

LEARNTEC '03 – WORKSHOP

Aufbau und Einsatz von Begriffsnetzen zur semantischen Suche von Wissensmaterialien

Moderation:

Dr. Christian Mayerl

Fakultät für Informatik
Universität Karlsruhe (TH)
Am Fasanengarten 5, 76131 Karlsruhe

Tel: +49 (721) 608-6390 Fax: +49 (721) 697760

E-Mail: mayerl@ira.uka.de

Web: www.ira.uka.de

Zusammenfassung

Ziel des Workshops war ein lebendiger Erfahrungsaustausch bzgl. Aufbau und Einsatz semantischer Wissensnetze in der Aus- und Weiterbildung. Als Themenschwerpunkte wurden Anwendungsszenarien von Wissensnetzen, deren Einordnung in die Autoren-, Dozenten- und Lernprozesse, der Mehrwert im Hinblick auf die Qualität der Lehre, Aufwand für Aufbau und Pflege sowie Nutzen, Technologien und Lösungen (best practices, Produkte etc.) diskutiert. Weitere Informationen sind unter <http://alfi.ira.uka.de/learntec03/> zu finden.

Schlagworte

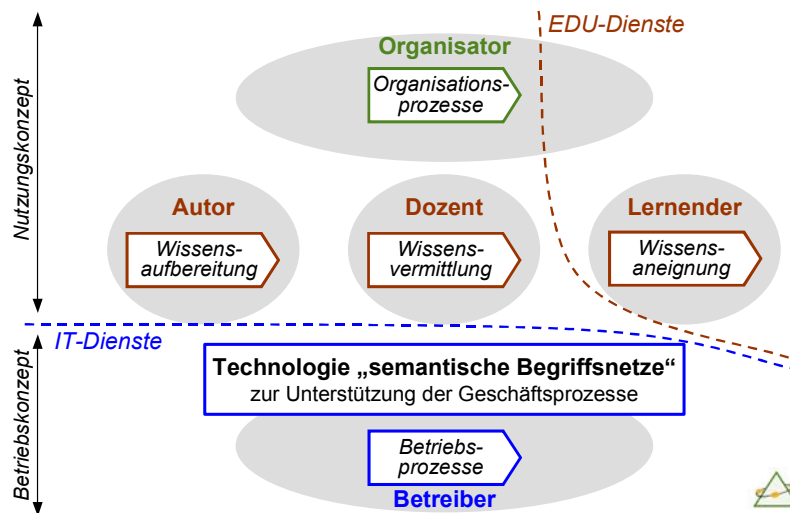
Szenario, Wissensnetz, Prozesse, Aufwand, Nutzen, Lehre, Technologie, Produkt

Einführung

Dr. Christian Mayerl
Universität Karlsruhe (TH)
mayerl@ira.uka.de

Der Workshop wurde von der Arbeitsgruppe Lehrunterstützung der Fakultät für Informatik (Leitung: Dr. Christian Mayerl) zusammen mit der Forschungsgruppe Cooperation & Management (Leitung: Prof. Dr. Sebastian Abeck) des Instituts für Telematik der Universität Karlsruhe (TH) veranstaltet. Die Themen, die während des Workshops diskutiert wurden, wurden anhand eines Beispielszenarios motiviert. Die folgende Abbildung zeigt hierzu den Einsatz von Technologien –

insbesondere von semantischen Begriffsnetzen – zur Unterstützung der Aus- und Weiterbildung.



Beispielszenario: Aus- und Weiterbildung

Für einen erfolgreichen Einsatz von semantischen Begriffsnetzen stellt sich die Frage, welche Anforderungen jeweils seitens der Nutzer, wie z.B. Autoren, Dozenten und Lernende, an Anwendungen zum Aufbau und zur Nutzung von Begriffsnetzen gestellt werden. Neben dem Nutzen galt es vor allem auch den mit derartigen Aufwand z.B. für den Betrieb zu diskutieren und Erfahrungen über erfolgreiche Projekte auszutauschen. Für die Diskussion wurden die folgenden Workshop-Beiträge eingereicht:

- Wissensnetz-basierte Suchanfragenoptimierung
- Semantische Wissensnetze reichen nicht aus
- Begriffsnetz für medizinisches Online-Lernprojekt
- Aufbau einer Topic Map
- DCC MindMap: Eine neuartige Such- und Darstellungstechnologie zum Aufdecken von Sinnzusammenhängen

Die einzelnen Beiträge werden im Folgenden zusammenfassend vorgestellt.

Wissensnetz-basierte Suchanfragenoptimierung

Karsten Krutz, Sebastian Abeck,
Universität Karlsruhe (TH)
 {krutz, abeck}@ira.uka.de

Zusammenfassung

Die Forschungsgruppe Cooperation & Management hat in den vergangenen Jahren ein Internetbasiertes Wissenstransfer-System entwickelt, das sie erfolgreich im täglichen Ausbildungsbetrieb einsetzt. Anhand dieser Lösung und den damit gesammelten Erfahrungen wird die Nutzung semantischer Netze zur Unterstützung der Suche nach Schulungsmaterial aufgezeigt.

