

Automatische Erstellung individualisierter, domänenspezifischer Topic-Maps zur nachhaltigen Nutzung von Projektdokumentationen

**Dipl.-Wirtsch.-Inf. Lutz Maicher,
Prof. Dr. Gerhard Heyer**

Universität Leipzig, Institut für Informatik,
Abteilung Automatische Sprachverarbeitung

Augustusplatz 10-11
04109 Leipzig

E-Mail: {maicher|heyer}@informatik.uni-leipzig.de

Dipl.-Inf. Karsten Böhm

Universität Leipzig, Institut für Informatik,
Lehrstuhl für Anwendungsspezifische Informationssysteme

Augustusplatz 10-11
04109 Leipzig

E-Mail: boehm@informatik.uni-leipzig.de

Dipl.-Wirtsch.-Inf. Olaf Grahn

BearingPoint GmbH

Münzgasse 2,
04107 Leipzig

E-Mail: olaf.grahn@bearingpoint.com

Automatische Erstellung individualisierter, domänenspezifischer Topic-Maps zur nachhaltigen Nutzung von Projektdokumentationen

Abstract

In diesem Beitrag wird die Entwicklung einer IT-Applikation zur Steigerung der Produktivität der Erstellung und Nutzung von Projektdokumentationen durch konsequente Automatisierung am Beispiel eines realisierten Projektes der BearingPoint GmbH beschrieben. Die Applikation basiert auf dem vorgestellten Konzept des Wissensraums, als ein Gespräch der Nutzer dieses Raums über Informationen, inklusive der Dokumentation ihrer kooperativen Prüf- und Kombinationsoperationen. Zu Beratungsprojekten wird eine Vielzahl von Dokumenten erstellt und genutzt, welche bereits einen Großteil der organisationsinternen Kommunikation über dieses Projekt darstellt und zumeist entsprechend eines organisationalen Ordnungsschemas (Vorgehensmodell) gespeichert wird. Um diese Kommunikation wiederzuverwenden, und Unternehmensressourcen nachhaltig zu nutzen, fehlen themenbasierte, für den Kontext des Konsumptionszeitpunkts strukturierte Zugriffsmethoden. Mit einer durch Methoden der automatischen Sprachverarbeitung erstellten domänenspezifischen Topic-Map kann dieses Problem gelöst werden. Topic-Maps erlauben die Erstellung themenorientierter Sichten, die vielfältig visualisiert und in Applikationen integriert werden können. Diese Topic-Maps können als Grundgerüst für die umfassende Implementierung des Konzepts der Wissensräume genutzt werden.

1 Automatisierung in wissensintensiven Prozessen

Die Notwendigkeit der Steigerung der operationalen Effizienz in wissensintensiven Prozessen und Dienstleistungen, wie z. B. Beratungsprojekten, steht außer Frage. Es werden somit Werkzeuge benötigt, die deren Produktivität massiv erhöhen können. Exemplarisch für diese Problemstellung wird in diesem Beitrag die Fragestellung der Produktivität der Erstellung und Nutzung der Dokumentation von Beratungsprojekten für den unternehmensinternen Gebrauch aufgegriffen, um ein teilautomatisiertes und kooperatives Wissensmanagement zu realisieren.

Produktivität kann einerseits bei der Erstellung der Dokumentation und andererseits bei der Projektarbeit durch die Nutzung von bestehender Projektdokumentation gesteigert werden. Aus betriebswirtschaftlicher Sicht bedeutet dies, dass ei-

nerseits die *Produktionskosten* für die Projektdokumentation bei gleichbleibender oder sogar steigender Qualität gesenkt werden. Diese Qualität impliziert sinkende *Transaktionskosten* bei der Informationsbeschaffung während der Nutzung der Projektdokumentation. Deren Wiederverwendung führt darüber hinaus zu nachhaltiger Nutzung von bereits im Unternehmen produziertem Wissen, ergo zu einer Erhöhung der Rentabilität der in der Vergangenheit eingesetzten Ressourcen.

Das am häufigsten genutzte Medium für die Projektdokumentation ist die natürliche Sprache, so dass von einem ‚Wissensrohstoff Text‘ gesprochen werden kann. Die Projektdokumentation für den internen Gebrauch ist eine Kollektion von Dokumenten folgender Klassen. Die *Ergebnisdokumente* der Projektarbeit beschreiben die Prozesse und ihre Ergebnisse in Organisationen auf vielfältige Art und Weise. Ebenfalls circa 80% der strategisch relevanten Informationen liegen in Form von Texten vor, z. B. als Berichte, Gutachten, Produktbeschreibungen, Patente etc.([Geri00]). Diese bilden die Klasse der *entscheidungsrelevanten, projekt-externen Dokumente*. Darüber hinaus wird während der Projektarbeit eine Vielzahl von Entscheidungsprozessen realisiert. Die Ergebnisse dieser Prozesse sind implizit in den Ergebnisdokumenten gespeichert, doch für eine Wiederverwendung des angewandten Methodenwissens ist die Nutzung der *Dokumentation der Entscheidungsfindungsprozesse* notwendig. Diese ist z. B. indirekt in den E-Mails zur Kommunikation zwischen den Projektteilnehmern realisiert.

In diesem Beitrag wird die auf einer solchen Dokumentenkollektion und Methoden der automatischen Sprachverarbeitung (ASV) basierende Erstellung und Nutzung einer individualisierten, domänenspezifischen Topic-Map beschrieben. Am Beispiel dieser werden die Möglichkeiten für die Entwicklung von IT-Systemen diskutiert, um Projektdokumentationen für den internen Gebrauch teilautomatisiert und kooperativ mit den gewünschten Kostendegressionen zu erstellen und zu nutzen. Das im folgenden vorgestellte Konzept des Wissensraums lässt die Produktion und Nutzung solcher Dokumentationen zu einem aktiven Wissensmanagement verschmelzen.

2 Projektdokumentation und Wissensmanagement

Neben der Erfüllung interner und externer Dokumentationspflichten soll die Projektdokumentation einen substantiellen Beitrag für das unternehmensinterne Wissensmanagement liefern. In diesem Abschnitt werden eingangs die Aufgaben der Projektdokumentationen in der Praxis skizziert und die bestehenden Herausforderungen an die genutzten Wissensmanagementsysteme (WMS) herausgearbeitet. Anschließend wird das Konzept des Wissensraums als Möglichkeit der Implementierung eines aktiven Wissensmanagements eingeführt. Abschließend wird der Aufbau dieses Beitrags beschrieben, um den Weg von der Theorie in die Praxis zu skizzieren.

