationen aus früheren Business Maps zum entsprechenden Themengebiet bestehen.

Ein weiteres Konstrukt der Topic Maps ist der *Scope*. Ein Scope ist ein Gültigkeitsbereich oder Kontext, in dem eine Aussage gültig ist. Werden gar keine Scope-Attribute explizit definiert, so liegt der *unconstrained Scope* (der Scope, der alle Topics einer Topic Map beinhaltet) vor. Der Zweck des Scope ist es, dem Autor einer Topic Map die Möglichkeit zu geben, zu definieren in welchem Rahmen seine Aussagen oder charakteristischen Zuweisungen Gültigkeit besitzen. [2]

Nach dem Vereinigen zweier Business Maps kann man nun den Scope benutzen, um die Business Prozesse heraus zu filtern, an denen man interessiert ist.

Da Business Maps auf dem Konstrukt von Topic Maps beruhen, hat man eine einfache, flexible Möglichkeit, das Wissen in bereits existierenden Mappings wieder zu verwenden.

3 Ausblick -- Zukunft der Business Maps

Wünschenswert wäre ein akzeptierter Standard für Business Maps, um den Austausch von *Mappings* – Abbildungen zwischen verschiedenen Ontologien – mit allen Unternehmen zu ermöglichen. Dies hat nichts mit dem ehrgeizigen Ziel zu tun eine einzige einheitliche Ontologie (*Single Ontology*) zu kreieren! Um Business Maps noch schneller zu etablieren, wäre Software zum Erstellen, Filtern, Importieren/Exportieren, Editieren und Abfragen der Maps von Vorteil. Berücksichtigt wird auch der Fakt, dass es Menschen sind und selten Maschinen, die Abbildungen erstellen, weshalb Business Maps, da einfache Topic Map Konstrukte, eine große Erleichterung darstellen.

“All in all, Business Maps could provide for a huge facilitation of human-mediated ontology mapping.” [1]

Referenzen / Weiterführende Literatur

1. de Graauw, Marc : Business Maps : Topic Maps Go B2B, xml.com (2002)
2. Steve Pepper, Graham Moore: XML Topic Maps :
<http://www.topicmaps.org/xtm/1.0>
3. de Graauw, Marc : What is 'is' ? , marcdegrauw.com (2001)
4. What Topic Maps are: <http://www.easytopicmaps.com>
5. Beispielcode für Business Maps in XTM:
<http://www.marcdegrauw.com/itm/businessmap1.xtm>
6. Zum generellen Überblick in Webtechnologien: Tim Berners-Lee, James Hendler, Ora Lassila: The Semantic Web, <http://www.sciam.com>

Anhang

Ontologie (Begriffsklärung) :

Formal definiertes System von Dingen oder Konzepten und Relationen zwischen diesen Dingen, sowie Regeln. Als Beispiel lässt sich eine Datenbank anführen: die Struktur der Datenbank und der Inhalt (Daten) bilden ein Ganzes. Sprachen zur Beschreibung von Ontologien sind zum Beispiel RDF, F-Logic oder die vom World Wide Web Consortium als Sprache des *Semantic Web* propagierte Ontology Web Language (OWL).

Ontologien in der Informatik versuchen das Wissen eines bestimmten Bereiches in semantischen Konzepten und Beziehungen zwischen ihnen abzubilden. Universell einsetzbare Ontologien (*Single Ontologies*) scheitern jedoch an Problemen, wie mangelnde Akzeptanz und wenigen Anwendungsmöglichkeiten, da sie künstlich geschaffen und nur von wenigen benutzt werden. [6]

XML (Extensible Markup Language) :

Standard zur Definition von Auszeichnungssprachen, die zur Beschreibung von Informationen oder des Verfahrens oder der Schritte dienen, die zur Darstellung nötig sind.

Vereinfachte Teilmenge von SGML und verwandt mit HTML. Die Grundidee hinter XML ist, Daten und ihre Repräsentation zu trennen.

Semantik (Bedeutungslehre):

Lehre von der inhaltlichen Bedeutung einer Sprache; Programmiersprache in der Informatik. Sie legt die Bedeutung einzelner Sprachelemente fest.