Einleitung

Mit dem zunehmenden Angebot in der rechnergestützten Aus- und Weiterbildung wird der effizienten Suche zur Auswahl interessanter Materialien ein immer größerer Stellenwert beigemessen.

Den technischen Problemen durch unterschiedliche e-learning-Systeme und heterogenen Plattformen versucht man durch Standards zur Beschreibung von Wissensmaterial wie IMS[1] oder LOM[2] und Ablage dieser Daten im XML-Format Herr zu werden. Auf Basis dieser Informationen lässt sich eine System- und plattformübergreifende Suchdatenbasis erzeugen, die im Gegensatz zum Internet eine semantische Suche ermöglicht.

Diese Datenbasis ermöglicht zwar eine sehr umfassende und, durch Nutzung der Semantik, zielgerichtete Suche, doch auch bei diesen Suchansätzen scheitert man oft an dem Problem, die „richtige Suchanfrage“ zu finden. Die Suchergebnisse bei der Suche nach einem Begriff sind oft ungeeignet, da sie z.B. nicht aus dem gewünschten semantischen Kontext stammen. Eine Suche nach „Architektur“ kann Ergebnisse zur Architektur von Informatiksystemen liefern, aber auch aus dem Bereich des Bauingenieurwesens.

Für die Modifikation der Suchanfrage zum Erzielen besserer Suchergebnisse werden Informationen über die Begriffe des Kontextbereichs benötigt. Ein Wissensnetz zu diesem Kompetenzbereich kann diese Informationen liefern und somit eine Unterstützung bei der Suche bieten.

Anhand des in der Forschungsgruppe Cooperation & Management entstandenen ed.tec-Systems (educational technologies) wird die im Rahmen dieses Systems konzipierte und implementierte Wissensnetz-basierte Suchunterstützung vorgestellt.

Suchanfragenoptimierung durch Wissensnetze

Das zur Modifikation der Suchanfrage benötigte Wissen muss in geeigneter Form vorliegen, um es automatisch auswerten zu können.

Wissensrepräsentierende Datenstrukturen speichern Informationen in

maschinenlesbarer Form und stellen somit die Basis für rechnerunterstütztes Wissensmanagement dar. Neben dem im Folgenden vorgestellten Modell der semantischen Wissensnetze gibt es weiterhin eine Vielzahl anderer Möglichkeiten, Wissen zu repräsentieren. Für einen ausführlichen Überblick über die gesamte Komplexität des Bereichs der Wissensrepräsentation sei auf [3] hingewiesen.

Die semantischen Wissensnetze setzen einzelne Begriffe (in diesem Zusammenhang häufig als Konzepte bezeichnet) in Relationen. Diese Relationen können gerichtet oder ungerichtet sein, was durch Pfeile bzw. Linien repräsentiert wird. Die Wissensmaterialien (Inhalte) sind über ihre semantischen Beschreibungen (Schlüsselworte) mit den Begriffen des Wissensnetzes verbunden. Indem man die Beziehungen zwischen den Begriffen verwendet, kann man Suchanfragen (z.B. Suche nach dem Begriff A) effizienter gestalten.

Durch die Verwendung des AND-Operators ($A \text{ AND } B$) kann die Anzahl der Suchergebnisse reduziert und bei geeigneter Wahl von B die Qualität der Ergebnisse erhöht werden. Dazu muss B den gleichen Kontext wie A haben oder anders ausgedrückt A und B müssen semantisch nahe beieinander liegen. Dies ist z.B. bei hierarchischen Relationen wie „Ist-Teil-von“ oder „Ist-Verallgemeinerung-von“ gegeben.

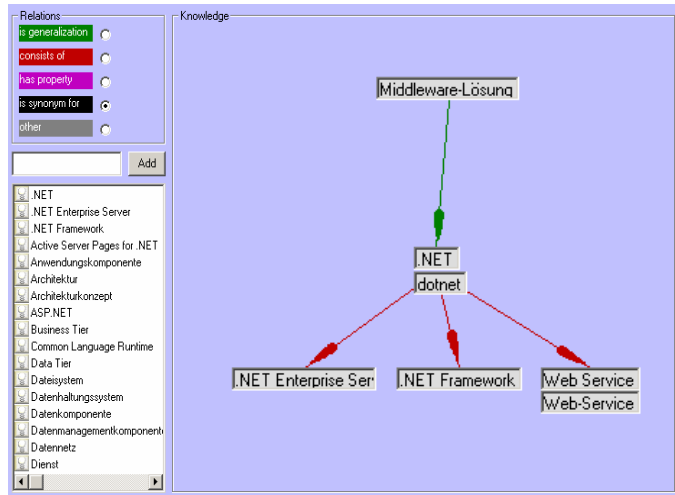
Durch die Verwendung des OR-Operators hingegen ($A \text{ OR } B$) kann die Anzahl der Suchergebnisse gesteigert werden. Hierbei sollte der zweite Begriff B eine sehr ähnliche Bedeutung wie der Begriff A besitzen. Daher bietet sich an, den Begriff B so zu wählen, dass er in Synonymrelation zu A steht.

Ausführlichere Überlegungen zur Verbesserung der Suchfunktionalität durch Suchanfragenoptimierung lassen sich unter [4] nachlesen.

