

*G. Szwillus, J. Ziegler (Hrsg.): Mensch & Computer 2003: Interaktion in Bewegung.
Stuttgart: B. G. Teubner, 2003, S. 99-112*

Verbindung heterogener Experten-Communities durch die Entdeckung, Visualisierung und Nutzbarmachung von stillem Wissen – das AWAKE Projekt

Jasminko Novak, Monika Fleischmann, Wolfgang Strauss
Fraunhofer Institut für Medienkommunikation, MARS Exploratory Media Lab

Michael Wurst, Katharina Morik
Universität Dortmund, Abteilung Künstliche Intelligenz

Christoph Kunz, Jürgen Ziegler
Fraunhofer Institut für Arbeitswirtschaft und Organisation, Competence Center Human-Computer
Interaction

Zusammenfassung

Dieser Beitrag beschreibt einen Ansatz zur empirischen Untersuchung der Frage, wie implizite Wissensstrukturen in unterschiedlichen Experten-Communities entdeckt, visualisiert und nutzbar gemacht werden können. Personalisierte Zugänge und kollaborative Strukturierung von Informationsräumen sind die praktischen Anwendungskontexte. Der Schwerpunkt liegt auf dem entwickelten Modell und dem System zur Erstellung und kollaborativer Nutzung von personalisierten Wissenskarten und kollaborativen Begriffsnetzen. Der entwickelte Prototyp verbindet semantische Textanalyse mit maschinellem Lernen und mit Interfaces zur Visualisierung von Kontexten und Beziehungen sowie zum Aufbau neuer Wissensstrukturen. Zu den Anwendungsszenarien gehören die automatische Generierung von personalisierten Wissensportalen, kollaborative semantische Exploration von komplexen Informationsräumen und der Aufbau von gemeinschaftlichen Ontologienetzwerken. Die Internetplattform netzspannung.org, ein Wissensportal das digitale Kunst, Kultur und Informationstechnologie verbindet, dient als Entwicklungskontext und empirische Testumgebung.

1 Einleitung

Der Ausgangspunkt unseres Ansatzes ist, dass implizite Wissensstrukturen (stilles Wissen), die von einer Nutzergruppe geteilt werden, maßgeblich für die Kommunikation und die gemeinschaftliche Nutzung von Wissen sind. Als entscheidendes Element für die Entwicklung eines Modells zum Aufbau einer Wissensgemeinschaft, die Experten aus unterschiedlichen Fachgebieten verbindet, ergibt sich daraus die folgende Fragestellung:

Wie können bereits bestehende, aber nicht explizit formulierte Wissensstrukturen einer bestimmten Experten-Gemeinschaft entdeckt, visualisiert und für die kooperative Entdeckung und Konstruktion von Wissen in heterogenen Informationspools nutzbar gemacht werden?

Bei der Formulierung eines praktischen Ansatzes zu dieser Fragestellung stellen wir die folgenden Definitionen und Einschränkungen auf. Wir setzen das Konzept der "Wissensfindung" (Fayyad et al 1996) in Bezug zur Unterstützung der Entdeckung von semantischen Zusammenhängen und Beziehungen in einem Informationspool, der entweder 1) zu umfangreich ist oder zu schnell

wächst, um manuell gescannt und klassifiziert zu werden, oder 2) aus zu heterogenen Inhalten besteht, um mit einer bestimmten Klassifizierung belegt zu werden, oder 3) unterschiedliche Nutzergruppen mit heterogenen Interessen bedient.

In diesem Beitrag stellen wir ein Modell zur Entdeckung und Sichtbarmachung impliziter Wissensstrukturen einzelner Nutzer und Nutzer-Gruppen vor, und diskutieren die Möglichkeiten der Nutzung dieser Wissensstrukturen zur semantischen Navigation und Auffindung von Beziehungen in heterogenen Informationsräumen. Wir beschreiben die implementierten Prototypen und stellen die Möglichkeiten zur kollaborativen Nutzung des Systems dar. Ebenso wird die Anwendung des entwickelten Systems zum Aufbau eines Wissensportals, das unterschiedliche Experten-Gemeinschaften verbindet, anhand des Beispiels „netzspannung.org“ erläutert.

2 Grundkonzept: Personalisierte lernende Wissenskarten

Unser Ausgangspunkt kann auf die Sichtweise von Nonaka & Takeuchi bezogen werden, welche argumentiert dass Wissen vor allem aus persönlichen, schwer artikulierbaren und teilweise unbewussten Komponenten besteht, die in der Regel als implizites oder stilles Wissen bezeichnet werden (Nonaka & Takeuchi 1995). Demnach liegt der Schlüssel zur Kommunikation und Nutzung des bereits vorhandenen Wissens sowie zum Aufbau von neuem Wissen, in der Externalisierung von implizitem Wissen in erkennbare und von anderen nutzbare Strukturen (Kodifizierung).