2.1 Projektdokumentation in der Beratungspraxis

Die Nutzung von Projektdokumentationen im Beratungsprozess ist Wiederverwendung von produziertem Wissen, ergo von eingesetzten Ressourcen. Folgende Tätigkeiten in der Beraterpraxis profitieren von solcher Wiederverwendung:

- *Abgrenzung und Vorstrukturierung des Analyseprozesses in Projekten ähnlicher Domänen.* Dieser Prozess wird zusätzlich durch bereits vorhandene Informationen zu speziellen Produkten, Branchen und Unternehmen im Markt, sowie durch die aus dem Rechercheprozess resultierenden Impulse und Anregungen unterstützt.
- *Nutzung von Methodenwissen.* Die Projektdokumentation ist die Dokumentation der Anwendung von Methodenwissen in speziellen Domänen (z. B. Einführungsprozesse, Managementprozesse etc.).
- *Expertenlokalisierung.* Zentrales Moment der Arbeit in einem Beratungsunternehmen ist die Lokalisierung unternehmensinterner, domänenspezifischer Experten und Nutzung deren Wissens.
- *Vokabular der Domäne.* Klärung der Bedeutung domänenspezifischer Fachbegriffe, Anglizismen und Abkürzungen.
- *Dokumentationspflichten.* Projekte unterliegen unternehmensinternen und -externen Dokumentationspflichten, deren Ziel z. B. die Unterstützung des Controlling oder der unternehmensexternen Kommunikation ist.

Es wurden zur Unterstützung dieser Tätigkeiten in den Unternehmen der Beratungsbranche WMS eingeführt, welche bereits zu deutlichen Verbesserungen des Wissensmanagements im Beratungsprozess beigetragen haben, jedoch folgende, der Produktivität abträgliche, Eigenschaften besitzen:

- *Mitarbeiter entscheiden über die Gestaltung der Projektdokumentation für die interne Nutzung.* Die Principal-Agent-Theorie lässt vermuten, dass die Zielsysteme von Mitarbeitern und Unternehmen kollidieren, insbesondere wenn die Dokumentation nur für eine spätere Nutzung außerhalb des aktuellen Projektes erstellt werden soll.
- *Redaktionelle Bearbeitung* führt zu mehr Systematik, jedoch zu überproportional steigenden Kosten durch Einsatz von qualifiziertem Personal und sinkender Qualität durch partielle Informationsdefizite der Projektexternen.
- *Expertenwissen wird in Skilldatenbanken gespeichert* und z. B. als interne Gelbe Seiten publiziert (vgl. [Sihn⁺02]). Diese Datenbanken werden durch die Mitarbeiter mit hohem Personalaufwand manuell gepflegt. Die Mitarbeiter besitzen graduelle Gestaltungsspielräume bezüglich der über sie publizierten Expertisen, was potenziell zu Fehlallokationen bei der Mitarbeiterausstattung von Projekten führen kann.

2.2 Das Konzept des Wissensraums - Dokumentation als aktives Wissensmanagement

Wissensintensive Organisationen grenzen sich durch ständige, orts- und zeitunabhängige Kommunikation, basierend auf vergangene Entscheidungen, in einem permanenten Prozess von ihrer Umwelt ab. Wir wollen Wissen als eine Struktur verstehen, die es ermöglicht und erleichtert, mit Informationen umzugehen, das heißt, Informationen als „neu“ zu akzeptieren oder als „irrelevant“ abzulehnen und Informationen mit anderen zu vergleichen und zu dritter zu kombinieren ([BECK03]). Aus unserer Sicht ist Wissensmanagement die Organisation des Zusammenspiels dieser kooperativen Prüf- und Kombinationsoperationen.

Das vordergründige organisatorische Problem des Wissensmanagements ist somit die Bereitstellung von Systemen, mit denen dieser ständige Kommunikationsprozess adäquat zu den Zielen der Organisation gestaltet werden kann. Erst die permanente, auf die bisherige basierende, reflexive und selbstreferenzierende Kommunikation innerhalb der Organisation führt zu, auf dem Wissen der Organisation basierende, Entscheidungen. Dies bedeutet, Fundament des Wissensmanagements ist der Zugriff auf die bisherige Kommunikation innerhalb der Organisation.

In wissensintensiven Organisation kann jede erstellte Information für spätere Entscheidungen bedeutend sein, denn zum Zeitpunkt der Produktion besteht Unbestimmtheit, welche potenziellen Nachfrager zukünftig Interesse an dieser Information besitzen könnten. Erst in der aktuellen Perspektive (Kontext) zum Zeitpunkt der Konsumtion unterzieht ein Nachfrager die Information einer Bewertung bezüglich einer kontextspezifischen Relevanz. Dabei sollte er durch eine für den Kontext der Konsumtion spezifische Strukturierung der Texte der bestehenden Kommunikationen und durch die Bereitstellung von bereits in der Organisation realisierten, kontextspezifischen Prüfoperationen zu der Information des Interesses unterstützt werden.

Ein Wissensraum ist die vertrauenswürdige, kontextspezifische Sichtweise auf eine Menge von Informationsobjekten. De facto ist ein Wissensraum somit ein Gespräch der Nutzer dieses Raums über diese Objekte, inklusive der Dokumentation ihrer kooperativen Prüf- und Kombinationsoperationen ([HEYE⁺03]). Aufgabe eines IT-Systems, welches das Konzept des Wissensraums umsetzt, ist die Unterstützung der Nutzer im Moment des Erkenntnisgewinns mit Methoden für die aktive Produktion dieser „Gespräche“ (vgl. „*cognitive web*“ in [THOM02]), d. h. der Etablierung eines aktiven Wissensmanagements.

In dem Kontext dieses Beitrags kapselt ein Wissensraum die Dokumentation eines Beratungsprojektes und die Dokumentation über die im Laufe der Arbeit geleisteten Prüf- und Kombinationsoperationen. Unser langfristiges Ziel ist es, ein IT-System zu entwickeln, welches die Projektdokumentation zur internen Verwendung entsprechend des Konzepts des Wissensraums erstellt und nutzt, so dass Produktionskosten und Transaktionskosten im Vergleich zum heutigen IST-

Zustand signifikant gesenkt werden können. Um die gewünschten Verbesserungen zu erreichen, sollte das System die folgenden drei Prinzipien realisieren:

- *Schlankheit.* Der Wissensraum basiert nur auf Texten und Informationen, die während des Projektes *für* die Projektarbeit erstellt oder genutzt wurden. Eine manuelle Produktion zusätzlicher Texte *allein* zum Zweck der *Organisation* der internen Projektdokumentation führt zu Mehrarbeit der Berater, die zu vermeiden ist.
- *Kooperation.* Projekte sind kooperative Entscheidungs- und Produktionsprozesse, so dass die Ergebnisse der Prüf- und Kombinationsoperationen kooperativ *während* dieser Prozesse dokumentiert werden sollen.
- *Automatisierung und menschliche Intervention.* Die Methoden des Wissensraums werden durch automatisierte Verfahren so weit wie möglich unterstützt. Es besteht jedoch zu jedem Zeitpunkt die Möglichkeit menschlicher Intervention, um die Akzeptanz des Systems zu gewährleisten.

2.3 Von der Theorie in die Beratungspraxis

Die aus den vorangegangenen Überlegungen abgeleitete Fragestellung ist, wie das Konzept des Wissensraums für die Verbesserung des internen Gebrauchs der Projektdokumentation von Beratungsprojekten mit realen IT-Systemen umgesetzt werden kann. In Kooperation mit der BearingPoint GmbH wurde die Dokumentation eines abgeschlossenen, mittelgroßen Projektes ausgewählt. In einem Workshop bestand die Möglichkeit gemeinsam mit Mitarbeitern von BearingPoint den gewählten Ansatz und die entwickelte Lösung zu diskutieren. Die Ergebnisse dieses Workshops sind in Abschnitt 4.1 dokumentiert.