Implementierter Lösungsansatz: der ed.search_supporter

Im Rahmen des an der Forschungsgruppe entwickelten ed.tec-Systems wurde ein Werkzeug, der ed.search_supporter, entwickelt, das den Anwender sowohl bei der Erstellung und Pflege als auch bei Nutzung von Wissensnetzen zur Suchanfragenoptimierung unterstützt.

In der folgenden Abbildung ist der ed.search_supporter mit einem Ausschnitt eines Wissensnetzes dargestellt. Die Visualisierung der Relationen wird durch verschiedenfarbige Pfeile, analog der links oben angezeigten Legende realisiert. Synonyme werden als direkt untereinander angeordnete Begriffe dargestellt.



Erstellung und Pflege eines Wissensnetzes mit dem ed.search_supporter

Zur Suche gibt der Anwender ein Schlüsselwort ein, wie er es von vielen Internetsuchmaschinen gewohnt ist. Der ed.search_supporter generiert daraus eine Menge von Suchanfragen für die verschiedenen Optimierungskriterien, aus denen eine geeignete ausgewählt und direkt an eine WWW-Suchmaschine (z. B. Google) abgesetzt werden kann.



Suchanfragenoptimierung mit dem ed.search_supporter

Ausblick

In diesem Beitrag wurde eine semantische Suchunterstützung im Rahmen des Internet-basierten Wissenstransfers vorgestellt und am Beispiel des ed.tec-Systems verdeutlicht. Hierbei wurden die Vorzüge gegenüber einer Suche ohne Unterstützung der Suchanfragengenerierung durch Wissensnetze aufgezeigt.

Die vorgestellte Lösung ist bereits einsatzbereit und wird in der Lehre der Forschungsgruppe verwendet, lässt sich aber noch weiter ausbauen. In einem weiterführenden Schritt wird die Funktionalität des Werkzeugs Web-basiert angeboten und eine engere Integration der Such- und Navigationsfunktionalität in die Lernumgebung realisiert.

Dabei wird den Lernenden die Möglichkeit geboten werden, Suchanfragen aus dem Kontext der Schulung über das bereitgestellte Wissensnetz optimieren zu lassen. Diese Suchanfragen können dann zur Suche weiterführenden Materials im Internet oder anderen Schulungsmaterialablagen genutzt werden.

Literatur

1. Instructional Management Standard - Global Learning Consortium, Designing the Future of Global Distributed Learning, <http://www.imsproject.org/>
2. Learning Technology Standards Committee, <http://grouper.ieee.org/p1484>
3. Reimer, U., Einführung in die Wissensrepräsentation - netzartige und schema-basierte Repräsentationsformate. Leitfäden der angewandten Informatik, ed. W. Stucky. 1991, Stuttgart: Teubner.
4. Eck, M., Wissensnetze zur Navigation und Semantischen Suche in der Internetbasierten Aus- und Weiterbildung. 2001: Diplomarbeit an der Universität Karlsruhe in der Forschungsgruppe "Cooperation & Management" (Prof. Abeck).

Semantische Wissensnetze reichen nicht aus

Khaldoun Ateyeh, Michael Klein, Birgitta König-Ries, Jutta Mülle
Universität Karlsruhe (TH)

[ateyeh, kleinm, koenig, muelle}@ira.uka.de](mailto:{ateyeh, kleinm, koenig, muelle}@ira.uka.de)

Semantische Begriffsnetze dienen der inhaltlichen Strukturierung von Fachgebieten. Sie können daher verwendet werden, um Lehr-/Lernmaterialien zu annotieren. Ein Zusammenhang, in dem dies von großer Bedeutung ist, ist die Wiederverwendung existierender Materialien für neue Kurse – ein Ziel, das sich zahllose existierende Plattformen auf die Fahnen geschrieben haben und das fast durchweg durch Begriffsnetze realisiert werden soll. Unsere These ist nun: Die heute üblichen semantischen Begriffsnetze, die lediglich die Inhalte der Lernmaterialien erfassen, reichen hierzu bei Weitem nicht aus. Benötigt werden zusätzliche Begriffsnetze, die andere, ebenso wichtige zum Inhalt orthogonale Dimensionen der Lernmaterialien beschreiben. Erst eine solche mehrdimensionale Beschreibung ermöglicht eine effektive Wiederverwendung von Lernmaterialien.

Die folgenden Ausführungen sollen diese These belegen: Betrachtet man eine beliebige Sammlung von Lernmaterialien zu einem gegebenen Thema, so wird man feststellen, dass es, trotz des inhaltlichen Zusammenhangs eben nicht möglich ist, unterschiedliche Materialien einfach zu kombinieren und so einen neuen Kurs zu erstellen. Stammt ein Beispiel etwa von einem anderen Autor als eine theoretische Einführung, so werden in aller Regel weder das Layout, noch die Notation, weder die didaktische Herangehensweise noch der Detaillierungsgrad zueinander passen. Will man eine einfache Kombination dieser Materialien ermöglichen, so muss man den Inhalt von den anderen Aspekten trennen und jeden Aspekt getrennt auffindbar und manipulierbar machen. Unserer Erfahrung nach sind die Aspekte Inhalt, Struktur und Präsentation die Aspekte, die im Falle ihrer Vermischung massive Probleme bei der Anpassung der Lehrmaterialien und damit bei ihrer Wiederverwendung bereiten. Dies ist allerdings eine echte Herausforderung: Zum Einen behindern existierende Tools die Aspekttrennung eher, als das sie sie unterstützen, zum anderen sind die Aspekte nicht hundertprozentig orthogonal und zum Dritten erfordert die Aspekttrennung ein generelles Umdenken schon bei der Kurserstellung.