Als konkreten Untersuchungsrahmen für diese Fragen nehmen wir den Kontext der Informationssuche und der Exploration eines Informationsraumes. Dieser kann als ein Prozess verstanden werden, in dem der Nutzer durch seine Interaktion mit Informationen sein vorhandenes Wissen widerspiegelt und neue Wissensstrukturen entwickelt.

Um die Wissensstrukturen des Nutzers in Bezug zu einem Informationsraum zu stellen, ziehen wir die Metapher einer Wissenskarte heran. In unserem Ansatz beschreibt sie ein Modell zur Darstellung von Informationsräumen, in dem die Einzelinformationen nach ihrer möglichen Bedeutung und semantischen Beziehung strukturiert sind. Die Leitfrage ist, wie solche Karten aufgrund von Nutzerinteraktionen mit Informationen aufgebaut und für eine kollaborative Auffindung und gemeinsame Nutzung von Wissen angewandt werden kann.

Als eine konkrete Lösung haben wir das folgende Modell entwickelt. Als Ausgangspunkt für seinen Umgang mit den Informationen, wird dem Nutzer eine semantische Strukturierung des Informationsraumes angeboten, die autonom durch das System erstellt wird (z.B. durch die Verwendung von Dokument-Clustering Methoden). Dies ist die systemgenerierte Wissenskarte.

Der Nutzer kann diese vorgeschlagene Struktur nutzen, um im Informationspool zu navigieren und einen Überblick über vorhandenen Themen, Inhalte und deren Beziehungen zu gewinnen. Im Rahmen dieser Interaktionen entdeckt der Nutzer bestimmte Informationen die sein Interesse erwecken, setzt diese in den Kontext seines Informationsbedürfnisses und stellt damit bestimmte inhaltliche Verbindungen zwischen unterschiedlichen Informationen her. Er entwickelt somit eine persönliche Interpretation der Bedeutung der Informationen und ihrer Zusammenhänge.

Diese persönliche Perspektive kann der Nutzer durch das Umordnen der systemgenerierten Struktur in Form einer persönlichen Karte ausdrücken (z.B. durch das Auswählen relevanter Informationsobjekte, durch das Bewegen von Objekten zwischen Objektgruppen, die Erstellung neuer Gruppen, das Hinzufügen von Beziehungen etc.).

Da die Nutzeraktionen in Bezug auf eine bestehende semantische Struktur gesetzt sind, ist ihre Bedeutung kontextualisiert und kann vom System als solche entnommen werden. Dementspre-

chend können Nutzeraktionen als Ausgangspunkt für den Aufbau einer neuen Wissensstruktur herangezogen werden, welche die persönliche Perspektive des Nutzers widerspiegelt.

Die nutzerdefinierte Struktur kann so von dem System erlernt und formalisiert werden (z.B. durch Anwendung überwachter Lernmethoden), so dass sie als nutzerspezifische Vorgabe für die semantische Strukturierung von beliebigen Informationsräumen angewandt werden kann. Auf diese Weise wird das implizite Wissen eines Nutzers, das sich durch seine Aktionen in der ursprünglichen semantischen Struktur spiegelt, erfasst und formalisiert. In dieser Form kann es dann visualisiert und auf neue Situationen angewandt werden.

Die Entwicklung einer konkreten Realisation des beschriebenen Modells umfasst die Integration und Erweiterung von Methoden aus mehreren Forschungsfeldern. Zur autonomen, maschinellen Erstellung von Wissenskarten aufgrund semantischer Beziehungen zwischen Text-Dokumenten verweisen wir auf Methoden des Dokumenten-Clustering und der semantischen Netzwerke (Lin et al. 1991, Kohonen et al. 2000, Sack 2000, Becks et al. 2000). Hinsichtlich bestehender Ansätze zur Extraktion der Semantik aus Nutzerinteraktionen mit Informationen, zum Lernen nutzerspezifischer Profile und zu kollaborativen Filtertechniken, siehe Resnick et al. 1994, Herlocker et al. 2000, Chalmers 2001, Joachims et al. 1997. Zur Visualisierung großer Informationsbestände gehen wir von Fokus+Kontext-Techniken aus (Robertson & Mackinlay 1993, Sarkar et al. 1993, Bederson et al., 1996).

