

Zusammenfassung des Beitrags zur KnowTech 2003

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

Lutz Maicher, Karsten Böhm, Prof. Dr. Gerhard Heyer, Universität Leipzig
Kontakt: {maicher|boehm|heyer@informatik.uni-leipzig.de}

Olaf Grahn, BearingPoint
Kontakt: olaf.grahn@bearingpoint.com

Das Problem

Die meisten Lösungen zum Wissensmanagement setzen an der Kategorisierung von Informationsobjekten entsprechend eines manuell erstellten Klassifikationsschemas an. Dies bedeutet, es wird ein unternehmensweit gültiges Begriffsnetz in einem Beratungsprozess manuell erstellt und jeder Mitarbeiter muss seine Informationen manuell den entsprechenden Begriffen zuordnen. Dies hat folgende Nachteile:

- *Kosten.* Die manuelle Erstellung dieser Netze ist teuer, deren Pflege (insbesondere in Branchen mit dynamischer Entwicklung) aufwändig und unbezahlbar.
- *Unwirtschaftlichkeit.* Die erkaufte Präzision ist für die meisten Anwendungsfälle nicht notwendig.
- *Unflexibilität.* Unternehmensweite Ontologien sind für ein personalisiertes Wissensmanagement zu unflexibel.

Die Idee

Jede inhaltlich hinreichend homogene Zusammenstellung von Texten beschreibt eine bestimmte Fachdomäne. Diese besitzt ihre eigenen, spezifischen Fachbegriffe und Zusammenhänge. Als Wissensraum bezeichnen wir die Gesamtheit einer solchen Dokumentenkollektionen, das zugehörige Begriffsnetzwerk und alle zusätzlichen Bemerkungen, die dessen Nutzer bei der Arbeit hinzugefügt haben. Mit Methoden der automatischen Sprachverarbeitung können diese individuellen Begriffsnetzwerke kostengünstig extrahiert und gleichzeitig die Informationen automatisch kategorisiert werden. Somit können durch das Zusammentragen von beliebigen Texten verschiedene, persönliche Wissensräume erstellen werden. Um die Akzeptanz des Ansatzes zu erhöhen, besteht immer die Möglichkeit einer individuellen Verbesserung der Ergebnisse entsprechend der persönlichen Sichtweise. Da wissensintensive Prozesse zumeist kooperativ sind, können auch Gruppen gemeinschaftliche Wissensräume erstellen. Unser Ansatz erlaubt es, die Produktivität dieser Prozesse durch Automatisierung der Erstellung von Wissensräumen für beliebige Fachgebiete zu erhöhen. Die Wirtschaftlichkeit wird zudem durch Unterstützung einer nachhaltigen Wiederverwendung bereits produziertem Wissens im Unternehmen erhöht.

Die Umsetzung

Für die automatische Extraktion der Begriffe und ihrer Relation werden Daten und Verfahren der automatischen Sprachverarbeitung von der gleichnamigen Abteilung am Institut für Informatik an der Universität Leipzig genutzt. Deren Kern ist die Referenzwortschatzanalyse. Als Grundlage dient eine umfangreiche Textdatenbank, in die in den letzten 8 Jahren 35 Millionen Sätze mit 500 Millionen Wortvorkommen aus allgemein verfügbaren Quellen des Deutschen integriert worden sind (vgl. <http://wortschatz.uni-leipzig.de>). Im Vergleich dazu wird das Vorkommen aller Worte

in dem Wissensraum gezählt. Jeder Begriff, der gegenüber dem gesamten deutschen Wortschatz überproportional häufig vorkommt, ist ein Fachwort in der aktuellen Domäne. In der Umgebung (Satz bzw. Absatz) dieser Fachworte in den Dokumenten des Wissensraums ist zudem das meiste domänenspezifische Fachwissen zu diesen Begriffen beschrieben. Das Verfahren der Referenzwortschatzanalyse funktioniert (sprachunabhängig) für beliebige Sammlungen von Dokumenten.

Die erstellten Begriffsnetzwerke sind Topic-Maps. Diese sind durch die ISO standardisiert und können mit XML eine von den Dokumenten getrennte Informationsschicht beschreiben, d. h. eine Topic-Map kann beliebige Dokumente referenzieren und jedes Dokument kann von beliebigen Topic-Maps referenziert werden. Darüber hinaus erlauben Topic-Maps das Erstellen von kontextspezifischen Sichten. Außerdem wird ihr Zusammenführen unterstützt, so dass individuelle Wissensräume zu gemeinschaftlichen verschmolzen werden können.

Beispielanwendung: Projektdokumentation

In Zusammenarbeit mit BearingPoint (<http://www.bearingpoint.de>) wurde ein individueller Wissensraum für ein abgeschlossenes Projekt erstellt. Alle aus Sicht eines Mitarbeiters relevanten Dokumente dieses Projektes, die Ergebnisdokumente der Projektarbeit, zusätzliche externe Informationen, z. B. Gutachten, und die persönliche, projektbezogene E-Mail-Korrespondenz waren die Basis für den Wissensraum. Das automatisch erstellte Begriffsnetzwerk setzt die relevanten Fachbegriffe innerhalb des Projektes in Beziehung. Somit kann der Mitarbeiter in zukünftigen Projekte gezielt auf alle Informationen des abgeschlossenen Projektes zugreifen.

UNIVERSITÄT LEIPZIG

Projektwissensraum
in Kooperation mit

KONZEPT → BEGRIFF → ZITAT

... Anbietern FileNet (DMS- ...

Visualisierung

Begriff zu diesem Zitat

■ FileNet

Zitat: Die erforderlichen technischen Komponenten wurden von den Anbietern FileNet (DMS- und Archiv-System), Fujitsu Siemens (Server-Hardware), Microsoft (Datenbank-System) und Headway (Kodak-Scanner) zur Verfügung gestellt

Quelldokument des Zitats