Ein Begriffsnetz für ein medizinisches Online-Lernprojekt

Andreas Faatz, Cornelia Seeberg, Ralf Steinmetz
Technische Universität Darmstadt
 {afaatz, seeberg, steinmetz}@kom.tu-darmstadt.de

Zusammenfassung

Innerhalb des Projektes k-Med entsteht ein Begriffsnetz, welches zur Verwaltung, Beschreibung und Anzeige multimedialer Unterrichtsmaterialien dient. Das folgende Papier zeigt, wie im Projekt k-Med ein Erstellungsprozess für das Begriffsnetz definiert wurde.

Einleitung

k-Med (knowledge-based Multimedia Medical Education) ist ein Verbundprojekt im Rahmen des Förderprogrammes „Neue Medien in der Bildung“.

Innerhalb des Projektes k-Med entsteht ein medizinisches Begriffsnetz, auch Ontologie genannt, welches zur Verwaltung, Beschreibung und Anzeige multimedialer Unterrichtsmaterialien dient. Dem Begriffsnetz kommt durch die besondere Anforderung der Modularisierung des multimedialen Unterrichtsmaterials zentrale Bedeutung zu. Das folgende Papier zeigt, wie im Projekt k-Med ein Erstellungsprozess für das Begriffsnetz definiert wurde. Wir wenden dabei die Methodik von Holsapple und Joshi [4] und erklären deren Umsetzung in den einzelnen Erstellungsphasen. Zusätzlich beschreiben wir, dass automatische Verfahren diesen Prozess unterstützen können.

Überblick

Der vorliegende Abschnitt hat das Ziel, einen kurzen Überblick über den Ontologierstellungsprozess in k-Med zu liefern.

Das Begriffsnetz im Projekt k-Med besteht aus medizinischen Fachbegriffen, die durch sinnträchtige Relationen miteinander in Bezug gebracht werden. Der Grund für die Einführung einer solchen Datenstruktur liegt darin, dass eine möglichst vielfältig einsetzbare Wissensbasis, bestehend aus den k-Med-Modulen und ihren semantischen Beschreibungen, entstehen soll - ein Vorgehen, das die Wiederverwendung der Unterrichtsmaterialien ermöglichen wird. Die Begriffe eines solchen Netzes sind verbunden durch Ober-/Unterbegriffsrelationen und fachspezifische Relationen.

Das Begriffsnetz wird von den medizinischen Autoren gemeinsam erstellt. Dieses Vorgehen nennen wir *Kollaboration*.

Zentral hierbei ist ein abgestimmtes Vorgehen der medizinischen Autoren im Projekt. Die Abstimmung betrifft die Festlegung eines gemeinsamen Vokabulars, die Festlegung der verbindenden Relationen, den Detaillierungsgrad des Begriffsnetzes, die Auflösung von Konfliktfällen bei sich überschneidenden Wissensgebieten, die Rolle eines Moderators bei der Erstellung.

Wir beziehen uns von nun an auf die typischen Kollaborationsphasen nach Holsapple und Joshi [4].

Ablauf

Die *Vorbereitungsphase* erfolgte durch eine Koordination der Bedürfnisse der k-Med Autoren. Wir halten hiermit die Ergebnisse der *Vorbereitung* fest:

Als *Quelle* der Begriffe und der Hierarchie des Begriffsnetzes dient demnach zunächst der jeweilige Gegenstandskatalog des jeweiligen medizinischen Faches. Gegenstandskataloge existieren für alle vorklinischen Fächer der Medizin und umfassen die verbindlich zu unterrichtenden Themengebiete. Vorteilhaft hieran ist, dass die Gegenstandskataloge Medizinstudierenden oftmals als Anhaltspunkt der Stoffmenge bei Prüfungsvorbereitungen dienen. Für den *Detaillierungsgrad* der Begriffe im Begriffsnetz besteht folgende Vereinbarung:

Außer den aus dem Gegenstandskatalog stammenden Begriffen wird es lediglich diejenigen spezielleren Begriffe im Begriffsnetz geben, welche nötig sind, um den Inhalt eines Moduls zu charakterisieren. Hierzu ein Beispiel aus der Pharmakologie: Aus dem Gegenstandskatalog und seiner Untergliederung gewinnen wir die hierarchische Beziehung 'Antidepressiva'---> ist Unterbegriff von---> 'Eingriffe in das zentrale Nervensystem'---> ist Unterbegriff von ---> 'Allgemeine Pharmakologie und Toxikologie' --->ist Unterbegriff von---> 'Pharmakologie'.