Dem Workshop ging die Entwicklung des IT-Systems voraus. Abschnitt 3.1 beschreibt die Grundlage für die Entscheidung, Topic-Maps als Wissensrepräsentationssprache zu wählen. Die mit einer Topic-Map beschriebene thematische Struktur der Projektdokumentation wurde automatisch mit den in Abschnitt 3.2 vorgestellten Methoden der ASV erstellt.

Ziel der in diesem Beitrag beschriebenen Forschungsarbeit war die Erstellung eines *statischen* Wissensraums zur Dokumentation eines *abgeschlossenen* Projektes, welches als Demonstrationslösung für die Evaluierung seitens der Nutzer und als Ausgangspunkt für weiterführende Entwicklungen genutzt werden soll. In Abschnitt 3.3 werden die entsprechenden technologischen, funktionalen und ergonomischen Designentscheidungen diskutiert. In den Abschnitten 4.2 und 4.3 werden abschließend die Entwicklungsmöglichkeiten des gewählten Ansatzes und die dabei bestehenden technologischen Limitationen erörtert.

3 Methodischer Hintergrund

In diesem Abschnitt werden die Methoden zur Realisierung des skizzierten Ansatzes beschrieben. Ausgangspunkt ist eine Analyse von Topic-Maps als Wissensrepräsentationssprache in dem Kontext der Wissensräume, darauf aufbauend wird detailliert die Erstellung von domänenspezifischen Topic-Maps mit Methoden der ASV beschrieben und abschließend die Architektur des entwickelten IT-Systems vorgestellt.

3.1 Topic-Maps als Sprache der Wissensrepräsentation

Wie bereits in [HEYE⁺03] diskutiert, eignen sich Topic-Maps als Wissensrepräsentationssprache, respektive Topic-Map-verarbeitende Applikationen als Werkzeuge, für die Erstellung und Nutzung von Wissensräumen. Die ISO-Standardisierung und die Nutzung einer XML-Syntax machen Topic-Maps zu einer nachhaltigen Technologieentscheidung für Unternehmen. Auf eine ausführliche, grundlegende Einführung zu Topic-Maps soll mit Verweis auf [PEPP00] als weiterführende Literatur an dieser Stelle verzichtet werden.

Da die Wissensrepräsentation mit Topic-Maps nicht die zugrunde liegenden Basisdokumente verändert, kann unabhängig von der expliziten Speicherung und Organisation der Dokumentation die Kommunikation aus den Projekten in vielfältigen, subjektiven Metainformationsschichten beschrieben werden.¹ „One advantage of Topic Maps is that they add semantics to data without modifying it. Moreover, one Topic Map may describe several information pools and several Topic Maps may apply to one single information pool.” ([LEGR⁺01]). Das gewählte Ordnungsschema bei der Produktion (Vorgehensmodell) wird durch die Nutzung von Topic-Maps nicht verändert, sondern ergänzt. Darüber hinaus können Topic-Maps vielfältig visualisiert werden ([LEGR⁺01]) und z. B. in unternehmensweite Wissensmanagementportale oder andere Applikationen integriert werden.

Besonders interessant ist die Nutzung des *Scope-Konzeptes*. Mit Hilfe dieses Konzeptes können kontextabhängige Sichten auf Topic-Maps definiert werden. Dies erlaubt vielfältige Anwendungen des themenabhängigen Filterns, Rankens und Auswählens und somit eine adäquate Unterstützung der kontextspezifischen Kommunikationsprozesse [Pepp01]. Daneben war das *Merge-Konzept* entscheidendes Kriterium für die Technologieentscheidung. Mit dessen Hilfe können verschiedene Topic-Maps automatisch zusammengeführt werden. Dies ist z. B. interessant, wenn verschiedene Personen Topic-Maps zum selben Diskursbereich erstellen und diese Arbeit in einem gemeinsamen Projekt integriert werden soll.

¹ Es besteht jedoch das Problem der „Unstetigkeit der realen Welt“, d. h. wenn Informationsobjekte, zu denen Aussagen in Topic-Maps gespeichert sind, modifiziert oder gelöscht werden, so steht auch die Gültigkeit der Aussage in Frage.

3.2 Automatische Erstellung domänenspezifischer Topic-Maps mit Methoden der ASV

In [BÖHM⁺02] wurden Methoden der ASV nutzend Topic-Map-basierte Themen-netzwerke aus großen Textkorpora erstellt, welche gleichzeitig einen konzeptbasierten Index und, durch die Nutzung von semantischen Relationen, zusätzliches Wissen über den Diskursbereich der Texte darstellen (vgl. auch [HEYE⁺02]). Grundlage dieser Arbeit ist das Wortschatzprojekt an der Universität Leipzig (<http://wortschatz.informatik.uni-leipzig.de/>). Diese Topic-Maps erleichtern den themen- respektive konzeptorientierten Zugriff auf große Textkorpora enorm und verbessern somit die Kommunikation über abgeschlossene Projekte in zwei Dimensionen: einerseits können die kontextrelevanten Texte oder Textstücke identifiziert werden, so dass frühere Argumentationen zu den betreffenden Themen leichter auffindbar sind, und andererseits ist die Topic-Map selbst (Meta-) Information über das abgeschlossene Projekt.

Insbesondere bei dem letztgenannten Aspekt geht die umgesetzte Lösung über die klassischen Ansätze des Information Retrieval hinaus, da neben dem Auffinden der entsprechenden Belegstellen in den Ursprungsdokumenten auch die in der Kollektion enthaltenen inhaltlichen Zusammenhänge extrahiert und dargestellt werden können. Im Gegensatz zu vielen anderen Metainformationen werden diese Zusammenhänge automatisch erstellt und können daher kostengünstig und effizient aus der bestehenden Datenbasis gewonnen werden.

Knoten dieser Netzwerke repräsentieren automatisch extrahierte Konzepte, mit denen die entsprechenden Dokumente oder Dokumententeile extern annotiert werden. Konzepte können dabei z. B. in Prozessen behandelte Probleme („Versionsmanagement“, „Teilprojekt XY“) oder beteiligte Personen sein. Die Relationen zwischen den Konzepten repräsentieren zusätzliche Informationen, welche durch Kollokationen gewonnen werden [Böhm⁺02]. Aus Sicht des Einzelkonzepts kann so dessen semantische Umgebung visualisiert werden und erlaubt dem Benutzer ausgehend von einem Einstiegsbegriff ein exploratives, assoziatives Navigieren im Begriffsnetz (vgl. z. B. Abbildung 6). Insgesamt ergibt sich so eine domänenspezifische Sicht auf die inhaltlichen Konzeptzusammenhänge im analysierten Datenbestand.

