

Freie wissenschaftliche Arbeit zur Erlangung  
des Grades eines Diplom-Wirtschaftsinformatikers  
mit dem Thema:

# **Einsatzmöglichkeiten von Topic Maps zur flexiblen Navigation in elektronischen Dokumenten**

Eingereicht am: 07.05.2001

Eingereicht bei: Prof. Dr. Eric Schoop  
Lehrstuhl für Wirtschaftsinformatik,  
insbesondere Informationsmanagement  
Fakultät Wirtschaftswissenschaften  
Technische Universität Dresden

von Ronald Heckel  
aus Rodewisch  
geb. am 31. August 1976

## Zusammenfassung:

In der Arbeit wird der noch junge Standard der Topic Maps vorgestellt. Es wird ein Einblick in den inneren Aufbau konformer Dokumente gegeben. Mit XML Topic Maps und RDF werden Weiterentwicklungen bzw. ergänzende Konzepte vorgestellt. Im zweiten Teil werden denkbare Anwendungsmöglichkeiten des Standards mit ihren Vor- und Nachteilen beschrieben. Abschließend wird ein Vorgehensmodell zur Modellierung mit Topic Maps vorgeschlagen und dieses prototypisch für das Projekt ELBE umgesetzt.

## Schlagworte:

Topic Maps, ISO 13250, XTM, XLink, XML, semantisches Netz, RDF, ELBE, Information Mapping®

# Inhalt

<b>1</b>	<b>Einleitung und Aufbau der Arbeit .....</b>	<b>1</b>
<b>2</b>	<b>Topic Maps .....</b>	<b>2</b>
2.1	ISO/IEC 13250 .....	2
2.1.1	Das Topic Map Element.....	2
2.1.2	Das Topic Element.....	2
2.1.3	Das Occurrence Element (Vorkommnis).....	5
2.1.4	Das Element Assoc (Assoziation) .....	7
2.1.5	Zusätzliche Elemente .....	8
2.1.6	Weiterführende Arbeiten .....	9
2.2	XML Topic Maps (XTM).....	9
2.2.1	Änderungen in Bezug auf ISO 13250.....	10
2.3	Vergleich von ISO 13250 mit XTM .....	12
<b>3</b>	<b>Topic Map Software .....</b>	<b>13</b>
3.1	Die Topic Map Engine .....	14
3.1.1	Der Topic Map Graph.....	14
3.1.2	Merging .....	14
3.1.2.1	Subject-based Merging .....	14
3.1.2.2	Topic Naming Constraint-based Merging.....	15
3.1.3	Relationen .....	15
3.1.4	Überprüfung der Gültigkeit .....	16
3.1.5	Anfragesprache .....	16
3.1.5.1	XML Query .....	17
3.1.5.2	TMQL .....	17
3.1.5.3	Engine-spezifische API .....	17
3.1.6	Weitere Anforderungen .....	18
3.1.7	Bisher verfügbare Topic Map Engines .....	18
3.2	Topic Map Design .....	18
3.2.1	Das Topic Map Template (Vorlage) .....	19
3.2.2	Gültigkeitsprüfung .....	20
3.2.3	Versionskontrolle.....	20
3.2.4	Nutzerrollen und –rechte .....	20

3.2.5	Zugriff auf Informationsressourcen .....	20
3.2.6	Testumgebung .....	20
3.3	Topic Map Browser.....	21
<b>4</b>	<b>Einsatzmöglichkeiten von Topic Maps .....</b>	<b>23</b>
4.1	Wissensmanagement im Unternehmen .....	23
4.1.1	Dokumentenmanagement.....	24
4.1.1.1	Vorteile .....	26
4.1.1.2	Nachteile .....	26
4.1.2	Einsatzmöglichkeiten von Topic Maps .....	26
4.1.2.1	Verwendung unstrukturierter Dokumente.....	26
4.1.2.2	Suchbegriffe .....	27
4.1.2.3	Nutzung von Assoziationen.....	27
4.1.2.4	Fazit .....	28
4.2	Suchmaschinen im Internet.....	28
4.2.1	Themenkataloge .....	29
4.2.2	Suchmaschinen.....	29
4.2.2.1	Vorteile .....	30
4.2.2.2	Nachteile .....	30
4.2.3	Einsatzmöglichkeiten von Topic Maps .....	31
4.2.3.1	Einsatz in Suchmaschinen .....	31
4.2.3.2	Einsatz in Themenkatalogen.....	31
4.2.3.3	Fazit .....	32
4.2.4	RDF als Alternative .....	32
4.2.4.1	Grundlagen .....	33
4.2.4.2	Vergleich mit Topic Maps.....	34
4.2.4.3	Bewertung und Ausblick.....	35
4.3	Unterstützung für elektronisches Lernen.....	35
4.3.1	Bisheriger Einsatz elektronischer Hilfsmittel in der Lehre .....	35
4.3.2	Anforderungen an Lernumgebungen .....	36
4.3.3	Einsatzmöglichkeiten von Topic Maps .....	37
4.3.3.1	Vorteile .....	37
4.3.3.2	Nachteile .....	38
4.4	Katalogisierung großer Dokumentenbestände .....	39
4.4.1	Diplomarbeitsverzeichnis .....	40
4.4.1.1	Einheitlicher Begriffskatalog.....	40
4.4.1.2	Vereinfachte Erstellung eines Glossars .....	40

4.4.1.3	„Referenzbuchliste“ .....	41
4.4.1.4	Leichtereres Auffinden von Literatur zum Thema .....	41
4.4.1.5	Leichtere Navigation zu Quellen im Internet .....	41
4.4.1.6	Fazit .....	41
4.4.2	MSDN Library .....	41
4.4.2.1	Bessere Unterteilung der Dokumentation .....	42
4.4.2.2	Verbesserte Nutzung der Beispielprogramme .....	42
4.4.3	Technische Dokumentation .....	43
4.4.3.1	Vorteile .....	43
4.4.3.2	Nachteile .....	44
4.4.4	Lexika und Enzyklopädien .....	44
4.5	Container für Metadaten .....	44
<b>5</b>	<b>Modellierung von Topic Maps .....</b>	<b>45</b>
5.1	Grundlagen semantischer Netze .....	45
5.1.1	Modellierung mit semantischen Netzen .....	46
5.1.2	Vergleich mit Topic Maps .....	47
5.2	Modellierungskriterien .....	48
5.2.1	Syntaktische und semantische Richtigkeit .....	48
5.2.2	Relevanz .....	49
5.2.3	Klarheit und Verständlichkeit .....	49
5.2.4	Wirtschaftlichkeit .....	49
5.2.5	Strukturierung und Integration von Konzepten .....	50
5.2.6	Vergleichbarkeit .....	50
5.3	Vorschlag eines möglichen Vorgehensmodells .....	50
5.3.1	Schritt 1: Abgrenzung des Diskursbereiches und Festlegen der Zielstellung .....	51
5.3.2	Schritt 2: Aufstellen der Ontologie .....	51
5.3.3	Schritt 3: Aufstellen der Topics .....	52
5.3.4	Schritt 4: Zuordnen der Informationsressourcen .....	52
5.3.5	Schritt 5: Aufstellen der Assoziationen .....	52
5.3.6	Schritt 6: Abgleichen mit der Zielstellung .....	52
5.4	Prototypische Umsetzung für das Projekt ELBE .....	53
5.4.1	Besonderheiten für die Umsetzung .....	53
5.4.2	Schritt 1: Abgrenzung des Diskursbereiches und Festlegen der Zielstellung .....	53
5.4.3	Schritt 2: Aufstellen der Ontologie .....	54

## Inhalt

---

5.4.4	Schritt 3: Aufstellen der Topics.....	55
5.4.5	Schritt 4: Zuordnen der Informationsressourcen .....	57
5.4.6	Schritt 5: Aufstellen der Assoziationen.....	57
5.4.7	Schritt 6: Abgleichen mit der Zielstellung .....	57
5.4.8	Probleme bei der Umsetzung.....	57
5.5	Offene Fragen im Vorgehensmodell .....	58
5.5.1	Zuordnung von Gültigkeitsbereichen.....	58
5.5.2	Vorgehensweise bei der Modellierung von Assoziationen .....	59
5.5.3	Integration des Wirtschaftlichkeitsprinzips .....	59
<b>6</b>	<b>Zusammenfassung .....</b>	<b>60</b>
	<b>Abkürzungen .....</b>	<b>61</b>
	<b>Abbildungen .....</b>	<b>63</b>
	<b>Tabellen.....</b>	<b>64</b>
	<b>Glossar.....</b>	<b>68</b>
	<b>Anhang.....</b>	<b>71</b>

# 1 Einleitung und Aufbau der Arbeit

Durch die ständig anwachsende Fülle an verschiedenartigen Informationen ist es wichtig, diese in einer geeigneten Weise ordnen und katalogisieren zu können, um nicht den Überblick über die tatsächlich benötigten Informationen zu verlieren.

In der vorliegenden Arbeit soll mit dem Topic Map-Standard eine Möglichkeit vorgestellt werden, diese Aufgabe zu bewältigen. Da es sich hierbei um einen noch jungen Standard handelt, liegen noch ungenügend Arbeiten von dritten Autoren vor. So soll zum Beispiel erst in diesem Jahre das erste Buch zu diesem Thema erscheinen. Die Quellen, die zum Erstellen dieser Diplomarbeit hinzugezogen werden, beschränken sich deshalb häufig auf Arbeiten der Entwickler des Standards, die auf einschlägigen Konferenzen vorgestellt wurden, um Topic Maps der breiteren Öffentlichkeit vorzustellen.

Das folgende Kapitel befasst sich mit den Grundlagen von Topic Maps. Es gibt einen verkürzten Einblick in die Entwicklungsgeschichte und stellt die wichtigsten Elemente vor. Anschließend wird ein Einblick in die aktuellen Weiterentwicklungen des Standards getan. Bei der Beschreibung der XML Topic Maps (XTM) verzichtet der Autor bewusst auf eine detaillierte Beschreibung der einzelnen Elemente der hier zugrunde liegenden Dokumenttypdefinition (DTD). Die entscheidende Ursache für dieses Vorgehen wird mit der äußerst dynamischen Entwicklung an dieser Weiterentwicklung begründet, die zu häufigen Änderungen an der Spezifikation führt.

Im Kapitel 3 stehen die Anforderungen an entsprechende Verarbeitungssoftware im Mittelpunkt der Ausführungen. Die Verfügbarkeit solcher Computerprogramme wird einen wesentlichen Anteil an der Durchsetzung von Topic Maps ausmachen. Bei der Beschreibung wird dabei auf die jeweiligen Phasen der Erstellung einer Topic Map eingegangen und die sich daraus ergebenden Anforderungen abgeleitet.

Das Hauptaugenmerk der vorliegenden Arbeit liegt auf den potentiellen Einsatzgebieten dieser neuen Technologie. Der Autor beschreibt jeweils die herkömmlichen Ansätze und zeigt deren entscheidende Nachteile auf. Anschließend werden die Chancen aufgezeigt, die durch den Einsatz von Topic Maps entstehen. Es werden aber auch Defizite der neuen Technik angesprochen und soweit möglich Vorschläge zur Lösung dieser Probleme unterbreitet. Nach Möglichkeit wird versucht, eine Integrationsmöglichkeit zu bestehenden Standards herzustellen.

Im letzten Kapitel wird der Versuch unternommen, ein Vorgehensmodell zur Modellierung von Topic Maps vorzustellen und dieses auf das Projekt ELBE prototypisch anzuwenden. Die vom Autor dazu entwickelte Software wird im Anhang beschrieben.

## 2 Topic Maps

Im folgenden Kapitel werden Teile des grundlegenden Standards ISO/IEC 13250 beschrieben. Zudem werden weiterführende Entwicklungen und die Arbeit mit XML Topic Maps (XTM) vorgestellt.

### 2.1 ISO/IEC 13250

Der Standard ISO/IEC 13250 wurde im Zeitraum von 1996 bis 2000 entwickelt. Die Anfänge gehen jedoch schon bis an den Anfang der neunziger Jahre zurück. Die sogenannte Davenport Group beschäftigte sich mit Anwendungsmöglichkeiten des Hyperlinkings im *HyTime*-Standard. Die Gruppe teilte sich später und eine daraus resultierende Teilgruppe beschäftigte sich mit dem Problem des Zusammenführens von Indexen, welches ein Hauptanwendungsgebiet von Topic Maps ist. Indexe eines Buches erlauben dem Leser einen direkten Zugriff auf die gewünschte Information ohne langes aufwendiges Suchen. Eine solche Eigenschaft kann mit Hilfe von Topic Maps auch für elektronische Dokumente erreicht werden. Diese und weitere Einsatzmöglichkeiten werden im Kapitel 4 beschrieben.

ISO/IEC 13250 wurde auf Basis von architektonischen Formen als Meta-DTD modelliert. Die nachfolgenden Abschnitte behandeln die Kernkomponenten eines Topic-Map-Dokumentes und vernachlässigen Konstrukte, die aus dem HyTime-Standard resultieren. Zu einer Einführung in die architektonischen Formen wird das Buch von Lobin empfohlen ([Lobi00], S.86ff). Ein kürzerer Abriss findet sich auch in einer Seminararbeit des Autors ([Heck00], S.11f).

#### 2.1.1 Das Topic Map Element

Dieses Element stellt aus Sicht der DTD das Dokumentenelement einer gültigen Topic Map dar ([Biez+00], Abschnitt 5.1). Es kann als Unterelemente Topics, Assoziationen, Facetten und sogenannte „Themes to be added“ haben. Die Elemente bosspec und TMBrid stammen von den Anforderungen an HyTime-konforme Dokumente und werden hier nicht berücksichtigt. Die möglichen Unterelemente sind kontextfrei, das heißt, dass sie unabhängig vom jeweiligen Diskursbereich der Topic Map auftreten können.

#### 2.1.2 Das Topic Element

Topics bilden den zentralen Bestandteil in ISO 13250. Ein *Topic* kann für jedes beliebige „Ding“ erstellt werden, für jeden Gegenstand, jedes Substantiv oder was auch immer. Es ist eine Art Container, der erst mit Inhalt gefüllt werden muss. Dieser Inhalt charakterisiert

dann das Topic und ist dann anwendungsspezifisch. Je nach Anwendungsgebiet kann ein Topic dann einen Begriff (z.B. in einem Lexikon), eine Komponente (z.B. in der technischen Dokumentation), ein Projekt (z.B. im innerbetrieblichen Wissensmanagement) beschreiben. Zu diesen Charakteristiken gehören die Namen des Topics, Verweise auf Informationsressourcen und die Rollen, die das Topic in Assoziationen einnimmt ([Biez+00],S.5). Der Standard unterscheidet *Topic Names* nach ihrer Verwendung. Es wird dabei unterschieden in die sogenannten „Base Names“ zur allgemeinen Benennung des Topics sowie die optionalen „Sort Names“ als Sortierschlüssel und „Display Names“ als alternative Anzeigemöglichkeit im Anwendungsprogramm ([Biez+00],S.12ff.). Ein Topic muss dabei aber mindestens ein Element „Base Name“ besitzen, um gemäß der DTD gültig zu sein. Der Standard sieht vor, für ein und dasselbe Topic ein oder mehrere Topic Names vergeben zu können. Dies kann genutzt werden, um bei der Modellierung der Topic Map zum Beispiel die Mehrsprachigkeit zu berücksichtigen. Aber auch andere Einsatzmöglichkeiten sind denkbar (vgl. [Pepp00], S. 7). So ist beispielsweise auch die Berücksichtigung von Adelstiteln in einer Geschichtsdatenbank möglich.

Innerhalb der *Topic Map* muss jedes Topic durch ein ID-Attribut eindeutig gemacht werden. Topics mit ein und dem selben ID-Attribut müssen von der Anwendung zu einem Topic zusammengefasst werden. Dieser Vorgang wird als *Merging* bezeichnet und an späterer Stelle detaillierter beschrieben. Problematisch wird dies jedoch, wenn die beiden Topics tatsächlich unterschiedliche Dinge beschreiben. Als Beispiel für diesen Fall sei ein Topic „Mondaufgang“ benannt. Es beschreibt zum ersten ein Gemälde von Caspar David Friedrich und könnte aber auch zur Beschreibung des Mondaufgangs in einem astronomischen Programm dienen. Zur Unterscheidung bietet ISO 13250 das Konzept des *Scope* (Gültigkeitsbereich) an. Durch dieses Attribut kann angegeben werden, für welche Bereiche ein Topic Gültigkeit besitzt. Diese Eigenschaft könnte nach Ansicht des Autors besonders gut bei Suchmaschinen im Internet genutzt werden. Durch den Einsatz eines Suchraumes wird die Trefferanzahl mitunter drastisch reduziert. Das folgende Beispiel demonstriert die Einsatzmöglichkeiten des Attributs „scope“.

```
<topic id="Mondaufgang" scope="Gemälde">
  <topname>
    <basename>Mondaufgang</basename>
  </topname>
</topic>

<topic id="Mondaufgang" scope="Astronomie">
  <topname>
    <basename>Mondaufgang</basename>
  </topname>
</topic>
```

Abb. 2.1: Nutzung des Attributes "scope"

Ein Gültigkeitsbereich kann jedoch nicht nur für ein Topic angegeben werden. Es ist ebenfalls möglich, die Gültigkeit ganzer Namensangaben (das Element „topname“) und / oder einzelne Namen (für Sortierung, Anzeige,...) einzuschränken. Das nahe liegendste Anwendungsbeispiel ist die bereits oben erwähnte Mehrsprachigkeit. Dazu müsste nur die jeweilige Landessprache im Attribut „scope“ angegeben werden.

Durch die Angabe von optionalen Typen (*Topic Types*) können auf einfache Art und Weise Klasse-Instanz-Beziehungen modelliert werden. Die Typen stellen ihrerseits ebenso ein Topic dar, welches ebenfalls von einem bestimmten Typ sein kann, usw. Andere Relationen lassen sich durch den Einsatz von *Assoziationen* darstellen. Sie werden im Abschnitt 2.2.4 dieses Kapitels und im Kapitel 3 beschrieben.

```
<topic id="Person">
  <topname>
    <basename>Person</basename>
  </topname>
</topic>

<topic id="Künstler" types="Person">
  <topname>
    <basename>Künstler</basename>
  </topname>
</topic>

<topic id="Maler" types="Künstler">
  <topname>
    <basename>Maler</basename>
  </topname>
</topic>

<topic id="CDFriedrich" types="Maler">
  <topname>
    <basename>Caspar David Friedrich</basename>
    <sortname> Friedrich, Caspar David</basename>
    <dispname>Caspar D. Friedrich</basename>
  </topname>
</topic>
```

Abb. 2.2: Beispiel für den Einsatz von Topic Names und Topic Types

Die im obigen Beispiel angegebenen Beziehungen sind aber trotzdem nur Klasse-Instanz-Beziehungen. ISO 13250 legt fest: „The topics relationships established by the types attribute are not superclass-subclass relationships. They are only class-instance relationships“ ([Biez<sup>+</sup>00], §5.2.1). Um die eigentlich beabsichtigte Aussage „Ein Maler ist ein Künstler“ abbilden zu können, müssen entsprechende Assoziationen zwischen diesen Topics aufgebaut werden. Für eine solche Assoziation werden aber noch eine Reihe weiterer Topics benötigt, die im Grunde genommen nichts mit der eigentlichen Aussage zu

tun haben. Sie werden lediglich als eine Art Vorlage für gleichartige Assoziationen genutzt. Der „Quelltext“ der obigen Aussage könnte dann wie folgt aussehen:

```
<topic id="tt-klasse">
  <topname><basename>Topic Klasse</basename></topname>
</topic>

<topic id="tt-superkl">
  <topname><basename>Superklasse</basename></topname>
</topic>

<topic id="tt-subkl">
  <topname><basename>Subklasse</basename></topname>
</topic>

<topic id="at-superklasse-von">
  <topname><basename>Superklasse-von</basename></topname>
</topic>

<assoc type="at-superklasse-von">
  <assocrl type="tt-subkl">Maler</assocrl>
  <assocrl type="tt-superkl">Künstler</assocrl>
</assoc>
```

Abb. 2.3: Umwandlung in eine Superklasse-Subklasse-Relation (in Anlehnung an [Rath99])

### 2.1.3 Das Occurrence Element (Vorkommnis)

Jedes Topic kann mit einer beliebigen Anzahl von Informationsressourcen verknüpft werden. Diese Ressourcen sind in ihrer Art durch den Standard nicht näher spezifiziert ([Biez<sup>+</sup>00], S.15). Dadurch könnten zum Beispiel auch Dokumente, die nicht in elektronischer Form vorliegen, referenziert werden. Diese Eigenschaft lässt sich besonders gut für einen schrittweisen Umstieg im innerbetrieblichen Wissens- und Dokumentenmanagement verwerten, da auch Papierdokumente über ein geeignetes Adressierungsverfahren in der Topic Map über einen entsprechenden Verweis abgelegt werden können. Durch die Angabe eines Rollennamens wird die Kategorie festgelegt, zu der die Informationsressource gehört. Das optionale Typ-Attribut verweist auf ein Topic der Map und beschreibt dadurch den Typ der Ressource näher. Für den Adressierungsmechanismus werden durch den Standard ebenfalls keine Vorgaben gemacht. Bei der Adressierung elektronischer Dokumente wird sich nach Meinung des Autors XLink (nach der Standardisierung durch das W3-Consortium) auf lange Sicht durchsetzen können. Derzeit gibt es jedoch noch nicht genügend Software, die solche Dokumente verarbeiten kann. Eine Implementation von XLink wird von Empolis ([www.empolis.co.uk](http://www.empolis.co.uk)) mit dem Programm X2X angeboten. Die gleiche Firma arbeitet auch an einer Topic Map Engine. Für die

Adressierung nicht-elektronischer Informationsressourcen müsste eine anwendungsspezifische Auszeichnung dieser Ressourcen gefunden werden (zum Beispiel würde sich für eine Bibliothek die übliche Standortsignatur anbieten). Die folgende Abbildung zeigt eine mögliche Anwendung.

```
<topic id="Gemälde">
  <topname><basename>Gemälde</basename></topname>
</topic>

<topic id="Mondaufgang" types="Gemälde">
  <topname>
    <basename>Mondaufgang am Meer</basename>
    <sortname>Mondaufgang</sortname>
    <dispname>Mondaufgang am Meer</dispname>
  </topname>
  <occurs
    occr1="Bild von C.D. Friedrich"
    type="Abbildung"
    quelle="Das große Buch des Allgemeinwissens"
    seite="740"
    standort="abcd"
  />
</topic>
```

Abb. 2.4: Vorschlag zur Adressierung nicht-elektronischer Dokumente

Durch die Angabe von Rollen und Typen bieten Topic Maps eine deutlich erweiterte Funktionalität gegenüber herkömmlichen Indexen in Büchern ([Pepp99], S.10). Viele Bücher erweitern Indexe durch typographische Besonderheiten (z.B. Fettdruck der Seitenangabe, wenn zu einem Stichwort eine Abbildung existiert). Durch den Einsatz von Topic Maps sind solche Erweiterungen ebenfalls möglich, bieten aber dem Nutzer weitaus mehr Möglichkeiten. Durch die Angabe der *Occurrence Types* erhält der Nutzer zum Beispiel mehr Informationen zum Kontext des aufgeführten Eintrags im Index. Er kann dadurch erkennen, ob es sich bei der angegebenen Informationsressource um ein Video, einen Text, ein Bild oder ähnliches handelt. Der Rollenname gibt Auskunft darüber, in welchem Zusammenhang die Ressource mit dem gewählten Thema steht.

Die nachfolgende Abbildung ist einem Einführungstext zum Thema von Steve Pepper entnommen und soll die bis hierher aufgeführten Eigenschaften des Standards kurz verdeutlichen. In der Grafik werden die Topics durch die sechs Kreise versinnbildlicht. Die untere Ellipse stellt den Informationspool dar. Die Vorkommnisse (*Occurrences*) werden aufgrund ihrer unterschiedlichen Typen durch verschiedenartige Linien verdeutlicht. Aus der Abbildung wird auch die Trennung zwischen Topics und dem Informationspool deutlich. Diese Trennung ist einer der Schlüssel zur Mächtigkeit des Standards ([Pepp99], S.9). Dieser Sachverhalt wird in den nächsten Abschnitten noch ausführlicher erläutert.

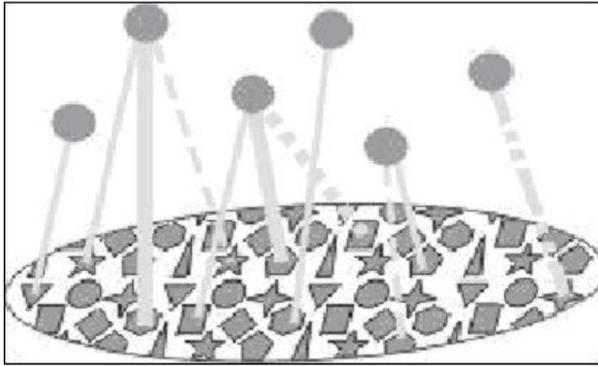


Abb. 2.5: schemenhafte Darstellung der vorgestellten Konzepte ([Pepp99], S.9)

### 2.1.4 Das Element Assoc (Assoziation)

Die bisher vorgestellten Konzepte von ISO 13250 ermöglichen dem potentiellen Nutzer lediglich eine lose Sammlung von Beschreibungen von Topics sowie die Angabe von wichtigen Informationsquellen. Aber erst durch die Zuordnung von Beziehungen zwischen einzelnen Topics eröffnen sich viele neue Anwendungsmöglichkeiten. Im Abschnitt 2.1.2. wurde eine Möglichkeit gezeigt, wie man mit Hilfe der Topic Types einfache Hierarchien im Sinne einer is-a Relation aufbauen kann. Mit Hilfe von Assoziationen können jedoch weit mehr Informationen in eine Topic Map gepackt werden. Zum Beispiel:

- Das Gemälde „Mondaufgang am Meer“ ist in Berlin ausgestellt
- Berlin ist die Hauptstadt der Bundesrepublik Deutschland
- Deutschland ist Mitglied der europäischen Union

Durch die Angabe von Rollen werden die einzelnen Teile der Assoziation näher beschrieben. Die optionalen Typangaben verweisen ihrerseits ebenfalls wieder auf ein entsprechendes Topic in der Map. Mit diesen Zusätzen könnte die erste Aussage der Aufzählung wie folgt modelliert werden.

```
<assoc type="ausgestellt-in">
  <assocrl anchrole="Ausstellungsort" type="Ort">
    Berlin
  </assocrl>
  <assocrl anchrole="Ausstellungstück" type="Gemälde">
    Mondaufgang
  </assocrl>
</assoc>
```

Abb. 2.6: Nutzung von Assoziationen

Die Angabe der Rollen, die einzelne Topics in einer Assoziation einnehmen, ist äußerst wichtig. Verständnisprobleme können zum Beispiel dann auftreten, wenn mehrere Topics in einer Assoziation vom gleichen Typ sind. Es ist unzureichend zu wissen, das ein Topic

„Max Maier“ mit dem Topic „Paul Müller“ in einer „ist-Mitarbeiter-von“ Beziehung steht. Die Rolle der jeweiligen Person ist ungeklärt (wer ist Mitarbeiter?, wer Vorgesetzter?). Deshalb besitzen die Rollenangaben eine große Bedeutung für den Nutzer einer Map.

Topics mit ihren zugehörigen Assoziationen sind gut vergleichbar mit semantischen Netzen. Dabei stellen die Topics die Knoten im Netz dar, während die Assoziationen die Kanten repräsentieren. Durch die Attribute kann ebenfalls der Typ der Kanten abgebildet werden ([Free00], S.13). Topic Maps besitzen durch die bereits oben angesprochene Trennung von den Informationsressourcen einen hohen Informationswert und können als eine Art portables semantisches Netz weitergegeben werden. Ein anderer Bearbeiter dieser Map ist dann in der Lage, auch andere Informationsquellen entsprechend seinen Bedürfnissen zu nutzen (z.B. könnte eine Topic Map alle Werke eines Künstlers referenzieren, während eine zweite alle Arbeiten von Dritten (Kritikern, o.ä.) referenziert). Die Verbindungen zwischen den einzelnen Topics bleiben dabei jedoch erhalten. Die folgende Abbildung soll diesen Fakt noch einmal verdeutlichen.

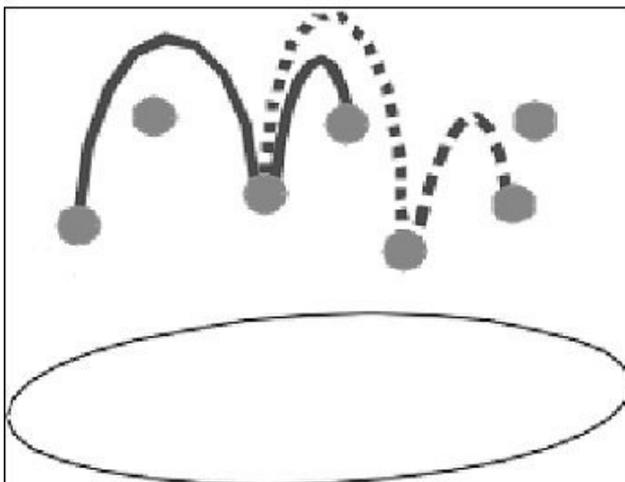


Abb. 2.7: Topic Map als portables semantisches Netz ([Pepp00], S.14)

### 2.1.5 Zusätzliche Elemente

Über das Element „addthms“ können einzelne Topics einer externen Topic Map zu den Gültigkeitsbereichen der „internen“ Topics zugefügt werden. Welchem Gültigkeitsbereich sie genau zugefügt werden, legen die Attributparameter „cassign“ bzw. „tmdocs“ fest (siehe Abbildung 2.8). Mit Hilfe von Facetten ist es möglich, der Topic Map Eigenschafts-Wert-Paare zuzufügen. Dabei können die Typen der Eigenschaften wieder auf Topics innerhalb der Map verweisen. Facetten können zum Beispiel benutzt werden, um Meta-informationen über die Topic Map abspeichern zu können. Damit könnte beispielsweise die Art einer Assoziation näher beschrieben werden (bei einer nicht symmetrischen Assoziation ändert sich die Bedeutung, wenn die Betrachtungsweise geändert wird

(„wurde-gemalt-von“ wird zu „hat-gemalt“). Auf verschiedene Assoziationsarten wird im Abschnitt 3.1.3. detaillierter eingegangen.

```
<addthms addthems="Zusatz" cassign="basename"/>

<topic id="irgendwas">
  <topname>
    <basename>Irgendetwas</basename>
  </topname>
</topic>

wird umgewandelt in:

<topic id="irgendwas">
  <topname>
    <basename scope="Zusatz">Irgendetwas</basename>
  </topname>
</topic>
```

Abb. 2.8: Verwendung von „Added Themes“

### 2.1.6 Weiterführende Arbeiten

Da ISO 13250 als Meta-DTD auf Basis von architektonischen Formen modelliert wurde, bildeten sich im Laufe der Zeit einige abgeleitete DTD's heraus, die in ihrer Struktur geringfügig verschieden sind, aber alle die grundlegenden Prinzipien abdecken. Eine Aufzählung verfügbarer DTD's findet sich unter <http://www.doctypes.org/xtm>. Der vom Autor entwickelten Designsoftware für Topic Maps liegt ebenfalls nur eine abgeleitete Version dieser DTD zugrunde. Sie wird im Anhang der Arbeit beschrieben.