Die Projektteilnehmer aus der Pharmakologie können diese Vorgabe des Gegenstandskataloges nun als hinreichend zur Beschreibung ihrer Module akzeptieren oder für ein Modul zum Thema 'Prozac' noch den Unterbegriff 'Prozac'---> ist Unterbegriff von 'Antidepressiva' einführen. Dieser *Detaillierungsgrad* ist allerdings nur dann erwünscht, wenn tatsächlich ein solches Modul zum Thema 'Prozac' in k-Med oder in einem darauf aufbauenden

Folgeprojekt produziert wird. Wir raten aus Gründen der Übersichtlichkeit und der strukturierten Suchanfragen dringend davon ab, den *Detaillierungsgrad* des Begriffsnetzes über den Gegenstandskatalog hinauszuführen, wenn kein Modul zum detaillierten/speziellen Fachbegriff (in unserem Beispiel 'Prozac') erstellt wird oder bereits vorliegt. Umgekehrt sollte zu jedem Modul, das für k-Med produziert wird, ein ausreichend genauer Fachbegriff im Begriffsnetz enthalten sein. In diesen Fällen wäre das Begriffsnetz von den Fachautoren zu erweitern. Obwohl das Werkzeug zum Anlegen des Begriffsnetzes, der Knowledge Builder der Firma intelligent views [5], eine freie Definition von Relationen und Begriffen zulässt, wird es auch Konventionen in Form eines vom Moderator vorgegebenen initialen, durch die medizinischen Autoren erweiterbaren Begriffsnetzes und eines nur durch den Moderator erweiterbaren Basissatzes an Relationen geben. Als *Relationen* können von den Autoren neben der Ober-Unterbegriffsrelation etwa 30 weitere verwendet werden.

Jede *Relation* besitzt eine definierte umgekehrte Richtung (beispielsweise 'ist Oberbegriff von' für 'ist Unterbegriff von').

Die Erstellung der Relationsnamen erfolgt durch den *Moderator*. Die Autoren verwenden ausschließlich solche Relationen, die der *Moderator* zuvor benannt hat. Eine Besonderheit der *Kollaboration* ist die mögliche *Überschneidung der verwendeten Terminologie*. Verwenden zwei oder sogar mehrere beteiligte Fachgebiete den gleichen Begriff, so müssen die betroffenen Autoren eine Vereinbarung darüber treffen, ob der Begriff verschmolzen wird, das heißt, einunderselbe Begriff von beiden Autoren benutzt und gepflegt wird, oder ob der Begriff unterschieden wird. Für den letzteren Fall gibt es eine Namenskonvention, die vorsieht, bei Unterscheidung hinter dem eigentlichen Namen noch das Fachgebiet anzugeben. Beispiel: verwenden sowohl die Biochemie als auch die Pharmakologie den Begriff 'Säure', so sollten sie gemeinsam (nach Absprache!) den Begriff weiterpflegen, wenn sein Sinngehalt aus dem Blickwinkel der beiden Fachgebiete so einheitlich ist, daß eine gemeinsame Nutzung widerspruchsfrei verläuft. Wäre allerdings Säure für den Biochemiker etwas, das von seiner Definition und Verwendung her in einem fundamentalen Gegensatz zur Definition und Verwendung in der Pharmakologie stünde, so müßte 'Säure' zu den beiden Begriffen 'Säure (Biochemie)' und 'Säure (Pharmakologie)' aufgelöst werden, um Konflikte bei der Erweiterung des Begriffsnetzes zu vermeiden. Grundsätzlich ist beim Neuanlegen eines Begriffes anhand der alphabetischen Liste des Erstellungswerkzeuges zu prüfen, ob der Begriff schon existiert. Der Moderator wird allerdings an dieser Stelle regelmäßig eine Überprüfung der vorhandenen Begriffe vornehmen und betroffene Autoren zur Begriffsklärung veranlassen.

In der *Verankerungsphase* wird für die k-Med Autoren der Gegenstandskatalog nebst seiner Hierarchie elektronisch umgesetzt, sodass er zur Erweiterung zur Verfügung steht. Insofern bildet der Gegenstandskatalog das initiale Begriffsnetz. In der Phase der *iterativen Verbesserung* wird das initiale Begriffsnetz erweitert und gepflegt. Dies obliegt vollständig den Fachautoren unter Anleitung des Moderators. Erweiterungen bestehen aus Einführungen speziellerer Begriffe gemäß der Verabredungen zum Detaillierungsgrad und einer Quervernetzung des zunächst nur hierarchisch (in Form von Ober-/Unterbegriffsbeziehungen)

gegliederten Begriffsnetzes durch die in der Vorbereitungsphase verabredeten Relationen. Pflegeschritte bestehen in der Identifikation von Überschneidung und deren Handhabung, wie sie in der Vorbereitungsphase verabredet wurde. Die bereits im Gegenstandskatalog vorhandenen terminologischen Überschneidungen werden hierbei vom Moderator ermittelt und den Autoren mitgeteilt. Die *Anwendungsphase* umfasst zum einen sowohl Definition von Anfragen an das Begriffsnetz als auch die Pflege und Beschreibung von k-Med-Modulen.