Da die eingesetzten statistischen Verfahren, die unter Berücksichtigung von Wortfrequenzen und Kollokationsmaßen signifikante Konzepte und deren Beziehungen untereinander berechnen, a priori nicht zwischen häufig vorkommenden Allgemeinbegriffen und hochfrequenten Fachbegriffen unterscheiden können, ist hier eine weitere Differenzierung notwendig. Das eingesetzte Verfahren der *Referenzwortschatzanalyse* ermöglicht die Filterung domänenspezifischer Konzepte und basiert dabei auf der Gewichtung der Begriffe des Fachwortschatzes gegenüber einem allgemeinsprachlichen Referenzwortschatz, z. B. dem Korpus des Wortschatzprojektes. Diese umfangreiche Sammlung von über 9 Millionen verschiede-

nen Wortformen aus dem allgemeinsprachlichen Bereich repräsentiert dabei näherungsweise das Allgemeinwissen, das auch menschliche Experten zusätzlich zu ihrem Fachwissen besitzen.

Die Wortfrequenzen der Begriffe im Fach- und im Referenzwortschatz werden zunächst unter Berücksichtigung der Größe der verwendeten Korpora normiert und dann zueinander in Beziehung gestellt. Die sich daraus ergebende Kennzahl spiegelt die Fachspezifik des betrachteten Begriffs wider. Bei Überschreiten eines einstellbaren Mindestfaktors wird von einer domänenspezifischen Verwendung des Fachbegriffs ausgegangen. Werden beispielsweise die Konzepte „Versionsmanagement“ und „Teilprojekt XY“ innerhalb der Projektdokumentation signifikant häufiger als im deutschen Wortschatz gebraucht, so ist dies ein Indiz für eine fachspezifische Verwendung und führt zur Übernahme der Konzepte in die Topic-Map. Die aus der Fachtextanalyse stammende Kollokation zwischen den beiden Begriffen motiviert zur Annahme einer semantischen Nähe und wird in der Topic-Map durch eine assoziative Beziehung ausgedrückt.

Das eingesetzte Verfahren ist dabei insbesondere unabhängig von einer festen Sachgebietsklassifikation. Solche Informationen können zwar vom System mit berücksichtigt werden, allerdings ist in sich schnell entwickelnden Domänen (z. B. IT- oder Biotechnologiebranche) nur selten eine vollständige Sachgebietsklassifikation vorhanden. Gerade in diesen Bereichen ist aber die schnelle Visualisierung domänenspezifischer Zusammenhänge von tragender Bedeutung. Das Verfahren der Referenzwortschatzanalyse kann hierbei nicht nur für die Filterung allgemeinsprachlicher Begriffe (auch in mehreren Sprachen), sondern auch zur Filterung domänenfremder Begriffe verwendet werden (z. B. Eliminierung IT-spezifischer Formulierungen in einer Projektdokumentation einer bestimmten Anwendungsdomäne). Voraussetzung ist jeweils das Vorhandensein eines Wortschatzes, der die Begrifflichkeit der Domäne ausreichend abdeckt.

In Ergänzung der so gewonnenen Roh-Topic-Map kann durch eine semi-automatische Überarbeitung auch nicht direkt aus dem Dokumentenkörper extrahierbares Wissen integriert und genutzt werden. Ist z. B. die Personalstruktur eines Unternehmens in elektronischer Form verfügbar, kann diese automatisch in eine Topic-Map transformiert werden. Das Zusammenführen der beiden Topic-Maps kann zu deutlichen Verbesserungen bei der Expertenlokalisierung führen.

Für die Nutzung der Verfahren zur Generierung von Roh-Topic-Maps zum Aufbau von Wissensräumen im Bereich der Projektdokumentation wurde eine Gliederung der Topic-Map in drei Ebenen vorgenommen:

- Die *Zitatebene* fasst die aus dem Dokumentenbestand extrahierten Belegstellen zusammen und spiegelt so die Verwendung eines Begriffs über die gesamte Datenkollektion wider. Dies entspricht oft einer impliziten Definition eines Begriffs, die sich den Benutzer über die Begriffsebene leicht erschließt.

- Die *Begriffsebene* umfasst alle domänenspezifischen Wortformen des analysierten Korpus auf der Ebene von Einzelworten und Vollformen. Sie verbindet die Begriffe untereinander entsprechend der ermittelten Kollokationen aus der Textanalyse und kennzeichnen so deren semantische Nähe zueinander. Das so entstehende Begriffsnetz stellt damit eine erste Abstraktionsstufe zu der Dokumentensammlung dar.
- Auf der *Konzeptebene* werden die verschiedenen Ausprägungen der einzelnen Begriffe zusammengefasst, wenn damit das gleiche mentale Konzept oder dasselbe physikalische Objekt referenziert wird. Dies betrifft beispielsweise die Zusammenfassung verschiedener Vollformen oder Schreibvarianten zu der korrespondierenden Grundform, oder die Abbildung von Synonymiebeziehungen zwischen Konzepten. Im Rahmen der vorliegenden Arbeit konnte nur ein Teil der Konzeptaggregation realisiert werden, deren Nutzenpotenziale sind jedoch enorm (vgl. [SCHA02], [LENZ⁺02]).

Die entstehende Topic-Map nach XTM-Standard fasst diese drei Ebenen in einer einzelnen Datei zusammen und übergibt sie zur Visualisierung für den Benutzer an die im folgenden Abschnitt beschriebene Webapplikation. Die Vorgehensweise für die Erstellung eines Wissensraums aus einer bestehenden Projektdokumentation gliedert sich im Wesentlichen in die beiden Hauptverarbeitungsschritte Aufbereitung und Visualisierung (siehe Abbildung 1).

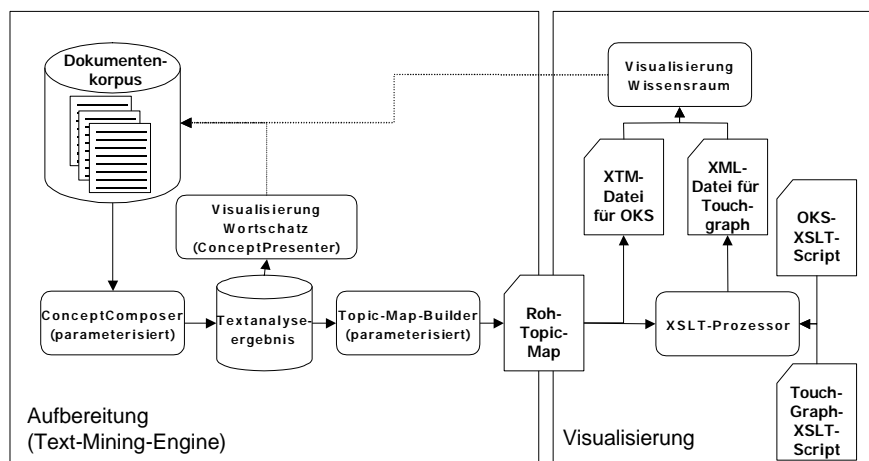


Abbildung 1: Vorgehensweise zur Erstellung des Wissensraums aus dem Dokumentenkorpus

Als Schnittstelle zwischen den beiden Verarbeitungsschritten dient die Roh-Topic-Map im XTM-Format. Bei der Erzeugung dieser Topic-Map wird zunächst die Projektdokumentation mit Hilfe der Text-Mining-Software (ConceptComposer) analysiert. Dabei werden neben der Dokumentkonvertierung und der Satz- und

Wortsegmentierung auch die Kollokationsberechnung und die Referenzwortschatzanalyse durchgeführt, sowie ein Volltextindex aufgebaut. Die Ergebnisse werden in einer Datenbank abgelegt und stehen dort für die weitere Verarbeitung zur Verfügung. Eine Visualisierung der Konzeptzusammenhänge ist in dieser Phase bereits mit dem ConceptPresenter möglich. Die eigentliche Erzeugung der Roh-Topic-Map erfolgt mit einem weiteren Werkzeug – dem TopicMapBuilder – der nach entsprechender Parametrisierung die relevanten Konzepte und Kollokationen aus dem Textanalyseergebnis extrahiert und template-basiert eine XTM-Datei generiert, die dann an die Präsentationsschicht übergeben wird.

