



Topic Maps im industriellen Einsatz Erste Erfahrungen

Stefan Mintert

Sörman Information GmbH

Zusammenfassung Dieser Artikel beschreibt in einem Erfahrungsbericht den Einsatz von Topic Maps in einem industriellen Pilotprojekt. Neben der Beschreibung des Projekts und dessen Realisierung wird auch eine Bewertung der gewählten Techniken abgegeben.

1 Einleitung

Der Begriff der Topic Map erfährt seit circa zwei Jahren eine moderate, aber deutlich wachsende Beachtung in der Fachöffentlichkeit. Viele Anwendungen zeichnen sich aber durch ihren Demonstrationscharakter aus oder sind akademischer Natur. In diesem Beitrag geht es um ein Pilotprojekt in der Luftfahrtindustrie, das der Verfasser im Rahmen seiner Tätigkeit für die Sörman Information GmbH durchgeführt hat. Nach einem kurzen Überblick über Topic Maps folgt eine ausführliche Projektbeschreibung.

1.1 Topic Maps

Topic Maps dienen dazu, Informationen über Beziehungen, die zwischen sogenannten Topics bestehen, auszudrücken. Als "Topic" wird dabei irgendein *Ding* verstanden, über das die Topic Map eine Aussage macht. Topics können typisiert sein (Topic-Typ).

Ein gern gewähltes Beispiel kennt die Topic-Typen *Stadt* und *Person*. Es wird dann weitere Topics geben, etwa *Hamburg*, *Berlin*, *Köln*, *München* (alle vom Typ *Stadt*) sowie konkrete Personen. Zunächst besteht keine Verbindung zwischen den Topics. Solche Verbindungen werden "Associations" genannte. Beispielsweise kommen die folgenden in Frage: *geboren-in*, *gestorben-in*, *größer-als*. Die ersten beiden Associations bestehen zwischen Personen und Städten, die dritte zwischen Städten.

Topic-Typen, Topics und Associations sind Bestandteile der Topic Map. Ein Bezug zu einem realen Objekt (außerhalb der Topic Map) würde über eine "Occurrence" modelliert werden. Eine Occurrence einer Stadt ist zum Beispiel ein Online-Stadtplan im WWW. Er würde über einen URL in der Occurrence adressiert werden. Occurrences von Personen könnten ihre Lebensläufe im WWW sein; oder auch eine in Buchform erschienene Biographie, die per ISBN (als Occurrence-Adresse) referenziert wird.

Leser, die tiefer in das Thema einsteigen möchten, finden in [TAO] eine lesenswerte Einführung.





1.2 Projektumfeld und Voraussetzungen

Bei der hier diskutierten Anwendung geht es um die Dokumentation von Flugzeugen. Als Spezialfall eines *Aircraft Operating Manual* (AOM) handelt es sich hier um die Flugzeugdokumentation für das Flugpersonal, das *Flight Crew Operating Manual* (FCOM). Das FCOM liegt im SGML-Format vor (Standard Generalized Markup Language, ISO 8879) und besitzt eine in der Luftfahrt übliche Struktur (Dokumenttyp-Definition, DTD). Bei der Struktur handelt es sich im wesentlichen um eine klassische Buchstruktur mit Kapiteln, Abschnitten und so weiter. Die SGML-Elemente tragen jedoch nicht die Namen "Kapitel", "Abschnitt" usw., sondern domänenspezifische Bezeichner wie "Aeroplane-systems".

Aufgrund der Größe von mehreren MBytes pro Dokument, liegt es nahe, FCOMs nicht im Dateisystem abzulegen. Neben Schwierigkeiten mit gängigen SGML-Editoren bei der Handhabung solcher Dokumentgrößen, besteht anwenderseitig in der Regel der Wunsch, nur auf Teilen eines Dokuments zu arbeiten. Damit wird es möglich, dass mehrere Autoren parallel an einem Dokument schreiben.