3 Das Wissenskartensystem

Das beschriebene Modell der personalisierten, selbstlernenden Wissenskarten wird auf folgende Weise realisiert. Eine Wissenskarte besteht aus den Elementen ContentMap, ConceptMap und OntologieMap (Abbildung 1). Die ContentMap zeigt einen Überblick des inhaltlich strukturierten Informationsraumes. Sie visualisiert Gruppen von thematisch in Beziehung stehenden Dokumenten und bietet Einblicke in deren impliziten, auf deren Inhalt beruhenden Beziehungen. Die ConceptMap visualisiert ein Begriffsnetz das zuerst aus dem Dokumentenpool extrahiert wird und danach, ausgehend von Interaktionen seitens des Nutzers mit den Dokumenten, in Bezug auf die persönliche Sichtweise eines Nutzers, sowie auf die Sichtweisen anderer Nutzer gesetzt wird (siehe Kap. 3.1). Diese Begriffsstruktur wird möglichst einfach gehalten und beinhaltet nur unbenannte Beziehungen, um dem Benutzer eine übersichtliche und klare Aussage zu vermitteln. Die Aufgabe solch einer ConceptMap ist es, eine begriffliche Navigationsstruktur bereitzustellen und eine Einsicht in die Kriterien der Dokumentengruppierung in unterschiedlichen ContentMaps, bzw. ihre interaktive Parametrisierung zu ermöglichen. Im Gegensatz dazu stellt die OntologieMap eine viel komplexere Struktur dar, die viele unterschiedliche Arten von benannten Beziehungen beinhaltet und hierarchische Klassifizierung und Navigation des Informationsraumes ermöglicht (siehe Kap 3.4.).

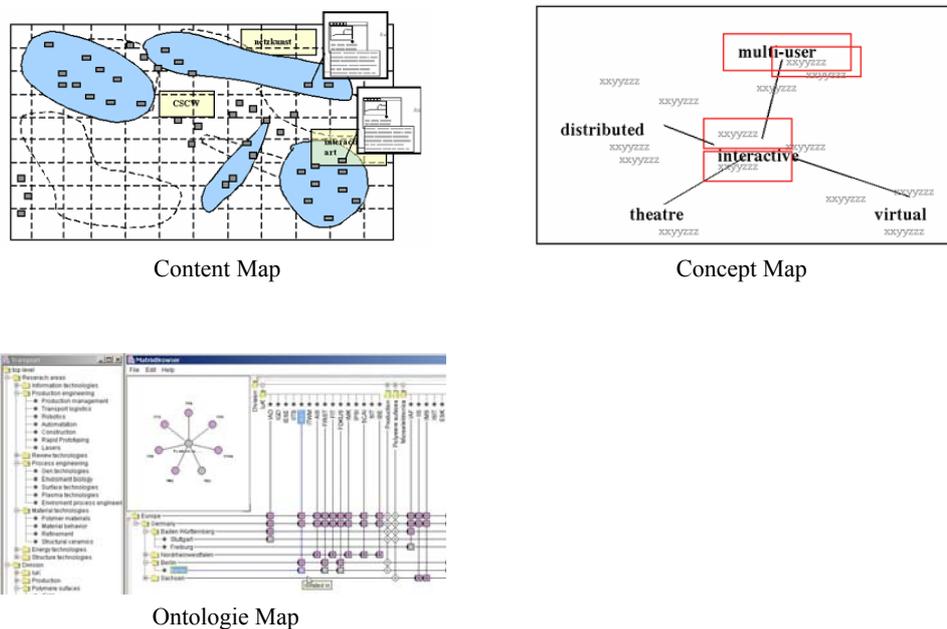


Abbildung 1: Hauptelemente einer Wissenskarte

Zur Realisierung dieses Modells wurde ein agenten-basiertes System entwickelt, das nachfolgende Komponenten beinhaltet:

- Ein Textanalysewerkzeug repräsentiert die bedeutungstragenden Merkmale von Dokumenten durch Wordvektoren im Vektorraummodell.
- Ein Gruppierungsagent erstellt basierend auf der Methode der selbstorganisierenden neuronalen Netze (Kohonen et al. 2000) eine ContentMap und eine erste ConceptMap. Durch diese Technik liegen semantisch ähnliche Dokumente auf der zweidimensionalen Dokumenten-Karte nahe beieinander und Wörter, die in ähnlichen Kontexten vorkommen, befinden sich auf der ConceptMap nahe beieinander.
- Ein Klassifizierungsagent ermöglicht es sowohl nutzereditierte Wissenskarten zu lernen und auf beliebige Mengen von Dokumenten anzuwenden als auch systemgenerierte ConceptMaps mit personalisierten Begriffsstrukturen an zu bereichern.
- Ein Ontologieagent generiert und verwaltet ein gemeinsames Begriffsnetz aus der Überlagerung von persönlichen Wissenskarten, welches mit ontologischen Informationen angereichert werden kann.
- Den Knowledge Explorer (Novak et al. 2003, Novak & Wurst 2003) bietet ein integriertes Werkzeug für die Erstellung, die Visualisierung und das Explorieren der personalisierten Wissenskarten und Begriffsstrukturen.
- Den MatrixBrowser (Ziegler et al. 2002) bietet ein Visualisierungs- und Abfragewerkzeug für ontologische Strukturen auf Nutzerseite, sowie für Erstellung der zugrunde liegenden Ontologie auf Administratoreseite.