In letzter Zeit erschienen mehrere Arbeiten, um die unterschiedlichen Entwicklungsschienen zu vereinheitlichen und die neuesten technischen Standards besser auszunutzen. Dazu zählt die Veröffentlichung eines XML-Schemas für ISO 13250 ([Brya00]). Es nutzt, soweit möglich, die Fähigkeiten von XLink.

## 2.2 XML Topic Maps (XTM)

Die neueste und auch am weitesten fortgeschrittene Arbeit stellt die Entwicklung von *XTM* (XML Topic Maps) dar. Die Version 1.0 wurde im Dezember des Jahres 2000 der Öffentlichkeit vorgestellt. XTM soll dazu dienen, die Such- und Managementfähigkeiten für Informationen des Topic Map Standards über das Internet zu verbreiten ([XTM], S4). Es soll eine Vielzahl von Anwendungen unterstützt werden, wobei die Kompatibilität zu bereits existierenden Standards gewährleistet bleibt. XTM soll leicht implementierbar sein und zu XTM konforme Dokumente sollen sich leicht erstellen lassen. Dies waren einige

wichtige Ziele, die sich die Entwicklungsgruppe gestellt hatte ([XTM], S.4). XML Topic Maps machen starken Gebrauch von XLink-Mechanismen. Aus der DTD wird ersichtlich, dass aber nur die einfachen Linktypen genutzt werden dürfen. Zur Beschreibung der Linktypen wird auf die XLink-Spezifikation verwiesen ([DeRo+00], §5.2).

Die DTD zu XTM ist auf der Seite von [www.topicmaps.org](http://www.topicmaps.org) erhältlich. Da die Arbeit an der endgültigen Spezifikation noch nicht abgeschlossen ist, können sich die nachfolgend verwendeten Dokumente inhaltlich noch ändern. Als einzige Konstanten bleiben vorerst die DTD, die Konformitätsregeln und die *published subject indicators* (PSI). Auf deren Bedeutung wird in den folgenden Abschnitten eingegangen. Zum „Lieferumfang“ von XTM gehört weiterhin auch das Processing Model. Darin wird beschrieben, wie eine Anwendung – speziell eine Topic Map Engine – ein zu XTM konformes Dokument zu verarbeiten hat ([XTMP]). Auch dieses Dokument kann bis zur endgültigen Verabschiedung noch inhaltliche Änderungen erfahren. Aufgrund dieser äußerst dynamischen Entwicklung werden nur die wichtigsten Neuerungen aufgezeigt und auf eine detaillierte Beschreibung einzelner Elemente der Struktur verzichtet.

### 2.2.1 Änderungen in Bezug auf ISO 13250

Beim näheren Betrachten der XTM DTD lassen sich zunächst keine prinzipiellen Änderungen zu ISO 13250 feststellen. Die Elemente „topic“, „association“ und „occurrence“ stellen weiterhin die wichtigsten Grundelemente dar. Oberflächlich betrachtet könnte man die Änderungen an der DTD als rein syntaktischer Natur auffassen. Auffällig ist die nach Ansicht des Autors verbesserte Namensgebung für die Elemente in der DTD. Aus dem Namen wird jetzt deutlicher ersichtlich, welche Rolle das angegebene Element bei der Verarbeitung der Topic Map spielt. Die wichtigste Veränderung hat das in ISO 13250 häufig verwendete Attribut „type“ bzw. „types“ erfahren. Es wurde in ein Element umgewandelt und erhielt den Namen „instanceOf“, der die eigentliche Verwendung - die Abbildung einer Klasse-Instanz-Beziehung zwischen zwei Topics - einem menschlichen Leser<sup>1</sup> deutlicher werden lässt. Des Weiteren wurde für sämtliche Attribute und Elemente, die auf ein Topic verweisen, ein Element „topicRef“ eingeführt. Erst in diesem Element wird dann der endgültige Verweis über einen einfachen Link angegeben.

```
<!ELEMENT topicRef EMPTY >
<!ATTLIST topicRef
  id ID #IMPLIED
  xlink:type NMTOKEN #FIXED 'simple'
  xlink:href CDATA #REQUIRED
>
```

Abb. 2.9: Definition des Elementes „topicRef“ ([XTM], §4.2.1)

---

<sup>1</sup> einem verarbeitenden Programm sollte die Namensgebung von Elementen egal sein

Eine Neuerung ist für Assoziationen möglich. Hier können jetzt auch Ressourcen eine bestimmte Rolle innerhalb der Assoziation einnehmen ([XTM], §4.8.2). Bislang konnten in ISO 13250 nur Topics dafür verwendet werden. Auch für den Bereich der Vorkommnisse (Occurrences) wurde eine Erweiterung vorgenommen. Eine Ressource muss sich jetzt nicht mehr außerhalb der Topic Map befinden, sondern kann über ein Element „resourceData“ direkt in die Map eingebunden werden. Diese Möglichkeit beschränkt sich aber auf reine Textdaten, was nach Meinung des Autors auch sinnvoll ist (eine Topic Map soll nicht als Container für irgendwelche Bilder, große Texte oder ähnliches missbraucht werden, sondern sie nur adressieren und einen schnellen Zugriff auf die Ressourcen ermöglichen). Erwähnenswert ist auch die Verwendung von XML Base. Ein entsprechendes Attribut wird im Dokumentenelement angegeben und legt eine URI fest, auf die sich dann nachfolgende URI's beziehen. Somit kann innerhalb der Topic Map eine komplette Angabe des Pfades zu der gewünschten Ressource vermieden werden. Zur Verwendung von XML Base wird auf die Dokumentation des W3C verwiesen. Zum besseren Verständnis soll aber die folgende Abbildung dienen.

```
<?xml version="1.0"?>
<doc xml:base="http://example.org/today/"
     xmlns:xlink="http://www.w3.org/1999/xlink">
  <head>
    <title>Nachrichten</title>
  </head>
  <body>
    <paragraph>Ticker
      <link xlink:type="simple"
            xlink:href="Ticker.xml">Was gibt's
      neues</link>?
    </paragraph>
    <paragraph>Die letzten Meldungen</paragraph>
    <olist xml:base="/news/">
      <item>
        <link xlink:type="simple"
              xlink:href="news1.xml">
          News 1</link>
        </item>
      </olist>
    </body>
  </doc>
```

Abb. 2.10: Nutzung von XML Base (in Anlehnung an [XML Base], §3)

In diesem Beispiel werden die URI's wie folgt aufgelöst:

- „ticker.xml“ wird zu <http://example.org/today/ticker.xml> und
- „news1.xml“ wird zu <http://example.org/news/news1.xml> aufgelöst.

Eine Veränderung die sowohl Vorteile als auch Nachteile mit sich bringt, bezieht sich auf die Benennung von Topics. Bislang waren in ISO 13250 dazu drei verschiedene Elemente vorhanden (Base name, Sort name und Display name). Diese wurden für XTM auf eine

Zeichenkette für den Basisnamen beschränkt. Zusätzliche Angaben können aber über ein Element „variant“ gemacht werden. Dies ermöglicht noch Namensangaben mit zusätzlichen Informationen. Da diese Struktur aber rekursiv aufgebaut ist, geht nach Meinung des Autors sehr viel an Übersichtlichkeit verloren. Für ein sehr umfangreiches Beispiel wird auf die Dokumentation zu XTM verwiesen ([XTM], §4.7.3). In Anbetracht der vorwiegend maschinellen Verarbeitung und Erstellung von Topic Maps ist diese komplizierte Modellierung aber zu verschmerzen. Zusätzlich zur DTD und den Konformitätsregeln wurden von der Autorengruppe noch die sogenannten „published subject indicators“, kurz PSI, veröffentlicht. Sie dienen zur Identifikation von Topics zum gleichen Thema. In ISO 13250 wurde dafür das Attribut „identity“ verwendet. Die Angabe dieser PSI's spielt eine große Rolle bei der Verarbeitung einer Topic Map. Werden zwei oder mehrere Topics mit einem identischen PSI gefunden, müssen diese zu einem einzigen Topic zusammengefasst werden. Dieses Topic besitzt am Ende die Vereinigungsmenge der Charakteristiken der Ausgangstopics. Zu der Charakteristik eines Topics gehören die Namen und alle Vorkommnisse Occurrences. Dieser Vorgang heißt *Merging* und wird zum Beispiel beim Zusammenführen zweier Topic Maps durchgeführt ([XTMP],§1.3).

## 2.3 Vergleich von ISO 13250 mit XTM

Zum Vergleich der beiden Arbeiten sollen die folgenden Kriterien herangezogen werden:

- DTD-Form (Meta-DTD oder standalone)
- Verständlichkeit (für einen menschlichen Leser)
- Verarbeitungsregeln (wie wird eine Topic Map verarbeitet)
- Ausdrucksmächtigkeit (welche Dinge können modelliert werden)
- weitere Besonderheiten

Die nachfolgende Tabelle gibt einen kurzen Überblick über die Unterschiede und Gemeinsamkeiten.

Tab. 2.1: Gegenüberstellung von ISO13250 und XTM

Kriterium	ISO 13250	XTM
DTD-Form	Meta-DTD	DTD (standalone)
Verständlichkeit	gut, aber teilweise umständliche Namensgebung	gut verständlich durch verbesserte Namensgebung, aber unübersichtlich durch die hohe Anzahl an Tags und die rekursiven Definitionen

Verarbeitungsregeln	nicht festgelegt ([Biez <sup>+</sup> 00], S.26 „Note 50“)	im Processing Model geregelt
Ausdrucksmächtigkeit	hoch	hoch
weitere Besonderheiten	Möglichkeit über Facetten Metainformationen anzugeben	Nutzung von XLink Ressourcen in Assoziationen möglich Ressourcen innerhalb der Topic Map als Text zulässig

Als Vorteil für ISO 13250 bleibt die Modellierung als Meta-DTD. Daraus ergibt sich eine erleichterte Umwandlung für bereits – nach einem den Topic Maps ähnlichen Prinzip - strukturierte Datenbestände. Den bisherigen Tagnamen müsste dann nur die entsprechende architektonische Form zugewiesen werden und mit einem entsprechendem Parser verarbeitet werden. Auf diese Art und Weise sind zumindest eine große Anzahl der Topics zu übernehmen. XTM hat den Vorteil, dass konforme Dokumente mit jedem XML-Parser ausgewertet werden können (bezüglich ihrer syntaktischen Richtigkeit – für die Navigation wird eine speziellere Anwendung benötigt). Die Nutzung von XML als Notationssyntax fördert auch die Verbreitung über das Internet. Die Verwendung von XLink wird dazu ebenfalls mit beitragen.

XTM und ISO 13250 regeln den syntaktisch richtigen Aufbau eines konformen Dokumentes. XML Topic Maps beschreiben zusätzlich über das Processing Model die Verarbeitung durch eine Topic Map Engine. Im nachfolgenden Kapitel sollen die grundlegenden Anforderungen an ein solches Anwendungsprogramm vorgestellt werden. Zudem wird auf Hindernisse bei der Modellierung von Topic Maps eingegangen und Möglichkeiten gezeigt, wie diese durch eine entsprechende maschinelle Unterstützung beseitigt werden können.

### 3 Topic Map Software

Eine Topic Map durchläuft bis zur ihrer Herausgabe an den Endkunden alle Teilprozesse des Elektronischen Publizierens. Dazu gehören das Design, die Erstellung, die Verwaltung und Pflege sowie die Verteilung und letztendlich die Nutzung der Map ([Rath99]). Um alle diese Teilprozesse computerunterstützt ablaufen lassen zu können, ist eine Vielzahl unterschiedlicher Software nötig. Das Vorhandensein solcher Software ist ebenfalls ein Schlüssel für die breite Anwendung dieser neuen Technologie. Im Folgenden sollen die Anforderungen an solche Topic Map Software skizziert werden. Das Hauptaugenmerk liegt dabei auf dem Verarbeitungsprogramm – der *Topic Map Engine*.

## 3.1 Die Topic Map Engine

Eine spezielle Verarbeitungsvorschrift liegt derzeit nur für XTM vor. Das entsprechende Dokument befindet sich aber noch im Prozess laufender Änderungen. Die folgenden Ausführungen greifen auf den Stand vom 4. Dezember 2000 zurück ([XTMP]).

### 3.1.1 Der Topic Map Graph

Der Verarbeitung einer Topic Map geht die Analyse des Dokumentes mit dem „topicmap“ Element voraus. Dabei wird der sogenannte Topic Map Graph aufgebaut. Dieser besteht aus unterschiedlich getypten Knoten (sogenannte t-nodes für Topics, a-nodes für Assoziationen und s-node für den Gültigkeitsbereich (scope)) und den Kanten zwischen diesen Knoten. Das XTM Processing Model (XTMP) beschreibt die Behandlung der einzelnen Knoten und Kanten. Jeder Topic Map Engine ist es aber erlaubt, noch zusätzliche beliebige Verarbeitungsschritte durchzuführen ([XTMP],§2.2.1).

### 3.1.2 Merging

Nach dem Aufbauen des Topic Map Graphen muss die Anzahl der im Topic Map Dokument aufgeführten Anzahl nicht unbedingt exakt mit der Anzahl der t-nodes im Graphen übereinstimmen. Ein Grund dafür kann im Merging zweier oder mehrerer Topics zu einem einzigen t-node liegen. Im XTM Processing Model werden zweierlei Verfahren des Mergings beschrieben.

#### 3.1.2.1 Subject-based Merging

Durch die Veröffentlichung eines sogenannten PSI (*published subject indicator*) und die Referenzierung in einem Topic wird diesem ein „subject“ (Thema, Organisationsprinzip von Topics) zugeordnet. Es gibt drei Möglichkeiten, einem Topic ein entsprechendes „subject“ zuzuordnen ([XTM],§1.3):

- über die Angabe eines beschreibenden Topics mit dem Element „topicRef“
- durch die Angabe einer beschreibenden Ressource mit dem Element „subjectIndicatorRef“
- durch Angabe einer Ressource mit dem Element „resourceRef“, die das Thema selbst repräsentiert

Für ein entsprechendes Beispiel mit Erläuterung wird auf die Spezifikation von XTM verwiesen ([XTM],§4.6).

Die Zuordnung eines „subject“ geschieht in ISO 13250 durch das Attribut „identity“ eines Topics.

Die Regel besagt dann, dass alle Topics mit identischem „subject“ zu einem t-Knoten im Topic Map Graph zusammengefasst werden müssen.

### 3.1.2.2 Topic Naming Constraint-based Merging

Diese Regel besagt, dass ein Merging aufgrund der namentlichen Gleichheit von Topics vorgenommen werden soll. Die Gleichheit bezieht sich dabei auf die Angabe des „basename“-Elementes ([XTMP],§2.3.2). Diese Regel ist aber zur Zeit Gegenstand heftiger Diskussionen der Arbeitsgruppe und es wurde von einigen Mitgliedern sogar vorgeschlagen, diese Regel aus der offiziellen Dokumentation zu entfernen und sie lediglich als optionales Feature für Topic Map Engines zu deklarieren. Sie sei deshalb nur der Vollständigkeit halber erwähnt.

### 3.1.3 Relationen

Die Assoziationen zwischen den einzelnen Topics stellen eines der wichtigsten Konstrukte der Standards dar. Erst durch sie können zusätzliche Informationen abgebildet werden. Für die Verarbeitung in einer Topic Map Engine sind speziell aber die Eigenschaften dieser Relationen von Bedeutung. Zu den Eigenschaften von Relationen zählen zum Beispiel die Symmetrie, die Reflexivität, die Transitivität und Vollständigkeit. Anhand von diesen Eigenschaften kann eine Topic Map Engine weitere Aussagen herleiten, die nicht in der Topic Map modelliert werden müssen. Das reduziert die Komplexität dieser Map und erspart Zeit beim Erstellen der Map ([Rath00], S.10). Zudem verringert sich die Anzahl der Fehlerquellen bei der Modellierung. Das nachfolgende Beispiel wurde aus einem Text von Hans Holger Rath von der Firma STEP übernommen ([Rath00], S.10): Anhand der Assoziationen

- Pennsylvania is-in USA
- Philadelphia is-in Pennsylvania
- Pittsburgh is-in Pennsylvania

kann die Topic Map Engine zusätzlich folgende Aussagen ableiten:

- Philadelphia is-in USA
- Pittsburgh is-in USA
- USA is-not-in Pennsylvania
- Philadelphia is-not-in Philadelphia usw.

Für die Ableitung dieser Aussagen muss die Topic Map Engine die folgenden Eigenschaften der Assoziation vom Typ „is-in“ kennen: transitiv, nicht symmetrisch.

Eine nützliche Eigenschaft eines Verarbeitungsprogramms für Topic Maps ist also die Fähigkeit, die Relationseigenschaften der Assoziationstypen festlegen zu können und entsprechend auszuwerten. Wichtig ist dabei auch die Möglichkeit der Angabe von weiteren Assoziationstypen durch den Benutzer, denn zum Beispiel soll die Engine wissen, dass aus der Assoziation „(Bild) wurde-gemalt-von (Maler)“ der Name der Assoziation zu „(Maler) hat-gemalt (Bild)“ bei der Änderung der Anfragestellung wird. Solche zusätzlichen Informationen sollte eine Topic Map Engine ebenfalls verarbeiten können.

### 3.1.4 Überprüfung der Gültigkeit

Eine weitere wichtige Anforderung ist die Wahrung der Konsistenz einer Topic Map. Diese Eigenschaft kommt besonders zum Tragen, wenn die Map aufgrund der Anzahl ihrer Topics und Assoziationen kaum oder nicht mehr für einen einzelnen menschlichen Leser überschaubar ist. Aufgrund der Fülle an Informationen, die zum Beispiel in einem Lexikon enthalten ist, dürfte der oben geschilderte Fall sehr häufig eintreten.

Für die Überprüfung der Gültigkeit müssen der Topic Map Software wieder zusätzliche Informationen zur Verfügung gestellt werden. Diese Informationen unterscheiden sich aber in Abhängigkeit vom Typ (Assoziation, Topic, Occurrence) der zu überprüfenden Einheit (s.a. [Rath00], S.14f).

Für ein Topic könnte zum Beispiel die Angabe des Topicstyps überprüft werden (zum Beispiel sollten für eine Firmenkontaktbörse alle Topics vom Typ Firma oder Person (für den Ansprechpartner) sein).

Vorkommnisse (Occurrences) werden ebenfalls durch ihren Typ näher bestimmt. Es könnte zusätzlich nur die Angabe bestimmter Rollen zugelassen werden (z.B. Webseite als Rollename und im Inhalt des Elementes eine gültige Webadresse über Textmuster).

Für Assoziationen sollten noch mehr Angaben gemacht werden können. Es müsste möglich sein, den Typ der beiden Endpunkte (Rollen) festzulegen (z.B. „(Maler) hat-gemalt (Bild)“ oder entsprechend andere gültige Kombinationen (z.B. „(Maler) hat-gemalt (Bild | Portrait | Stilleben)“).

### 3.1.5 Anfragesprache

Einen erheblichen Teil der Nutzung bilden die Anfragen, die an eine Topic Map gestellt werden. Eine Topic Map Engine muss demnach eine Anfragesprache dem Nutzer zur Verfügung stellen. Weder durch ISO 13250 noch durch XTM werden nähere Erläuterungen hierzu gegeben. Prinzipiell könnten nach Meinung des Autors zwei Konstrukte dafür zur Verfügung stehen.

### **3.1.5.1 XML Query**

XML Query befindet sich derzeit noch im frühen Anfangsstadium der Entwicklung (<http://www.w3.org/XML/Query>) und es ist deshalb noch nicht absehbar, inwieweit diese Anfragesprache auch auf Topic Maps angewendet werden kann. Nach Meinung des Autors wäre XML Query nur mit erheblichen Leistungseinbußen auf Topic Maps zu übertragen. Das betrifft zum Beispiel die oben angesprochene Möglichkeit, weitere Assoziationen aus gegebenen Zusammenhängen abzuleiten, da sich XML Query ja nur auf konkrete Teile eines XML-Dokumentes bezieht. XML ist für Topic Maps nach Ansicht des Autors aber nur eine Syntaxvorschrift für das Austauschformat. Die interne – in der Engine vorgenommene – Verwaltung der Topics und Assoziationen wird aber nicht festgelegt. So wäre es zum Beispiel auch möglich, die Daten in einer herkömmlichen relationalen oder objektorientierten Datenbank abzulegen. Dafür müsste dann die entsprechende Anfragesprache (vornehmlich SQL) genutzt werden.

### **3.1.5.2 TMQL**

Der Vorschlag einer eigenen, an SQL angelehnten, Anfragesprache wurde von Rafal Ksiezyc (ebenfalls STEP) auf der XML Europe 2000 ([Ksie00], S.11ff) gemacht. Eine Arbeitsgruppe ist derzeit damit beschäftigt, diesen Vorschlag weiter auszubauen. Allgemein ist zu sagen, dass dieser Entwurf den besonderen Anforderungen der Topic Maps deutlich mehr gerecht wird als eine Anfragesprache, die generell auf XML Dokumente bezogen ist. Die Anlehnung an SQL ist ebenfalls von Vorteil. Aus den aufgeführten Beispielen ist aber eine deutliche Komplexität der Anfragen zu erkennen, die es dem Nutzer erschwert, eine umfangreiche Anfrage an ein solches System zu formulieren.

### **3.1.5.3 Engine-spezifische API**

Diese Möglichkeit wird sicherlich in der ersten Zeit die vorherrschende Art und Weise sein, auf die man eine Anfrage an eine Topic Map Engine stellt. Doch auch hier gibt es einige Anfragetypen, die auf jeden Fall von allen Programmen beherrscht werden müssen. Dazu gehören zum Beispiel:

- „suche das Topic mit der speziellen Kennung (ID)“
- „stelle den Typ des Topics fest“ oder „ist das Topic vom Typ x“
- „ermittle die Rolle in der Occurrence“ usw.

Diese Anfragen haben die Ermittlung der einzelnen Eigenschaften der Topics als Grundlage. Sie werden sich lediglich in ihrer Syntax unterscheiden und mit der Standardisierung einer Anfragesprache nach Meinung des Autors hauptsächlich nur noch

für programmspezifische Erweiterungen (z.B. Add-Ins für die Topic Map Engine oder einen Browser) genutzt werden.

### 3.1.6 Weitere Anforderungen

Zu den weiteren Anforderungen an eine Topic Map Engine gehören zudem noch eine Nutzerverwaltung mit Rechtvergabe. Damit kann sichergestellt werden, dass während der Erstellungsphase der Topic Map nur bestimmte Nutzer Topics anlegen, ändern und löschen dürfen. Es wäre auch denkbar, bereits validierte Topics, die zudem nicht weiter bearbeitet werden müssen, mit einer Sperre zu belegen. Alle diese Maßnahmen würden dazu beitragen, mögliche Fehler in der Topic Map zu verringern (so kann zum Beispiel dann kein Topic gelöscht werden, das im Typ-Attribut anderer Topics vorkommt usw.).

Die Implementierung von Linkmechanismen gehört ebenfalls mit zu den Anforderungen der Engine. Sie muss in der Lage sein, angegebene Adressen zu analysieren und gegebenenfalls von einer externen Quelle (außerhalb der aktuellen Topic Map) nachladen zu können. Zu den Linkmechanismen gehören zum Beispiel in erster Linie XLink, aber auch HyTime ([Gold+97]) bietet viele Möglichkeiten, Informationsobjekte zu adressieren.

### 3.1.7 Bisher verfügbare Topic Map Engines

Die folgenden Topic Map Engines sind bisher am Markt erhältlich. Für einige Produkte sind Evaluationsversionen kostenlos erhältlich. TmProc und TM4J sind kostenlose Tools.

Tab. 3.1: Topic Map Engines

Firma	Produkt	Webseite
	TmProc	<a href="http://www.ontopia.net">http://www.ontopia.net</a>
	TM4J	<a href="http://www.techquila.com">http://www.techquila.com</a>
	jTME	<a href="http://thinkalong.com/jtme/jTME.html">http://thinkalong.com/jtme/jTME.html</a>
Ontopia	Free Navigator (Demo)	<a href="http://www.ontopia.net">http://www.ontopia.net</a>
Mondeca	Topic Navigator	<a href="http://www.mondeca.com">http://www.mondeca.com</a>
Empolis	K42	<a href="http://www.empolis.co.uk">http://www.empolis.co.uk</a>

## 3.2 Topic Map Design

Neben dem eigentlichen Verarbeitungsprogramm wird auch Software benötigt, die den eigentlichen Erstellungsprozess einer Topic Map unterstützt. Für diese Programme ist die Konsistenzprüfung eine der wichtigsten Aufgaben, insbesondere wenn berücksichtigt wird, dass die Erstellung der Topic Map zumeist in Teamarbeit vorgenommen wird. Daraus

ergeben sich wieder eine Reihe von Anforderungen, die nachfolgend dargelegt werden sollen.

### 3.2.1 Das Topic Map Template (Vorlage)

Der Topic Map Standard legt fest, dass auch Objekte, welche die Topics in einer Map näher beschreiben, ihrerseits ebenfalls Topics sind. Dazu gehören zum Beispiel die Topic Typen, Rollentypen, Assoziationstypen usw. Aufgrund dieser „rekursiven“ Definitionen (ein Topic Typ ist ebenfalls ein Topic) ist es oft schwer, zwischen deklarativen und „normalen“ Topics – die den eigentlichen Inhalt der Topic Map darstellen – zu unterscheiden ([Rath00],S.4). Deshalb wäre es wichtig, eine Trennung zwischen diesen beiden Teilen zu haben. Eine Möglichkeit, eine solche Trennung herbeizuführen, ist ein sogenanntes Topic Map Template. Es handelt sich dabei ebenfalls um eine Topic Map, die alle Topics enthält, die in der zugeordneten Topic Map als Topic Typen, Rollentypen, Assoziationstypen usw. genutzt werden. Das Template spiegelt sozusagen das Organisationsprinzip der Topic Map wieder und befindet sich auf einer höheren Abstraktionsebene. Eine solche Vorlage kann man sich wie folgt verdeutlichen.

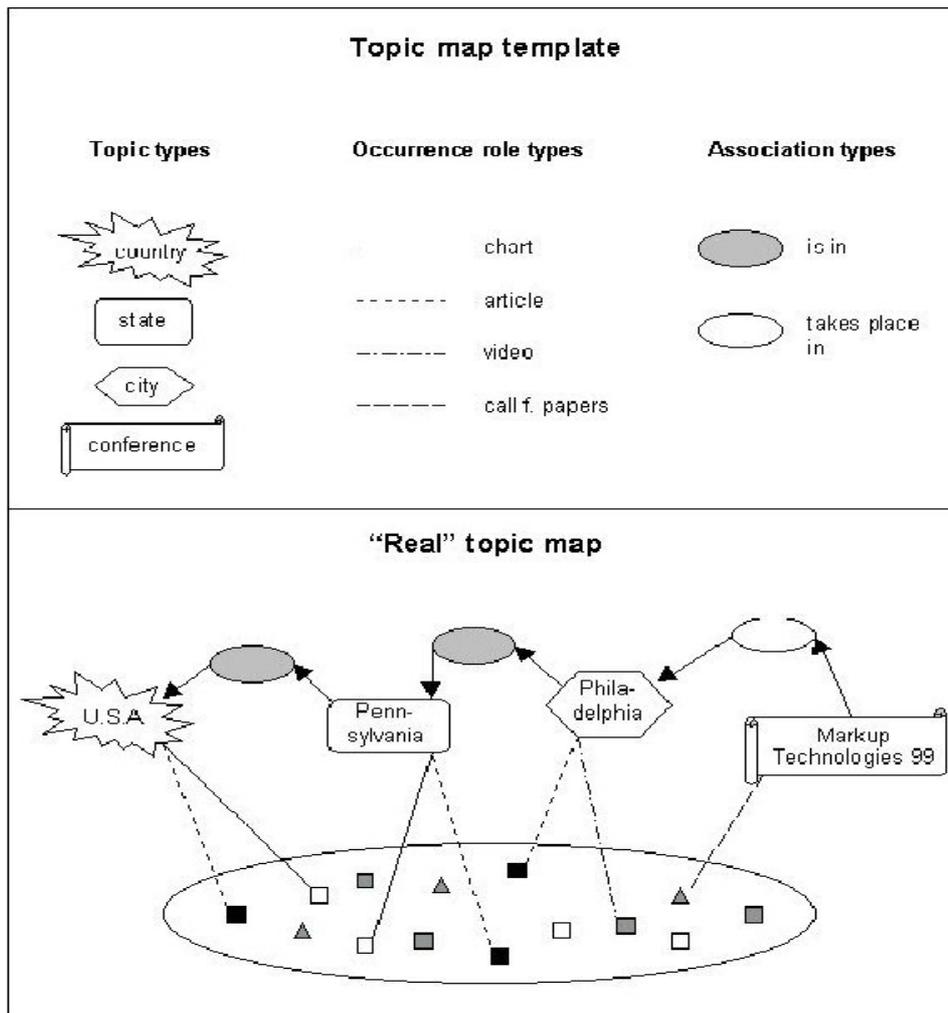


Abb. 3.1: Topic Map Template ([Rath00],S.6)

Auf die Nutzung eines solchen Templates wird im letzten Kapitel der Arbeit noch ausführlicher eingegangen.

### **3.2.2 Gültigkeitsprüfung**

Software, die zur Erstellung von Topic Maps dient, muss ebenfalls Mittel zur Überprüfung der Gültigkeit der angegebenen Topics bereitstellen. Sie sollte in gleicher Weise in der Lage sein, festzustellen, ob die modellierten Sachverhalte einem benutzten Topic Map Template entsprechen. Zu diesem Punkt wird auch auf Abschnitt 3.1.4 verwiesen.

### **3.2.3 Versionskontrolle**

Aufgrund der zu erwartenden Häufigkeit von Änderungen an einem Topic und der Arbeit im Team, sollte die Software ebenfalls Hilfen zur Versionskontrolle ganzer Topic Maps, aber auch einzelner Topics und Assoziationen anbieten.

### **3.2.4 Nutzerrollen und –rechte**

Bei der Benutzung eines Templates können die Aufgaben zur Erstellung einer Topic Map nochmals untergliedert werden. Ein „Designer“ erstellt das Template mit allen entsprechenden Topics. Nur er soll in der Lage sein, diese zu erstellen, zu ändern und zu löschen. Der „Redakteur“ erhält nur lesenden Zugriff auf die Topics der Vorlage und verwendet diese dann als entsprechende Typen in den „realen“ Topics.