Zur Datenhaltung wird im vorliegenden Fall die Sörman Content Management Suite (CMS), ein SGML/XML-fähiges System, verwendet, das den gewünschten feingranularen Zugriff und Locking unterstützt.

1.3 Aufgabenbeschreibung

Ausgehend von der zuvor beschriebenden Situation war es das Ziel des Pilotprojekts, das FCOM in einer Form aufzubereiten, die es erlaubt, das Dokument als Basis für eine webgestützte Pilotenschulung zu verwenden. Zu diesem Zweck sollte die klassische, sequentielle Sortierung durch eine geeignetere Navigationsstruktur ergänzt werden.

Neben der stark domänenspezifischen Frage, was eine "geeignete" Navigation ist, sind folgende Randbedingungen wichtig: (a) Weder das Ursprungsdokument noch seine Struktur durften verändert werden. (b) Es sollte keine redundante Datenhaltung eingeführt werden. Die in der Sörman CMS gespeicherten Daten stellen die einzige Dokumentkopie dar.

2 Beschreibung der Lösung

2.1 Lösungsansatz

Zur Lösung der Aufgabe wurde eine Realisierung über Topic Maps gewählt. Die Gründe für diese Entscheidung waren folgende:

- Die Repräsentation von Topic Maps in einer SGML/XML-Form ist die "normale" Darstellung. Dadurch lassen sich vorhandene Werkzeuge und existierendes Know How einsetzen.
- Sollte die Notwendigkeit auftreten, können Topic Maps im vorgegebenen Content Management System in ihrer SGML/XML-Form abgelegt werden.
- Durch den Einsatz von Topic Maps müssen die Ursprungsdaten nicht modifiziert werden. Die Topic Map beschreibt den Zusammenhang der Dinge und verweist per Occurrence auf die Stelle ihres Vorkommens in den Ursprungsdaten.





Zur Umsetzung mit Topic Maps sollte in einem ersten Schritt eine domänenspezifische Ontologie manuell erstellt werden. In der Ontologie stehen die Topic-Typen, die Association-Typen sowie die Occurrence-Typen. Im nachfolgenden Schritt sollte die Topic Map für das FCOM regelbasiert generiert werden. Die darin enthaltenen Topics werden über die in der Ontologie deklarierten Typen typisiert. Als dritte Teilaufgabe sollte dann die Darstellung und Navigation der Topic Map im Web-Browser realisiert werden. Abschließend sollte eine Integration mit der Sörman CMS geschaffen werden, damit (a) Anwender, die das Dokument über die Topic Map navigieren, seinen Inhalt betrachten können und (b) damit die Autoren das Dokument ändern können und aktuelle Daten "live" über die Web-Oberfläche verfügbar sind.

2.2 Lösungsarchitektur

Die Lösung wurde um den zentralen Datenbestand herum entworfen (vgl. Abbildung 1). Im Zentrum steht das Content Repository in dem das FCOM gehalten wird. Über eine direkte Kopplung zwischen SGML-Editor und CMS arbeiten die Verfasser des Dokuments auf diesem Datenbestand.



Aus der CMS heraus wird die Topic Map generiert. Sie dient als Grundlage zur Navigation und stellt per Occurrence eine Verknüpfung zum Inhalt des FCOM her.

2.3 Implementierung

Die Ontologie wurde manuell im XML-Format erstellt. An diesem Prozess waren Personen mit Topic-Map-Know-How sowie Kenner der Luftfahrtdokumentation beteiligt. Es



wurden hier auch erste Regeln formuliert, die später zur automatischen Generierung der Topic Map dienen.