Automatische Unterstützung des Autorenprozesses

Neben der Anpassung des Ontologieerstellungsprozesses an die Gegebenheiten des Projektes k-Med stellt sich, auch aufgrund erster Erfahrungen mit der Motivation der beteiligten Autoren, die Frage, an welchen Stellen des Verfahrens eine automatische Unterstützung der Autoren vollzogen werden kann. [1] und [AF01] stellen zwei Varianten eines sogenannten Anreicherungsverfahrens dar, welches prinzipiell von der in [2] getroffenen Annahme ausgeht, dass ähnliche sprachliche Verwendungskontexte von Fachbegriffen in Fachtexten auch eine Ähnlichkeit der Begriffe an sich implizieren. Neuartig an der vorgeschlagenen Methode ist die Optimierung der Einflußgrößen. Beispielhaft für die resultierenden Ähnlichkeitsdefinitionen ist dabei die Umsetzung folgender Ausgangssituation:

Wenn in einer großen Anzahl von Fachtexten der Begriff „Kolibakterium“ häufig im gleichen Satz mit dem Begriff „Darminfektion“ oder „Infektionskrankheiten“ auftaucht, so kann dies ein Hinweis darauf sein, dass andere Fachbegriffe aus der gleichen Textmenge, die ebenfalls mit „Darminfektion“ oder „Infektionskrankheiten“ verwendet werden, einen ähnlichen Sinngehalt wie „Kolibakterium“ besitzen. In der Praxis sammeln sich häufig mehrere Hundert Kontextmerkmale an, so dass die Frage an Bedeutung gewinnt, welche davon ausschlaggebend für die Ähnlichkeitsdefinitionen sind, und welche in den Berechnung sogar vernachlässigt werden können. Hierfür wird in [1] ein Optimierungsproblem zur Auswahl relevanter Merkmale abgeleitet. Das Resultat des automatischen Anreicherungsverfahrens sind Begriffsvorschläge, die jeweils automatisch an eine möglichst sinnvolle Stelle des bestehenden Begriffsnetzes platziert werden. Bei ersten Experimenten mit dem Ansatz gelang es, eine Gruppe von Symptomen automatisch zu identifizieren und als zu ergänzende Begriffe in einer Testontologie über Durchfallerkrankungen vorzuschlagen.

Ausblick

Ein vereinheitlichter Erstellungsprozess vereinfacht wesentlich die Arbeit bei der Modellierung von Begriffsnetzen in der Lehre. Die Autoren in einem Projekt von der Größenordnung, die bei k-Med vorliegt, brauchen klare Handlungsanweisungen. Daher werden sich unsere zukünftigen Arbeiten mit einer verbesserten Integration von automatischen Anreicherungsverfahren und dem kollaborativen Erstellungsprozess befassen.

Literatur

- [1] Andreas Faatz, Stefan Hoermann, Cornelia Seeberg and Ralf Steinmetz. Conceptual Enrichment of Ontologies by means of a generic and configurable approach. In In Proceedings of the ESSLLI 2001 Workshop on Semantic Knowledge Acquisition and Categorisation, August 2001.
- [2] 1968a. Mathematical Structures of Language New York: Interscience Publishers John Wiley & Sons, 1968
- [3] Stefan Hoermann and Ralf Steinmetz. k-MED: Kurse gestalten und adaptieren mit rhetorisch-didaktischen Relationen. In Rechnergestützte Lehr- und Lernsysteme in der Medizin. J. Bernauer and M. R. Fischer and F. J. Leven and F. Puppe and M. Weber, April 2002.
- [4] Clyde W. Holsapple and K.D. Joshi: A collaborative approach to ontology design, Communications of the ACM, Vol. 45, 2, February 2002
- [5] www.intelligent-views.com

Aufbau einer Topic Map

Frank Schuster
Universität Karlsruhe (TH)
schuster@ira.uka.de

In dem von mir geleiteten Praktikum „Web-Technologien“ wird in diesem Semester ein System realisiert, mit dem alle Dokumente (HTML-Seiten und PDF-Dokumente) der Universität Karlsruhe erfasst werden.

Das System besteht aus mehreren Modulen. Das erste Modul generiert und verwaltet einen Katalog aller Dokumente. Darauf baut die Metadatenverwaltung auf, welcher vorhandene Daten übergeben werden.

Ebenso können neue Daten manuell eingepflegt werden. Das dritte Modul bildet eine Volltextsuche, bei der einfache statistische Bewertungen vorgenommen werden.

Auf dieser Grundlage werden weitere Module entwickelt:

- Ein Topic Map Modul für die Verwendung von semantischen Netzen.
- Erzeugung von Topic Maps beispielsweise aus RDF-Ontologien.
- Ein Modul zur Gewinnung von unscharfen (Fuzzy) Dokumentmerkmalen, welche unscharfe Assoziationen zu Topic Maps ermöglichen.
- Ein Bewertungsmodul, das mit neuronalen Ansätzen die Zugehörigkeit eines Dokuments zu vorhandenen Bereichen (Scopes) einer Topic Map bewertet.
- Erweiterung der Suche über Topic Maps.
- Ein Modul zur Reihung und Cluster-Bildung von Suchergebnissen.