### **3.2.5 Zugriff auf Informationsressourcen**

Weil ein großer Teil aller Topics immer auch mit entsprechenden Informationsressourcen verknüpft ist, sollte ein Redaktionswerkzeug einen leichten Zugang zu diesen Ressourcen zur Verfügung stellen. Dies könnte beispielsweise durch die Integration eines Webbrowsers oder eines Dokumentenmanagementsystems geschehen.

### **3.2.6 Testumgebung**

Schon während der Erstellungsphase der Topic Map sollte das Testen unter gleichen Bedingungen wie beim Endkunden stattfinden. Aus diesem Grund muss die Software geeignete Konzepte zur Verfügung stellen, die einen solchen Test durchführen. Zu einem solchen Test gehören zum Beispiel die einfache Navigation innerhalb der Map, das Stellen von einfachen und komplexen Suchanfragen und das Anzeigen von typischen Informationsressourcen (vgl. [Rath99]). Aufgrund dieser Tatsache muss das Redaktionswerkzeug Möglichkeiten anbieten, die auch dem Endbenutzer zur Verfügung stehen.

### 3.3 Topic Map Browser

Der Topic Map Browser dient dem Nutzer zur schnellen Navigation zwischen einzelnen Topics und zum „Entlangwandern“ an den Assoziationen. Er muss ebenfalls in der Lage sein, die verknüpften Informationsressourcen selbst darzustellen oder zumindest ein entsprechendes Anwendungsprogramm aufzurufen. Für die Benutzeroberfläche sind mehrere Möglichkeiten denkbar. Zum ersten ist hierbei die klassische Baumdarstellung zu nennen. Sie bietet dem Nutzer einen einfachen Einstieg, da solche Komponenten in vielen Anwendungsprogrammen eingesetzt werden. Die folgende Abbildung zeigt eine mögliche Benutzeroberfläche.

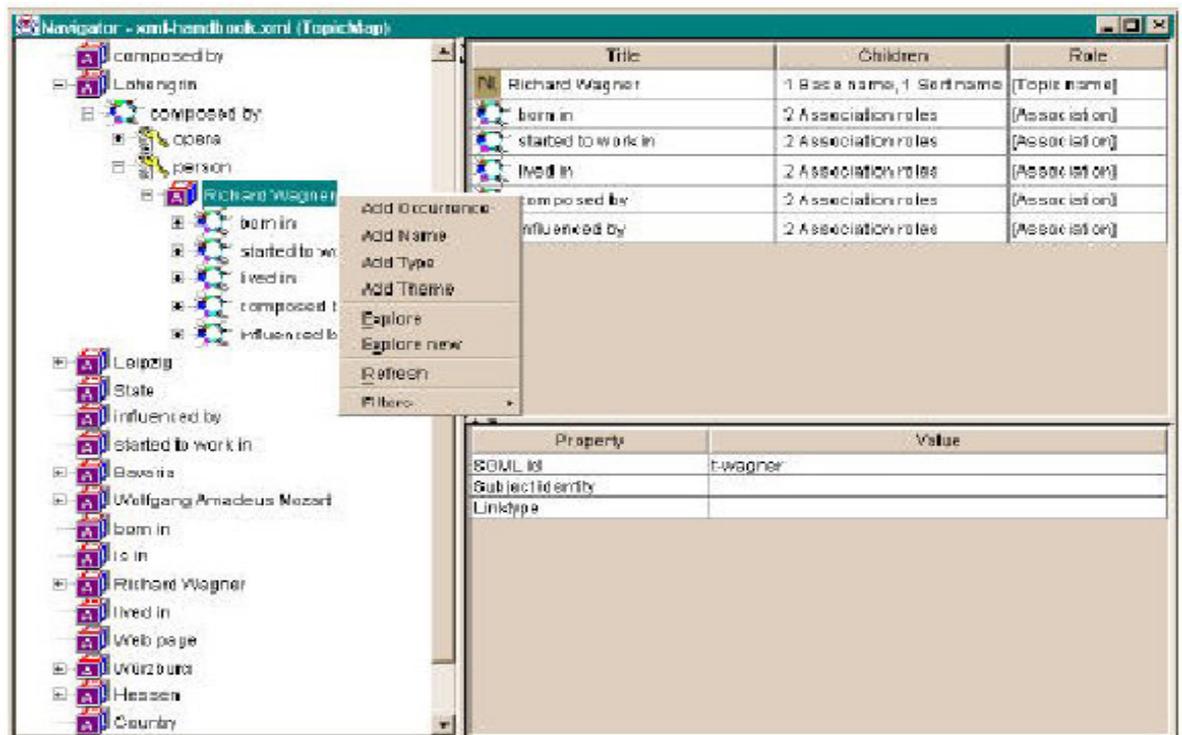


Abb. 3.2: Baumdarstellung einer Topic Map ([Rath99])

Nachteilig ist zu bemerken, dass mit steigender Zahl von Topics und Assoziationen die Navigierbarkeit drastisch nachlässt. Probleme treten weiterhin auf, wenn viele gleichartige Assoziationen auftreten. Beim Aufklappen eines Zweiges im Baum wird der Nutzer von der Vielzahl der Einträge fast „erschlagen“. Trotzdem ist die Baumdarstellung eine akzeptable Lösung zur Visualisierung einer Topic Map.

Eine weitere Möglichkeit bietet ein Vorgehen, wie es die Firma I-Views auf ihrer Webseite <http://www.i-views.de> als Demonstration zeigt. Hier werden einzelne Topics als Punkte mit den zugehörigen Assoziationen angezeigt. Problematisch ist hierbei aber, nach Ansicht des Autors, ebenfalls wieder die Unübersichtlichkeit bei einer großen Anzahl an Topics. Sollten zudem zu einem Topic noch sehr viele Assoziationen vorhanden sein, wirkt sich dies ebenfalls nachteilig auf die Darstellung aus. Zudem bereitet die programmier-

technische Implementierung einige Schwierigkeiten. Bei der Demonstration auf der Webseite handelt es sich um eine graphische Animation – die einzelnen Zweige des Baumes sind dabei unregelmäßig angeordnet. Der Programmierer müsste erst einen Algorithmus finden, die einzelnen Topics auf eine solche Art und Weise anzuordnen. Prinzipiell bietet der gezeigte Ansatz aber eine bessere Anschaulichkeit der Zusammenhänge in einem solchen Netz.

Neben diesen Möglichkeiten der Darstellung von Topic Maps wurde von TheBrain® (<http://www.thebrain.com>) eine weitere Art der Navigation vorgestellt.



Abb. 3.3: Navigation mit TheBrain®

Entsprechend dem ausgewählten Eintrag ändert sich die Darstellung und wird durch eine Animation „untermalt“. Der Vorteil bei dieser Ansicht liegt darin, dass nur die jeweils in Frage kommenden nächsten Zweige angezeigt werden. Problematisch ist der verloren gegangene Gesamtüberblick. Dieser könnte aber in einem weiteren Fenster zusätzlich angezeigt werden.

Eine weitere Möglichkeit wäre die dreidimensionale Darstellung mit den gleichen Vor- und Nachteilen.

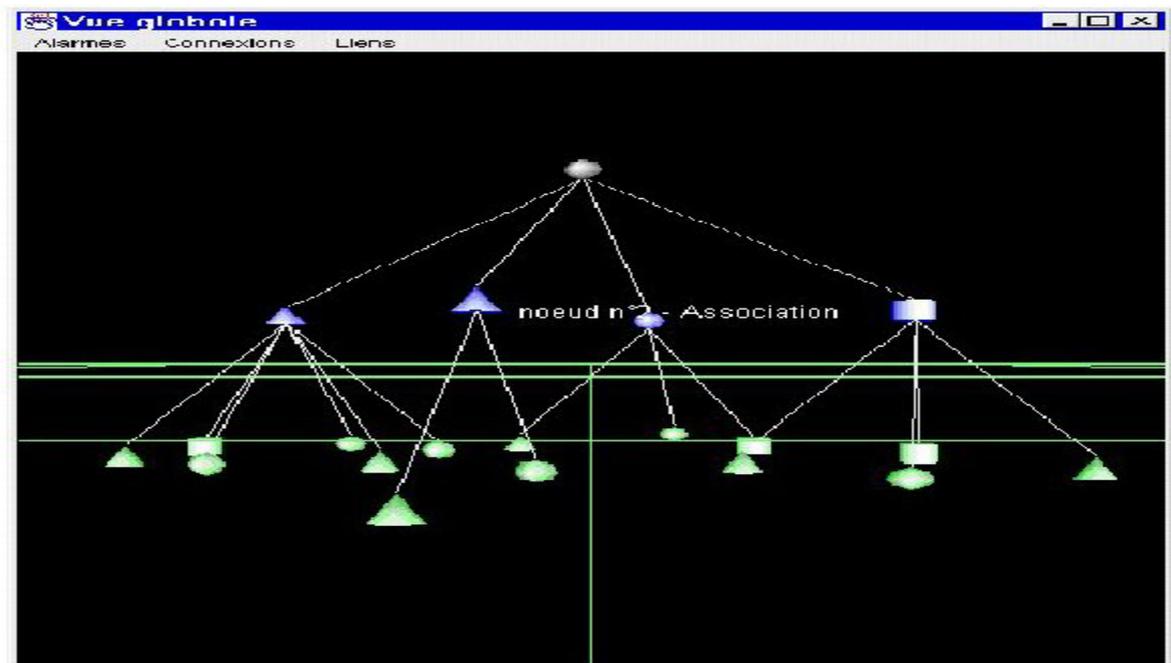


Abb. 3.4: dreidimensionale Darstellung von Topic Maps ([Gran<sup>+</sup>00], S.8)

## 4 Einsatzmöglichkeiten von Topic Maps

Um die Durchsetzung einer neuen Technologie in einem möglichst breiten Rahmen sicherzustellen, ist es notwendig, viele Anwendungsfelder ausfindig zu machen und Einsatzgebiete beschreiben zu können. In den folgenden Abschnitten soll eine Reihe möglicher Einsatzgebiete für Topic Maps aufgeführt werden. Zusätzlich zu den Potentialen sollen aber auch die Probleme und die Nachteile aufgezeigt werden, auf die man bei der Umstellung zu dieser neuen Technologie stoßen kann.

### 4.1 Wissensmanagement im Unternehmen

Ein erstes Hauptanwendungsgebiet wird in der Unterstützung des innerbetrieblichen Wissensmanagement gesehen. Aufgrund der Tatsache, dass die Unternehmensressource Wissen mehr und mehr zu einem zentralen Wettbewerbsfaktor wird, spricht der Autor hier bewusst vom innerbetrieblichen Wissensmanagement. Es kann sich für ein Unternehmen äußerst nachteilig auswirken, wenn internes Wissen Dritten zur Verfügung steht. Eine Ausnahme bilden hier sicherlich Unternehmensverbände, wobei es aber aufgrund von Informationsasymmetrien zur Benachteiligung einzelner Unternehmen kommen kann. Wissensmanagement soll im Folgenden als Leitungshandeln verstanden werden, das auf die Unternehmensressource Wissen ausgerichtet ist ([Scho99a], S.1). Es soll dazu dienen, die organisatorische Wissensbasis auf- bzw. auszubauen, das vorhandene Wissen der Unternehmung besser zu nutzen und verstecktes Wissen (implizites Wissen) einzelner für alle Mitglieder der Organisation zugänglich zu machen. Der Aufbau der organisatorischen Wissensbasis soll mit der folgenden Abbildung noch einmal verdeutlicht werden.

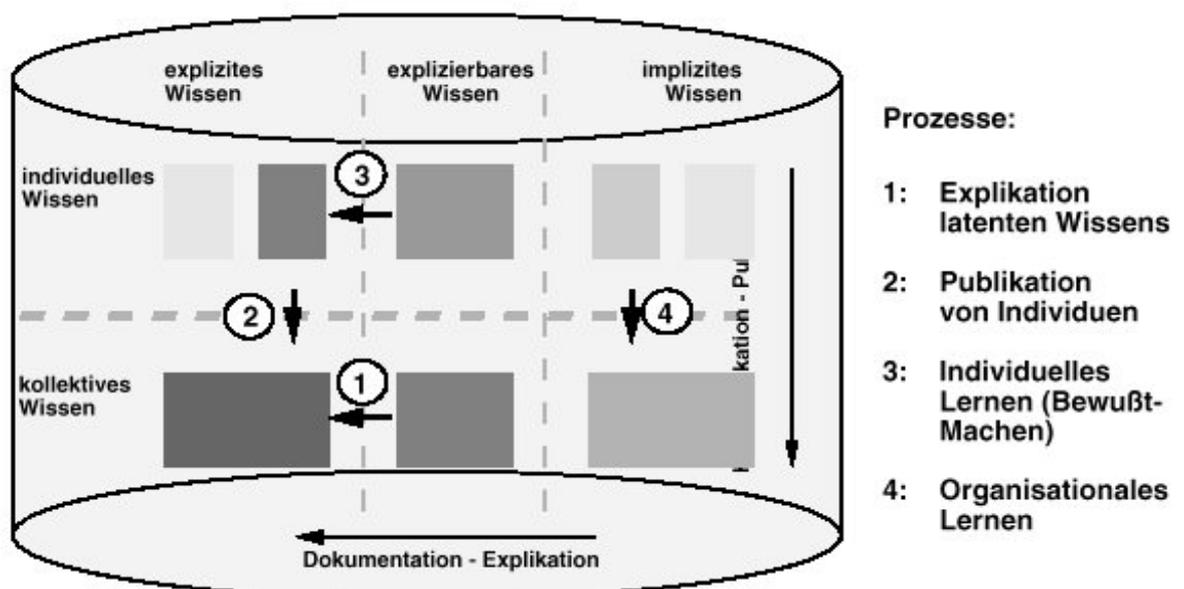


Abb. 4.1: Elemente der organisatorischen Wissensbasis ([Scho99a], S.3)

Aus dem Schema wird die Unterteilung in explizites, explizierbares und implizites Wissen deutlich. Ziel der Bemühungen ist die Vergrößerung des expliziten, kollektiven Wissens. Aus dem Schema ist ersichtlich, dass dieses Ziel durch verschiedene Prozesse erreicht werden kann. Eine erste Möglichkeit ist die Publikation individuellen Wissens durch menschliche Wissensträger. Diese stellen ihr persönliches Wissen zur Lösung bestimmter Problemstellungen allen Organisationsmitgliedern zur Verfügung. Meist wird dieses Wissen in Form von Dokumenten abgelegt. Diese können sowohl in elektronischer Form als auch auf Papier vorliegen. In gleicher Weise geschieht dies bei der Explikation von verstecktem Wissen.

Zu den Arbeitsfeldern im Wissensmanagement gehören unter anderem:

- die Identifikation
- der Erwerb
- die Entwicklung
- die Bewahrung
- die Verteilung und
- die Nutzung von Wissen.

Da der Begriff Management die Funktionen Planung, Steuerung und Kontrolle zur Grundlage hat, werden in einem größeren Rahmen auch die Bildung von Wissenszielen (welches Wissen benötigt unsere Unternehmung für das Projekt?) und die Wissensbewertung (reicht das vorhandene Wissen aus, um die Problemstellung zu lösen? und wie ist unser Wissen qualitativ einzuordnen?) als weitere Aufgaben angeführt (siehe Abbildung in [Scho00], Bausteine des Wissensmanagements). Nach dieser grundlegenden Einführung sollen ein bisheriger Ansatz und dessen Nachteile beschrieben werden. Abschließend werden die Einsatzpotentiale, aber auch Probleme beim Einsatz von Topic Maps besprochen.

### **4.1.1 Dokumentenmanagement**

Ein möglicher Ansatz zum Wissensmanagement führt über die Verwendung von Dokumentenmanagementsystemen (DMS). In solchen Systemen werden elektronische Dokumente durch die Vergabe von Indizes und Schlagwörtern näher beschrieben und im System abgelegt. Auch nicht in elektronischer Form vorliegende Dokumente können durch die Verwendung von Scannern im System gespeichert werden. Mit Hilfe von Suchanfragen mit entsprechenden Angaben der Schlagwörter werden die Dokumente wieder auffindbar gemacht. Diese Vorgehensweise birgt aber auch Nachteile in sich. So ist zum Beispiel zum Auffinden eines bestimmten Dokumentes die Kenntnis der Schlagwörter notwendig. Wenn

das System aber nur eine bestimmte Anzahl dieser Schlagwörter zulässt, so erhöht sich die Wahrscheinlichkeit, dass mehrere Dokumente mit der gleichen Beschreibung ausgestattet sind und somit wächst bei einer Suche mit den angegebenen Schlagwörtern auch die Trefferanzahl. Ein weiterer Nachteil liegt in den Dokumenten selbst. Jedes Dokument enthält nahezu immer mehr Informationen als für den eigentlichen Verwendungszweck prinzipiell benötigt werden. So enthält eine Rechnung beispielsweise nicht nur den zu zahlenden Endbetrag (der eigentliche Verwendungszweck) sondern noch die zugehörige Auftragsnummer, den Bearbeiter, das Erstelldatum, die Rechnungspositionen usw. Diese zusätzlichen Informationen sind notwendig, um Beziehungen zu zugehörigen Dokumenten herstellen zu können. Zusätzliche Schwierigkeiten bereitet die Extraktion von Informationen, die in Fließtexten „untergebracht“ sind. Alle diese Informationen sind im Dokument selbst „versteckt“ und die referenzierten Dokumente werden oft nur durch weitere Suchläufe gefunden. Der Nachteil liegt also in der „losen“ Speicherung einzelner Dokumente ohne erkennbaren Zusammenhang. Besonders deutlich wird dieser Nachteil bei der Verwendung eingescannter Dokumente, aus denen sich diese Informationen entweder gar nicht oder nur mit hohem Aufwand für eine maschinelle Verwertung gewinnen lassen. Des weiteren ist es erforderlich, die im Dokument enthaltenen Informationen für einen Leser schnell erkenntlich zu machen. Um dieser Problematik zu begegnen, wird von Schoop ein Lösungsvorschlag unterbreitet. Dabei handelt es sich um eine medienneutrale Speicherung und Distribution von Wissen in Form modularer elektronischer Dokumentbausteine ([Scho99a], S.2). Zur Lösung des Problems der Informationsfilterung fordert er den „Einsatz einer akzeptierten Methode zur

- zielorientierten
- konsistenten
- aufgaben- und nutzergerechten

Informationsmodellierung“. Es wird hierzu die Anwendung der Information Mapping® Methode vorgeschlagen ([Scho99a], S.2).

Die Lösung für den Einsatz von Dokumentbausteinen wird im Einsatz von einer inhaltsorientierten Dokumentauszeichnung auf Basis von SGML/XML gesehen. Dabei werden den unterschiedlichen Dokumentarten spezifische Strukturen zugeordnet, die für alle Dokumente dieser Art gelten. Eine solche Strukturbeschreibung heißt DTD – Document Type Definition. Der Inhalt eines jeden Dokumentes wird dann dem entsprechenden Feld der Struktur zugeordnet. Über entsprechende Software kann dann auch in großen Datenbeständen inhaltsorientiert gesucht werden.

### 4.1.1.1 Vorteile

Der oben beschriebene Ansatz bietet eine Lösung für die Erschließung des Inhalts von bisher unstrukturierten Dokumenten. Aufgrund der Trennung von Inhalt und Darstellung der im Dokument enthaltenen Daten ist eine Transformation in verschiedene Layouts möglich. Entsprechende Technologien stehen mit XSL bzw. XSLT zur Verfügung.

### 4.1.1.2 Nachteile

Im Wesentlichen sieht der Autor zwei Nachteile eines rein strukturorientierten Dokumentenmanagements. Die Einführung eines solchen Konzeptes ist mit hohen Kosten verbunden. Zum einen muss bei konsequenter Einführung für jede Dokumentenart eine spezifische DTD entworfen werden. Zudem müssen entsprechende Arbeitsplätze mit geeigneter Softwareausstattung bereitgestellt werden. Die Schulung und Einführung der Mitarbeiter ist mit zusätzlichen Kosten verbunden. In allen diesen Kosten ist der Aufwand, bereits existierende Dokumente umzustellen, noch nicht enthalten. Des Weiteren bietet dieser Ansatz keine Lösung für die vom Autoren angesprochene „zusammenhanglose“ Abspeicherung von Einzeldokumenten. Es wird lediglich die Extraktion der benötigten Information (z.B. der zugehörigen Rechnungspositionen) stark vereinfacht.

## 4.1.2 Einsatzmöglichkeiten von Topic Maps

Gemeinsamkeiten zur Speicherung von Dokumenten in herkömmlichen Dokumentenmanagementsystemen mit dem Prinzip der Topic Maps sind auf den ersten Blick leicht feststellbar. Es wird beiderseits ein Informationsressourcenpool aufgebaut, der durch einzelne Schlagwörter referenziert wird. Zur Veranschaulichung wird noch einmal auf die Abbildung 2.5. verwiesen. Topic Maps bieten zusätzlich weitere Möglichkeiten. Es ist jetzt auf einfache Art und Weise möglich, Zusammenhänge zwischen einzelnen Dokumenten zu modellieren. Dabei wird jedes Dokument durch ein Topic mit dem zugehörigen Verweis auf die Informationsressource repräsentiert. Danach wird eine entsprechende Assoziation zwischen diesen Topics gebildet. Trotz allem bietet dieses Vorgehen nicht nur Vorteile. Eine ausführliche Besprechung der Konsequenzen des Einsatzes von Topic Maps wird nachfolgend vorgenommen.

### 4.1.2.1 Verwendung unstrukturierter Dokumente

Ein erhebliches Einsparpotential gegenüber einem rein strukturorientierten Dokumentenmanagement ergibt sich aus der Tatsache, dass es möglich ist auch unstrukturierte Dokumente als Vorkommnis (Occurrence) zu referenzieren. Nachteilig wirkt sich die fehlende Struktur jedoch auf die Erschließung des Dokumentinhaltes aus. Ähnlich einem herkömmlichen DMS muss das Dokument näher beschrieben werden. Topic Maps bieten

hierfür das Konzept der Topic Names an. Ein Vorteil ist hierbei wiederum die unbegrenzte Nutzung solcher Namen. Auch kann im Element „Occurrence“ eine Typangabe zur näheren Klassifikation der Informationsressource erfolgen. Trotzdem ist für den Aufbau von Assoziationen die genaue Analyse des Dokumentinhaltes unerlässlich.

### 4.1.2.2 Suchbegriffe

Für die Recherche im Informationspool müssen ebenfalls die Schlagwörter bekannt sein. An diesem Punkt bieten Topic Maps keine wesentlichen Vorteile gegenüber herkömmlichen DMS. Für den Einsatz unstrukturierter Dokumente stellen Topic Maps nur unzureichende Beschreibungsparameter zur Verfügung. Alle Informationen müssen in den Attributen „Role“ und „Type“ untergebracht werden. Vorteilhaft gestaltet sich jedoch eine Suche entlang der Assoziationen zwischen den einzelnen Topics.

### 4.1.2.3 Nutzung von Assoziationen

Den größten Vorteil verschaffen Topic Maps bei dem Einsatz von Assoziationen. Damit lassen sich Zusammenhänge zwischen einzelnen Dokumenten herstellen. Es ist durch den Standard und seine Abkömmlinge jedoch nicht vorgesehen, Assoziationen direkt zwischen Occurrence-Elementen modellieren zu können. So müssen zum Beispiel für verschiedene Rechnungen auch verschiedene Topics angelegt werden, die dann jeweils den Typ „Rechnung“ zugewiesen bekommen. Die folgende Abbildung verdeutlicht dieses Vorgehen.

```
<topic id="Kd4711" types="Kunde">
  <topname>
    <basename>Kd4711</basename>
  </topname>
  <occurs occr1="Stammdaten zu Kunde 4711"
    type="Stammdaten"
    xlink:type="locator"
    xlink:href="http://server/kd/kd4711.xml"/>
</topic>

<topic id="Meier" types="Bearbeiter">
  <topname>
    <basename>Meier</basename>
    <dispname>Hans Meier</dispname>
  </topname>
</topic>

<topic id="Re1234" types="Rechnung">
  <topname>
    <basename>Re1234</basename>
    <dispname>Rechnung 13.08.2000</dispname>
  </topname>
  <occurs type="Rechnungsdokument"
    xlink:type="locator"
    xlink:href="http://server/re/re13082000.xml"/>
</topic>

<assoc type="wird-betreut-von">
  <assocc1 anchrole="Kunde"
    type="Kunde">Kd4711</assocc1>
  <assocc1 anchrole="Bearbeiter"
    type="Betreuer">Meier</assocc1>
</assoc>

<assoc type="hat-Rechnung">
  <assocc1 anchrole="Kunde"
    type="Kunde">Kd4711</assocc1>
  <assocc1 anchrole="Rechnung"
    type="Rechnung">Re1234</assocc1>
</assoc>
```

Abb. 4.2: Nutzung von Assoziationen

Es ist auf diese Art und Weise leicht möglich, auch entsprechende Verbindungen zu den Dokumentbearbeitern, zu Projektteilnehmern usw. zu ziehen. Mit Hilfe von entsprechenden Suchanfragen und/oder über eine geeignete Nutzeroberfläche kann dann leicht in der Topic Map navigiert werden.

### 4.1.2.4 Fazit

Im Abschnitt 4.1.1. wurden 3 Punkte erwähnt, denen die Informationsmodellierung für die betriebliche Wissensbasis genügen sollte. Die Punkte Zielorientierung und eine entsprechende Aufgaben- und Nutzerorientierung kann mit Topic Maps durch die Verwendung von Gültigkeitsbereichen (scopes) realisiert werden. Eine weitere Möglichkeit ist die Modellierung von mehreren Maps für die unterschiedlichen Zwecke. Der Punkt Konsistenz kann nach Ansicht des Autors auf zwei verschiedene Weisen gesehen werden. Zum ersten spielt die „innere“ Konsistenz eine Rolle - alle Informationen sollen in ein und derselben Art und Weise modelliert werden. Ein zweiter Punkt ist aber auch das „äußere“ Erscheinungsbild dieser Informationen dem Nutzer gegenüber. Beim Einsatz von Topic Maps ist sofort die innere Konsistenz gewährleistet. Alle Informationen werden über Topic-Elemente abgebildet. Wie im Kapitel 2 erwähnt sind Assoziationstypen und Occurrence-Typen ebenfalls Topics. Für die Sicherung der äußeren Konsistenz ist die von Schoop vorgeschlagene Methode des Information Mapping® geeignet, da sie auch auf die gleichartige Darstellung aller Informationen abzielt ([Horn89], S.85).

Topic Maps sind also für das betriebliche Wissensmanagement einsetzbar. Durch die Modellierung der Assoziationen bieten sie erweiterte Möglichkeiten gegenüber von herkömmlichen DMS. Trotzdem müssen auch die beschriebenen negativen Konsequenzen berücksichtigt werden. Nach Meinung des Autors stellt der Einsatz von Topic Maps einen kostengünstigen Zwischenschritt bei der Migration zu einem strukturorientierten Dokumentenmanagement dar. Topic Maps könnten herkömmliche Systeme um eine semantische Komponente erweitern, die Zusammenhänge zwischen den einzelnen Dokumenten erschließt. Zusätzlich könnte ihr Anwendungsgebiet beim Austausch zwischen verschiedenen Wissens- und Dokumentenmanagementsystemen liegen.

## 4.2 Suchmaschinen im Internet

Aufgrund des exponentiellen Wachstums der permanent am Internet angeschlossenen Rechner steigt auch die Anzahl der zur Verfügung stehenden Dokumente im Internet an. Im Januar 2001 waren laut DENIC im Zuständigkeitsbereich etwa 12,7 Millionen Rechner ans Internet angeschlossen. Die letzte weltweite Zählung datiert zurück auf Juli des Jahres 2000. Dort wurden in etwa 93 Millionen Rechner gezählt. Durch die Vielzahl der auf diesen Rechnern abgelegten Dokumente ist es ohne geeignete Hilfsmittel nahezu un-

möglich, die gewünschte Information ausfindig zu machen. Erschwerend wirkt dabei die Tatsache, das es sich hierbei um einen völlig unstrukturierten und heterogenen Datenbestand handelt. Zusätzlich unterliegt eine große Anzahl an Dokumenten ständigen Veränderungen und Aktualisierungen. Einen entscheidenden Beitrag zur Lösung des oben genannten Problems sollen Suchmaschinen und Themenkataloge leisten. In den folgenden Abschnitten wird kurz deren prinzipielle Wirkungsweise erklärt und sich daraus ergebende Probleme vorgestellt. Es wird weiterhin gezeigt, wie Topic Maps Antworten auf diese offenen Fragen bieten und mit dem Resource Description Framework (RDF) wird eine zusätzliche Alternative aufgezeigt.

### **4.2.1 Themenkataloge**

Themenkataloge sind von menschlichen Redakteuren zusammengestellte Sammlungen von Links. Dabei werden die Links entsprechend ihrer Themengebiete in Kategorien unterteilt. Aus diesen Kategorien werden wiederum Hierarchien mittels Unterkategorien gebildet (vgl. [Woll01], Folie 7). Aufgrund dieser Verfeinerungen kann der Nutzer schnell zu den entsprechenden Seiten gelangen. Der bekannteste Dienst dieser Art ist vielleicht das Portal Yahoo.

Die manuelle Klassifikation schlägt sich in einer hohen Qualität der im Portal aufgenommenen Seiten nieder, geht aber deutlich zu Lasten des Umfangs der zur Verfügung stehenden Seiten. Unterliegen die Seiten dann noch dynamischen Veränderungen (z.B. durch eine interne Umstrukturierung seitens des Inhaltsanbieters) kann es zum Auftreten von „toten“ Links kommen. Zudem können solche Kataloge nicht mit der rasanten Entwicklung des Internets mithalten und beschränken sich deshalb auf Spezialgebiete (z.B. [www.cetus-links.org](http://www.cetus-links.org)). Themenkataloge werden auch häufig in Portalen als „Einstiegspunkt“ ins World Wide Web eingesetzt.

### **4.2.2 Suchmaschinen**

Suchmaschinen bieten den Nutzern den Zugriff auf weitaus mehr Seiten, als dies bei Themenkatalogen der Fall ist. Sie bestehen aus mehreren verschiedenen Komponenten und verfolgen zur Aktualisierung ihres Datenbestandes einen vollständig automatischen Ansatz. Die folgende Abbildung gibt einen prinzipiellen Überblick.