Bei der Umsetzung des Projektes mit konkreten Daten wurde der Textanalyseprozess insbesondere für den Bereich der Konzeptaggregation angepasst. Durch den modularen Aufbau des ConceptComposers ist die anwendungsspezifische Integration von spezifischen Verarbeitungsschritten jederzeit möglich. Im TopicMapBuilder kann auf die aggregierten Konzepte des Analyseprozesses einfach zugegriffen werden, da auch diese in der Datenbank abgelegt sind. Durch die Verwendung von Templates reduziert sich der Aufwand für die notwendigen Anpassungen auf die Erstellung der Zugriffsmodule für die neu zu integrierenden Daten.

3.3 Die realisierte IT-Lösung

Im folgenden Abschnitt wird der Entwicklungsprozess für die Erstellung der Webapplikation zur Visualisierung dokumentiert. Ausgangspunkt war die im vorangegangenen Abschnitt beschriebene Anforderungsdefinition an die zu erstellende Topic-Map. Darauf aufbauend wurden entsprechend „overview first, detail later“ [LeGr⁺01] Anforderungen an deren Visualisierung formuliert. Als Überblicksicht wurde eine Graph-basierte Variante und als Detailsicht eine Topic-basierte Variante der Repräsentation gefordert. In ersterer wird jede Topic durch einen Knoten und die entsprechenden Assoziationen als Kanten repräsentiert (vgl. Abbildung 6), in letzterer erhält jede Topic eine eigene Webseite, in der alle für diese Topic spezifischen Informationen repräsentiert werden (vgl. Abbildung 3).

Für die Realisation stand die Dokumentation eines Projektes von BearingPoint zur Einführung eines DMS im Beschaffungswesen als Kollektion von 5 MB Ergebnisdokumenten und 5 MB entscheidungsrelevanten, externen Dokumenten zur Verfügung. In weiterführenden Arbeiten ist zusätzlich die projektinterne E-Mail-Kommunikation als Dokumentation der Entscheidungsfindungsprozesse zu integrieren. In die entwickelte Demonstrationslösung für dieses abgeschlossene Projekt wurden keine Möglichkeiten der Dokumentation von Prüf- und Kombinationsoperationen integriert, da diese kooperativ *während* der Projektarbeit in den Wissensraum integriert werden sollten. Dies ist jedoch ebenfalls in einer weiterführenden Arbeit zu realisieren.

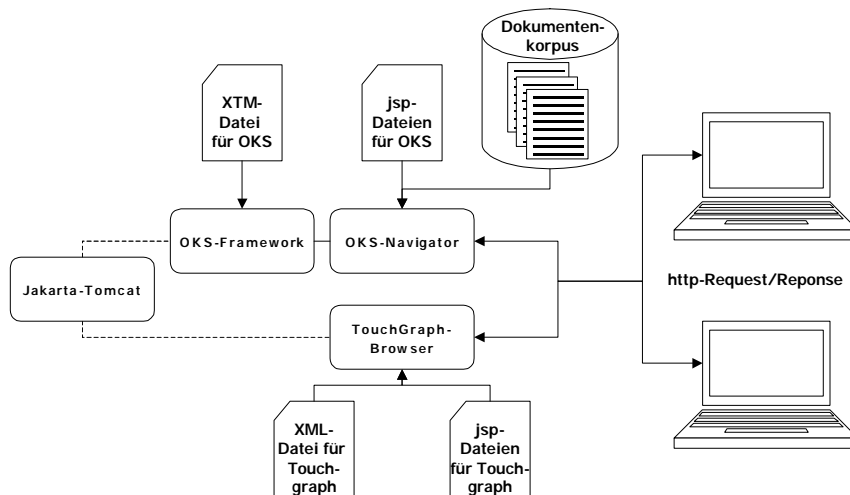


Abbildung 2: Architektur der erstellten Applikation

In Abbildung 2 ist die entwickelte Architektur für die Applikation abgebildet. Die erstellte Topic-Map (XTM-Datei für OKS) besteht aus etwa 600 Begriffen, welche in 7000 Zitaten referenziert werden (rund 23.000 TAOs). Die einzusetzende Software muss in der Lage sein, Topic-Maps mit dieser Größe in hoher Geschwindigkeit zu verarbeiten. Für die Topic-basierte Visualisierung wurde von der Firma Ontopia das OKS-Framework zur Verfügung gestellt, welches diese Anforderung sehr gut erfüllt. Für jeden Webseitentyp (z. B. für einen Begriffstopic) muss eine JSP-Datei erstellt werden, die die durch das OKS-Navigator-Framework zur Verfügung gestellte Funktionalität nutzen kann.

Die Graph-basierte Visualisierung wird mit dem Touchgraph-Browser realisiert, einer Open-Source-Applikation für die Repräsentation von Graphen, welche als Java-Applet im Webbrowser genutzt werden kann. Die Touchgraph-Applikation benötigt eine alternative XML-Notation des zu visualisierenden Graphen, so dass die erstellte Topic-Map mit Hilfe eines XSLT-Prozessors entsprechend transformiert werden muss. Touchgraph und das OKS-Framework nutzen eine Apache Tomcat-Engine, auf der auch der Dokumentenkörper verfügbar gemacht wurde, um ein dezentrales Öffnen der Dokumente zu ermöglichen.

4 Projektergebnisse

Im folgenden Abschnitt werden die Ergebnisse der Arbeit an der Realisierung zusammengefasst. Dabei werden zuerst die Ergebnisse des Evaluierungsworkshops bei BearingPoint beschrieben, bei dem die entwickelte Applikation gemein-

sam mit Beratern diskutiert wurde. Daran anschließend werden die bei der Anwendung und Entwicklung extrahierten technologischen Limitation beschrieben und Lösungsmöglichkeiten skizziert. Unabhängig von diesen bestehenden Herausforderungen werden abschließend die Weiterentwicklungspotenziale des entwickelten Konzeptes diskutiert.