Zur eigentlichen Topic-Map-Generierung wurde XSLT (XSL Transformations) verwendet. XSLT ist eine Sprache zur Beschreibung von Transformationen von einem XML-Format in ein anderes XML-Format. Die Ursprungs-SGML-Daten ließen sich dank eines CMS-API wie XML-Daten behandeln. So war es möglich, die XML-Topic-Map mit XSLT zu erzeugen. Außerdem sorgt diese Vorgehensweise dafür, dass die Daten in der CMS geändert werden können und sich die Topic Map automatisch ableiten lässt.

Für den Aufbau einer Web-Oberfläche wurde die Topic-Map-Software der norwegischen Firma Ontopia verwendet. Ontopia bietet eine Topic-Map-Engine, komfortable APIs zum Aufbau einer Topic Map, sowie Tag-Libraries für den Einsatz in JSPs (Java Server Pages) an. Letzteres stellt die Grundlage für die FCOM-Web-Oberfläche dar: Ontopias Software liest die durch XSLT erzeugte Topic Map ein und zeigt eine auf Tag-Libraries basierende Web-Navigation im Browser an.

Die Web-Schnittstelle zum Content Management System wurde eigens auf Basis der Open-Source-Software "Cocoon" implementiert.

2.4 Projektergebnis

Die im Pilotprojekt implementierte Anwendung lässt sich naturgemäß nur unzureichend in schriftlicher Form vermitteln. Erste öffentliche Demonstrationen fanden im Mai 2002 auf der Internationalen Luftfahrt-Ausstellung (ILA) in Berlin und während der 16. DFN-Arbeitstagung in Düsseldorf statt. Zur besseren Veranschaulichung sollen im folgenden einige Screenshots beitragen.

Es wurden (ohne Anspruch auf Vollständigkeit) zunächst sieben Topic-Typen in der Domäne der FCOMs identifiziert. Dazu zählen "normal Procedures" und "abnormal/Emergency Procedures". Dies sind Handlungen, die in bestimmten "Phases of Flight" durchzuführen sind. Des Weiteren wurden die technischen Systeme ("ATA-System") und ihre "Components" als Topic-Typen festgelegt (vgl. Abbildung 2 und 3).

Diese Topics sind über drei Typen von Associations miteinander verknüpft. Als erste ist die triviale Association zu nennen, die besagt, dass ein technisches System aus seinen Komponenten besteht. In Abbildung 3 ist die Association dargestellt, die eine Prozedur (z.B. "Landen") mit einzelnen Schritten (z.B. "Fahrwerk ausfahren") in Beziehung bringt. Für Kenner der Materie am wertvollsten ist die Association, deren Übersicht in Abbildung 4 zu sehen ist. Sie beschreibt, welche Systemkomponente in welcher Handlung zum Einsatz kommt. Das ist deshalb so interessant, weil das FCOM Abschnitte enthält, die *entweder* system-orientiert *oder* handlungs-orientiert sind, aber nicht beides. Die hier gefundene Association schließt diese Lücke.

Die vorgestellten Topics und ihre Associations erlauben eine effiziente, nicht-sequentielle Navigation des FCOM-Datenbestandes. Neben der leichteren Erschließbarkeit des Textmaterials sollte im Pilotprojekt aber auch die Basis für ein Pilotentraining gelegt werden. Hierzu wurde auf Basis der Topic Map ein Testfragengenerator realisiert. Dieser Generator ist unabhängig von der Domäne und funktioniert mit beliebigen Topic Maps. Im