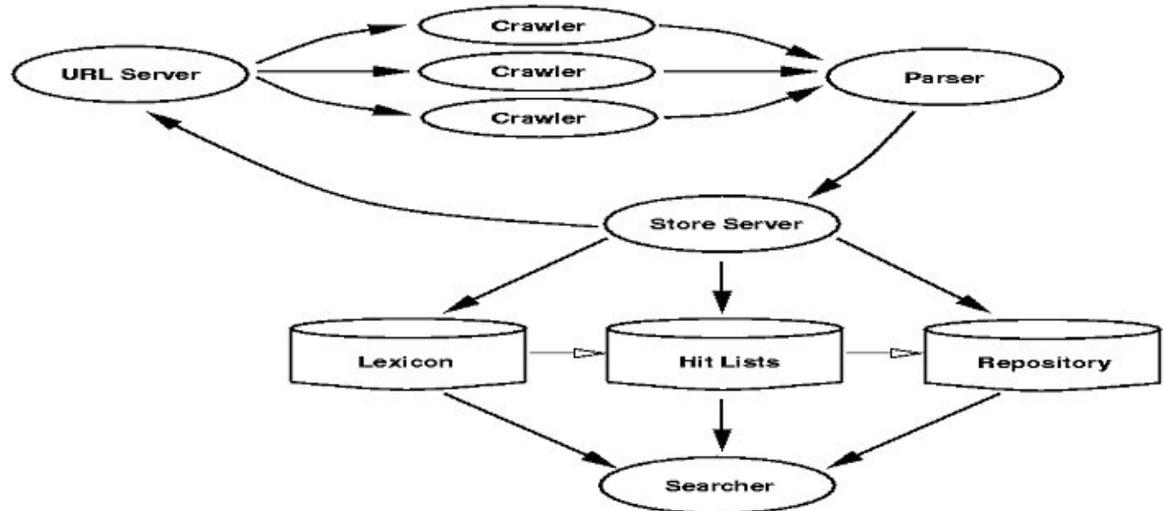


Abb. 4.3: Komponenten einer Suchmaschine ([Woll01])

Der URL-Server verwaltet sämtliche Adressen, die noch nicht indiziert worden sind. Über die Anmeldeformulare der Suchmaschine können eigene Seiten dort registriert werden. Die URLs werden an den oder die Crawler weitergegeben. Dieser stellt eine Verbindung mit der angegebenen Seite her und übergibt diese – bei erfolgreicher Übertragung – dem Parser. Der Parser analysiert die Seite und reicht die erhaltenen Informationen an den Store Server weiter. Dort werden die in der Seite enthaltenen Links zur weiteren Indizierung an den URL-Server zurückgegeben. Die Seite wird weiterhin nach neuen Wörtern durchsucht, mit denen das Lexikon erweitert wird. Zu jedem Wort wird in der Hit List die Häufigkeit des Auftretens vermerkt, und im Repository werden entweder Teile der Seite (z.B. der Titel, Hauptüberschriften,...) oder sogar die ganze Seite als Volltext gespeichert. Der Searcher stellt das Frontend der Suchmaschine dar und stellt die Anfragemaske zur Verfügung. Für eine ausführlichere Beschreibung wird auf die Arbeiten von Wichmann und Wollenweber verwiesen ([Wich01], [Woll01]).

#### 4.2.2.1 Vorteile

Der große Vorteil von Suchmaschinen liegt in der vollautomatischen Aktualisierung ihres Datenbestandes. Eingriffe von menschlicher Seite sind nur bei der Anmeldung neuer Seiten oder Domains nötig. Positiv ist auch die hohe Anzahl an indizierten Seiten, welche die von Themenkatalogen um ein Vielfaches übersteigt.

#### 4.2.2.2 Nachteile

Nachteilig wirkt sich die oftmals hohe Trefferquote bei geringer Relevanz aus. Als eine Ursache ist hierfür die Volltextindizierung und ein ungenügendes Ranking-System zu sehen. Kommt in einem Text ein Wort sehr häufig vor, so wird vom System her eine hohe Relevanz angenommen, obwohl dies aber nicht zwingend der Fall sein muss. Selbst wenn ein Wort nur einmalig auftritt, wird durch die Volltextindizierung zusätzlich ein Treffer

angezeigt, der nur eine untergeordnete Wichtigkeit besitzt. Es ist ebenfalls nicht möglich, entlang der Links von einer Seite zur anderen zu suchen. So ist zum Beispiel die Suche nach den grundlegenden Arbeiten zum Topic Maps Standard (siehe Einleitung zu Kapitel 2) nicht von Erfolg gekrönt, wenn der Sucher nicht den Begriff „Davenport Group“ kennt. Als ebenfalls schwierig erweist sich für den durchschnittlichen Benutzer der Gebrauch der Abfragesprache mit allen Möglichkeiten der Angabe von Operatoren.

### **4.2.3 Einsatzmöglichkeiten von Topic Maps**

Um die möglichen Potentiale eines Einsatzes besser untersuchen zu können, ist, wie aus den vorangehenden Gliederungspunkten ersichtlich, eine Unterscheidung zwischen Suchmaschinen und Themenkatalogen vorzunehmen. Für beide Einsatzgebiete ergeben sich zwangsläufig unterschiedliche Vor- und Nachteile.

#### **4.2.3.1 Einsatz in Suchmaschinen**

Es ist prinzipiell vorstellbar, dass alle indizierten Seiten durch eine Suchmaschine in Form einer Topic Map abgespeichert werden. Dazu müsste zu jedem Wort im Lexikon ein entsprechendes Topic angelegt und mit den Einträgen aus dem Repository mit den Vorkommen (Occurrences) versehen werden. Diese Schritte werden vom Autoren als vollautomatisch realisierbar angesehen. Problematisch sieht er allerdings dabei die Tatsache, dass damit die Potentiale von Topic Maps bei weitem noch nicht ausgeschöpft werden. Bei der Angabe der Vorkommen eines Topics sollte auf alle Fälle auch der Typ des Vorkommens (z.B. Biographie, Abbildung, Konferenztext,...) mit angegeben werden. Um diese Aufgabe automatisch vom Rechner durchführen zu lassen, müssten Verfahren zur Inhalts- und Sinnerkennung der bearbeiteten Texte zur Verfügung stehen. Ob zum Beispiel ein technischer Artikel auf einer Entwicklerkonferenz diskutiert wurde oder lediglich auf der Webseite eines Herstellers publiziert wurde, ist dem Text selbst nicht zu entnehmen. Es muss hierzu der Kontext, in dem der Artikel veröffentlicht wurde, betrachtet werden. Diese Aufgabe ist nach dem Dafürhalten des Autors sehr schwierig automatisch zu lösen. Suchmaschinen können demzufolge also nur grundlegende Vorarbeiten für die Erstellung einer Topic Map leisten. Die Resultate müssen von menschlichen Redakteuren nachbearbeitet werden.

#### **4.2.3.2 Einsatz in Themenkatalogen**

Anders als bei Suchmaschinen werden Themenkataloge bereits von Beginn an von menschlichen Redakteuren erstellt. Aus diesem Grunde erhöht sich hier auch das Einsatzpotential von Topic Maps. Zu den vorliegenden Kategorien können entsprechende Topics erstellt werden. Die katalogisierten Seiten können diesen Topics als Vorkommen zugeordnet werden. Da bei Themenkatalogen die Seiten von den Redakteuren häufig

bewertet werden (z.B. Lycos) ist es denkbar, dass im Zuge der Bewertung der Kontext für die Angabe des Vorkommenstyps gleichzeitig mit erschlossen wird. Zusätzlich könnten einfache Angaben zum Text in Form von Assoziationen gemacht werden (wer hat den Text verfasst? in welchem Zusammenhang wurde er veröffentlicht? mit welchen weiteren Themen beschäftigt sich der Autor?). Es lassen sich ebenfalls Verbindungen zu verwandten Themen ziehen.

### 4.2.3.3 Fazit

Topic Maps sind für den Einsatz in Themenkatalogen sehr geeignet. Beim Einsatz in Suchmaschinen ergeben sich jedoch erhebliche Probleme. Wie gezeigt, können sie dort aber grundlegende Aufgaben für den Aufbau der Topics übernehmen. Eine wichtige Rolle wird in Zukunft nach Ansicht des Autors auch die Fähigkeit der Topic Map Engine beim Durchführen des Mergings spielen. Dann ist es nämlich denkbar, dass von einem leistungsfähigen Rechner die gesamten verfügbaren Themenkataloge zu einem einzigen vereinigt werden und damit auf einen größeren Informationsbestand zugegriffen werden kann. Um langfristig die Durchsetzung der neuen Technologie voranzutreiben, müssen noch weitere grundlegende Arbeiten durchgeführt werden. Dazu gehört die Entwicklung einer einfachen und trotzdem leistungsfähigen Anfragesprache und Verfahren für die zumindest halbautomatische Inhaltserschließung bereits katalogisierter Dokumente. Assoziationen zwischen einzelnen Topics werden aller Voraussicht nach trotzdem noch per Hand – wenn auch rechnergestützt – durchgeführt werden müssen. Als das große Ziel bleibt das „Semantic Web“ ([Bern00]). Um dieses Ziel zu erreichen, wurde mit dem Resource Description Framework (RDF) eine Initiative des World Wide Web Consortiums (W3C) gestartet, die nachfolgend vorgestellt werden soll.

### 4.2.4 RDF als Alternative

Alle Informationen im Internet in Form von Texten, Dokumenten, Bildern usw. sind für einen Rechner zwar lesbar, aber nicht auswertbar. Die Vision vom „Semantic Web“ greift diesen Misstand auf. Die Lösung wird in der Angabe von Metadaten über bestimmte Informationsressourcen gesehen ([Lass+99], §1). RDF legt die Grundlagen, um solche Metadaten verarbeiten zu können. Ziel ist „to define a mechanism for describing resources that makes no assumptions about a particular application domain, nor defines (a priori) the semantics of any application domain. The definition of the mechanism should be domain neutral, yet the mechanism should be suitable for describing information about any domain“ ([Lass+99], §1).

#### 4.2.4.1 Grundlagen

RDF basiert auf benannten Eigenschafts-wert-paaren und besteht aus folgenden 3 Objekttypen:

- **Ressource**  
Jedes beliebige Ding kann eine Ressource darstellen und wird durch einen URI adressiert. Ressourcen in RDF sind also mit Topics vergleichbar.
- **Property**  
Ein Property (Eigenschaft) beschreibt eine Ressource in irgendeiner Art näher.
- **Statement**  
Ein RDF-Statement ist eine Kombination aus Ressource mit einer benannten Eigenschaft (Property) und dem zugehörigen Wert der Eigenschaft.

Die folgende Abbildung zeigt die Zusammenhänge zwischen den einzelnen Objekttypen.

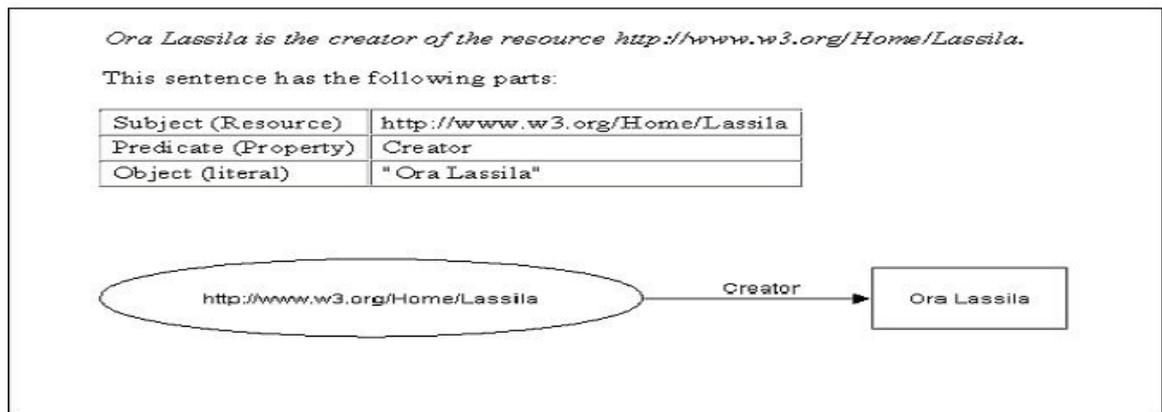


Abb. 4.4: Zusammenhänge der 3 Objekttypen in RDF ([Lass<sup>+</sup>99], §2.1.1.)

Als grundlegende Syntax wurde XML gewählt. Hiermit lässt sich wieder eine Gemeinsamkeit mit dem Topic Map Standard (und mit XTM in besonderer Weise) feststellen. Die folgende Abbildung zeigt die Verwendung der RDF-Syntax für das oben bereits vorgestellte Beispiel.

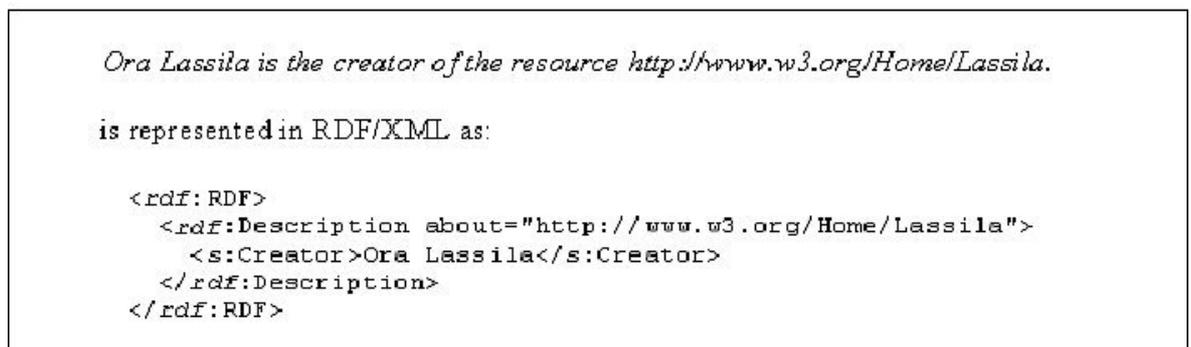


Abb. 4.5: Syntax von RDF ([Lass<sup>+</sup>99], §2.2.1)

Dabei stellt das „s“ vor dem Element „Creator“ lediglich den Verweis auf einen XML Namensraum dar, der die Definition des Elementes Creator enthält.

Aus der zugehörigen Grammatik zu RDF ([Lass+99], §2.2.1.) ist ersichtlich, dass die Property-Value-Paare wiederum ein Element „Description“ besitzen können. Durch diese Rekursion sind also „Beschreibungen über Beschreibungen“ möglich. Zusätzlich bietet RDF drei verschiedene Containerarten an, die für die nähere Beschreibung von mehreren Ressourcen (Ressourcensammlungen) genutzt werden können.

### 4.2.4.2 Vergleich mit Topic Maps

In der folgenden Tabelle sollen die Eigenschaften von RDF mit denen von Topic Maps verglichen werden.

Tab. 4.1: Vergleich RDF mit Topic Maps

Vergleichspunkt	RDF	Topic Maps
Syntax	XML	SGML, XML
Hauptelemente	Description mit zugehörigem Property-Value-Paar	Topic, Association, Occurrence
Zweck	nähere Beschreibung von Informationsressourcen	nähere Beschreibung von Dingen mit ihren Zusammenhängen zu anderen Dingen, zusätzlich ist die Angabe von relevanten Informationsressourcen möglich
Besonderheiten	Containerelemente zur Beschreibung von Reihenfolgebeziehungen zwischen Informationsressourcen	Benennung der Topics (Topic Names) für verschiedene Zwecke möglich  Angabe von Gültigkeitsbereichen (scopes)

### 4.2.4.3 Bewertung und Ausblick

RDF weist im Vergleich zu Topic Maps eine wesentlich einfachere Struktur auf. Aufgrund dessen ist es aber nicht in gleicher Weise so ausdrucksmächtig wie Topic Maps. RDF beschreibt zudem keine Zusammenhänge zwischen einzelnen Informationsressourcen. Dieses Problem wird erst durch RDF Schema gelöst ([Bric<sup>+</sup>00], §1). RDF Schema beschreibt grundlegende Typen von Properties. Damit ist es möglich, ähnlich mächtige Strukturen wie mit Topic Maps abzuleiten. Standardmäßig sind sogar Superklasse-Subklasse-Relationen vorgesehen. Über das Typ-Attribut kann ebenfalls eine Klasse-Instanz-Relation ausgedrückt werden. RDF stellt keine echte Alternative zu Topic Maps dar. Gründe hierfür werden in der fehlenden Möglichkeit, Assoziationen zwischen verschiedenen Informationsressourcen herzustellen, gesehen. Aufgrund der primären Ausrichtung auf die Beschreibung von Informationsressourcen bietet es aber sehr gute Ergänzungsmöglichkeiten zum Topic Map Standard, da Topic Maps bei den Vorkommnissen (Occurrences) nur dürftige Beschreibungsmöglichkeiten bieten. Durch die Einführung von Containerelementen zur Beschreibung von Reihenfolgebeziehungen könnte das Einsatzpotential von Topic Maps im Wissensmanagement noch gesteigert werden. Die beiden Entwicklergruppen haben bereits solche Synergieeffekte erkannt, und es bleibt abzuwarten, wie sich beide Konzepte weiterentwickeln.

## 4.3 Unterstützung für elektronisches Lernen

Die äußerst dynamische Entwicklung in allen Bereichen der Wissenschaft und Technik stellt hohe Anforderungen an die Qualifikation der Beschäftigten. Somit werden umfassende Fortbildungsmaßnahmen unumgänglich. Um die hohen Kosten (Reise-, Übernachtungs-, Verpflegungskosten usw.) der außerhalb des Unternehmens abgehaltenen Seminare zu senken, können zum Beispiel speziell zusammengestellte Lernprogramme auf CD-ROM oder über das Internet genutzt werden. In den folgenden Abschnitten soll kurz der bisherige Einsatz von elektronischen Hilfsmitteln in der Lehre vorgestellt werden. Aus den sich daraus ergebenden Problemstellungen lassen sich Anforderungen an komplexe elektronische Lernsysteme ableiten. Abschließend sollen die Einsatzpotentiale von Topic Maps und noch ungelöste Probleme erklärt und bewertet werden.

### 4.3.1 Bisheriger Einsatz elektronischer Hilfsmittel in der Lehre

Wenn vom Einsatz elektronischer Hilfsmittel in der Lehre die Rede ist, so bedeutet dies zumeist, dass entsprechend ausgearbeitete Folien mit einem Präsentationsgraphikprogramm mittels Videoprojektor an die Wand projiziert werden. Weitergehende Möglichkeiten stehen den Dozenten mit der Nutzung des Internets zur Verfügung. So können beispielsweise aktuelle Änderungen im Stundenplan, zusätzliche Angebote an

Vorträgen, Skripten, Übungsmaterial usw. auf einfache Art und Weise veröffentlicht werden. Email und Newsgroups dienen der Kommunikation zwischen Lehrenden und Lernenden. Die oben angesprochene Art und Weise des Vortrags per Videoprojektor hat aber auch entscheidende Nachteile. Zum einen muss der Vortragende auf herkömmliche Overheadprojektoren oder auf Kreide und Tafel zurückgreifen, sobald eingehende Veranschaulichungen notwendig sind, die nicht auf den Präsentationsfolien Platz finden. Dies gilt auch, wenn der Dozent auf spezielle Fragen seiner Schüler eingehen will. Dann ist die Entwicklung eines Tafelbildes „von Hand“ erforderlich (vgl. [Lovi01]).

### **4.3.2 Anforderungen an Lernumgebungen**

Aus den oben angeführten Schwächen lässt sich ein wichtiger Schluss ziehen: „Kommunikation ist der Schlüssel im Lernprozess!“ ([Chin00], Folie 22). Darunter fällt sowohl die Kommunikation zwischen dem Lehrenden und dem Lerner zum eigentlichen Wissenstransfer als auch die Kommunikation zwischen den Lernenden zum Austausch und der Vertiefung des Gelernten. Zusätzlich ist für den Lernenden die eigene Verarbeitung des Lernstoffs nötig. Aus diesem Grund müssen elektronische Lernsysteme geeignete Technologien einsetzen, welche die Kommunikation in der beschriebenen Art und Weise ermöglichen. Um den Lernstoff besser verinnerlichen zu können, ist es hilfreich, wenn sich der Schüler direkt an einen aufbereiteten Text eigene Anmerkungen machen kann. Die gleiche Möglichkeit sollte auch für Grafiken und Videos bestehen. Hier könnten dann Erklärungen zum Inhalt des Videos und Angaben zu den wichtigsten Stellen gemacht werden. Ein elektronisches Lernsystem sollte ebenso die Navigation zwischen artverwandten Themenpunkten anbieten. Um die unterschiedlichen Lern- und Lesegewohnheiten (vgl. [Horn98], S.52f) Einzelner zu berücksichtigen, sollte die Möglichkeit bestehen, unterschiedliche Lernpfade einzuschlagen. Ebenfalls könnte ein entsprechendes System den Lernstoff in unterschiedlicher Weise dem Schüler vermitteln. Zum einen kann dies geschehen, indem der Schüler selbst sich von Thema zu Thema bewegt, zum anderen könnte das System selbst die einzelnen Themen präsentieren. Letztere Vorgehensweise ermöglicht dann auch eine bessere Kontrolle der Lernzeiten. Nicht zuletzt sollte eine umfangreiche Suchfunktionalität zur Verfügung gestellt werden. Außerdem ist es wichtig, dass sich sowohl Lehrer als auch Lernende Informationen über den derzeitigen Lernfortschritt einholen können. Für die Tutoren ist eine Möglichkeit zur Bildung homogener Lerngruppen wünschenswert. Weiterhin müssen Module zur Erstellung der Inhalte und zum Zusammenstellen der Lerneinheiten vorhanden sein. Ein möglicher Aufbau eines elektronischen Lernsystems wird von China skizziert.

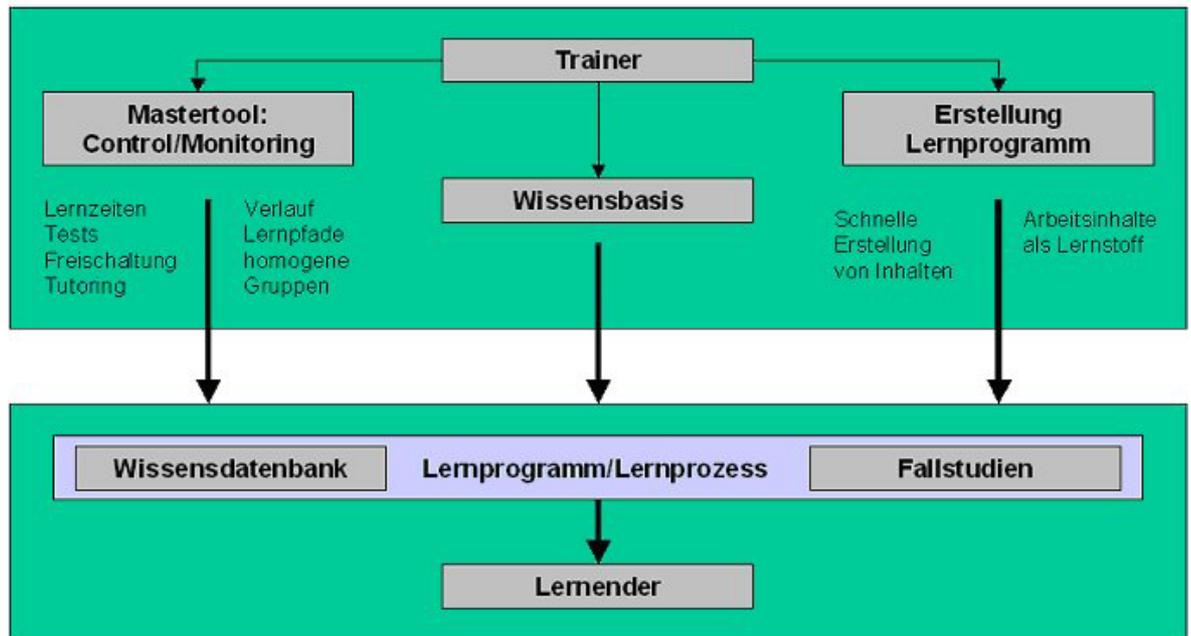


Abb. 4.6: Aufbau eines e-Learning-Systems ([Chin00], Folie 12)

Bei diesem Aufbau spielt die Wissensbasis die zentrale Rolle. Im folgenden Abschnitt soll gezeigt werden, welche Vor- und Nachteile sich ergeben, wenn die zentrale Wissensbasis mit Hilfe von Topic Maps aufgebaut wird.

### 4.3.3 Einsatzmöglichkeiten von Topic Maps

Aufgrund der Tatsache, dass Topic Maps zur Modellierung des Zusammenhangs zwischen einzelnen Begriffen und zur Verknüpfung von Informationsressourcen mit diesen Begriffen genutzt werden, ist für sie nur der Einsatz als Wissensbasis im vorgestellten Schema sinnvoll. Durch die Verwendung von Facetten (Element facet) in ISO 13250 als Container für zusätzliche Informationen ist es aber auch möglich, in diesen Elementen zum Beispiel Lernzeiten für die Lektionen und ähnliche Informationen abzulegen. Die angesprochenen Lernpfade können auf einfache Art und Weise durch die Verwendung des Attributes scope für die verschiedenen Topics angelegt werden. Dazu werden ausgewählten Topics entsprechend ihrer Gewichtung die Attribute „Erstlerner“, „Fortgeschrittener“ usw. zugeordnet. Gemäß ISO 13250 lassen sich auch die Verweise zu den Informationsressourcen mit der Angabe eines Gültigkeitsbereiches versehen. Somit wäre es denkbar, dass nur fortgeschrittene Lerner zum Beispiel komplexere Schaltbilder angezeigt bekommen oder nur für sie auch auf weiterführende Literatur verwiesen wird. Einem „Erstlerner“ werden dagegen nur die grundlegenden Informationen vermittelt.

#### 4.3.3.1 Vorteile

Einer der Vorteile, die sich aus dem Einsatz der Topic-Map-Technologie ergeben, ist zweifellos das standardisierte Format, in dem das gesamte Kursmaterial vorliegt. Es ist

damit denkbar, dass dieses Format auch beim Umstieg auf eine andere Lernplattform zum Datenaustausch zwischen den beiden Systemen dient. Auf diese Art und Weise vermindert der Käufer dieser Plattform sein Risiko – bei einem proprietären System müssen auch die Kurse auf das neue System angepasst werden. Für den Hersteller von Lernmaterial würde ein einheitliches Austauschformat die Herstellungskosten senken.

Ein weiterer Vorteil liegt in der Integrationsmöglichkeit multimedialer Inhalte. Wie im Kapitel 2 gezeigt wurde, werden keine genaueren Vorschriften über die Art der Informationsressourcen gemacht. Lediglich die Topic Map Engine muss in der Lage sein, entsprechende Verweise aufzulösen und die Darstellung im Browser zu veranlassen. Der Browser ist dann für die Ausgabe der verschiedenen Texte, Grafiken und Videos verantwortlich.

Die verbesserte Suchfunktionalität stellt einen weiteren Vorteil dar. Wie bereits erwähnt, kann durch die Verwendung von Topic Maps auch entlang den Verbindungen zwischen mehreren Begriffen gesucht werden.

Zusätzlich bietet der Einsatz von Topic Maps auch die Möglichkeit – auf standardisierte Art und Weise – die Wissensbasis durch den Lernenden erweitern zu lassen. Dabei können zusätzliche wichtige Begriffe zugefügt, neue Verbindungen vorgeschlagen und diskutiert und der Informationsressourcenpool vervollständigt werden. Dies hat eine Qualitätserhöhung des Lernmaterials zur Folge. Eventuell können auf diese Art und Weise auch Kosten für die Aktualisierung des Lernstoffs gesenkt werden.

### **4.3.3.2 Nachteile**

Ein grober Nachteil von Topic Maps liegt in der fehlenden Möglichkeit, Reihenfolgebeziehungen modellieren zu können. Es ist für das Sachverständnis beispielsweise hinderlich, wenn zuerst die Funktionsweise eines Gegenstands vor dem generellen Aufbau des Gegenstands erklärt wird.

Probleme bereitet auch die Integration multimedialer Inhalte. Auch hier gibt es keine Möglichkeit, angeben zu können, ob zum Beispiel zuerst das Video angezeigt wird und dann der Text oder umgekehrt. Bei Videos ist es zudem derzeit nicht möglich, den genauen Einstiegspunkt und Endpunkt zu adressieren (z.B. zur Definition eines Begriffs ab 1:54 Minuten bis 3:21 Minuten Laufzeit). Dieser Misstand sollte mit Entwicklung des HyTime-Standards behoben werden ([Gold+97]). Bis heute hat sich aber dieser Standard nicht durchsetzen können. Ein Grund hierfür ist die enorme Komplexität und Verständnisschwierigkeiten des prinzipiellen Anliegens und nicht zuletzt die daraus resultierenden Implementationsschwierigkeiten. Vielleicht könnten solche Adressierungsprobleme durch die Einführung von verschiedenen Locator-Typen mit entsprechenden Attributen im XLink-Standard gelöst werden.

Der Einsatz von Topic Maps hat auch keine Auswirkungen auf die Kosten für die Erstellung der multimedialen Inhalte. Dies ist wichtig, denn nicht die Topic Map an sich vermittelt den Lernstoff, sondern die Inhalte der Map. Topic Maps unterstützen lediglich die Autoren, Zusammenhänge zwischen einzelnen Begriffen des Lernstoffs deutlicher herausstellen zu können. Entscheidend für den Erfolg ist letztendlich die Aufbereitung dieser Inhalte. Als Anforderung für einen guten Lernprozess beschreibt China: „Erzähle es mir, und ich vergesse es; zeige es mir, und ich erinnere mich; beziehe mich ein, und ich verstehe es.“ ([Chin00], Folie 4). Die Interaktion und Darstellung wird durch das Rahmenprogramm – die Lernumgebung – geboten. Somit wird die Leistungsfähigkeit der Lernumgebung durch die Topic Map Engine und den Browser bestimmt, entscheidend bleibt aber die Qualität des Lehrmaterials.

### **4.4 Katalogisierung großer Dokumentenbestände**

Die ständig zunehmende Informationsfülle macht es notwendig, die anfallenden Informationen entsprechend ihrer Eignung zur Lösung auftretender Problemstellungen zu ordnen und wieder leicht auffindbar zu machen. Zu diesem Zweck können, wie bereits oben vorgestellt, Dokumentenmanagementsysteme eingesetzt werden. Oftmals sollen aber diese Dokumentenbestände in ihrer Gesamtheit auch dritten Personen zur Verfügung gestellt werden. In diesem Fall bieten sich Medien wie die CD-ROM oder DVD an. Zusätzlich können Abzüge dieser Daten auch im Internet publiziert werden. Um das Auffinden der gewünschten Informationen zu erleichtern, bieten die Anzeigeprogramme entsprechende Suchfunktionalitäten an. Zudem wird der Dokumentenbestand häufig in Kategorien unterteilt. Die Dokumente werden den einzelnen Kategorien zugeordnet. Trotzdem bereitet dieses Vorgehen einige Nachteile. Sehr oft kommt es vor, dass ein Dokument mehreren Kategorien zugeordnet werden kann. Je feiner dabei die Unterteilung der Kategorien ist, desto wahrscheinlicher ist dies der Fall. Die korrekte Eingliederung ist sehr zeitaufwendig und damit kostspielig. Für die Einordnung muss wieder der Inhalt der Dokumente erfasst und richtig bewertet werden. Bei umfangreichen Dokumentenbeständen gehen dabei oft Informationen verloren, die nicht der Hauptzweck bei der Erstellung der Dokumente waren. Ein Beispiel hierzu wird im Abschnitt 4.1. vorgestellt. Es wurde dort auch gezeigt, wie sich diese Situation durch den Einsatz von Topic Maps umgehen lässt. Würden diese „Nebeninformationen“ bei der Katalogisierung berücksichtigt, so würde ein Zusatznutzen für den Endnutzer entstehen. Nachfolgend sollen für diesen Sachverhalt einige konkrete Anwendungsfälle entwickelt werden.