- Die bereits im Projekt Solaris (Neugestaltung von www.hfg-karlsruhe.de) verwendeten Benutzer-Links sollen übernommen und auf die Topic-Maps abgebildet werden.

Wenn die Entwicklungen abgeschlossen sind, werden zwei Anwendungsszenarien umgesetzt. Zum einen soll eine Organisationsstruktur in der Topic Map modelliert werden und mit den Dokumenten der Universität Karlsruhe verknüpft werden. Ziel dabei ist die Verbesserung von Arbeitsabläufen durch eine gezielte Kontaktaufnahme zu den jeweiligen Verantwortlichen. Im zweiten Szenarium sollen vor allem wissenschaftliche Arbeiten (PDF) indexiert werden. Hierbei soll eine interdisziplinäre Suche von Dokumenten ermöglicht werden.

DCC MindMap : Eine neuartige Such- und Darstellungstechnologie zum Aufdecken von Sinnzusammenhängen

Dr. Thomas Luba
Universität Hildesheim / IMBIT
thomas.luba@web.de

Ausgangslage

Ausgangspunkt für die nachfolgenden Betrachtungen sei der folgende situative Kontext:

- Informationsbedarf: Der Benutzer führt eine spezifische, problemorientierte Recherche durch, wobei wegen der Neuartigkeit der Problemstellung die Suche durchaus explorativen Charakter annehmen kann. In der Praxis ergibt sich diese Situation beispielsweise bei arbeitsplatzorientiertem Lernen.

An Materialien stehen zur Verfügung:

- Sammlungen narrativer Texte wie Zeitschriften, Protokolle: Typischerweise unterschiedliche Autorenschaft und Zielrichtung, wechselnde Terminologie und mehrdeutige Möglichkeiten der Kategorisierung.
- Spezifische Werke wie Studienbriefe und Handbücher, Nachschlagewerke: Typischerweise als linearer Lehrgang innerhalb einer stark hierarchischen, häufig organisatorischen Anordnung (Handbuch für den Administrator, für den Operator ...). Vertiefende Materialien meist abseits als Textsammlung.

Zusammenfassend lässt sich beobachten:

- Externe redaktionelle Eingriffe bzw. Regelsysteme sind häufig nicht transparent, Zusammenhänge sind weder beliebig pflegbar noch in ihrer Relevanz vollständig vorhersehbar (Mangel an Querverweisen).

- Erfahrungswissen liegt meist schlecht aufbereitet vor, Lehrsammlungen selten problemspezifisch.

Lösungsansatz

Im vorliegenden Fall fiel die Wahl auf die Extraktion von gedanklichen Assoziationen aus den Materialien selbst.

Hierzu werden die Kontexte von Worten in Texten betrachtet [1]. Zum Beispiel:

- VW Tochter Skoda wächst stürmisch Prag dpa Der tschechische **Autohersteller** Skoda legt für ... eine Rekordbilanz vor. Der Umsatz stieg ...
- Skoda wächst weiter AUTOBRANCHE Der tschechische **Automobilhersteller** Skoda hat ... den Absatz gesteigert. Die Tochter des Wolfsburger Volkswagen Konzerns ...

Diese Text-Fragmente legen allein aufgrund psychologischer Aspekte ohne vorgegebenen Wortschatz oder vorgegebene Terminologie folgenden Zusammenhang nahe:

- Skoda ist mit Autohersteller bzw. Automobilhersteller assoziiert.
- Autohersteller und Automobilhersteller sind assoziiert.
- Die Art der Assoziation ist ohne linguistisches Wissen nicht zu klären, erschließt sich aber hier unmittelbar durch Nachlesen der Fundstelle bzw. Allgemeinwissen des Lesers. ('is-a' Beziehung, Synonymie).
- Es entsteht eine 'horizontale' Ebene zur 'vertikalen' redaktionellen Ebene

Umsetzung mittels vollautomatisch generierter Mindmaps

Die DCC (Didactic Content Concept) Mindmap liefert wie folgt ein für den Informationsbedarf spezifisches Angebot:

- Assoziationsstärke wird durch Abstände angedeutet. Visuelle Clusterbildung ermöglicht die Interpretation der Reichhaltigkeit und Bewertung der Beziehungen.
- Beziehungen werden durch Dokumente unterlegt, die als Quellennachweis dienen. Dabei können Dokumente auch Autoren etc. repräsentieren.

Suchprozess: Informationsbedarf interaktiv spezifizieren

Als Datenbestand dienen ca. 27000 Artikel der Rubrik 'Wirtschaft' einer überregionalen Zeitschrift aus einem Zeitraum von 10 Jahren, insgesamt rund 50MB an Textmaterial. Es werde Information zum Thema 'Chipkarte' gesucht. Je häufiger der Begriff, desto mehr generalisiert die Map, insbesondere tritt Unterspezifizieren nicht mehr auf, also direkt mit 'Chipkarte' starten (siehe Abbildung).