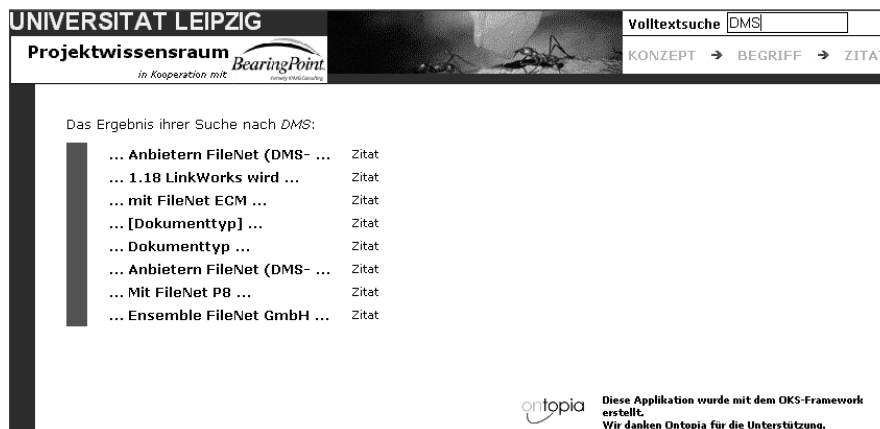
4.1 Ergebnisse des Workshops

In einem Workshop wurde gemeinsam mit Mitarbeitern von BearingPoint das entwickelte System evaluiert. Ausgangspunkt war die Einführung des theoretischen Hintergrunds. Anschließend wurde das im folgenden beschriebene Anwendungsszenario² diskutiert.

Datenbasis des erstellten Wissensraums war die Dokumentation eines Projektes zur Einführung eines DMS im Beschaffungswesen. Im skizzierten Anwendungsfall nutzt ein Manager in der Phase der Angebotserstellung für ein ähnliches Projekt diesen Wissensraum für die Beantwortung der folgenden Fragestellungen:

- Für welche DMS besteht in unserem Haus Praxiserfahrung im Anwendungsfall „Warenwirtschaft“?
- Welche Mitarbeiter besitzen die entsprechende Praxiserfahrung?

Um die gestellten Fragen zu beantworten, wählt der Manager den erstellten Projektwissensraum, da in diesem die genutzten Systeme und die involvierten Mitarbeiter bei der Einführung eines DMS im Beschaffungswesen dokumentiert sind. Als Ausgangspunkt nutzt der Manager die integrierte Volltextsuche und sucht nach dem Begriff „DMS“. Die Ergebnisliste ist die Abbildung 3.



UNIVERSITÄT LEIPZIG

Projektwissensraum *BearingPoint*
in Kooperation mit

Volltextsuche

KONZEPT → BEGRIFF → ZITAT

Das Ergebnis ihrer Suche nach *DMS*:

... Anbietern FileNet (DMS- ...	Zitat
... 1.18 LinkWorks wird ...	Zitat
... mit FileNet ECM ...	Zitat
... [Dokumenttyp] ...	Zitat
... Dokumenttyp ...	Zitat
... Anbietern FileNet (DMS- ...	Zitat
... Mit FileNet P8 ...	Zitat
... Ensemble FileNet GmbH ...	Zitat

ontopia

Diese Applikation wurde mit dem OKS-Framework erstellt.
Wir danken Ontopia für die Unterstützung.

Abbildung 3: Ergebnisliste der Volltextsuche nach Begriff "DMS"

² Die Namen der am Projekt beteiligten Mitarbeiter von BearingPoint wurden geändert.

Der Manager wählt das Zitat „...Anbietern FileNet (DMS- ...“ und gelangt zu der Webseite entsprechend Abbildung 4. Das gewählte Zitat beschreibt die Systemlandschaft in dem aktuellen Projekt, inklusive dem eingesetzten DMS.

Abbildung 4: Webseite des Zitats "...Anbieter FileNet (DMS-..."

Der Manager kann das Quelldokument zu diesem Zitat betrachten und weiterführende Informationen extrahieren. Er kann jedoch auch die in dem Wissensraum gespeicherten Informationen zu dem Begriff „FileNET“ durch Wahl des entsprechenden Links betrachten. Er gelangt zu der Webseite in Abbildung 5.

Abbildung 5: Webseite des Begriffs "FileNET"

Auf dieser Webseite sind auf der einen Seite alle Begriffe zu sehen, welche in dem Kontext des aktuellen Projektes eine inhaltliche Nähe zu dem Begriff „FileNET“ besitzen und andererseits eine Liste aller Zitate, in denen dieser Begriff genutzt

wird. In der Liste der inhaltlich nahen Begriffe sind drei Eigennamen zu lokalisieren, wobei nur Herr Homdberg und Herr Taleske als Mitarbeiter von BearingPoint zu identifizieren sind. Der Manager kann antizipieren, dass diese Mitarbeiter über spezifisches Wissen in dem Kontext des aktuellen Projektes verfügen. Durch Wahl des Begriffs „Friedland“ gelangt der Manager zu der Webseite des Begriffs und kann schnell erfassen, dass Herr Friedland ein potenzieller Ansprechpartner bei FileNET für ein solches Projekt ist. Für die weitere Recherche wählt der Manager die Graph-basierte Visualisierung und gelangt zu der Abbildung 6.

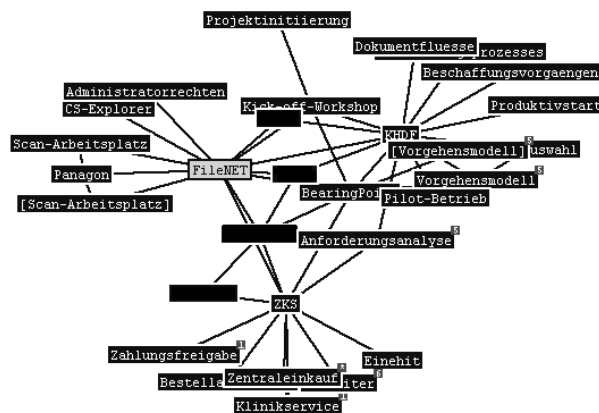


Abbildung 6: Graph-Repräsentation der Umgebung des Begriffs "FileNET"³

In dieser Darstellungsform kann der Manager relativ zügig einen Überblick über das Projekt erlangen und evaluieren, welche Potenziale zur Wiederverwendung von bereits im Unternehmen vorhandenen Wissen für das neue Projekt existieren (z. B. Wiederverwendung eines Vorgehensmodells; Expertise zum Produktivstart).

Bei der parallelen Diskussion mit den Workshopteilnehmern wurde die Frage nach der Möglichkeit der Produktivitätssteigerung von Erstellung und Nutzung der Projektdokumentation durch das Konzept automatisch erstellter Wissensräume sehr positiv beantwortet. Dabei sind die folgenden Punkte aus der Diskussion hervorzuheben:

- *Akzeptanz durch hohen Automatisierungsgrad.* Zusätzliche Tätigkeiten (z. B. manuelle Klassifikation), die nicht unmittelbar der Projektarbeit dienen, wurden skeptisch beurteilt. Dies impliziert gleichzeitig die Notwendigkeit einer intuitiven Handhabbarkeit des zu erstellenden Systems. Die dadurch realisierbare Senkung der Produktionskosten wurde sehr positiv bewertet.

³ Durch Wahl von Begriffen, welche mit einer Zahl annotiert sind, erweitert sich der repräsentierte Graph. Die Namen der Projektbeteiligten sind geschwärzt.