#### **4.4.1 Diplomarbeitsverzeichnis**

In einem solchen Verzeichnis könnten alle Diplomarbeiten eines Lehrstuhls abgebildet werden. Mittels geeigneter Assoziationen zwischen den einzelnen Arbeiten können die Antworten auf folgende Fragen modelliert werden:

- In welchen Arbeiten werden Ideen das erste Mal erwähnt bzw. weiterentwickelt?
- Welche Arbeiten von anderen Studenten werden zitiert?
- Welche Randgebiete des Themas werden gestreift?
- Wo liegen Verbindungen zu Themen anderer Lehrstühle?
- Wo finde ich dazu Quellen im Internet?

Doch nicht nur der Titel der Arbeit, der Verfasser und eine Zusammenfassung sollten in der Topic Map modelliert werden. Es besteht die Möglichkeit, auch die in der Arbeit vorgenommenen Definitionen, das Glossar und das Literaturverzeichnis in einen solchen Katalog mit aufzunehmen. Aus einer solchen Vorgehensweise würden sich nach Ansicht des Autors einige Vorteile ergeben. Dazu gehören zum Beispiel:

##### **4.4.1.1 Einheitlicher Begriffskatalog**

Durch die Abbildung von Kernbegriffen in der Topic Map ist es für den Verfasser einer Diplomarbeit möglich, ein verstärktes Augenmerk auf das eigentliche Thema der Arbeit zu legen und eine umständliche Definitionsentwicklung für einen Begriff zu umgehen. Er könnte dann einfach die entsprechenden Definitionen übernehmen und weiterarbeiten. Oftmals ist aber eine Definitionsneuentwicklung unumgänglich und sogar gewünscht, wenn eine bereits vorgenommene Definition nicht alle in Frage kommenden Aspekte berücksichtigt. Durch das Vorliegen von vielen Definitionen zu einem Thema wird eine Weiterentwicklung unterstützt. Zudem ist später für Dritte diese Entwicklung nachvollziehbar.

##### **4.4.1.2 Vereinfachte Erstellung eines Glossars**

Aufgrund der Verfügbarkeit von Begriffsdefinitionen wird die Erstellung des Glossars erleichtert, indem der Verfasser nicht immer von vorn anfangen muss und „das Rad zum zweiten Mal erfindet“. Bei der Aufstellung kann sich somit voll auf neue Begriffe konzentriert werden, die für seine Arbeit von Bedeutung sind. Diese erweitern wiederum den Begriffskatalog und stehen auch anderen Studenten zur Verfügung.

#### **4.4.1.3 „Referenzbuchliste“**

Aus den Quellenangaben ist nachvollziehbar, welche Bücher, Diplomarbeiten und Artikel zu welchem Thema häufig herangezogen werden. Mit diesen Informationen ist es möglich festzustellen, ob die verwendeten Bücher in den Bibliotheken in ausreichendem Maße zur Verfügung stehen. Zudem können diese Quellen auch in den Vorlesungen als Literaturempfehlung weitergegeben werden.

#### **4.4.1.4 Leichteres Auffinden von Literatur zum Thema**

Wenn eine solche Literaturliste existiert, wird für den Verfasser einer Diplomarbeit die Literaturrecherche erleichtert. Er kann sich dann ganz auf Spezialwerke zum Thema seiner Arbeit konzentrieren. Außerdem wird sichergestellt, dass der Katalog dann für jede Quelle eine einheitliche Quellenangabe enthält.

#### **4.4.1.5 Leichtere Navigation zu Quellen im Internet**

Durch den Topic Map Browser ist es möglich, leicht zu den angegebenen Quellen im Internet zu navigieren. Wenn zudem noch die Diplomarbeiten in elektronischer Form vorliegen würden, wäre es einfach, zu den gewünschten Arbeiten, Definitionen, Quellenangaben usw. zu wechseln. Dies würde die Suchzeit gegenüber den Papierversionen stark verkürzen.

#### **4.4.1.6 Fazit**

Es bleibt also festzuhalten, dass ein solches Verzeichnis dem Verfasser einer Diplomarbeit erlauben würde, sich mehr auf das wesentliche seiner Arbeit zu konzentrieren und somit mehr Raum für eigene Gedanken zum Thema besteht. Die Nebenarbeiten, wie das Erstellen des Glossars und des Literaturverzeichnisses, werden verringert – ein Wegfallen ist nicht möglich und auch nicht erwünscht (es gehört auch zum Nachweis, dass der Student gelernt hat, wissenschaftlich zu arbeiten).

Ein zusätzliches Potential ergibt sich aus der Möglichkeit des Mergings (des Zusammenfassens) mehrerer solcher Topic Maps von unterschiedlichen Lehrstühlen. Dies ist gerade für die Wirtschaftsinformatik als Schnittstelle zwischen der Informatik und der Betriebswirtschaft von Bedeutung, weil hier zwangsläufig auch Begriffe aus den angesprochenen Disziplinen verwendet werden.

### **4.4.2 MSDN Library**

Das Microsoft Developer Network ist eine Sammlung zahlreicher Ressourcen, die für die Programmierung unter der Windows™-Plattform notwendig und hilfreich sind. Dazu gehört unter anderem die vollständige Dokumentation des SDK sowie umfangreiche

Beispielprogramme, die im Quelltext vorliegen. Zusätzlich befinden sich in dieser Sammlung etliche elektronisch aufgearbeitete Bücher und Artikel aus regelmäßig erscheinenden Magazinen. Die Auslieferung erfolgt vierteljährlich sowohl auf CD-ROM und DVD. Unter der Adresse <http://msdn.microsoft.com> wird der kostenlose Zugriff auf die Online-Variante geboten. Bei vollständiger Installation beträgt der Umfang ungefähr 1,3 Gbyte (ohne Beispielprogramme). Dies zeigt die enorme Fülle dieser Dokumentationsammlung. So verwundert es wenig, dass bei Suchanfragen häufig 500 und mehr angebliche Treffer angezeigt werden. Problematisch ist auch die Suche in der sogenannten Knowledge Base. Hier werden sämtliche bekannte und behobene Fehler, Umgehungsmöglichkeiten, Besonderheiten der verschiedenen Versionen und Anleitungen zur Programmierung für viele Produkte angeboten. Es finden sich sehr häufig noch Verweise zu Online-Ressourcen, die auf CD-ROM nicht mit ausgeliefert werden konnten. Durch den Einsatz von Topic Maps könnten auch hier nach Meinung des Autors viele Verbesserungen erreicht werden.

### **4.4.2.1 Bessere Unterteilung der Dokumentation**

Die Hauptkategorien werden auf niedrigerer Ebene noch einmal in die verschiedenen Programmiersprachen unterteilt. Zu den Hauptkategorien zählen technische Artikel, die eigentliche Produktdokumentation, Hintergrundartikel, Beispielprogramme, Bücher usw. Durch das Anlegen von Begriffen in der Topic Map wäre der Entwickler (als Hauptnutzer) aber in der Lage, von diesem Thema – das ihn eigentlich interessiert – auf die verschiedenartigen Ressourcen zuzugreifen. Er muss dann nicht mehr den umgekehrten Weg gehen und in den verschiedenen Kategorien nach seinem Thema suchen.

### **4.4.2.2 Verbesserte Nutzung der Beispielprogramme**

Ein wichtiges Hilfsmittel im MSDN sind die mitgelieferten Beispielprogramme. Ihr volles Potential kann aber im derzeitigen Zustand noch nicht voll ausgeschöpft werden. Bei der Beschreibung der Beispiele werden häufig nur die verwendeten API-Rufe und die grundlegend genutzte Technologie als Suchwort festgelegt. Zu den Technologien gehören zum Beispiel ADO, COM(+), XML, ASP. Die Programme enthalten aber beispielsweise auch Vorlagen zur Lösung von Problemen bei der Oberflächenprogrammierung, der Netzwerkprogrammierung usw.. Sie werden aber nur mit den Suchwörtern ihres ursprünglich gedachten Zwecks in der Sammlung abgelegt. Somit können vielleicht wichtige Informationen nur mehr oder weniger „per Zufall“ gefunden werden. Die Abbildung solcher Informationen in einer Topic Map könnte die Nutzbarkeit der Beispielprogramme somit noch erhöhen. Die nachfolgende Abbildung soll die Vorgehensweise verdeutlichen.

```
<topic id="xyz" types="ListBoxCtrl">
  <topname>
    <basename>ListBox-Steuererelement</basename>
  </topname>
  <occurs
    occtl="Beispielprogramm"
    scope="Netzwerkprogrammierung"
    type="Programm"
    xlink:type="locator"
    xlink:href="NetMonitor.cpp"
  />
  <occurs
    occtl="Artikel"
    scope="Oberflächenprogrammierung"
    type="technischer Artikel"
    xlink:type="locator"
    xlink:href="WindowsUI.html"
  />
</topic>
```

Abb. 4.7: Verweise auf unterschiedliche Quellen

### 4.4.3 Technische Dokumentation

Aufgrund umfangreicher Qualitätsvorschriften wird die Dokumentation eines Produktes immer umfangreicher. Entsprechend komplexer wird auch die Informationsfülle in den enthaltenden Einzeldokumenten des Produktes. Auch hier können Topic Maps die Suche nach einzelnen Informationen vereinfachen. Voraussetzung ist aber das Vorliegen der gesamten Dokumentation in elektronischer Form.

#### 4.4.3.1 Vorteile

Aufgrund des Einsatzes einer Topic Map kann jeder einzelne Bestandteil des Produktes detailliert beschrieben werden. Dazu muss für jede Komponente ein zugehöriges Topic angelegt und mit einem Identifikator versehen werden. Dazu würden sich zum Beispiel die Schlüsselnummern aus dem Materialkatalog eignen. Zu jedem Topic werden die zugehörigen Dokumente als Informationsressourcen angegeben und über das Rollenattribut näher beschrieben. Da wie bereits mehrfach angesprochen keine Vorschriften bezüglich der Adressierung gemacht werden, können auch nicht elektronisch vorliegende Dokumente referenziert werden. Über die Assoziationen kann der Aufbau des Produktes abgebildet werden. Dazu sind vor allem Assoziationen wie „ist-Teil-von“ bzw. „besteht aus“ notwendig. Über den Einsatz von Gültigkeitsbereichen wird die Trennung in die einzelnen Bestandteile der Dokumentation wie Nutzerhandbuch, Qualitätsnachweise, Servicehandbuch usw. ermöglicht, ohne dass dazu jedes Mal ein neues Topic angelegt werden muss. Es sind also alle Informationen zu einem Teil des Produktes an einer

zentralen Stelle – der Topic Map – zusammengefasst. Über geeignete Browser-Software kann auch auf einfache Art und Weise ein Online-Hilfesystem aufgebaut werden. Dazu müssen nur die Topics bzw. Vorkommnisse mit entsprechenden Gültigkeitsbereichen abgearbeitet werden.

### **4.4.3.2 Nachteile**

Die Erstellung einer vollständigen Produktdokumentation ist eine zeit- und kostenintensive Angelegenheit. Der Aufbau einer Topic Map erfordert zusätzliche Arbeiten wie Pflege und Aktualisierung der Daten. Zudem ist es nicht ausreichend, die Dokumentation nur auf elektronischen Medien auszuliefern. Zugehörige Kontrollnachweise und Bestätigungen bedürfen meist der Schriftform. Somit wären dann zwei Datenbestände aktuell zu halten und folglich steigt der Zeit- und/oder Personalbedarf und damit die Kosten. Wenn es allerdings gelänge, aus dem Autorensystem der Topic Map die gewünschte Papierdokumentation sich auf Knopfdruck erzeugen zu lassen, können wieder Zeit und Kosten gespart werden. Um dies zu erreichen muss zuerst aber hochspezialisierte Software zur Verfügung stehen. Da es sich jedoch um einen noch sehr jungen Standard handelt, wird nach dem Dafürhalten des Autors bis zur Verfügbarkeit solcher Software noch einige Zeit vergehen. Prinzipiell sind aber solche Anwendungsfälle denkbar.

### **4.4.4 Lexika und Enzyklopädien**

Die Anwendung von Topic Maps für solche Nachschlagewerke liegt aufgrund der Konzeption des Standards nahe und soll deshalb auch weniger ausführlich behandelt werden. Solche Werke sind schon von ihrer internen Struktur her in einzelne Begriffe gegliedert. Die benötigten Verweise auf Bilder und Texte lassen sich einfach vornehmen. Die Verbindungen zu anderen Themen ist durch die Umwandlung bestehender Querverweise in analoge Assoziationen möglich. Die Qualität Topic Map-basierter Nachschlagewerke wird nach Ansicht des Autors stark von den verwendeten Beschreibungen der Assoziationen und Vorkommnisse abhängen, weil die Namensgebung dieser zentralen Bestandteile bei einer Suche entscheidend ist.

## **4.5 Container für Metadaten**

Allgemein können Topic Maps zum Kapseln von Metadaten in beliebigen XML-Dokumenten genutzt werden. Sie könnten dort zum einen die Struktur der zugrunde liegenden DTD dokumentieren und entsprechende Verweise auf die Elemente setzen. Zum anderen können sie auch genutzt werden, um bestimmte Formatinformationen für die einzelnen Elemente der DTD zu beschreiben. Der Begriff Format – im Sinne einer Vorlage für Zeichenketten – muss hier aber deutlich vom Layout - für die Darstellung der Daten – unterschieden werden. Das beschriebene Vorgehen wird jedoch nicht benötigt, wenn die

Dokumentstruktur mit Hilfe von XML-Schemata beschrieben wird. Dort sind solche Formatangaben über reguläre Ausdrücke bereits im Standard vorgesehen ([Biro<sup>+</sup>00], §4.2.4). Das Verfahren bleibt aber für SGML-Dokumente anwendbar.

Ein zusätzliches Anwendungsgebiet sieht der Autor noch in einer Art „Bildbeschreibungssprache“ für Scalable Vector Graphics (SVG) ([Ferr01]). Dort könnten mit Hilfe von Topic Maps dann die einzelnen Bildelemente mit einem ID-Attribut versehen werden, das seinerseits wieder auf das zugehörige Topic in der Topic Map verweist. Über das Topic kann dann auf erklärende Informationsressourcen verwiesen werden. Nach Ansicht des Autors könnte ein solches Verfahren besonders gut für elektronisch aufbereitetes Lehrmaterial verwendet werden – insbesondere für die Erklärung von Schaubildern, mit den Verweisen auf die entsprechende Literatur und verwandte Konzepte. SVG bieten die Unterstützung von Metainformationen an ([Ferr01], §21), es fehlt aber die Verknüpfung zu den einzelnen Bildelementen, so dass das Bild nur in seiner Gesamtheit zu beschreiben ist. Für komplexe Abbildungen (z.B. Schaltpläne oder schematische Darstellungen von Maschinen in der elektronischen Produktdokumentation) ist dies aber unzureichend.

Einen weiteren interessanten Einsatzpunkt könnte eine solche Bildbeschreibungssprache auch bei der Aufbereitung von Inhalten für blinde oder sehbehinderte Menschen haben. Durch unterschiedlich getypte Informationsressourcen, die einem Bildausschnitt zugeordnet werden können, ist auch die Nutzung auf entsprechenden Ausgabegeräten möglich.

## 5 Modellierung von Topic Maps

Die Modellierung von Topic Maps dient der Abbildung von Wissensstrukturen in einem für den Computer verständlichen und auswertbaren Format. Wissen wird beschrieben als zweckorientierte Vernetzung von Informationen ([Scho<sup>+</sup>99], S.2). Dabei werden die einzelnen Topics mit entsprechenden Informationsressourcen verknüpft und anschließend ein inhaltlicher Zusammenhang durch die Modellierung von Assoziationen zwischen den Topics hergestellt. Im nächsten Abschnitt wird ein kurzer Einblick in die Modellierung mit Hilfe semantischer Netzwerke gegeben und nachfolgend wichtige Modellierungskriterien vorgestellt. Mit diesem theoretischen Grundgerüst möchte der Autor dann versuchen, einen Vorschlag für ein mögliches Vorgehensmodell zur Modellierung von Topic Maps zu unterbreiten.

### 5.1 Grundlagen semantischer Netze

Semantische Netze sind neben Logik- und Frame-basierten Ansätzen ein Repräsentationsformat für Wissen. Eine Repräsentation liegt vor, „wenn zusätzlich zu einer Menge von

Repräsentationsstrukturen Angaben dazu vorliegen, wie die Strukturen der Repräsentation auf die Angaben der repräsentierten Welt abzubilden sind. Diese Angaben stellen die Interpretationsvorschrift dar.“ ([Reim91], S.10). Die Repräsentationsstrukturen stellen dabei eine Menge von Sachverhalten dar und sind ein Modell für die betrachtete Welt ([Reim91],S.9).

Semantische Netze entstanden auf Modellen des menschlichen Gedächtnisses der Kognitionspsychologie. Dargestellt werden diese Netze durch Knoten und Kanten. Dabei stehen die Knoten für beliebige Begriffe (Konzepte) aus der Modellwelt. Die Kanten repräsentieren Verbindungen (Assoziationen) zwischen den Begriffen. Semantische Netze besitzen die Eigenschaft typisierter Kanten, d.h. zu den vorliegenden Kanten ist die Angabe einer näheren Beschreibung möglich. Die Netze erhalten somit viel mehr Informationen als Netze mit nur einem (untypisierten) Kantentyp ([Reim91], S.79).

Die Kanten in semantischen Netzen sind gerichtet. Es ist unsinnig, aussagen zu wollen, dass das Gemälde „Mondaufgang am Meer“ die Stadt Berlin ausstellt. Ausnahmen stellen aber symmetrische Beziehungen dar. Als Beispiel sei dazu eine Assoziation „ist-verheiratet-mit“ genannt. In diesem Fall wird eine ungerichtete Kante verwendet, um diese Aussage deutlich zu machen. Die folgende Abbildung stellt ein kleines semantisches Netzwerk da und verwendet dabei das Beispiel aus Kapitel 2.

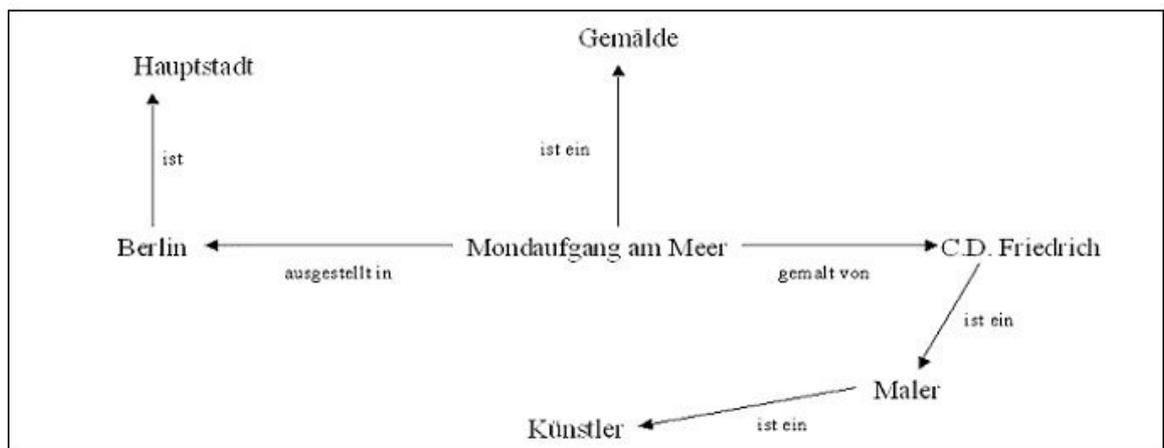


Abb. 5.1: ein einfaches semantisches Netzwerk

Einen weiteren Sonderfall bilden die sogenannten transitiven Kanten. Durch die Aussagen „Caspar David Friedrich ist ein Maler“ und „ein Maler ist ein Künstler“ kann sofort auf die Aussage „C.D. Friedrich ist ein Künstler“ geschlossen werden. Solche Schlüsse müssen durch die verarbeitende Software unterstützt werden.

### 5.1.1 Modellierung mit semantischen Netzen

Zur Modellierung mit Hilfe von semantischen Netzen werden folgende Vereinbarungen getroffen (s. [Reim91],S.82):

- In einem semantischen Netz werden zwei Knoten mit unterschiedlicher Beschriftung als verschieden angesehen.
- In einem semantischen Netz treten keine zwei Knoten mit gleicher Beschriftung auf.

Ein semantisches Netzwerk beschreibt eine Anzahl von zweistelligen Relationen über einer Anzahl von Knoten. Jede mehrstellige Relation kann in mehrere zweistellige Relationen zerlegt werden ([Ksie00], S.2; [Reim91], S. 100ff).

Semantische Netze lassen es zu, jedem Konzeptknoten Eigenschaften zuzuordnen. Dabei wird die Eigenschaft auch durch einen Knoten repräsentiert. Zwischen Ursprungsknoten und Eigenschaft wird eine Kante eingezeichnet, die mit der Eigenschaftsklasse bezeichnet wird (zum Beispiel eine schwarze (Eigenschaft) Tasche (Begriff, o. Konzept) mit der Eigenschaftsklasse Farbe). Beziehungen zwischen zwei Konzepten<sup>2</sup> werden durch eine entsprechende Kante modelliert. Die Beschriftung gibt dabei den Kantentyp an ([Reim91], S. 83).

### 5.1.2 Vergleich mit Topic Maps

Topic Maps weisen starke Gemeinsamkeiten aber auch Unterschiede zu semantischen Netzwerken auf. Grundlegend nutzen beide Verfahren das Prinzip eines Graphen mit Knoten und Kanten. Gemeinsam ist auch die Nutzung von unterschiedlich getypten Kanten. Semantische Netze erlauben zudem noch die Möglichkeit der Modellierung von Eigenschaften zu Objekten ([Reim91], S.83). Bei Topic Maps ist dieses ohne weiteres nicht möglich. Um eine Eigenschaft abbilden zu können, muss hier eine zusätzliche Assoziation erstellt werden, die auf ein entsprechendes Topic zeigt (zum Beispiel „Ferrari (Topic) hat-Farbe (Assoziation) rot (Topic)“). Im Quelltext sollte dies in etwa wie folgt aussehen.

```
<topic id="Ferrari" types="Auto">
  <topname><basename>Ferrari</basename></topname>
</topic>

<topic id="rot" types="Farbe">
  <topname><basename>Farbe rot</basename></topname>
</topic>

<assoc type="hat-Farbe">
  <assocrl type="Gegenstand">Ferrari</assocrl>
  <assocrl type="Farbe">rot</assocrl>
</assoc>
```

Abb. 5.2: Zuordnung von Eigenschaften über Assoziationen

---

<sup>2</sup> Die Bezeichnung „Konzept“ ist in der KI gebräuchlich und wurde vom englischen „concept“ abgeleitet und hat dieselbe Bedeutung wie „Begriff“ ([Reim91], S.16)

Ein weiterer Unterschied liegt in den Fragen der Vererbung. Bei semantischen Netzen erben alle abgeleiteten Knoten (über die „ist-ein“-Beziehung) die Eigenschaften der Elternknoten ([Reim91], S.89ff). Für Topic Maps wurden solche Verhaltensweisen nicht explizit festgelegt. Da Eigenschaften bei Topic Maps, wie oben gezeigt wurde, nur über Assoziationen modelliert werden können, erheben sich dann bei einer Vererbung einige wichtige Fragen. Welche Assoziationen bezeichnen Eigenschaften? Sollen alle Assoziationen eines Topics weitervererbt werden und wenn nicht, welche werden ausgeschlossen? Die erste Frage könnte mit dem Einsatz von PSI's beantwortet werden. Dazu müssten die gewünschten Topics nur mit einem entsprechenden PSI als „Eigenschafts-Topic“ gekennzeichnet werden. Eine Lösung der zweiten Frage kann nur mit Hilfe geeigneter Software beantwortet werden. Ein solches Verhalten würde aber wie erwähnt über den Standard hinausgehen und dementsprechend von jeder Topic Map Engine anders behandelt werden.

Ein weiterer Unterschied liegt nach Meinung anderer Autoren im Schwerpunkt der Aufgabe ihrer Grundeinheiten. Freese schreibt: „Topic Maps seems to focus more on the navigation between topics than associations. Semantic networks focus on the links between the nodes and the knowledge that is represented by the linked nodes ([Free00], S.9).

## 5.2 Modellierungskriterien

Modellierungskriterien legen einen groben Rahmen fest, welchen Eigenschaften ein fertiges Modell mindestens genügen muss. Diese Kriterien wurden zu den Grundsätzen ordnungsmäßiger Modellierung zusammengefasst. Zu diesen Grundsätzen gehören ([Schü98], S.111ff):

### 5.2.1 Syntaktische und semantische Richtigkeit

Dieses sind die zwei Grundanforderungen an jedes Modell. Syntaktische Vorgaben werden für XTM durch die DTD und für ISO 13250 durch die Meta-DTD gemacht. Eine gültige Topic Map entspricht entweder der XTM-DTD oder einer passenden Client-DTD im Falle des ISO-Standards. Die Korrektheit der Ableitung der Client-DTD muss dabei vom Designer der DTD sichergestellt werden ([Lobi00], S.87ff). Die syntaktische Richtigkeit kann durch den Einsatz von Prüfroutinen beim Import und Export der konformen Dokumente gewährleistet werden. Anhand der syntaktischen Strukturen kann dabei auf die Bedeutung der einzelnen Zeichen – deren Semantik – im Dokument geschlossen werden. So ist zum Beispiel festgelegt, dass alle Wörter (Zeichenfolgen, die durch Leerzeichen voneinander abgegrenzt sind) im Typattribut eines Topics auf andere Topic-elemente verweisen, die ein gleichlautendes ID-Attribut besitzen.

### **5.2.2 Relevanz**

Die zweckorientierte Abbildung des betrachteten Diskursbereiches auf das Modellsystem wird durch den Begriff der Relevanz beschrieben. Dabei geht es um die Frage, ob die betrachteten Objekte der Diskurswelt auch tatsächlich für das Modell dieser Welt von Bedeutung sind. Das Modell beschreibt ja lediglich ein Abbild der Wirklichkeit und vernachlässigt dabei unwichtige Details. Der Zweck des Modells wirkt sich dabei auch auf den Detaillierungsgrad aus. Soll zum Beispiel eine Topic Map zum Lebenswerk eines Malers erstellt werden, so sollten zumindest alle bekannten Werke erwähnt werden. Anders sieht dies aus, wenn über das Schaffen dieses Malers in einer bestimmten Epoche berichtet werden soll. In diesem Fall spielen dann nur die im entsprechenden Zeitraum entstandenen Werke eine Rolle.

### **5.2.3 Klarheit und Verständlichkeit**

Klarheit und Verständlichkeit des Modells bilden ebenfalls ein Modellierungskriterium. Dieses hat besonders großen Einfluss auf die Erweiterbarkeit und Änderbarkeit des Modells. Es müssen die Gesichtspunkte und Grundannahmen, nach denen das Modell erstellt wurde, erkennbar und nachvollziehbar sein. Um diese Verständlichkeit zu gewährleisten, ist es notwendig, eine entsprechende Dokumentation des Modellierungsvorganges anzufertigen. Nach Meinung des Autors sind Klarheit und Verständlichkeit besonders bei der Modellierung von Topic Maps zusätzlich auch aus der Sicht des Endnutzers zu betrachten. Dazu zählt auch die Fragestellung des Nutzers an die Topic Map, die eindeutige Namensgebung, die sinnvolle Zuweisung von Scopes zur Einschränkung des Gültigkeitsraumes, die verständliche Rollenbezeichnung in Assoziationen und so weiter. Fragen der Verständlichkeit betreffen unter anderem auch die Anzahl der Assoziationen. Diese Anzahl hat dabei wiederum Auswirkungen auf die Navigierbarkeit und den inhaltlichen Gesamtüberblick in einer stark oder weniger stark vernetzten Topic Map.

### **5.2.4 Wirtschaftlichkeit**

Das Erstellen jedes Topics in einer Map und aller zugehörigen Assoziationen ist mit einem Zeitaufwand verbunden. Daher ist es wichtig, den Detaillierungsgrad des Modells auch unter wirtschaftlichen Gesichtspunkten zu betrachten und dabei zu fragen: Was kostet die Erstellung eines Topics? Wie viel Geld muss an Lizenzgebühren für jede verwendete Informationsressource einkalkuliert werden? Muss unter solchen Aspekten das kleine Zitat von Herrn X. noch mit abgebildet werden? Solche Entscheidungen werden je nach Interessenlage des Herstellers einer Topic Map unterschiedlich ausfallen. Es wird dabei sicherlich zwischen Qualität und Herstellungskosten abgewogen werden. Zur Abschätzung

des Zeit- und Geldaufwandes könnte das Ausmaß an Informationen berücksichtigt werden, die zur zweckmäßigen Beschreibung eines Topics erforderlich sind. Dazu gehört die Anzahl an Informationsressourcen ebenso wie die Anzahl der Namen (Sortierung, Synonyme, Anzeigenamen) und die zu erwartende Zahl der Assoziationen zu anderen Topics. Aufgrund dieser Informationen kann das Topic dann auch hinsichtlich seiner Relevanz klassifiziert und entsprechend kalkulatorisch bewertet werden. Zur Klassifizierung könnte eine einfache Ordinalskala verwendet werden. Es ist aber gleichermaßen denkbar, ein Punktbewertungssystem einzuführen.

### **5.2.5 Strukturierung und Integration von Konzepten**

Bei der Integration von anderen Konzepten wird der Frage nachgegangen, inwieweit andere Sichten auf die gleiche Problemstellung in ein und dasselbe Modell eingefügt werden können. Bei der Modellierung von Informationssystemen seien als mögliche Sichten die Datensicht, die Prozesssicht und die Funktionssicht genannt. Ein integrierter Ansatz zur Modellierung von Informationssystemen ist beispielsweise ARIS ([Sche98]). Für das Projekt ELBE ist zum Beispiel die Information Mapping® Methode zu berücksichtigen, mit deren Hilfe die einzelnen Lehrveranstaltungen aufgearbeitet werden.

### **5.2.6 Vergleichbarkeit**

Um zwei Modelle vergleichen zu können, die mit Hilfe unterschiedlicher Methoden modelliert werden, muss ein gemeinsames Modell der beiden Ansätze ausgemacht werden. Dieses Modell befindet sich auf einer höheren Abstraktionsstufe und wird deshalb als Meta-Modell bezeichnet. Werden dann beide Ausgangsmodelle in dieses neue Modell transformiert, lassen sich diese beiden Modelle miteinander vergleichen. Ein solcher Vergleich ist wichtig, falls eine Unternehmung ihre Modellierungsmethode ändern will oder muss. Es ist zum Beispiel denkbar, dass sich ein Unternehmen, welches Expertensysteme entwickelt, nach neuen Möglichkeiten der Wissensrepräsentation umsieht. Bisher wurde im Unternehmen auf Logik-basierte Repräsentation zurückgegriffen. Um aber zum Beispiel über die Tauglichkeit von XTM für die Zwecke des Unternehmens befinden zu können, müssen beide Ansätze verglichen werden.