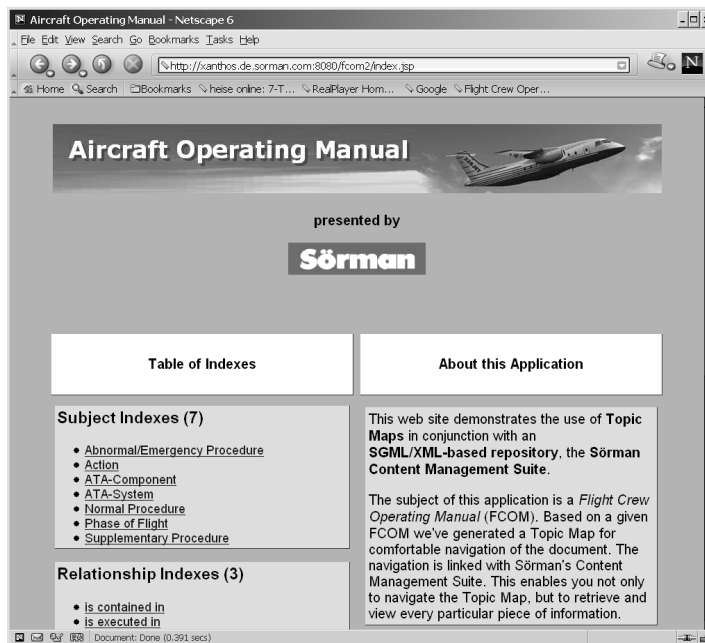


Abbildung 2: Web-Oberfläche mit den sieben Topic-Typen (links)

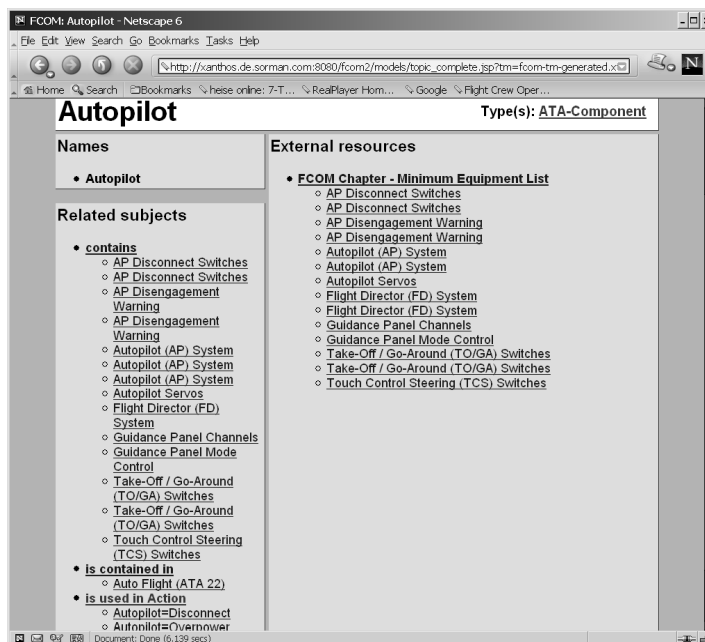


Abbildung 3: Der Autopilot, seine Bauteile und die Handlungen, in denen er zum Einsatz kommt

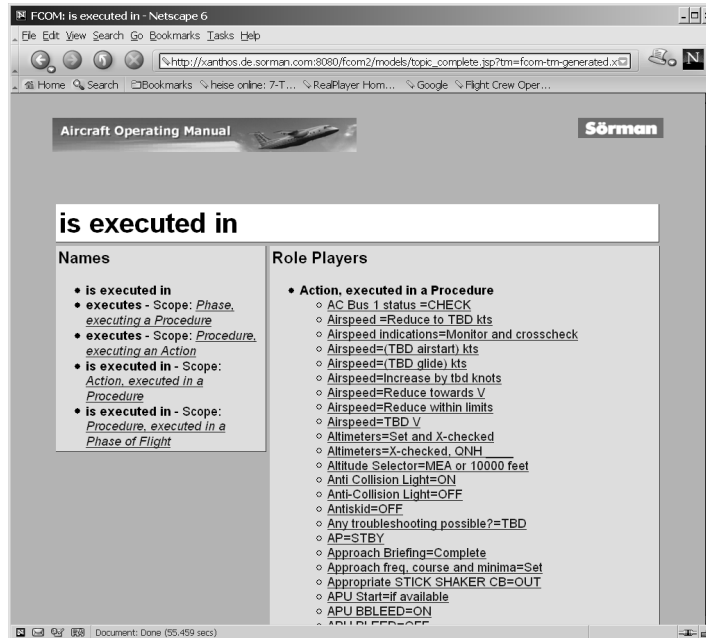


Abbildung 4: Association, die beschreibt welche Aktion in welcher Prozedur auszuführen ist