Anscheinend hat 'Chipkarte' im Datenbestand in drei Bereichen Bedeutung:

- als 'Zahlungsmittel'. Als Varianten finden sich 'Plastikkarte', 'Kreditkarten', 'Geldkarte', 'EC'-Karte'. Daneben wird anscheinend das Abheben an 'Automaten', 'Geldautomaten', 'Terminals' thematisiert und das Bezahlen in 'Restaurants' ...
- im Kontext von 'Krankenkasse', 'Arzt' und 'Patienten'. Ein Blick in das in diesem Teil der Map stehende Dokument '16567' beschreibt tatsächlich 'Mißbrauch' der Chipkarte in diesem Kontext.
- im Zusammenhang von 'Sicherheit' und 'Missbrauch', diffuser mit 'PC', 'Daten', 'Handel'. Klarheit verschafft ein Blick in das in diesem Teil der Map stehende Dokument '19807': Tatsächlich geht es um das verschlüsselte Bezahlen am PC via Internet.

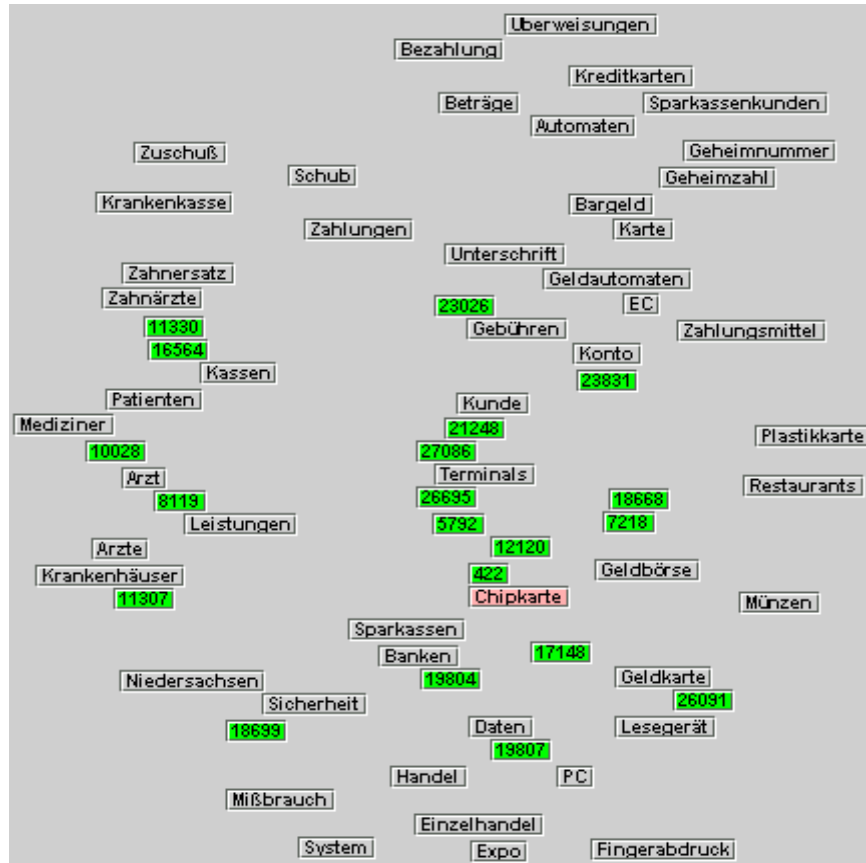
Durch sukzessives Hinzunehmen je eines Begriffs aus der Map ergibt sich eine zunehmende Fokussierung des Themas. Beispielsweise identifiziert die Anfrage 'Chipkarte', 'Missbrauch', 'Krankenkasse' genau einen Artikel, nämlich bereits oben erwähnten. Der Artikel findet sich dann mittig in der Map positioniert und die Begriffe herum lassen sich als Zusammenfassung der Themen, die im Kontext der Anfrage relevant sind, deuten.

Suchprozess: Dokumentliste 'fehlertolerant' vervollständigen

Neben der unklaren Terminologie ist oft die schiere Variation in den Schreibweisen, von Tippfehlern ganz zu schweigen, ein wesentliches Hindernis, um relevante Dokumente zu finden, in Fall 'Chipkarte' komplexere Formen wie

- 'Geldkartenchips'
- 'Euroscheck-Karten auch mit einem **Chip** ausgerüstet'

Für eine solche Expansion der Suchergebnisse mit Ähnlichkeitsbewertung wurde auf die bereits am IMBIT entwickelte Technologie der spärlich kodierten Assoziativspeicher (SpaCAM) zurückgegriffen [2],[3].



Suchanfrage 'Chipkarte' (gestauchte Darstellung)

Danksagung

Ich bedanke mich bei der Universität Hildesheim / IMBIT unter Prof. Bentz und bei Herrn Bernhard Rieke von der H.U.T GmbH, Hildesheim, für die herzliche Unterstützung.

Literatur

- [1] 'Statistische Korpusanalyse zum Extrahieren von semantischen Wortrelationen', Martin Ackermann, Dissertation, Universität Hildesheim, 2000
- [2] 'Einsatz der SpaCAM-Technik für ausgewählte Grundaufgaben der Informatik', Michael Heitland, Dissertation, Universität Hildesheim, 1994
- [3] 'Konkordanzen in langen Zeichenketten', Thomas Luba, Dissertation, Universität Hildesheim, 2001