## **5.3 Vorschlag eines möglichen Vorgehensmodells**

Die nachfolgenden Schritte sind nach dem Dafürhalten des Autors notwendig, um Wissen – im Sinne von vernetzter Information – mit Hilfe von Topic Maps zu modellieren, die auch den oben aufgeführten Modellierungskriterien genügen.

### 5.3.1 Schritt 1: Abgrenzung des Diskursbereiches und Festlegen der Zielstellung

Mit der Durchführung dieses Schrittes werden zwei Ziele verfolgt. Zum ersten werden dadurch Kriterien für die Relevanz eines Topics festgelegt und zum zweiten werden auch die Vorarbeiten für die Integration der Information Mapping® Methode ausgeführt. Horn drückt aus, dass der Diskursbereich die in der Methode verwendeten Blocktypen zu einem Teil im Voraus bestimmt ([Horn89], S.104ff). Dieser Zweck ist aber vorrangig für das Projekt ELBE von Bedeutung, auf das dieses Vorgehensmodell später noch prototypisch angewandt werden soll. Wird der Diskursbereich nicht deutlich genug abgegrenzt, kann es zu Problemen beim Modellieren der Topic Map kommen. Es müssen Entscheidungen getroffen werden, die über die Aufnahme eines Topics in die Map oder den Ausschluss aus derselben verfügen.

Die Zielstellung legt fest, welche Fragen und Problemstellungen der Endnutzer mit Hilfe der erstellten Topic Map lösen können soll.

### 5.3.2 Schritt 2: Aufstellen der Ontologie

*Ontologie* ist ein wichtiger Begriff in der Philosophie und in der künstlichen Intelligenz und stellt in letzterer die explizite Beschreibung eines Modells dar ([Stuc<sup>+</sup>01]). Es ist eine Beschreibung für ein Wissensorganisationsschema<sup>3</sup> (KOS) ([Hodg01], S.13). Ein KOS wird zur Organisation von verschiedenen Materialien wie zum Beispiel von Büchern, Dokumenten usw. und zum Zweck des schnellen und sicheren Wiederauffindens genutzt. Eine Ontologie regelt die Frage, wie einzelne Begriffe oder Dinge zueinander in Beziehung stehen. Sie stellt zudem sicher, dass sich alle Nutzer eines Systems auf die gleichen Begrifflichkeiten beziehen ([Endr00], S. 5). Für Topic Maps beinhaltet das Erstellen einer Ontologie die Suche nach allgemein gültigen Typen der vorkommenden Topics und Assoziationen. Für das Projekt ELBE ist ein solcher allgemein gültiger Typ zum Beispiel das Fachwort oder der Literaturverweis oder der Assoziationstyp „ist-Aufgabe-von“. Das entstehende Klassifikationsschema kann seinerseits wiederum als eigenständige Topic Map modelliert werden. Da Assoziationstypen selbst wieder auf Topics verweisen, muss für jeden Assoziationstyp ein entsprechendes Topic angelegt werden. Durch diesen Vorgang der Modellierung entsteht das bereits erwähnte Topic Map Template. Soweit notwendig kann dieses auch noch weiter verfeinert werden.

Für das Aufstellen einer Ontologie gibt es noch kein genaueres Vorgehensmodell im Rahmen des *Ontology Engineering*, da dieses Gebiet Gegenstand aktueller Forschungen ist

---

<sup>3</sup> aus dem Englischen Knowledge Organization System / Scheme

([Endr00], S.10). Es existieren vielmehr empirische Ansätze (als Beispiel siehe [Endr00], S.11). Bei diesem Ansatz wird durch die Formulierung von Suchszenarien die starke Einbeziehung des Nutzers hervorgehoben.

Das Aufstellen der Ontologie vereinfacht zudem grundlegende Plausibilitätstests der Topic Map und erleichtert die Wartung und Pflege. Kurze Beispiele hierzu finden sich in Abschnitt 3.1.4.

### **5.3.3 Schritt 3: Aufstellen der Topics**

Nachdem die Ontologie aufgestellt wurde, kann damit begonnen werden, die „realen“ Begriffe des Diskursbereiches in das Organisationsschema einzuordnen. Kann ein Begriff nicht zugeordnet werden, so muss entweder die Ontologie angepasst werden oder aber die Frage nach der Relevanz des einzuordnenden Begriffes für die Topic Map gestellt werden. Nach dem erfolgreichen Zuordnen kann der Begriff als Topic modelliert werden. Dabei ist auf die eindeutige Vergabe des zugehörigen ID-Attributes zu achten. Zudem muss mindestens ein Name (das Element *basename* in ISO 13250) angegeben werden.

### **5.3.4 Schritt 4: Zuordnen der Informationsressourcen**

Als nächster Schritt wird die Zuordnung der Vorkommnisse eines bestimmten Topics vorgeschlagen. Es ist ebenfalls darauf zu achten, dass die zuzuordnenden Typen der Vorkommnisse (*occurrence types*) gleichfalls in der Ontologie aufgeführt werden, andernfalls sind sie dort nachzutragen.

### **5.3.5 Schritt 5: Aufstellen der Assoziationen**

Jetzt können die Verbindungen zwischen den einzelnen Topics gemäß dem Topic Map Template erstellt werden. Die Beschränkung auf die vorher festgelegten Assoziationstypen verhindert, dass zu viele unterschiedliche Verbindungen die Map unübersichtlich und schwer navigierbar machen. Es ist gerade hier wichtig, genau auf die Bedürfnisse des Endnutzers einzugehen und nur das exakt benötigte Wissen – als vernetzte Information – abzubilden.

### **5.3.6 Schritt 6: Abgleichen mit der Zielstellung**

Nach dem Fertigstellen der ersten Version der Topic Map wird versucht die Fragestellung aus Schritt 1 durch „Umherwandern“ in der Topic Map oder durch das Stellen einer geeigneten Suchanfrage zu lösen. Tritt hierbei ein Problem auf, so müssen fehlende Topics und Assoziationen ergänzt werden, indem mit Schritt 3 fortgesetzt wird. Um Erweiterungen durchzuführen werden anschließend neue Fragestellungen entwickelt und ebenfalls mit Schritt 3 fortgesetzt.

## **5.4 Prototypische Umsetzung für das Projekt ELBE**

Die Abkürzung ELBE steht für „Elektronische LehrBuch Edition“ und beschreibt ein Projekt des Lehrstuhles für Wirtschaftsinformatik, insbesondere Informationsmanagement an der Technischen Universität Dresden. Im Rahmen dieses Projektes werden die Lehrmaterialien aus den Vorlesungen elektronisch überarbeitet bzw. neu erstellt. Anschließend werden diese Materialien entsprechend einer SGML/XML-DTD strukturiert. Bei dieser Strukturierung fließt auch die bereits mehrfach angesprochene Information Mapping® Methode mit ein. Die Resultate dieser Vorgehensweise sind einzelne Wissensbausteine, die dann entsprechend ihrer späteren Verwendung entweder auf Papier, CD-ROM oder über das Internet als Lernmaterial den Studenten zur Verfügung gestellt werden ([Scho99b]).

### **5.4.1 Besonderheiten für die Umsetzung**

Da die Quelldokumente für dieses Projekt im SGML- bzw. XML-Format vorliegen, besitzen die einzelnen Textteile oder Abbildungen schon eine durch die Struktur festgelegte Bedeutung (Semantik) für den Nutzer. Es gilt jetzt, diese Semantik auch in der zu erstellenden Topic Map abzubilden. Da es sich nicht um eine völlige Neuentwicklung der Topic Map<sup>4</sup> handelt, muss konkret auf die Gegebenheiten der Dokumente eingegangen werden. Aus diesem Grunde gestalten sich die einzelnen Schritte des Vorgehensmodells etwas anders, bleiben aber in ihrer Intention gleich.

### **5.4.2 Schritt 1: Abgrenzung des Diskursbereiches und Festlegen der Zielstellung**

Der Diskursbereich wird schon durch die Tatsache der Verwendung als Lehrmaterial auf das Gebiet des Informationsmanagements festgelegt. Will man diesen Bereich weiter einschränken, so bietet es sich an, für die einzelnen Lehrveranstaltungen spezielle Topic Maps mit dem Diskursbereich der jeweiligen Veranstaltung (z.B. Aufgaben und Ziele des Informationsmanagements, Sicherung und Schutz von Daten) anzufertigen. Aufgrund der Möglichkeit des Zusammenführens (Merging) der einzelnen Topic Maps, wird vom Autor dieser Weg gewählt. Vorteile werden in der besseren Überschaubarkeit der Teilbereiche gesehen. Nachteilig wirken sich die fehlenden Assoziationen aus, die nach einem Merging noch von Hand erstellt werden müssen, um Verbindungen zwischen den einzelnen Themen der Lehrveranstaltungen zu ziehen.

---

<sup>4</sup> der Autor verwendet immer die volle Bezeichnung Topic Map, um eine deutliche Unterscheidung zur Information Map der Information Mapping® Methode herzustellen

### 5.4.3 Schritt 2: Aufstellen der Ontologie

Durch die Verwendung bereits vorstrukturierter Datenbestände ist die Analyse dieser Struktur (in Form einer DTD) und die Analyse der Dokumente selbst eine Vorarbeit zum Aufstellen der Ontologie. Durch die Untersuchung der Elemente der DTD können bereits einige mögliche Topic-Typen abgegrenzt werden. Dazu gehören unter anderem:

- Fachbegriff
- Quelle
- Indexmarke
- Abbildung
- Verweis
- Erklärung
- Begriff (für Glossar)
- Schlagwort

Weitere mögliche Typen sind „Frage“, „Antwort“, „Fallstudie“. Aus der Analyse ist auch ersichtlich, dass einige Elemente der Struktur, entgegen der ursprünglichen Absicht, doch layoutorientierte Elemente darstellen. Dazu gehören zum Beispiel die Elemente „Hervorhebung fett“, „Hervorhebung kursiv“, „Hochgestellt“ und „FreieTabelle“. Die Verwendung des letztgenannten Elementes ist aber nahezu unumgänglich, wenn man ein allgemein gültiges Verfahren zum Erstellen einer Tabelle nutzen möchte. Ansonsten müsste man für jede Tabelle eine ihrem Zweck entsprechende Struktur erzeugen.

Die Elemente der Information Mapping® Methode „Map“ und „Block“ werden ebenfalls als Topic-Typen abgebildet. Einem von ihnen instanziierten Topic können zusätzlich weitere Typen zugeordnet werden, die dieses Topic näher beschreiben. Dazu gehört zum Beispiel der Typ „Einleitung“. Auf die Notwendigkeit eines solchen Typs kann man aus der Analyse der Quelldokumente schließen, weil ein Einleitungsblock sehr häufig verwendet wird.

Auf die Abbildung der Elemente „Aufzählung“ und „Nummerierung“ wird verzichtet, da Topic Maps – wie bereits erwähnt – keine Abbildung einer Reihenfolgebeziehung zulassen und jedes Unterelement einer Auflistung durch die oben aufgezählten Typen bereits klassifiziert ist.

Nachdem die allgemein gültigen Topic-Typen ausgemacht worden sind, kann in einem folgenden Schritt noch eine Verfeinerung vorgenommen werden. Hierbei kann zum Beispiel der Typ „Abbildung“ noch entsprechend der Struktur in „ergänzende Abbildung“ und „Strukturgraphik“ unterteilt werden.

Durch die Nutzung der vorliegenden Dokumentstruktur ist das Aufstellen der Topic Typen relativ einfach durchzuführen und auch für Dritte einfach nachzuvollziehen (Dokumentation des Modellierungsvorganges, siehe Abschnitt 5.2.3). Da aber in der Struktur keine Zusammenhänge zwischen einzelnen Elementen deutlich werden, gestaltet sich die Suche nach allgemein gültigen Assoziationen deutlich schwieriger. Aus der DTD ist zwar ableitbar, dass eine Frage entweder aus Aufzählungen, Tabellen und/oder freiem Text bestehen kann, notwendiger ist jedoch für den Nutzer die Erkenntnis, dass jede Frage eine Antwort hat. Daraus folgt die Existenz einer Assoziation „Frage (Topic) hat-Antwort (Assoziation) Antwort (Topic)“. Die Assoziationen sind demnach aus dem dem Nutzer präsentierten Dokumenten (in HTML, oder auf Papier), Text auszumachen. Der Autor verzichtet hier auf die Auflistung von verwendeten Assoziationstypen. Exemplarisch seien nur die Assoziation „wird-unterteilt-in“ und „gehört-zu“ genannt.

Nach dem Abschluss der Analyse können sämtliche gefundenen Typen als Topics im Topic Map Template abgebildet werden.

### 5.4.4 Schritt 3: Aufstellen der Topics

Nachdem das Topic Map Template erstellt wurde, können nunmehr die „realen“ Topics ermittelt werden. Da ein bereits vorstrukturierter Dokumentenbestand verwendet wird, ergeben sich auch hier wieder in der Art und Weise der Ausführung einige Unterschiede. Es bietet sich an, die XML-Dokumente über einen Transformationsmechanismus in eine andere Struktur umzuwandeln. Mit XSL bzw. XSLT stehen dazu die benötigten Technologien bereits zur Verfügung.

Wird eine solche automatisierbare Vorgehensweise gewählt, müssen einige zusätzliche Fragen beantwortet werden. Darunter fallen:

- Welche Elemente der Struktur sollen in ein Topic umgewandelt werden?
- Wie ist mit den Container-Elementen (Aufzählung, Information Map,...) zu verfahren? Stellen diese selbst ein Topic dar?
- Wie wird Klartext (#PCDATA in der DTD) behandelt?

Die Antwort auf die erste Frage wird je nach Struktur unterschiedlich ausfallen. Durch die Verwendung der Information Mapping® Methode können für jede Information Map und die Information Blocks entsprechende Topics angelegt werden. Diese Möglichkeit liegt im Chunking-Prinzip und Labeling-Prinzip (zusammenfassende Beschreibung einer Informationseinheit) der Methode begründet ([Horn89], S.86).

Für die dritte Frage stehen nach Ansicht des Autors zwei Möglichkeiten zur Verfügung. Zum einen kann der Text als Vorkommnis eines noch zuzuordnenden Topics modelliert werden. Die zweite Möglichkeit beinhaltet die manuelle Erstellung von Topics nach einer

semantischen Textanalyse durch den Wissensmodellierer (welche Bedeutung hat der Text für das Thema? Kommen weitere einzelne Begriffe vor, die wieder als Topics abgebildet werden können?). Die folgende Tabelle stellt die sich daraus ergebenden Konsequenzen der beiden Vorgehensweisen dar.

Tab. 5.1: Behandlung von Klartextelementen

	<b>als Vorkommnis</b>	<b>manuelle Extraktion weiterer Information</b>
<b>Vorteile</b>	einfach zu realisieren über XLink adressierbar automatisierbar	zusätzliche Ausdrucksstärke der Topic Map
<b>Nachteile</b>	„Verlust“ zusätzlicher Informationen in der Topic Map	zeitaufwendig automatische Verfahren zur semantischen Textanalyse noch nicht ausgereift (vgl. [Endr01])

Prinzipiell ist aber auch eine Kombination aus beiden Möglichkeiten denkbar.

Nachdem diese Vorarbeiten geleistet wurden, kann mit der Erstellung der XSLT-Stylesheets begonnen werden. Der Autor setzte bei der Entwicklung ein selbst entworfenes Programm ein, das im Anhang 4 vorgestellt wird. Es soll zumindest für jedes ausgewählte Element eine Art „Skelett“ des zugehörigen Topics erstellt werden. Die folgende Abbildung soll diesen Ausdruck etwas verdeutlichen.

```
<Fachbegriff>Datenschutz</Fachbegriff>  
  
soll zumindest umgewandelt werden zu:  
  
<topic id="Datenschutz">  
  <topname>  
    <basename>Datenschutz</basename>  
  </topname>  
</topic>
```

Abb. 5.3: "Skelett" eines Topics

So weit möglich können aber auch schon Informationsressourcen referenziert und eine Typzuordnung vorgenommen werden. Letztere wurde in der Abbildung vernachlässigt. Diese Möglichkeit besteht aber nicht sofort bei allen Elementen und somit ist häufig eine

Nachbearbeitung von Hand notwendig. Mit dieser automatisierten Vorgehensweise lässt sich die Zeit zur Erstellung der Topics jedoch drastisch verkürzen.

Auf die aufgetretenen Probleme soll in einem späteren Abschnitt eingegangen werden.

### **5.4.5 Schritt 4: Zuordnen der Informationsressourcen**

Nach dem Aufstellen der Topics werden jetzt die Informationsressourcen den Topics zugeordnet. Bei der Zuordnung ist vor allem auf die korrekte Festlegung des Typs der Ressource zu achten. Solche Typen können zum Beispiel eine Abbildung oder auch ein Quellenverweis sein. Es ist dabei auf die Vereinbarungen im Topic Map Template zu achten.

### **5.4.6 Schritt 5: Aufstellen der Assoziationen**

Gemäß dem Topic Map Template werden jetzt die Verbindungen zwischen den einzelnen Topics aufgebaut. Werden bei der Analyse des Dokumentes weitere Assoziationstypen ausgemacht, so empfiehlt sich eine Bewertung nach der Bedeutung für die Topic Map. Dadurch soll der von Horn angesprochenen Gefahr des „Overlinkings“ bzw. „Mislinkings“ entgegengewirkt werden (vgl. [Horn89], S.42). Zum Aufstellen der Assoziationen gehört zudem noch die Vergabe der Rolleninformation, welche die einzelnen Topics in der Assoziation einnehmen.

### **5.4.7 Schritt 6: Abgleichen mit der Zielstellung**

Abschließend wird die entstandene Topic Map auf ihre Tauglichkeit zur Lösung der im ersten Schritt formulierten Problemstellungen untersucht. Dazu können auch die im Dokument selbst vorhandenen Fragen (Element „Frage“) und Lernziele mit herangezogen werden. Treten hierbei Differenzen auf, so müssen die vorhandenen „Wissenslücken“ geschlossen werden, indem die fehlenden Topics und Assoziationen ergänzt werden. Werden neue Fragestellungen formuliert, setzt der Vorgang bei Schritt 3 fort.

### **5.4.8 Probleme bei der Umsetzung**

Folgende Probleme traten bei der Umsetzung für das Projekt ELBE auf:

- Strukturierte Dokumente enthielten auch Layoutinformationen.
- Durch grammatikalisch bedingte unterschiedliche Schreibweise wurden mehrere Topics für einen Begriff erzeugt. Dieser Misstand kann durch den Vorgang des Mergings der zwei Topics behoben werden.

- Entgegen der DTD besaßen nicht alle Information Blocks und Maps ein ID-Attribut. Dies hätte die automatische Erstellung der Topic-Skelette qualitativ hochwertiger gemacht.
- Hierarchiestufen von Begriffen müssen durch den Bearbeiter erkannt und festgelegt werden.

Aufgrund dieser Probleme ist ein vollautomatisiertes Erstellen von Topics nur sehr schwer möglich.

Die inhaltlichen Zusammenhänge zwischen einzelnen Begriffen ließen sich nur genauer aus der Papierversion (Ausdruck durch den Webbrowser) ermitteln. Bei der Analyse wurden die Begriffe entsprechend ihrer späteren Verwendung markiert (z.B. Topics mit rot, Assoziationen mit grün, Typen (für das Template) mit gelb). Auf Basis dieser Markierungen wurde dann die Topic Map erstellt.

### **5.5 Offene Fragen im Vorgehensmodell**

Das oben vorgestellte Vorgehensmodell stellt nur einen Vorschlag des Autors dar und bedarf noch weiterer Nacharbeit. Es ist allgemein gehalten, um unterschiedlichen Ausgangssituationen gerecht zu werden. Zu den nicht angesprochenen Themen gehören unter anderem:

#### **5.5.1 Zuordnung von Gültigkeitsbereichen**

Dieser Punkt besitzt eine hohe Bedeutung bei der Zusammenfassung mehrerer Topic Maps: Je größer die resultierende Topic Map ist, desto wichtiger wird eine genaue Zuordnung von Gültigkeitsbereichen zu den einzelnen Topics. Die genaue Klassifikation verhindert eine zu hohe Trefferquote bei Suchanfragen. Dabei ist noch einmal zu verdeutlichen, dass der Scope (Gültigkeitsbereich eines Elementes (topic, topname, ...)) der DTD nicht zu verwechseln ist mit dem „subject“ (dem Thema, über das ein Topic „spricht“) eines Topics. Zur Angabe des „Themas“ eines Topics werden für XTM die erwähnten published subject indicator (PSI) verwendet. ISO 13250 nutzt zum gleichen Zweck das Attribut „identifier“ im Element „topic“. Aufgrund dieser Trennung sollten auch verschiedene Konzepte zur Lösung verwendet werden. Zur Angabe eines Gültigkeitsbereiches ist nach Ansicht des Autors auch ein Namensraum-Konzept wie in XML verwendbar. Der Nutzen der Angabe eines Gültigkeitsbereiches lässt sich an dem oben vorgeschlagenen Diplomarbeitsverzeichnis noch einmal verdeutlichen.

Mehrere Diplomarbeiten unterschiedlicher Fakultäten verwenden die gleichen Begrifflichkeiten (z.B. den Begriff Ontologie). Ein Nutzer des Verzeichnisses möchte jetzt alle Arbeiten mit geisteswissenschaftlichem Hintergrund suchen. Dazu muss er in der Lage

sein, seinen Suchraum auf Topics vom Typ „Diplomarbeit“ mit dem Scope „Fakultät Geisteswissenschaften“ einzuschränken.

Je größer der Rahmen eines solchen Kataloges ist (Lehrstuhl, Fakultät, Uni,...) umso deutlicher müssen sich die Gültigkeitsbereiche einschränken lassen. Im Vorgehensmodell wird aber die Vergabe solcher Einschränkungen noch nicht berücksichtigt.

### **5.5.2 Vorgehensweise bei der Modellierung von Assoziationen**

Es fehlt eine detaillierte Beschreibung des Vorgehens bei der Modellierung von Assoziationen in der Topic Map. Prinzipiell sind hier wieder zwei Möglichkeiten durchführbar. Zum ersten werden alle gleichnamigen Assoziationen (d.h. alle Verbindungen gleichen Typs) der unterschiedlichen Topics eingearbeitet. Diese Typen sind durch das Aufstellen der Ontologie bekannt. Nachteilig ist ein solches Vorgehen, weil bei einer hohen Anzahl von Topics mit vielen Assoziationen leicht ein Topic übersehen werden kann. Die Fehlersuche gestaltet sich dann sehr aufwendig. Die zweite Möglichkeit besteht darin, alle Assoziationen eines Topics auf einmal in der Topic Map abzubilden. Der Nachteil dieser Vorgehensweise liegt darin begründet, dass eine Assoziation immer eine Verbindung zwischen zwei oder mehreren Topics herstellt. Somit muss dann für den Wissensmodellierer erkennbar sein, ob zum aktuell bearbeiteten Topic eine Assoziation schon modelliert wurde.

Weiterhin wird keine Aussage vorgenommen, wie nichtsymmetrische Assoziationen abgebildet werden sollen. Hierbei ändert sich die Bedeutung der Assoziation, wenn die Betrachtungsweise geändert wird. Die Angabe solcher zusätzlichen Informationen muss durch die Topic Map Engine unterstützt werden.

### **5.5.3 Integration des Wirtschaftlichkeitsprinzips**

Das angesprochene Wirtschaftlichkeitsprinzip ist noch nicht integriert. Als Grundlage für die Bewertung der Kosten müssen dazu die Erfahrungen aus mehreren Projekten vorliegen. Zudem lässt das Vorgehensmodell unterschiedliche konkrete Vorgehensweisen (z.B. bei der Erstellung der Topics im Projekt ELBE) zu. Dies kann wieder zu unterschiedlichen Kosten führen. So sinken zum Beispiel die Kosten für die manuelle Erstellung eines Topics, wenn ein strukturierter Datenbestand vorliegt, mit dessen Hilfe sich die Topics durch den Einsatz von XSL/T auf automatischem Wege erzeugen lassen. Hierbei treten aber Kosten für die Erstellung der Transformationsskripte auf (Entwicklungszeit, evtl. Softwarekosten). Bei Folgeprojekten sinken aber auch diese Kosten, wenn die Transformationen wiederverwendet werden können.

Aufgrund dieser vielfältigen Kombinationsmöglichkeiten ist eine konkrete Bewertungsvorschrift noch nicht eindeutig auszumachen.

## 6 Zusammenfassung

Topic Maps dienen zur Abbildung von Wissensstrukturen in einem für einen Computer lesbaren und zusätzlich auswertbaren Format. Sie sind gekennzeichnet durch eine Trennung der Wissensschicht (mit den Bestandteilen Topics und Assoziationen) und dem Informationsressourcenpool. Aufgrund des Einsatzes der Basistechnologie XML bzw. SGML wird das zur breiten Akzeptanz des Standards notwendige Kriterium der Plattformunabhängigkeit sichergestellt.

In der Arbeit wurden einige wichtige Einsatzmöglichkeiten angesprochen. Es wurden dabei aber auch Probleme und Konsequenzen aufgezeigt, die sich durch den Einsatz der neuen Technologie ergeben. Hauptprobleme liegen auf dem Gebiet der noch nicht in ausreichendem Maße verfügbaren bzw. fehlenden Software. Die bislang vorhandenen Topic Map Engines sollten in weiteren Arbeiten miteinander verglichen und bewertet werden. Weiterhin müssen geeignete Redaktionswerkzeuge und Browserprogramme geschaffen werden. Ein ebenfalls fehlender Punkt ist die Schaffung einer standardisierten Anfragesprache – es gibt mit der TMQL-Arbeitsgruppe bereits Initiativen zur Entwicklung eines solchen Konstruktes. Die Verfügbarkeit der angesprochenen Programme ist Voraussetzung, um die breite Durchsetzung des Standards zu ermöglichen. Zuarbeiten sind auch von anderen Forschungsgebieten notwendig. Dies betrifft vor allen Dingen das Ontology- und Knowledge Engineering, um detaillierte Herangehensweisen für die Modellierung der Wissensstrukturen zu schaffen. Aber auch Ansätze zum strukturierten Erstellen von Dokumenten, die im Informationsressourcenpool verwendet werden, sollten allgemein berücksichtigt werden, um die Qualität und Nutzbarkeit dieser Dokumente weiter steigern zu können. Die Integration der Information Mapping® Methode wurde exemplarisch für das Projekt ELBE durchgeführt.

Zusammenfassend lässt sich erkennen, dass die Entwicklung noch in den Kinderschuhen steckt. Eine Weiterentwicklung ist also unbedingt voranzutreiben. Der vielversprechendste Ansatz existiert derzeit mit XTM. Wichtig ist auch die Ausarbeitung weiterer konkreter Anwendungsszenarien.

Das größte Anwendungspotential liegt nach Ansicht des Autors im Einsatz von Topic Maps in Themenkatalogen und Suchmaschinen im Internet. Hier sollten aber die vorhandenen Synergieeffekte mit anderen Standards vertieft ausgenutzt werden. Es sei hier noch einmal auf die Möglichkeiten einer Bildbeschreibungssprache hingewiesen, welche die Chancen des Internets auch für blinde und sehbehinderte Menschen besser erschließen kann.

Nicht zuletzt durch neu entstehende Berufsbilder haben Topic Maps das Potential, die Informations- und Wissensgesellschaft nachhaltig zu prägen.