- *Geschwindigkeit der Informationsbeschaffung.* Es wurde konstatiert, dass die Umsetzung des Konzepts des Wissensraums die Geschwindigkeit der Beschaffung von kontextspezifisch relevanten Informationen deutlich erhöhen kann. Diese Möglichkeit der Senkung der Transaktionskosten wurde von den Workshopteilnehmern ebenfalls positiv bewertet.
- *Expertenlokalisierung.* Als eine Haupttätigkeit in der Projektarbeit wurde die Expertenlokalisierung im Unternehmen genannt, deren Unterstützung die Produktivität der Projektarbeit verbessern kann. Als interessanter Aspekt wurde herausgearbeitet, dass ein Wissensraum die Mitarbeiter dabei unterstützt, nach der Lokalisierung eines Experten das durch diesen in Dokumenten explizierte Wissen einfacher durchzuarbeiten, so dass eine Kontaktaufnahme nur bei neuartigen Fragestellungen notwendig ist. Diese Möglichkeiten wurden insbesondere für neue Projektmitarbeiter (z. B. Urlaubsvertretungen) als sehr positiv bewertet.
- *Akzeptanzprobleme bei vollkommener Automatisierung.* Den Mitarbeitern sollte immer die Möglichkeit der menschlichen Intervention gegeben werden. Dies erscheint wichtig, damit subjektiv als nicht korrekt erachtete Zusammenhänge revidiert werden können, insbesondere wenn diese unternehmenspolitische Fragestellungen tangieren.
- *Zusammenfassung mehrerer Projektwissensräume.* Für die Nutzung des Wissens abgeschlossener Projekte in neuen Projekten sollte die freie Kombinierbarkeit unterschiedlicher Projektwissensräume möglich sein. So können die Stärken des kontextbezogenen assoziativen Retrievals in Metadaten besonders gut genutzt werden.

Neben diesen konzeptuellen Bemerkungen wurden die folgenden, über die entwickelte Lösung hinausgehenden, Anforderungen an die Applikation gestellt:

- Kennzeichnung der Dokumentenklasse von Zitaten,
- Kennzeichnung der Stärke der Kollokationen,
- Klassifizierung bestimmter Begriffe, z. B. Mitarbeiter von BearingPoint und externer Unternehmen, Produkt- und Unternehmensbezeichnungen,
- Aufzählung und Beschreibung aller in der Projektdokumentation genutzten Abkürzungen.

4.2 Limitationen der genutzten Technologien

Während der Erstellung und Nutzung der entwickelten Lösung wurden die folgenden technologische Problemstellungen extrahiert:

- *Problem des kleinen Dokumentenkorpus.* Die genutzten Verfahren der ASV sind statistische Verfahren, welche für jede Dokumentensammlung

funktionieren, deren Güte jedoch mit der Größe des Korpus positiv korreliert.

- *Konvertierungsprobleme durch Dokumententypvielfalt.* Die genutzten Verfahren benötigen den Rohtext der Dokumente, so dass für jeden Dokumententyp ein entsprechendes Werkzeug zur automatischen Extraktion existieren muss. Dies ist insbesondere für proprietäre Formate aufwändig. Beispielsweise ist in PDF-Dokumenten die Trennung „hart“ gespeichert, d. h. ein Extraktionstool kann das getrennte Wort „Projektmanagement“ nur als die zwei Worte: „Projekt-“ und „management“ erkennen.
- *Grafische Repräsentation von Wissen.* Im Beratungsumfeld wird eine bedeutende Quantität von projektspezifischem Wissen in Form von grafischen Geschäftsprozessmodellen gespeichert. Die in diesen Abbildungen beschriebenen Zusammenhänge von Begriffen und Konzepten (z. B. „Aufgabe XY wird bearbeitet von Mitarbeiter A“) können von den eingesetzten Technologien nicht genutzt werden.
- *Untypisierte Assoziationen.* Die vorgestellten texttechnologischen Methoden erlauben nicht die Erzeugung von typisierten Assoziationen auf Basis der Kollokationen. Somit können Regeln im Sinne der deskriptiven Logik zur automatischen Extraktion von neuem Wissen nur dann formuliert werden, wenn diese Assoziationen manuell typisiert werden.
- *Erzeugung statischer Wissensräume.* Die vorgestellte Lösung ist nur für die einmalige, statische Erstellung von Wissensräumen nutzbar, z. B. als Methode für das Dokumentations-Reengineering von bereits abgeschlossenen Projekten. Für die produktive Nutzung in Unternehmen ist eine Erstellung des Wissensraums in Echtzeit notwendig, d. h. sobald ein neues Dokument von Interesse ist, sollte dies in den entsprechenden Wissensraum integriert werden können. Für diesen Anwendungsfall ist jedoch weiterführende Forschung zur Erweiterung der genutzten Methoden der ASV notwendig.
- *Erstellung der Wissensräume bedarf manueller Parametrisierung.* Es sind durch empirische Forschung Möglichkeiten der Automatisierung dieser Voreinstellungen zu entwickeln.
- *Referenzierung von Zitaten ist nicht möglich.* Ein Leistungsmerkmal des vorgeschlagenen Konzeptes des Wissensraums ist, dass nicht nur ganze Dokumente klassifiziert werden, sondern auch Zitate aus diesen. Dies ist konsequent, da in einem Dokument unterschiedliche Konzepte behandelt werden können. Technologisch problematisch ist, dass nur für XML-Dokumente eine standardisierte Methode zur Referenzierung beliebiger Dokumentausschnitte nutzbar ist (XLink).

4.3 Weiterentwicklungspotenziale

Die Möglichkeiten der Weiterentwicklung der erstellten Applikation sind vielfältig, wobei die Integration des Konzepts des Wissensraums in die bestehende IT-Landschaft eine zentrale Anforderung ist. Aus Anwendungssicht sind neben dem betrachteten IT-Projekt verschiedene wissensintensive Projekte, z. B. im Anlagenbau, in der Forschung, oder andere wissensintensive Szenarien mit der vorgestellten Technologie zu unterstützen. An dieser Stelle sollen erste Ideen für grundlegende technologische Weiterentwicklungen skizziert werden (vgl. [HEYE⁺03]).

Die realisierte prototypische Anwendung integriert noch keine Verfahren zur Erkennung von Eigennamen (named entities), die in Dokumentensammlungen sowohl in Form von Personennamen, als auch in Form von Produkt- und Firmenbezeichnungen häufig anzutreffen sind. Die vollständige Erfassung dieser Mehrwortbegriffe ist ohne die Unterstützung entsprechender, redaktionell zu erstellender Verzeichnisse, Yellow Pages o. ä., nicht trivial. Für die Aggregation solcher Namensbestandteile bietet sich die konzeptuelle Ebene des Systems an. In [BIEM⁺03] wird ein iteratives Verfahren beschrieben, das für den gegebenen Anwendungsbereich die Erkennung von Personennamen vollständig automatisieren kann. Eine Erweiterung auf Produkt- und Firmennamen ist hierbei ebenfalls unter Voraussetzung einer entsprechenden Startmenge von Namensbestandteilen denkbar. Diese kann entweder durch eine initiale Redaktion einer Liste von (kooperativ) generierten Vorschlägen erfolgen, oder unter Rückgriff auf bereits vorhandene Verzeichnisse. Da die Ausdrucksfähigkeit der Konzeptebene so erheblich gesteigert werden kann, ist die Integration der genannten Verfahren geplant.