wesentlichen arbeitet er so, dass er zufällig eine Association A ermittelt und ein damit verbundenes Topic T auswählt. Daraus lässt sich dann eine Testfrage der Form “Welche Topics stehen in Relation A zu Topic T?” ableiten. Konkret nimmt der Testfragengenerator die in Abbildung 6 gezeigt Form an.

3 Bewertung

Eine abschließende Zusammenfassung und Bewertung der gesammelten Erfahrungen soll nicht fehlen; sie differenziert sich nach der Topic-Map-Technik und den eingesetzten Werkzeugen.

Die Entscheidung, Topic Maps für dieses Pilotprojekt einzusetzen, kann rückblickend als richtig eingestuft werden. Sie erbrachte die angestrebten Vorteile, insbesondere Modellierung der Navigation außerhalb der Daten sowie flexible Restrukturierung der Daten. Es ist sehr stark deutlich geworden, dass die Modellierung mit Topic Maps neben dem Topic-Map-Kenner vor allem einen Experten der jeweiligen Anwendungsdomäne benötigt. Diese Erkenntnis ist jedoch wenig überraschend oder gar als negativ zu bewerten. Es verhält sich hier genauso wie bei der Modellierung von Dokumentstrukturen mit SGML/XML-DTDs.

Liegen die Ursprungsdaten bereits in SGML/XML vor, so lassen sich Topic Maps per Transformation mit XSLT generieren. Hierbei treten die Stärken von XSLT zur Behandlung von Dokumentstrukturen hervor. Anfänglich war geplant, nur eine erste Näherung

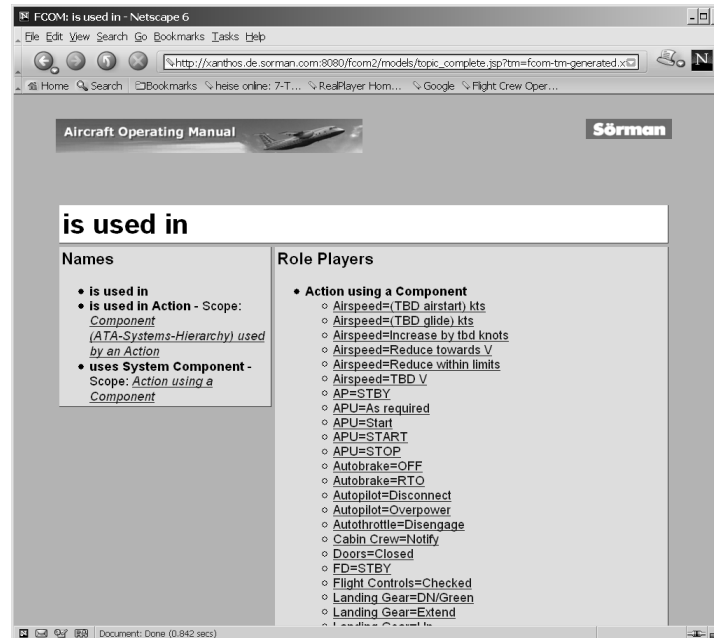


Abbildung 5: Association, die beschreibt welche Systemkomponente in welcher Aktion benutzt wird

der angestrebten Topic Map mit XSLT zu erzeugen und darauf aufbauend später andere Werkzeuge zu verwenden. Diese als Zwischenergebnis geplante Topic Map war jedoch qualitativ so gut, dass bis zum Ende ausschließlich XSLT zum Einsatz kam. Als Nachteile von XSLT sind zu nennen: (a) Die Ausgabe der Topic Map muss Stück für Stück in Eigenregie umgesetzt werden. APIs, wie sie die Ontopia-Software zur Verfügung stellt, erhöhen in diesem Punkt den Programmierkomfort nennenswert. (b) Die Fähigkeiten von XSLT sind auf Struktur und Volltextsuche beschränkt. Bei weniger oder schlechter strukturierten Ausgangsdaten ist ein Werkzeug zur Wissensentdeckung, Information Retrieval o.ä. sinnvoll.