## Abkürzungen

#PCDATA	:	Parsed Charakter Data (vom Parser erkannte Zeichenfolgen)
ADO	:	ActiveX Data Objects
API	:	Application Programming Interface
ASP	:	Active Server Pages
CD-ROM	:	Compact Disc – Read Only Memory
COM(+)	:	Component Object Model (erweiterte Version)
DMS	:	Dokumentenmanagementsystem
DOM	:	Document Object Model
DTD	:	Dokumenttypdefinition
DVD	:	Digital Versatile Disc
ELBE	:	Elektronische LehrBuch Edition
HTML	:	Hypertext Markup Language
HyTime	:	Hypermedia/Time-based Structuring Language (ISO 10744)
IEC	:	International Electrotechnical Commission
ISO	:	International Standards Organisation
KI	:	Künstliche Intelligenz
KOS	:	Knowledge Organization System/Scheme
MSDN	:	Microsoft Developer Network
PSI	:	Published Subject Indicator
RDF	:	Resource Description Framework
SAX	:	Simple API for XML
SDK	:	Software Developer Kit
SGML	:	Standard Generalized Markup Language
SQL	:	Structured Query Language
SVG	:	Scalable Vector Graphics

## Abkürzungen

---

TMQL	:	Topic Map Query Language
URI	:	Uniform Resource Identifier
URL	:	Uniform Resource Locator
W3C	:	World Wide Web Consortium
XML	:	Extensible Markup Language
XSL	:	Extensible Stylesheet Language
XSLT	:	XSL Transformations
XTM	:	XML Topic Maps
XTMP	:	XTM Processing Model

## Abbildungen

Abb. 2.1: Nutzung des Attributes "scope" .....	3
Abb. 2.2: Beispiel für den Einsatz von Topic Names und Topic Types .....	4
Abb. 2.3: Umwandlung in eine Superklasse-Subklasse-Relation (in Anlehnung an [Rath99]) .....	5
Abb. 2.4: Vorschlag zur Adressierung nicht-elektronischer Dokumente.....	6
Abb. 2.5: schemenhafte Darstellung der vorgestellten Konzepte ([Pepp99], S.8)..	7
Abb. 2.6: Nutzung von Assoziationen.....	7
Abb. 2.7: Topic Map als portables semantisches Netz ([Pepp00], S.14) .....	8
Abb. 2.8: Verwendung von „Added Themes“ .....	9
Abb. 2.9: Definition des Elementes „topicRef“ ([XTM], §4,2.1) .....	10
Abb. 2.10: Nutzung von XML Base (in Anlehnung an [XML Base],§3) .....	11
Abb. 3.1: Topic Map Template ([Rath00],S.6) .....	19
Abb. 3.2: Baumdarstellung einer Topic Map ([Rath99]) .....	21
Abb. 3.3: Navigation mit TheBrain® .....	22
Abb. 3.4: dreidimensionale Darstellung von Topic Maps ([Gran <sup>+</sup> 00], S.8) .....	22
Abb. 4.1: Elemente der organisatorischen Wissensbasis ([Scho99a], S.3) .....	23
Abb. 4.2: Nutzung von Assoziationen.....	27
Abb. 4.3: Komponenten einer Suchmaschine ([Woll01]).....	30
Abb. 4.4: Zusammenhänge der 3 Objekttypen in RDF ([RDF], §2.1.1.) .....	33
Abb. 4.5: Syntax von RDF ([RDF], §2.2.1) .....	33
Abb. 4.6: Aufbau eines e-Learning-Systems ([Chin00], Folie 12).....	37
Abb. 4.7: Verweise auf unterschiedliche Quellen .....	43
Abb. 5.1: ein einfaches semantisches Netzwerk.....	46
Abb. 5.2: Zuordnung von Eigenschaften über Assoziationen .....	47
Abb. 5.3: "Skelett" eines Topics.....	56

## Tabellen

Tab. 2.1: Gegenüberstellung von ISO13250 und XTM .....	12
Tab. 3.1: Topic Map Engines .....	18
Tab. 4.1: Vergleich RDF mit Topic Maps.....	34
Tab. 5.1: Behandlung von Klartextelementen .....	56

## Literatur

- [Bern00] Berners-Lee, T. Semantic Web,  
<http://www.w3.org/2000/Talks/1206-xml2k-tbl/>
- [Biez+00] Michel Biezunski, Martin Bryan, Steven R. Newcomb, Topic Maps:  
Information Technology – Document Description and Markup Languages, 3  
Dec 1999. <http://www.y12.doe.gov/sgml/sc34/document/0129.pdf> (PDF,  
332K, official text)
- [Biro+00] Biron, P. V., Malhotra, A. XML Schema Part 2: Datatypes – W3C Working  
Draft. World Wide Web Consortium, [http://www.w3.org/TR/2000/WD-  
xmlschema-2-20000407](http://www.w3.org/TR/2000/WD-xmlschema-2-20000407), 07.04.2000
- [Bric+00] Brickley, D., Guha, R.V., Resource Description Framework (RDF) Schema  
Specification 1.0 World Wide Web Consortium, [http://www.w3.org/TR/rdf-  
schema/](http://www.w3.org/TR/rdf-schema/)
- [Brya00] Bryan, M., XML Schema for ISO 13250 Topic Maps,  
<http://www.diffuse.org/TopicMaps/schema.htm>
- [Chin00] China, R., Wissen managen: So können Sie mit e-Learning die Leistungen  
in ihrem Unternehmen verbessern, Foliensatz zum Vortrag vom 14.Juli 2000  
gehalten an der TU Dresden
- [DeRo+00] DeRose, S., Maler, E., Orchard, D., XML Linking Language (XLink)  
Version 1.0, World Wide Web Consortium, <http://www.w3.org/TR/xlink>
- [Endr00] Endres-Niggemeyer, B., Bessere Information durch Zusammenfassen aus  
dem WWW,  
<http://www.ik.fh-hannover.de/ik/person/ben/WEBZusammenfassen.pdf>
- [Endr01] Endres-Niggemeyer, B., Human-style WWW summarization,  
<http://www.ik.fh-hannover.de/ik/person/ben/Human-styleSummaNew.pdf>
- [Ferr01] Ferraiolo, J., Scalable Vektor Graphics (SVG) 1.0 Specification, World  
Wide Web Consortium, <http://www.w3.org/TR/SVG/>
- [Free00] Freese, E., Using Topic Maps for the representation, management and  
discovery of knowledge,  
<http://www.gca.org/papers/xml europe2000/papers/s22-01.html>
- [Gran+00] Grand, Benedicte Le, Soto Michel, Information Management - Topic Maps  
visualisation,  
<http://www.gca.org/papers/xml europe2000/papers/s29-03.html>

- [Gold<sup>+</sup>97] Goldfarb, C. F., Newcomb, Steven R., Kimber, W. Eliot, Newcomb, Peter J. Information processing – Hypermedia/Time-based Structuring Language (HyTime) – 2nd edition, <http://www.ornl.gov/sgml/wg8/docs/n1920/>
- [Heck00] Heckel, R., Namespacekonzepte in XML und Abgrenzung zur Meta-DTD-Modellierung in SGML, Seminararbeit an der TU Dresden, SS 2000, Hauptseminar Dokumentenmanagement
- [Hodg01] Hodge, G., Systems of Knowledge Organization for Digital Libraries: Beyond traditional Authority Files, <http://www.clir.org/pubs/reports/pub91/pub91.pdf>
- [Horn89] Horn, R.E., Mapping Hypertext. The Analysis, Organization and Display of Knowledge for the Next Generation of On-Line Text and Graphics., Information Mapping Inc., Waltham MA.
- [Ksie00] Ksiezyk, Rafal, Answer is just a question (of matching Topic Maps), <http://www.gca.org/papers/xmleurope2000/papers/s22-03.html>
- [Lass<sup>+</sup>99] Lassila, O., Swick, R. R., Resource Description Framework (RDF) Model and Syntax Specification, World Wide Web Consortium, <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>
- [Lobi00] Lobin, H., Informationsmodellierung in XML und SGML. Springer-Verlag, Berlin, Heidelberg, New York 2000
- [Lovi01] Loviscach, J., Die elektronische Uni, in c't – Magazin für Computertechnik Heft 4/2001, S. 108-115
- [Pepp99] Pepper, S., Navigating Haystacks, Discovering Needles, <http://www.ontopia.net/topicmaps/materials/mlangart.pdf>
- [Pepp00] Pepper, S., The TAO of Topic Maps, <http://www.gca.org/papers/xmleurope2000/papers/s11-01.html>
- [Rath99] Rath, Hans Holger, Mit Topic Maps intelligente Informationsnetze aufbauen, in iX – Magazin für professionelle Informationstechnik Heft 12/99, S.149ff
- [Rath00] Rath, Hans Holger, Making Topic Maps more colourful, <http://www.gca.org/papers/xmleurope2000/papers/s29-01.html>
- [Reim91] Reimer, Ulrich, Einführung in die Wissensrepräsentation; Stuttgart, Teubner 1991

- [Sche98] Scheer, August-Wilhelm, Wirtschaftsinformatik: Referenzmodelle für industrielle Geschäftsprozesse, Studienausgabe, 2. durchgesehene Auflage, Berlin, Heidelberg, Springer, 1998
- [Scho99a] Schoop, E., Methoden und Standards des Dokumentenmanagement für den Aufbau einer organisatorischen Wissensbasis, TU Dresden
- [Scho99b] Schoop, E., Structured Documentation: An Approach to Identify an Model Knowledge Components for Learning Purposes, In: Gaul, W., Locarek-Junge, H. (Hrsg.), Studies in Classification, Data Analysis and Knowledge Organization, of the 22nd. annual conference of the GfKI „Classification in the Information Age“, Berlin, Heidelberg, Springer, 1999, S.410-417
- [Scho+99] Schoop, E., Strobel K., SGML als Grundbaustein für das betriebliche Wissensmanagement
- [Scho00] Schoop, E., Informationsmanagement – Wissensmanagement – 7.Semester, In: Wissensmanagement, Vorlesungsskript an der TU Dresden, 2000
- [Stuc+01] Stuckenschmidt, H., Ranze K.C., Herzog, O., Entwicklungskonzepte für die Wiederverwendung unsicherer Wissensmodelle,  
[http://www.informatik.uni-bremen.de/~kcr/paper/TZI/tzi\\_report.html](http://www.informatik.uni-bremen.de/~kcr/paper/TZI/tzi_report.html)
- [Schü98] Schütte, Reinhard, Grundsätze ordnungsmäßiger Referenzmodellierung: Konstruktion konfigurations- und anpassungsfähiger Modelle, Wiesbaden, Gabler, 1998
- [Wich01] Wichmann, A., Aufbau und Techniken von Suchmaschinen für das WWW,  
<http://www-student.cs.uni-bonn.de/~wichmann/writings/webcrawlers/index.html>
- [Woll01] Wollenweber, F., Suchmaschinen im WWW, <http://vsys-www.informatik.uni-hamburg.de/vir/vortraege/VIR2000-Wollenweber-suchmaschinen/sld001.htm>

## Glossar

**Added Theme** : ergänzt den Gültigkeitsbereich (→ scope) eines angegebenen Elementes der Topic Map DTD um ein oder mehrere Topics: S. 2, 8

**Assoziation** : stellt eine getypte Verbindung zwischen zwei oder mehreren Topics dar: S. 4f, 7f, 11, 15f, 27, 32, 44f, 47f, 55

**Base Name** : Namensbezeichnung für ein Topic; wird im Element basename angegeben: S. 3, 15

**Charakteristiken** : im Standard als topic characteristic bezeichnet; dazu gehören der Name, die Vorkommnisse und Rollen in → Assoziationen eines Topics: S. 3, 12

**DTD** : siehe Dokumenttypdefinition

**Display Name** : wird bei der Anzeige des → Topics von der → Topic Map Engine ausgegeben: S. 3

**Dokumenttypdefinition** : beschreibt die Struktur eines SGML- oder XML-Dokumentes; legt die Abfolge der einzelnen Elemente fest: S. 2, 9f, 48

**HyTime** : ISO Standard, der die Adressierung von verschiedenartigen Dokumenten erlaubt: S. 2, 18

**Information Mapping** : Methode zum strukturierten Schreiben von Dokumenten: 25, 28, 50f, 53ff

**Informationsressource** : eine beliebige adressierbare Ressource, die ein → Topic näher beschreibt; ihre Art wird durch den Standard nicht näher spezifiziert: S. 5, 20f, 26, 34, 38, 43, 45, 52

**Informationsressourcenpool** : Gesamtheit aller in der Topic Map verwendeten → Informationsressourcen: S. 20, 26, 38

**Merging** : Zusammenfassen zweier → Topics, die das selbe Thema beschreiben; wird von der → Topic Map Engine durchgeführt: S. 12, 14f

**Meta-DTD** : schreibt die Struktur von abgeleiteten DTD's vor, allerdings ist der konkrete Elementname in der abgeleiteten DTD dabei variabel: S. 2, 9, 13, 48

**Modell** : Abbild der Wirklichkeit, bei dem von unwichtigen Details abstrahiert wird: S. 46ff

**Occurrence** : beschreibt eine → Informationsressource, die als zum → Topic zugehörig betrachtet wird: S. 5ff, 31, 35

**Occurrence Type** : beschreibt den Typ einer → Informationsressource; der Typ selbst verweist auf ein gleichnamiges → Topic: S. 5f

**Ontologie** : Begriffswörterbuch, Wissensorganisationsschema: S. 51

**Ontology Engineering**: methodisches Vorgehen zum Erstellen einer Ontologie: S. 51f

**Parser** : Analyseprogramm für die interne Struktur von Dokumenten (hier XML-Dokumente oder Topic Maps): S. 13

**PSI** : siehe published subject identifier

**published subject indicator** : vordefinierte → Topics, die jederzeit in eigenen Dokumenten als Topic Typen auftreten (z.B. Assoziation, Subklasse, Länderkennungen,...); zwei Topic mit gleichem PSI beschreiben den gleichen Sachverhalt: S. 10, 12, 14, 48

**Relationen** : besitzen Eigenschaften, die für die Verarbeitung von Topic Maps von Bedeutung sind: S. 15f

**Resource Description Framework** : Standard des W3C; definiert die nähere Beschreibung von Informationsressourcen; verfolgt einen ähnlichen Ansatz wie → Topic Maps: S. 29, 32ff

**Scope** : Attribut für beinahe jedes Element der Topic Map → DTD; schränkt den Gültigkeitsbereich des entsprechenden Elementes ein: S. 3f, 28, 49

**Semantisches Netz** : eine Form zur Repräsentation von Wissen: S. 8, 45ff

**Sort Name** : gibt einen Namen an, der bei der Sortierung verwendet wird; ebenfalls ein Element in ISO 13250: S. 3

**Subject** : identifiziert das Thema eines Topics in der Topic Map; gleiche Subjects sind Grundlage für den Vorgang des → Mergings: S. 14

**Topic** : zentrales Element; fasst alle Informationen (Namen, Vorkommnisse) zu dem zugeordneten Begriff zusammen: S. 2ff

**Topic Map** : Dokument das dem Standard ISO 13250 oder XTM entspricht; Dokumentenelement beider Standards: S. 2ff

**Topic Map Browser** : Anzeigeprogramm für Topic Maps; stellt Navigations- und Suchmechanismen bereit und kommuniziert deshalb eng mit der → Topic Map Engine: S. 21f, 39, 41, 44

**Topic Map Engine** : Verarbeitungsprogramm für Topic Maps; verwaltet alle Topics und Assoziationen: S. 10, 13, 17f, 39

**Topic Map Graph** : interne Datenstruktur der → Topic Map Engine; besteht aus Knoten für → Topics, → Assoziationen und → Scopes: S. 14

**Topic Map Template** : beschreibt die grundlegenden Typen von Topics, die in einer Map verwendet werden; bildet eine Vorlage auf höherer Abstraktionsebene: S. 19, 51

**Topic Name** : besteht aus base name, sort name und display name; wird beschrieben durch das Element topname: S. 3, 26f

**Topic Type** : beschreibt den Typ des jeweiligen → Topics im Sinne einer Klasse-Instanz-Relation; die einzelnen Typen sind ihrerseits wieder Topics: S. 4, 19

**Vorgehensmodell** : legt die Abfolge der Anwendung von Techniken bei der Erstellung eines → Modells fest: S. 50ff

**Vorkommnis** : siehe Occurrence

**XLink** : Standard des W3C; definiert Adressierung und Verknüpfung von XML-Dokumenten: S. 5, 9f, 18, 38

**XML Topic Maps** : auf XML basierend; baut auf ISO13250 auf und erweitert Topic Maps um einige zusätzliche Möglichkeiten: S. 9ff, 48

**XML-Schema** : Standard des W3C; beschreibt die Struktur von XML-Dokumenten wie eine DTD, jedoch mit erweiterten Möglichkeiten und im XML-Format: S. 9

**XTM** : siehe XML Topic Maps

**XTM Processing Model** : regelt die Verarbeitung von zu XTM konformen Dokumenten in einer Topic Map Engine: S. 10, 13ff

## Anhang

Anhang 1:	Programmbeschreibung „Topic Map Designer“ .....	A-1
Anhang 2:	Die Klasse CTopic .....	A-7
Anhang 3:	Die Klasse CTMDesignerDoc.....	A-9
Anhang 4:	XSLT-Skripte.....	A-12

## **Anhang 1: Programmbeschreibung „Topic Map Designer“**

### **Konzeption**

Das Programm TMDesigner stellt noch kein voll ausgereiftes Werkzeug zur rechnergestützten Erstellung von Topic Maps dar. Es handelt sich dabei lediglich um einen Prototypen, an dem mögliche Darstellungsmöglichkeiten der Topic Map, einzelner Topics und der Assoziationen verdeutlicht werden sollen. Das Programm macht bei der Bedienung starken Gebrauch der Drag&Drop-Technologie. Außerdem enthält das Programm keine Topic Map Engine in ihrem eigentlich benötigten Umfang. Es wurden der Einfachheit halber nur solche Methoden implementiert, die zur Erstellung, zum Bearbeiten und Löschen von Topics und ihren Assoziationen benötigt werden.

### **Entwicklungswerkzeuge**

Das Programm wurde mit Visual C++ 6.0 der Firma Microsoft entwickelt. Es basiert auf der zugehörigen Klassenbibliothek MFC. Für die Unterstützung des Einlesens von XML-Dateien wurde der Parser MSXML 3.0 ebenfalls von Microsoft eingebunden. Zur Anzeige von HTML und XML-Dateien wird der Internet-Explorer verwendet. Bei der Programmierung wurde auch auf Beispielpprogramme aus dem MSDN (Microsoft Developer Network) zurückgegriffen. Um die Oberflächenprogrammierung zu vereinfachen wurde das Visual Framework – eine frei verfügbare Klassenbibliothek von Zoran M. Todorovic – benutzt. Ein Download ist unter <http://www.scasoftware.com> möglich. Aufgrund der Verwendung der Klassenbibliothek treten unter bisher noch nicht geklärten Umständen Probleme in den Dialogen zur Namensvergabe bei Topics auf. Zur Umgehung dieser Probleme wird die Nutzung der Maus anstelle der Tabulatortaste zum Wechseln der aktiven Eingabelemente empfohlen.

### **Konformität zu ISO 13250**

Aufgrund der dynamischen Entwicklung von XML Topic Maps (XTM) wurden diese bei der Unterstützung verschiedener Topic Map – Strukturen nicht berücksichtigt. Die zugrunde liegende DTD wird nachfolgend vorgestellt. Es werden dabei nicht alle von ISO 13250 genannten Elemente verwendet. So finden zum Beispiel die angesprochenen Facetten zur Modellierung von Eigenschafts-Wert-Paaren keine Berücksichtigung und werden daher auch vom Programm ignoriert, sollten sie in einem Dokument vorhanden sein. Ein weiterer Unterschied zum ISO-Standard ist auch das Fehlen der Attribute für die sogenannten „added Themes“ (Attribut „addthemes“ des Elementes Topicmap). Weiterhin dürfen Anzeige-, Sortier- und Basisname nur einmal in jedem Topicname-Element vorkommen. Alle diese Veränderungen stellen aber nach Ansicht des Autors keine

Verletzung des Standards dar und wurden nur zum Zwecke der Vereinfachung der Implementierung vorgenommen.

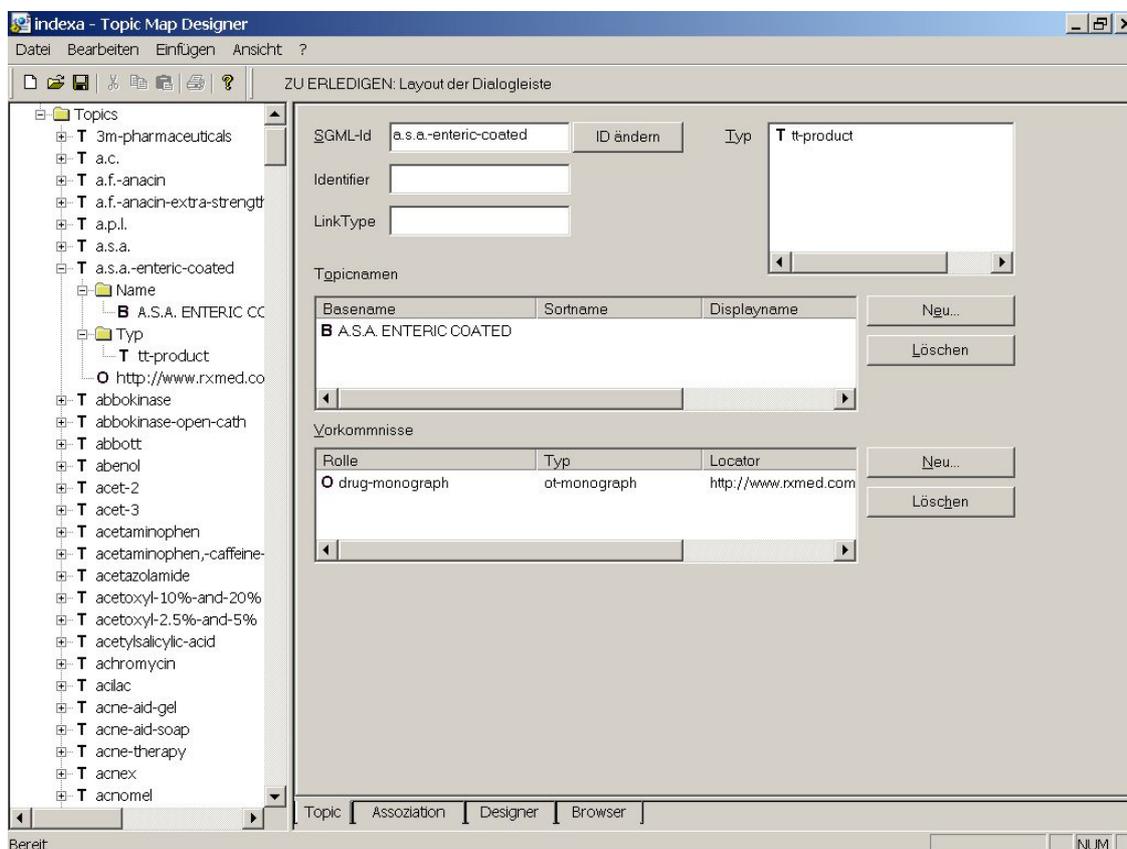
```
<!DOCTYPE topicmap [  
  
  <!ELEMENT topicmap (topic | assoc )* >  
  
  <!ELEMENT topic (topname | occurs)* >  
  <!ATTLIST topic  
    id          ID          #REQUIRED  
    types       CDATA      #IMPLIED  
    linktype    CDATA      #IMPLIED  
    identifier  CDATA      #IMPLIED  
  >  
  
  <!ELEMENT topname (basename, dispname?, sortname?)>  
  
  <!ELEMENT basename (#PCDATA) >  
  
  <!ELEMENT dispname (#PCDATA) >  
  
  <!ELEMENT sortname (#PCDATA) >  
  
  <!ELEMENT occurs EMPTY >  
  <!ATTLIST occurs  
    occr1       CDATA      #REQUIRED  
    type        'locator' #FIXED  
    href        CDATA      #REQUIRED  
  >  
  
  <!ELEMENT assoc (assocr1*) >  
  <!ATTLIST assoc  
    type        CDATA      #IMPLIED  
    linktype    CDATA      #IMPLIED  
  >  
  
  <!ELEMENT assocr1 (#PCDATA) >  
  <!ATTLIST assocr1  
    anchrole    CDATA      #REQUIRED  
    type        CDATA      #REQUIRED  
  >  
]>
```

### Programmarchitektur

Aufgrund der Verwendung der Klassenbibliothek MFC wurde auf die Dokument/View-Architektur zurückgegriffen. Dabei werden alle über die Oberfläche an den Daten vorgenommenen Änderungen an das zentrale Dokument weitergeleitet. Das Dokument (bzw. die Klasse, welche die Dokumentdaten kapselt) benachrichtigt alle zugehörigen Ansichten, die ihrerseits für die korrekte Darstellung der Daten zuständig sind.

## Programmfunktionalität

Nach dem Start des Programms und dem Öffnen einer Topic Map über den Menüpunkt „Datei/Öffnen...“ erscheint die folgende Nutzoberfläche.



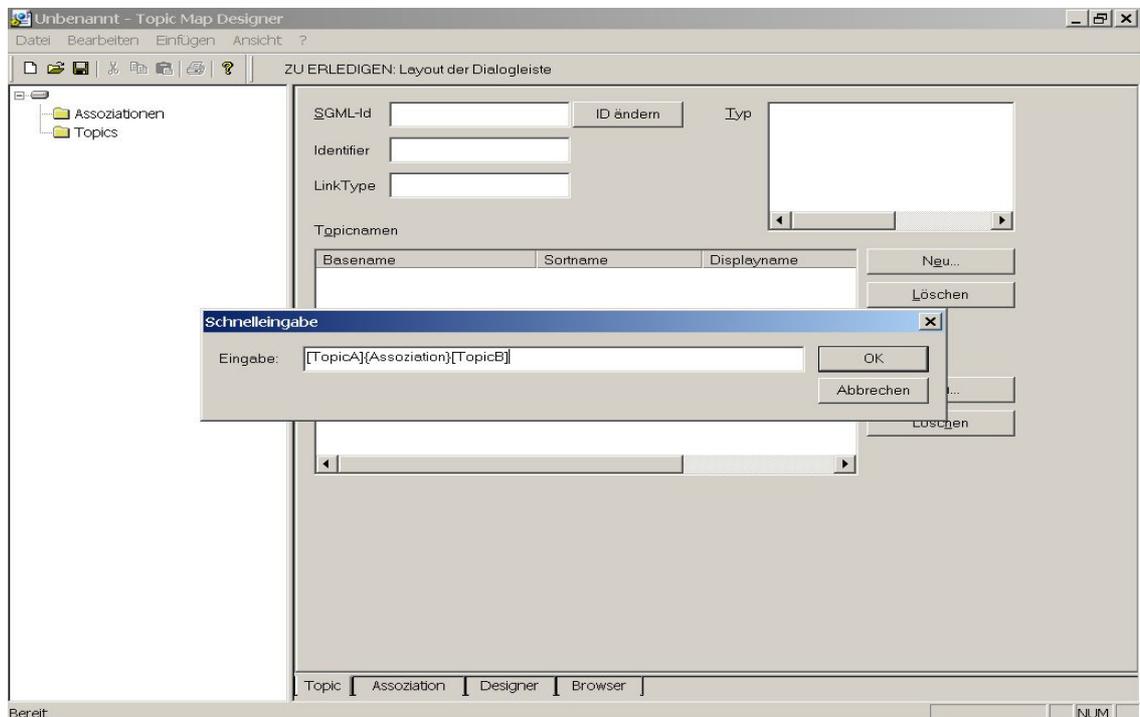
Bei der geöffneten Datei handelt es sich hier um die Topic Map zum Thema Medikamente. Sie wurde aus den Datenbeständen unter <http://www.rxmed.com> generiert und dient lediglich der Veranschaulichung. Der eigentliche Autor, der die Umwandlung aus dem HTML-Format vornahm, ist dem Autor dieser Arbeit nicht bekannt. Auch der Link zum Download konnte nicht mehr gefunden werden.

Am linken Rand befindet sich die Darstellung sämtlicher Topics und (im Bild nicht sichtbar) Assoziationen. Mittels Drag&Drop (siehe Abbildung) kann ein Topic aus der Baumdarstellung in den rechten Bildschirmbereich gezogen werden. Damit wird das Topic für die Bearbeitung aktiviert. Alle vorgenommenen Änderungen werden sofort in die programminterne Struktur übernommen und beim Speichern festgeschrieben. Über die Schalter „Neu“ und „Löschen“ werden die Namen und Vorkommen der Topics angelegt bzw. gelöscht. Eine Änderung ist über einen Doppelklick auf das gewünschte Element in der Liste möglich. Die Zuordnung eines Topicstyps wird ebenfalls per Drag&Drop realisiert. Dazu wird das gewünschte Topic einfach aus der Baumdarstellung in die Typliste gezogen. Zum Löschen wird der entsprechende Eintrag ausgewählt und im Kontextmenü (rechte Maustaste) die Option „Löschen“ gewählt. Der Autor ist sich einiger

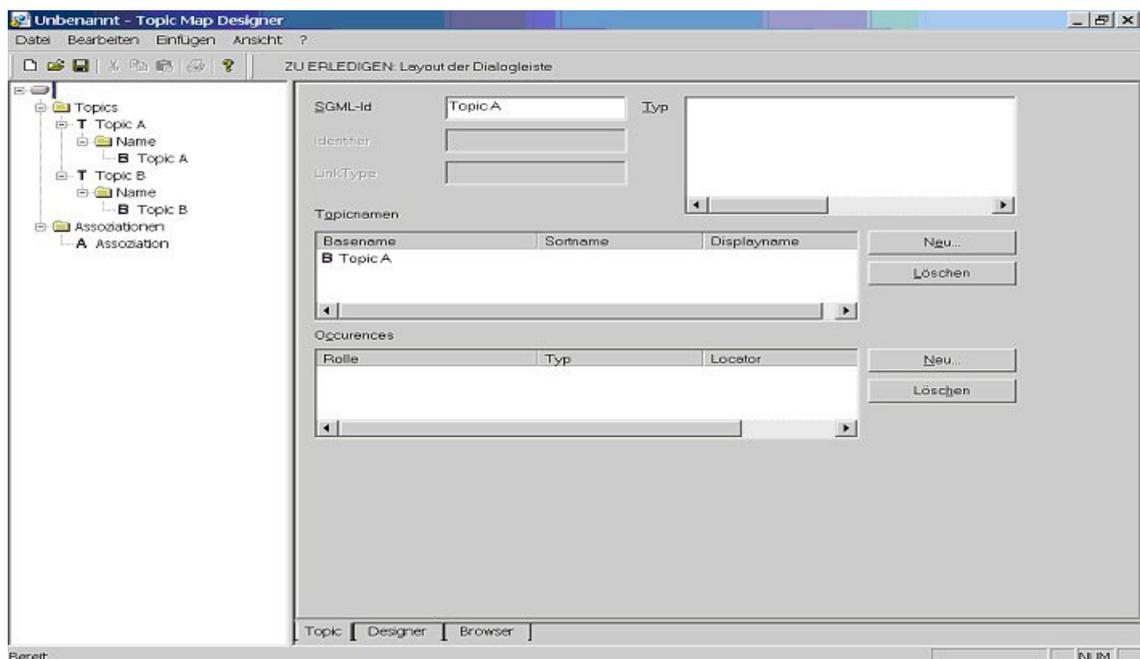
## Anhang

Inkonsistenzen in der Bedienung des Programms bewusst. Die umgesetzte Bedienung soll aber auch nur einige prinzipielle Möglichkeiten aufzeigen.

Ein Anlegen von Topics und Assoziationen ist über den Menüpunkt „Einfügen/Schnelleingabe“ oder jederzeit mit der Taste „Einfüg“ möglich. Es erscheint der nachfolgende Dialog.

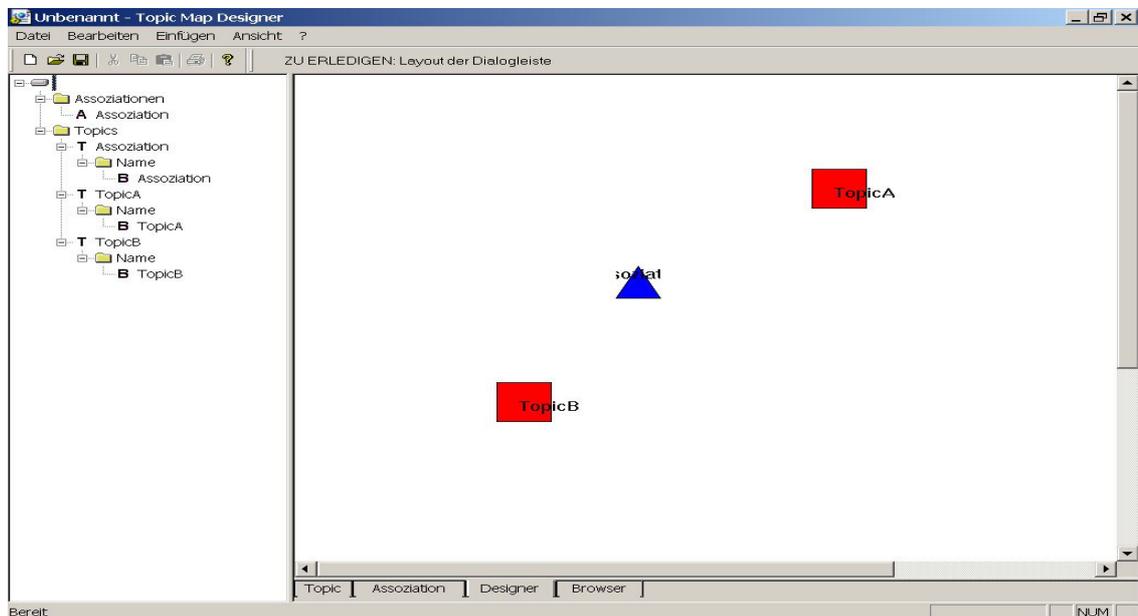


Topics werden durch die Eingabe einer Zeichenkette mit dem Format [Topic] – also ein Name in eckigen Klammer – erzeugt. Für Assoziationen werden geschweifte Klammern gebraucht. Die abgebildete Eingabe würde bei einer anfangs leeren Topic Map zu folgendem Resultat führen.