Die entwickelte Lösung konzentriert sich auf die statische Erstellung eines Wissensraums für ein abgeschlossenes Projekt. Langfristig sind folgende Möglichkeiten der (teil-)automatischen Dynamisierung der Wissensräume notwendig:

- Aufbau der Dokumentensammlungen durch kooperative Integration neuer Dokumente mit einem „Drag&Drop“-Verfahren (Ermöglichung konsequenter Individualisierung von Wissensräumen),
- Zusammenführen individueller Wissensräume zu gemeinschaftlichen⁴,
- Erstellung themenspezifischer Wissensräume aus bestehenden.

Für eine vollständige Umsetzung des Wissensraumkonzepts ist die in diesem Beitrag vollständig ausgeklammerte Integration der Dokumentation von Prüf- und Kombinationsoperationen notwendig. Dies kann, entsprechend der zugrunde liegenden Anwendungsfälle, folgendermaßen realisiert werden:

- manuelle, kooperative Erweiterung der automatisch erstellten Topic-Map mit neuen Begriffen, Assoziationen und Typisierungen,

⁴ Für diese Problemstellungen werden Überlegungen zur kooperativen Konstruktion von Ontologien an Bedeutung gewinnen (vgl. [Hols⁺02]).

- Integration der Möglichkeit manueller Annotationen (d. h. manuelle Zuordnung eines Dokumententeils zu einem Begriff, respektive Konzept),
- Integration von Rankings zur Vertrauensbildung,
- Integration von Diskussionsforen (und texttechnologische Nutzung des produzierten Texts) zu bestimmten Konzepten, Begriffen oder Zitaten.

Darüber hinaus sind durch die Nutzung des texttechnologischen Fundaments weiterführende Anwendungen denkbar, da der Kontext eines Wissensraum durch seine Dokumentenbasis spezifiziert ist. Dieser kann z. B. genutzt werden, um für die Besitzer des Wissensraums externe Dokumente automatisch zu recherchieren, welche in diesem Kontext relevant sein könnten.

5 Zusammenfassung und Ausblick

Das im Rahmen dieser Arbeit entwickelte IT-System wird als Keimzelle für die vollständige Umsetzung des Konzepts des Wissensraums in der Anwendungsdomäne der Projektdokumentation dienen. Von den notwendigen Eigenschaften der Themenorientierung, der Schlankeit, der Kooperation und der menschlichen Intervention wurden in der vorliegenden Arbeit die ersten beiden prototypisch umgesetzt. Technologisch wurde eine automatische Erstellung und Visualisierung einer individualisierten, domänenspezifischen Topic-Map zur nachhaltigen Nutzung von Projektdokumentation realisiert. Die Präsentation vor potenziellen Anwendern lässt eine entsprechende Weiterentwicklung des vorgestellten Konzepts als unbedingt angeraten erscheinen.

Das vorgeschlagene Konzept der Wissensräume scheint geeignet zu sein, um die Arbeit in wissensintensiven Prozessen durch Automatisierung und Kooperation zu unterstützen. Das im weiteren zu entwickelnde System sollte neben der Erweiterung um Komponenten der Kooperation und Intervention auch für beliebige Dokumentensammlungen in beliebigen Anwendungsfällen einsatzbereit sein. So wird es möglich sein, dass die Nutzer dieser Werkzeuge persönliche oder gemeinschaftliche, vertrauenswürdige, kontextspezifische Sichten auf beliebige Dokumentensammlungen erstellen. Die Entwicklung und Nutzung solcher Werkzeuge wird den Umgang mit der Masse der zur Verfügung stehenden Informationen enorm verändern und ermöglicht die Entwicklung neuartiger Geschäftsmodelle.

6 Literatur

- [BECK03] BECKER, Thomas: *Gezeitenstrom des Wissens*. In: Frankfurter Allgemeine Zeitung, 24.02.2003, 46 (2003), S. 20.

- [BIEM⁺03] BIEMANN, CHRISTIAN; QUASTHOFF, Uwe; BÖHM, Karsten; WOLFF, Christian: *Automatic Discovery and Aggregation of Compound Names for the Use in Knowledge Representations*. J.UCS - Journal of Universal Computer Science (Springer), Volume 9, Issue 6, S. 530-541.
- [BÖHM⁺02] BÖHM, Karsten; HEYER, Gerhard; QUASTHOFF, Uwe; WOLFF, Christian: *Topic Map Generation Using Text Mining*. J.UCS - Journal of Universal Computer Science (Springer), Volume 8, Issue 6, S. 623-633.
- [GERI00] GERICK, Thomas: *Topic Maps – der neue Standard für intelligentes Knowledge Retrieval*. In: Wissensmanagement 2 (2000), S. 8-12.
- [HEYE⁺02] HEYER, Gerard; QUASTHOFF, Uwe; WOLFF, Christian: *Möglichkeiten und Verfahren zur automatischen Gewinnung von Fachbegriffen aus Texten*. to appear in BULLINGER, H.-J. (ed.): *Tagungsband des Innovationsforums: „Content Management – digitale Inhalte als Bausteine einer vernetzten Welt“* Stuttgart.
- [HEYE⁺03] HEYER, Gerhard; MAICHER, Lutz: *Persönliche und gemeinschaftliche Wissensräume. Erfüllen Topic-Maps die technologischen Anforderungen?* Aus: FÄHNRIK, Klaus-Peter; HERRE, Heinrich (Hrsg.): *Content- und Wissensmanagement. Beiträge auf den LIT'03*. Leipzig (2003).
- [HOLS⁺02] HOLSAPPLE, Clyde W.; JOSHI, K. D.: *A Collaborative Approach to Ontology Design*. In: Communications of the ACM 45, 2 (2002), S. 42-47.
- [LEGR⁺01] LE GRAND, Bénédicte; SOTO, Michel: *Visualisation of the Semantic Web: Topic Map Visualisation*.
- [LENZ⁺02] LENZ, Eva Anna; BEIBWENGER, Michael; STORRER, Angelika: *Hypertextualisierung mit Topic-Maps – ein Ansatz zur Unterstützung des Textverständnisses bei der selektiven Rezeption von Fachtexten*. In: Vortragsband zu „XML Technologien für das Semantic Web – XSW 2002“, Springer (2002).
- [PEPP00] PEPPER, Steve: *The TAO of Topic Maps. Finding the way in the age of infoglut*. Verfügbar unter: <http://www.ontopia.net>
- [PEPP01] PEPPER, Steve: *Towards a General Theory of Scope*. Verfügbar unter: <http://www.ontopia.net>
- [SCHA02] SCHATZ, Bruce R.: *The Interspace: Concept Navigation Across Distributed Communities*. In: IEEE Computer 35, 1 (2002), S. 54-62.
- [SIHN⁺02] SIHN, Wilfried; HEEREN, Frank: *Xpertfinder – die Suchmaschine für Experten*. In: Tagungsband der KnowTech 2002.
- [THOM02] THOMPSON, Bryan: *The Cognitive Web. Presentation to the Semantic Web Interest Group*. Verfügbar unter: <http://www.cognitiveweb.org>