Die Bewertung der eingesetzten Programme fällt positiv aus. Die Sorman CMS unterstützte durch ihre umfangreichen APIs (hier: Java-API, W3C-DOM) die Verarbeitung der Ursprungsdaten. Ontopias Topic-Map-Engine und Tag-Libraries verstecken die Komplexität der Topic Map beim Aufbau einer Web-Anwendung. Das Navigationstool "Omnigator" erleichtert den Entwicklungsprozess der Topic Map erheblich. Das XML-Publishing-Framework Cocoon besitzt herausragende Möglichkeiten zum Web-Publishing auf XML-Basis. Dank des modularen, durchdachten Aufbaus war eine Anbindung an die Sorman CMS einfach zu implementieren.

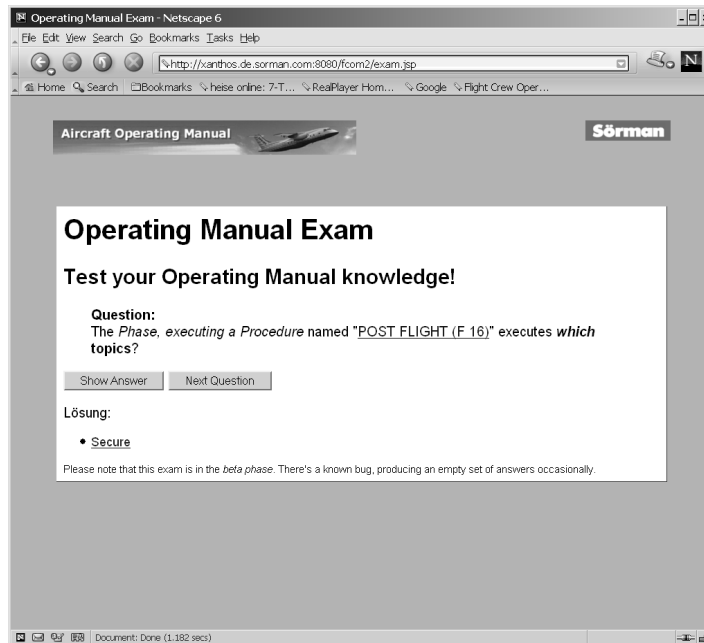


Abbildung 6: Aus der Topic Map generierte Testfrage

Literatur

- [XTM] *XML Topic Maps (XTM) 1.0*, TopicMaps.Org., 2001, <http://www.topicmaps.org/xtm/1.0/>
- [ISOTM] *ISO/IEC 13250: Information technology – SGML Applications – Topic Maps*. International Organization for Standardization, 2000
- [TAO] *The TAO of Topic Maps*, Steve Pepper, 2002, <http://www.ontopia.net/topicmaps/materials/tao.html>
- [XML] *Extensible Markup Language (XML) 1.0 (Second Edition)*, W3C, 6. Oktober 2000, <http://www.w3.org/TR/2000/REC-xml-20001006>
- [XSLT] *XSL Transformations (XSLT) 1.0*, W3C, 16. November 1999, <http://www.w3.org/TR/1999/REC-xslt-19991116>
- [SGML] *ISO 8879:1986(E): Information processing – Text and Office Systems – Standard Generalized Markup Language (SGML)*, International Organization for Standardization, 1986

3.1 Software

- Sörman CMS: www.sorman.com
- Ontopia: www.ontopia.net
- Cocoon: xml.apache.org