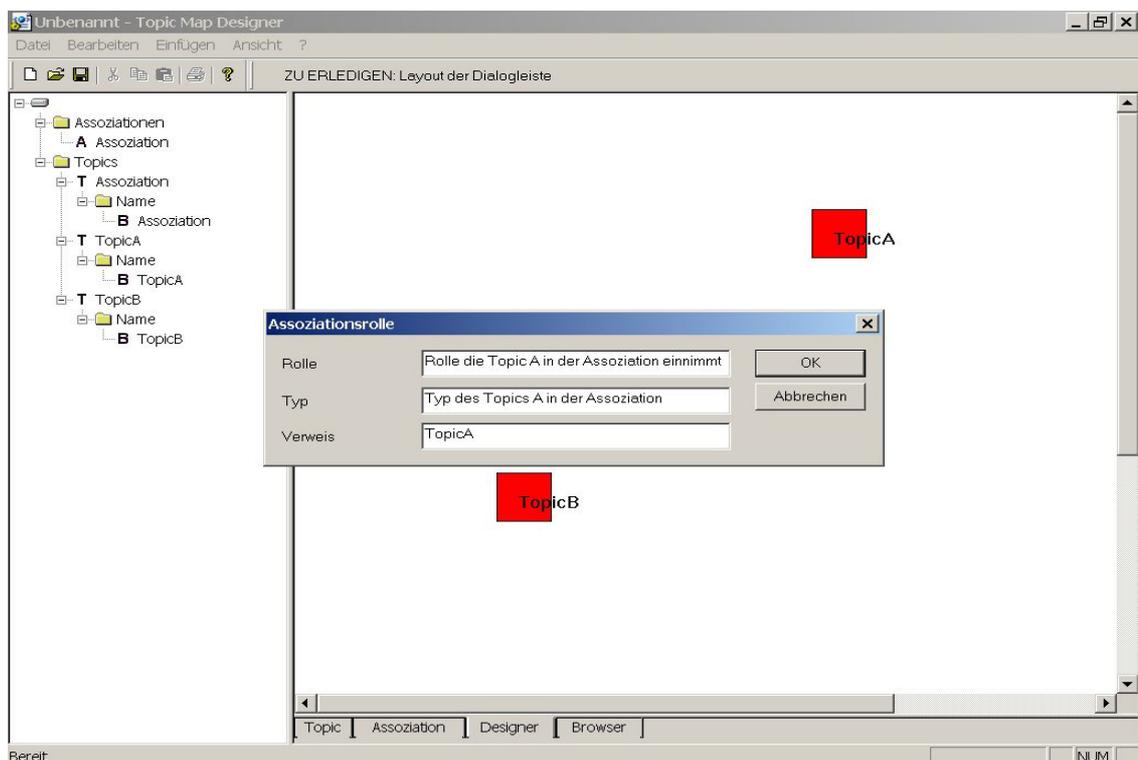


## Anhang

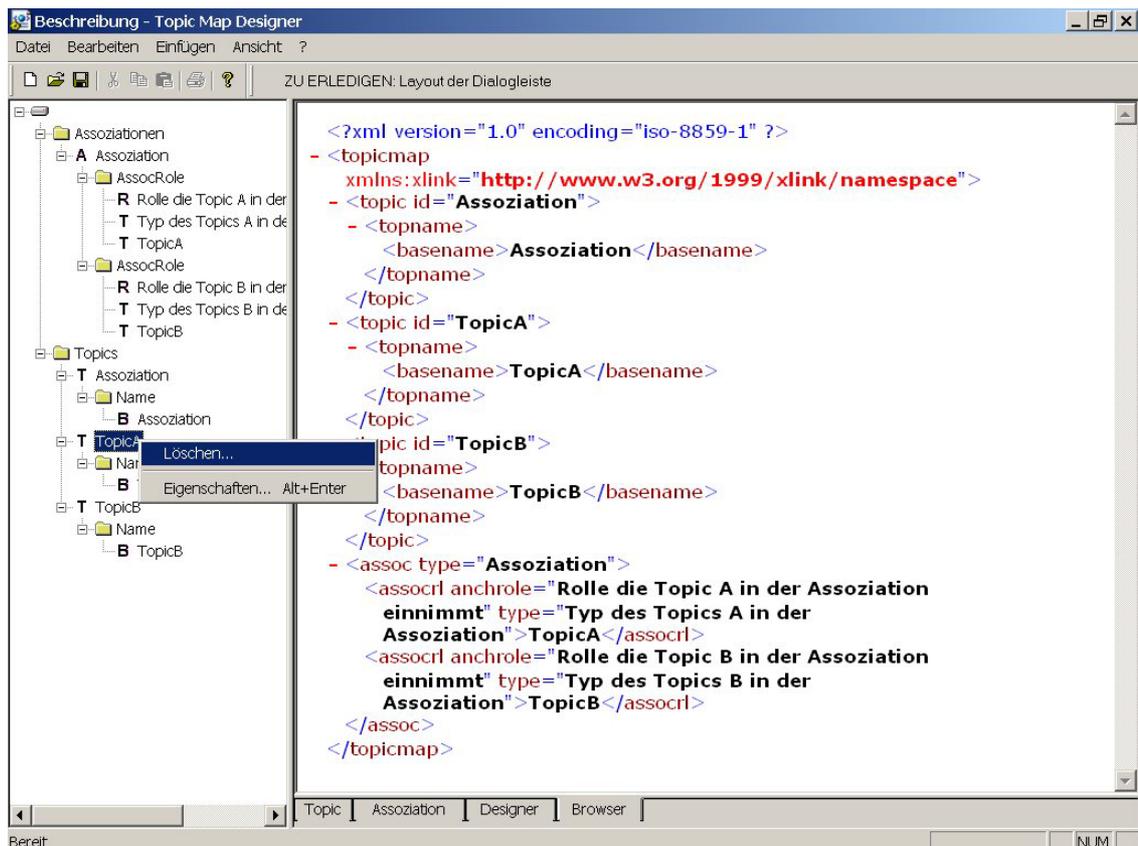
Zum Anlegen von Assoziationen ist ein Wechsel in die Designer-Ansicht notwendig. Auch hier werden über Drag&Drop die benötigten Elemente (Topics und Assoziationen) auf die Designerfläche gezogen.



Topics werden durch rote Vierecke und Assoziationen durch blaue Dreiecke dargestellt. Nicht benötigte Elemente lassen sich durch markieren und anschließendes Betätigen der „Entf“-Taste aus der Designerfläche entfernen. Sie werden dabei nicht aus der Topic Map gelöscht. Zur Verbindung von Topic und Assoziation wird das entsprechende rote Viereck auf das blaue Dreieck gezogen. Es erscheint dann der Dialog zur Eingabe der weiteren Daten. Alternativ ist eine Bearbeitung auch in der Assoziationsansicht möglich.



Die oben angegebenen Daten würden folgende Auswirkungen haben.



Für die Darstellung im Browser wurde die Datei vorher gespeichert und im Browserfenster neu geöffnet. Dieses Vorgehen ist nötig, weil der Browser in erster Linie zum Betrachten der Informationsressourcen (Occurrences) gedacht ist. Wird zu einem Topic ein neues Vorkommen einer Informationsressource angelegt, so wird vom Programm als Verweis automatisch die zuletzt besuchte Seite im Browser angegeben.

Ein Wechsel in die verschiedenen Ansichten ist zu jeder Zeit möglich. Topics und Assoziationen können über das Kontextmenü in der Baumdarstellung aus der Topic Map gelöscht werden.

### Fehlende Funktionen

Für ein vollständiges Autorensystem wäre es denkbar, noch eine Ansicht zum Erstellen von Informationsressourcen (wie Texte, Bilder...) einzufügen. Eine für große Maps bedeutungsvolle, komfortable Suchmöglichkeit fehlt ebenso. Im Programm wurde die Suche über die alphabetische Sortierung etwas erleichtert. Funktionen zum Merging sind nur in einfachem Maße vorhanden. Es fehlen schließlich noch die angesprochenen Möglichkeiten zum Testen der Map sowie umfangreichere Gültigkeitsprüfungen. Weiterhin sollte in nachfolgenden Arbeiten das Attribut „scope“ bei den entsprechenden Elementen mit berücksichtigt werden.

## Anhang 2: Die Klasse CTopic

```
// Topic.h: Schnittstelle für die Klasse CTopic.
//
/////////////////////////////////////////////////////////////////

#include "TopicName.h"
#include "TopicOccur.h"

class CTopic : public CObject
{
public:

    CTopic();
    CTopic(CString SgmlId, CString TopicBaseName);
    CTopic(CTopic &Src);

    DECLARE_DYNCREATE(CTopic)

    CTopic& operator=(CTopic &Src);

    void UnlockData();
    void LockData();
    BOOL IsValid();
    BOOL ToXML(CString &xml);
    void Init();
    void ClearAll();
    BOOL IsOfType(CString Type);
    int GetNamesCount();
    int GetOccursCount();
    int GetTypesCount();

    CString GetLinkType();
    void SetLinkType(CString NewLinkType);
    CString GetIdentifier();
    void SetIdentifier(CString NewIdentifier);
    CString GetSgmlId();
    void SetSgmlId(CString NewSgmlId);

    void AddTopicName(CTopicName *TopicName);
    CTopicName* GetCurrTopicName();
    void DeleteTopicName();
    void DeleteAllTopicNames();
    CTopicName* GetFirstTopicName();
    CTopicName* GetNextTopicName();
    CTopicName* GetPrevTopicName();
    void SetTopicName(int index, CTopicName* name);
    void DeleteTopicName(int index);
    CTopicName* GetTopicName(int index);

    void AddTopicOccur(CTopicOccurrence *TopicOccur);
    CTopicOccurrence* GetCurrTopicOccur();
    void DeleteTopicOccur();
    void DeleteAllTopicOccurs();
    CTopicOccurrence* GetFirstTopicOccur();
    CTopicOccurrence* GetNextTopicOccur();
    CTopicOccurrence* GetPrevTopicOccur();
```

```
void SetTopicOccur(int index, CTopicOccurrence* occur);
void DeleteTopicOccur(int index);
CTopicOccurrence* GetTopicOccur(int index);

void AddTopicType(CTopic* TopicType);
void AddTopicType(CString TopicType);
CString GetCurrTopicType();
void DeleteTopicType();
void DeleteAllTopicTypes();
CString GetFirstTopicType();
CString GetNextTopicType();
CString GetPrevTopicType();
void SetTopicType(int index, CString Type);
void DeleteTopicType(int index);
CString GetTopicType(int index);

virtual ~CTopic();

protected:
    CString          m_SgmlId;           // ID im SGML-Dokument
    CString          m_Identifier;       // „Thema“ (PSI)
    CString          m_LinkType;

    CSpecialObList  m_TopicNames;       // Namensliste
    CTopicName      *m_CurrTopicName;   // aktuelles Element
    POSITION          m_posTopicName;     // zum Durchlaufen
    CSpecialObList  m_TopicOccurs;      // Vorkommnisse
    CTopicOccurrence *m_CurrTopicOccur;
    POSITION          m_posTopicOccur;
    CSpecialStringList m_TopicTypes;    // Typen
    CString          m_CurrTopicType;
    POSITION          m_posTopicType;
};
```

Die Klasse CTopic stellt eine zentrale Klasse im Programm dar. Sie enthält als Attribute die entsprechenden Einträge aus dem ISO-Standard. Die Klasse besitzt Methoden zum Bearbeiten der einzelnen Attribute. Eine weitere wichtige Methode stellt ToXML() dar. Sie liefert ausgehend vom aktuellen Zustand eines Objektes die entsprechende Zeichenkette im XML-Format – entsprechend der oben vorgestellten DTD. Die Methode IsValid() untersucht, ob mindestens ein TopicName vergeben wurde und liefert einen entsprechenden Wert zurück.

Die Klassen CSpecialStringList und CSpecialObList sind von CStringList und CObList abgeleitet und definieren jeweils einen Zuweisungsoperator, der auch die in der Liste enthaltenen Elemente mit übernimmt. Dadurch vereinfacht sich die Programmierung.

Da die anderen Klassen der Topic Map Engine auf den gleichen Prinzipien beruhen, wird auf eine Darstellung dieser Schnittstellen hier verzichtet.

## Anhang 3: Die Klasse CTMDesignerDoc

```
// TMDesignerDoc.h : Schnittstelle der Klasse CTMDesignerDoc
////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////

#include "WidgBase.h"
#include "Topic.h"
#include "TopicName.h"
#include "TopicOccur.h"
#include "Association.h"
#include "AssocRole.h"

// Ereignisse für Views
#define UHINT_UPDATEALL 1 // neues Dokument geladen o.ä.
#define UHINT_ADDTOPIC 2 // neues Topic zugefügt
#define UHINT_DELETETOPIC 3 // Topic gelöscht
#define UHINT_UPDATETOPIC 4 // Topic geändert
#define UHINT_ADDASSOC 5
#define UHINT_DELETEASSOC 6
#define UHINT_UPDATEASSOC 7
#define UHINT_RECALC_LAYOUT 8

using namespace MSXML2;

class CTMDesignerDoc : public CDocument, public
SAXContentHandlerImpl
{
protected: // Nur aus Serialisierung erzeugen
    CTMDesignerDoc();
    DECLARE_DYNCREATE(CTMDesignerDoc)

    // Attribute für SAX-Parser
public:
    CTopic *m_pcurrTopic; // aktuelles Topic
    CAssociation *m_pcurrAssoc;
    CTopicOccurrence *m_pcurrOccur;
    CTopicName *m_pcurrTName;
    CAssocRole *m_pcurrAssocRole;

    //{AFX_VIRTUAL(CTMDesignerDoc)
public:
    virtual BOOL OnNewDocument();
    virtual void Serialize(CArchive& ar);
    virtual BOOL OnOpenDocument(LPCTSTR lpszPathName);
    virtual BOOL OnSaveDocument(LPCTSTR lpszPathName);
    //}AFX_VIRTUAL

    // Implementierung
public:

    int GetAssocCount();
    int GetTopicCount();
    void ClearAll();

    void AddTopic(CString id, CTopic *topic, BOOL &merged);
    void UpdateTopic(CString id, CTopic *topic);
    void DeleteAllTopics();
};
```

```
void DeleteTopic(CString id);
CTopic* FindTopic(CString id);
CTopic* GetFirstTopic();
CTopic* GetNextTopic();
void MergeTopics(CTopic* old, CTopic* topic,
                BOOL Manuell=FALSE);

void AddAssoc(CAssociation *assoc);
void UpdateAssoc(CAssociation *assoc);
void DeleteAssoc(CAssociation *assoc);
void DeleteAllAssocs();
CAssociation* GetFirstAssoc();
CAssociation* GetNextAssoc();

void SetMapDescription(CString NewDescription);
void SetTopicMapName(CString NewName);
CString GetMapDescription();
CString GetTopicMapName();
CString GetLastVisitedURL();
void SetLastVisitedURL(CString url);

// Widget bezogene Routinen
int AddWidget(int Type, LPCRECT pRect, CString Id,
             LPARAM data=NULL);
int AddWidget(CWidget* widget);
CWidget* GetWidget(int index);
BOOL DeleteWidget(int index);
int GetWidgetCount();
CWidget* FindWidgetById(CString id, int Type=0);
void MoveWidgets(int x, int y);

virtual ~CTMDesignerDoc();

protected:
    CString m_LastVisitedURL; // zuletzt besuchte Webseite
    CString m_TopicMapName; // Name der TopicMap
    CString m_TopicMapDescription; // Beschreibung
    CMapTopics m_Topics; // Liste aller Topics; es
                        // kann über die SgmlId gesucht werden
    POSITION m_posTopic;
    CMapAssocs m_Assocs;
    POSITION m_posAssoc;
    CObArray m_Widgets; // Widgets in der Designer-View

// Generierte Message-Map-Funktionen
protected:
    //{AFX_MSG(CTMDesignerDoc)
    afx_msg void OnInsertFast();
    afx_msg void OnFileMerge();
    afx_msg void OnUpdateFileMerge(CCmdUI* pCmdUI);
    afx_msg void OnFileProperties();
    //}AFX_MSG
    DECLARE_MESSAGE_MAP()

virtual HRESULT STDMETHODCALLTYPE startElement(
    /* [in] */ wchar_t __RPC_FAR *pwchNamespaceUri,
    /* [in] */ int cchNamespaceUri,
    /* [in] */ wchar_t __RPC_FAR *pwchLocalName,
```

```
        /* [in] */ int cchLocalName,
        /* [in] */ wchar_t __RPC_FAR *pwchQName,
        /* [in] */ int cchQName,
        /* [in] */ ISAXAttributes __RPC_FAR *pAttributes);
virtual HRESULT STDMETHODCALLTYPE endElement(
        /* [in] */ wchar_t __RPC_FAR *pwchNamespaceUri,
        /* [in] */ int cchNamespaceUri,
        /* [in] */ wchar_t __RPC_FAR *pwchLocalName,
        /* [in] */ int cchLocalName,
        /* [in] */ wchar_t __RPC_FAR *pwchQName,
        /* [in] */ int cchQName);
virtual HRESULT STDMETHODCALLTYPE startDocument();
virtual HRESULT STDMETHODCALLTYPE endDocument();
virtual HRESULT STDMETHODCALLTYPE characters(
        /* [in] */ wchar_t __RPC_FAR *pwchChars,
        /* [in] */ int cchChars);
    int currElement; // aktuelles Element im SAX-Parser
};
////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
```

Aus dem Umfang ist erkennbar, dass es sich hier um die wichtigste Klasse des Programms handelt. Sie enthält Methoden, um zusätzliche Informationen zur Topic Map zu speichern, und kümmert sich um das Laden und Speichern der Dateien.

Um den Ladevorgang einer XML-Datei zu beschleunigen, sollte die Nutzung des Document Object Models (DOM) vermieden werden. Stattdessen wird die SAX-Schnittstelle (Simple API for XML) des Parsers genutzt. Dabei wird für jedes vom Parser erkannte Element die Methode `startElement()` bzw. `endElement()` aufgerufen. Aus dem übergebenen Namen ist für das Programm ersichtlich, um welches Element (Topic, TopName,...) es sich handelt. Intern werden dann die zugehörigen Objekte erzeugt und am Ende über die Methoden `AddTopic()` bzw. `AddAssociation()` im Dokument eingefügt.

Beim Speichern wird lediglich für jedes Topic bzw. Assoziation die Methode `ToXML()` aufgerufen und die zurückgelieferte Zeichenkette in die Datei geschrieben.

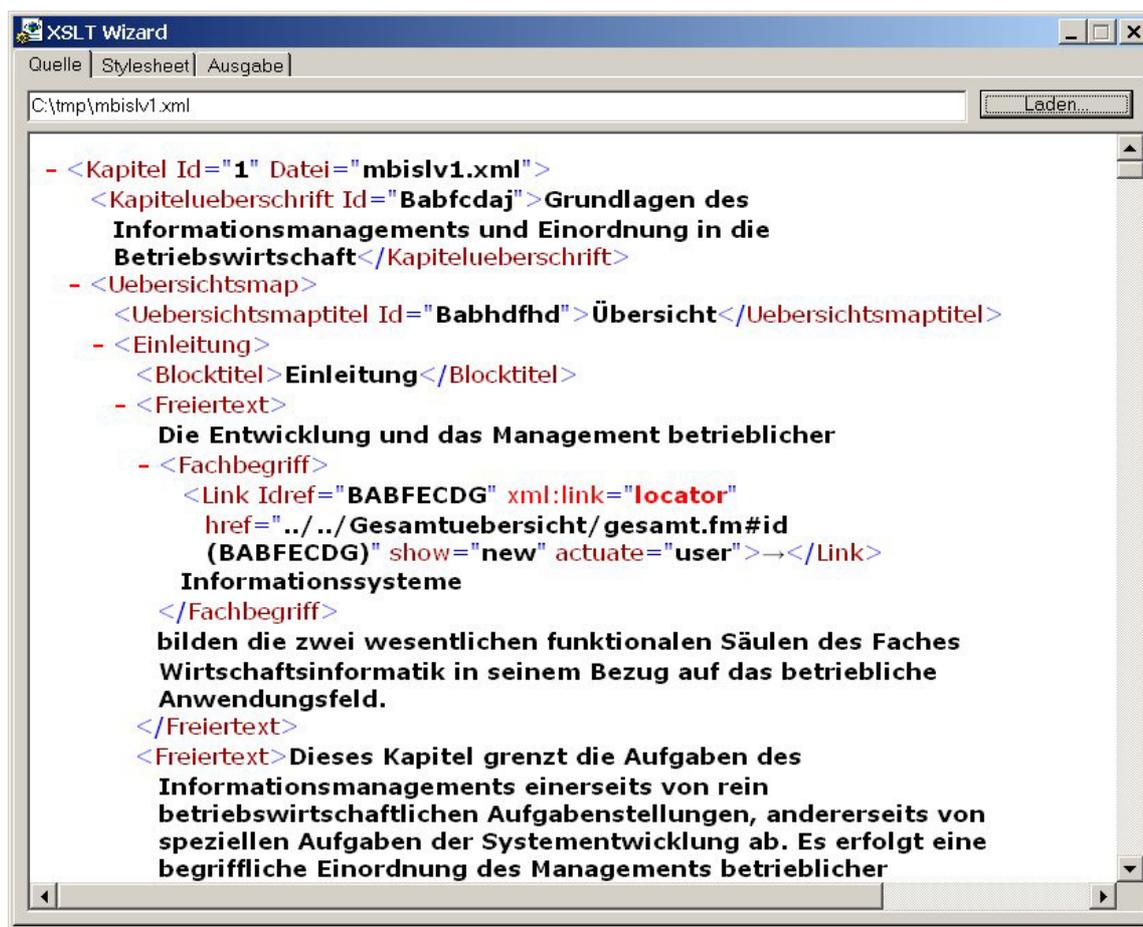
Als Widgets werden die Darstellungsobjekte im Designerfenster bezeichnet (die roten Vierecke bzw. blauen Dreiecke). Ihre Verwaltung obliegt ebenfalls dem zentralen Dokument.

Als interne Datenstrukturen werden für die Topics Strukturen verwendet, die zu jedem Topic eine zugehörige ID (im Programm die eindeutige SGML-ID) und einen Zeiger auf das konkrete Objekt im Speicher verweisen. Für Assoziationen wird eine einfache Liste genutzt.

In der Methode `MergeTopics()` wird ein Merging entsprechend des im Abschnitt 3.1.2.1. vorgestellten Subject-based Merging vorgenommen. Für tiefergehende Details verweist der Autor auf den Quelltext des Programms.

## Anhang 4: XSLT-Skripte

Zum Erzeugen der XSLT-Skripte wurde ein selbst entwickeltes Programm zu Hilfe genommen. Wie der Designer nutzt es die Funktionen des Internet-Explorers und des XML-Parsers von Microsoft. Das Programm wurde in Visual Basic erstellt. Die Benutzeroberfläche gestaltet sich wie folgt:



Die Registerkarte „Quelle“ ermöglicht das Laden und die Anzeige des XML-Quelldokumentes. Das Wechseln zur Registerkarte „Stylesheet“ bietet Optionen zum Laden, Bearbeiten und Anwenden der XSL(T)-Skripte auf das jeweils geladene Quelldokument. Vor dem Anwenden der Transformation wird nach dem Namen der Ausgabedatei gefragt. Diese wird nach einer erfolgreichen Umwandlung in der Registerkarte „Ausgabe“ angezeigt.

Der Stylesheet-Editor bietet noch keine Syntaxhervorhebung. Es fehlt eine Unterstützung zur Fehlersuche innerhalb der Skripte. Das Programm dient lediglich der Vereinfachung der einzelnen Arbeitsschritte unter einer einheitlichen Nutzeroberfläche.

```
<?xml version="1.0"?>
<xsl:stylesheet version="1.0"
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:strip-space elements="*" />

  <xsl:template match="/">
    <topicmap
      xmlns:xlink="http://www.w3.org/1999/xlink/namespace">
      <topic>
        <xsl:attribute name="id">
          <xsl:value-of select="/Kapitel/@Id"/>
        </xsl:attribute>
        <xsl:attribute name="types">Kapitel</xsl:attribute>
        <topname>
          <basename>
            <xsl:value-of select="/Kapitel/Kapitelueberschrift"/>
          </basename>
        </topname>
      </topic>
      <xsl:apply-templates select="Kapitel/Uebersichtsmap" />
      <xsl:apply-templates select="Kapitel/Abschnitt" />
    </topicmap>

  </xsl:template>
  <xsl:template match="Uebersichtsmap">
    <topic>
      <xsl:attribute name="id">
        <xsl:value-of select="Uebersichtsmaptitel/@Id"/>
      </xsl:attribute>
      <xsl:attribute name="types">Uebersicht</xsl:attribute>
      <topname>
        <basename>
          <xsl:value-of select="Uebersichtsmaptitel"/>
        </basename>
        <dispname>&#220;bersicht des Kapitels
          <xsl:value-of select="//Kapitelueberschrift"/>
        </dispname>
      </topname>
      <occurs>
        <xsl:attribute name="occr1">&#220;bersichtsmap
          </xsl:attribute>
        <xsl:attribute name="xml:type">locator</xsl:attribute>
        <xsl:attribute name="xml:href">
          <xsl:value-of select="/Kapitel/@Datei"/>
        </xsl:attribute>
      </occurs>
    </topic>
  </xsl:template>

  <xsl:template match="Aufzaehlung">
    <xsl:apply-templates select="Aelement"/>
  </xsl:template>

  <xsl:template match="Aelement">
    <xsl:apply-templates select="Fachbegriff"/>
  </xsl:template>
```

```
<xsl:template match="Uebersichtsmap" mode="Abschnitt">
  <topic>
    <xsl:attribute name="id">
      <xsl:value-of select="Uebersichtsmaptitel/@Id"/>
    </xsl:attribute>
    <xsl:attribute name="types">Uebersicht</xsl:attribute>
    <topname>
      <basename>
        <xsl:value-of select="Uebersichtsmaptitel"/>
      </basename>
      <dispname>&#220;bersicht des Abschnitts
        <xsl:value-of select="parent::*/*Abschnittsueberschrift"
          />
      </dispname>
    </topname>
    <occurs>
      <xsl:attribute name="occr1">
        &#220;bersichtsmap
      </xsl:attribute>
      <xsl:attribute name="xml:type">locator</xsl:attribute>
      <xsl:attribute name="xml:href">
        <xsl:value-of select="/Kapitel/@Datei"/>
        #id(<xsl:value-of select="Uebersichtsmaptitel/@Id"/>)
      </xsl:attribute>
    </occurs>
  </topic>
  <xsl:apply-templates select="Schlagworte/Schlagwort"/>
  <xsl:apply-templates select="Literatur"/>
</xsl:template>

<xsl:template match="Abschnitt">
  <topic>
    <xsl:attribute name="id">
      <xsl:value-of select="Abschnittsueberschrift/@Id"/>
    </xsl:attribute>
    <xsl:attribute name="types">Abschnitt</xsl:attribute>
    <topname>
      <basename>
        <xsl:value-of select="Abschnittsueberschrift"/>
      </basename>
    </topname>
  </topic>
  <xsl:apply-templates select="Uebersichtsmap"
    mode="Abschnitt"/>
  <xsl:apply-templates select="Map"/>
</xsl:template>
```

```
<xsl:template match="Schlagwort">
  <topic>
    <xsl:attribute name="id">
      <xsl:value-of select="."/>
    </xsl:attribute>
    <xsl:attribute name="types">
      <xsl:value-of select="name()" />
    </xsl:attribute>
    <topname>
      <basename><xsl:value-of select="."/></basename>
    </topname>
  </topic>
</xsl:template>

<xsl:template match="Map">
  <topic>
    <xsl:attribute name="id">
      <xsl:value-of select="Maptitel/@Id"/>
    </xsl:attribute>
    <xsl:attribute name="types">Map</xsl:attribute>
    <topname>
      <basename><xsl:value-of select="Maptitel"/></basename>
    </topname>
    <occurs>
      <xsl:attribute name="occr1">Beschreibung</xsl:attribute>
      <xsl:attribute name="xml:type">locator</xsl:attribute>
      <xsl:attribute name="xml:href">
        <xsl:value-of select="/Kapitel/@Datei"/>
        #id(<xsl:value-of select="Maptitel/@Id"/>)
      </xsl:attribute>
    </occurs>
  </topic>
  <xsl:apply-templates select="Schlagworte/Schlagwort"/>
  <xsl:apply-templates select="Thema"/>
  <xsl:apply-templates select="Block"/>
</xsl:template>

<xsl:template match="Fachbegriff">
  <topic>
    <xsl:attribute name="id">
      <xsl:value-of select="normalize-space(
        substring(current(),2))"/>
    </xsl:attribute>
    <xsl:attribute name="types"><xsl:value-of select="name()" />
    </xsl:attribute>
    <topname>
      <basename><xsl:value-of select="substring(current(),2)"/>
      </basename>
    </topname>
    <occurs>
      <xsl:attribute name="occr1">siehe</xsl:attribute>
      <xsl:attribute name="xml:type">locator</xsl:attribute>
      <xsl:attribute name="xml:href">
        <xsl:value-of select="Link/@href"/>
      </xsl:attribute>
    </occurs>
  </topic>
</xsl:template>
```

```
<xsl:template match="Freiertext">
  <xsl:apply-templates select="Fachbegriff"/>
</xsl:template>

<xsl:template match="Thema">
  <xsl:apply-templates select="Freiertext"/>
</xsl:template>

<xsl:template match="Block">
  <topic>
    <xsl:attribute name="id">
      <xsl:value-of select="normalize-space(Blocktitel)"/>
    </xsl:attribute>
    <xsl:attribute name="types">Block</xsl:attribute>
    <topname>
      <basename><xsl:value-of select="Blocktitel"/></basename>
    </topname>
    <occurs>
      <xsl:attribute name="occr1">siehe</xsl:attribute>
      <xsl:attribute name="xml:type">locator</xsl:attribute>
      <xsl:attribute name="xml:href">
        <xsl:value-of select="generate-id()"/>
      </xsl:attribute>
    </occurs>
  </topic>
  <xsl:apply-templates select="Freiertext"/>
  <xsl:apply-templates select="Aufzaehlung"/>
</xsl:template>

<xsl:template match="Literatur">
  <xsl:apply-templates select="Literatureintrag"/>
</xsl:template>

<xsl:template match="Literatureintrag">
  <topic>
    <xsl:attribute name="id">
      <xsl:value-of select="Literaturkuerzel/@Id"/>
    </xsl:attribute>
    <xsl:attribute name="types">Quelle</xsl:attribute>
    <topname>
      <basename>
        <xsl:value-of select="normalize-space(Literaturkuerzel)"/>
      </basename>
      <dispname>
        <xsl:value-of select="normalize-space(Literatur.lang)"/>
      </dispname>
    </topname>
  </topic>
</xsl:template>
</xsl:stylesheet>
```

## ERKLÄRUNG

Ich erkläre, dass ich die Arbeit selbständig verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und die diesen Quellen und Hilfsmitteln wörtlich oder sinngemäß entnommenen Ausführungen als solche kenntlich gemacht habe.

Auerbach, den 19. April 2001

Unterschrift mit Vornamen und Nachnamen