Das Zusammenspiel dieser Module ermöglicht 1) die automatische Erstellung einer Strukturierung eines größeren Dokumentenpools anhand der semantischen Analyse des Dokumenteninhalts,

2) die Visualisierung und Navigation dieser Struktur, 3) die Erfassung der persönlichen Perspektiven von unterschiedlichen Nutzern und 4) die Möglichkeit, diese in Form von persönlichen Karten anderen Nutzern zugänglich zu machen und in eine gemeinsame Ontologie einfließen zu lassen. Im folgenden Abschnitt werden die einzelnen Schritte dieses typischen Nutzungsszenarios beschrieben.

3.1 Exploration von Zusammenhängen und Beziehungen mittels systemgenerierten und personalisierten Wissenskarten

Durch das Aufrufen des Knowledge Explorers (Abbildung 2) steht dem Nutzer eine Visualisierung zur Verfügung, welche die in Gruppen zusammengefassten Dokumente darstellt. Die kritische Frage dabei ist, wie die Dokumente angeordnet werden sollen, um inhaltliche Zusammenhänge zwischen ihnen herauszustellen.

In einem ersten Schritt verwenden wir hierfür eine Kombination aus statistischer Textanalyse und auf Kohens SOM (Lin et al. 1991, Kohonen et al. 2000) basierenden Gruppierungsverfahren (Novak et al. 2003). Ähnliche Dokumente werden dabei nah zueinander auf einer zweidimensionalen Fläche angeordnet. Durch die Darstellung der Verteilung aller Dokumente über die Karte und ihrer Gruppierung in thematisch in Bezug stehenden Gruppen, wird dem Nutzer ein überblicksartiger Eindruck der jeweiligen Informationssammlung gegeben. Jede Gruppe wird zusätzlich mit einer Menge von Schlagwörtern beschrieben. Diese Schlagwörter werden zum einen aus explizit eingegebenen Schlagwörter der einzelnen Objekte extrahiert und zum anderen durch Textanalyse automatisch gewonnen. Beide Quellen werden durch ein Gewichtungsverfahren gemischt und ergeben die endgültige Schlagwortliste.

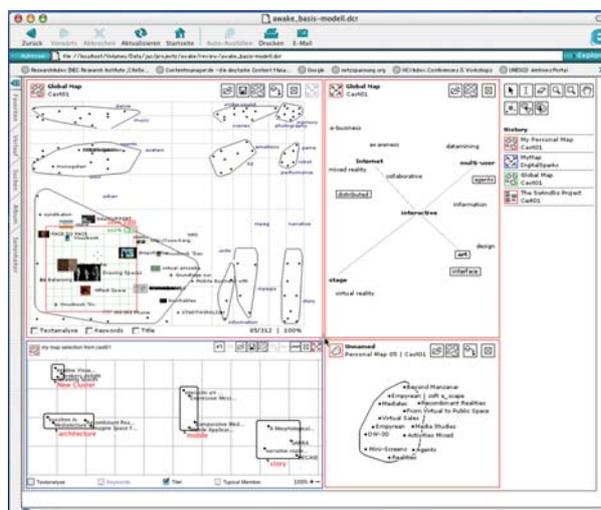


Abbildung 2: Das Knowledge Explorer Werkzeug

Die auf diese Weise entstandene Content Map wird durch eine Concept Map erweitert. Die ConceptMap gibt einen Einblick in die Kriterien, welche die semantische Struktur der ContentMap bedingen. Diese Kriterien bilden eine Art semantischer Achsen. Wird die ContentMap aufgrund der Analyse des Wortgebrauchs in Texten strukturiert, so ergibt sich ein spezieller Typ von ConceptMap, die WordMap. Sie zeigt diejenigen Wörter an, welche den Beziehungen zwischen den Dokumenten zu Grunde liegen. Auch zwischen diesen Wörtern selbst können Beziehungen herge-

stellt werden, indem zu jedem Wort, alle häufig im Kontext dieses Wortes auftretenden Wörter als ähnlich betrachtet werden. Dies ermöglicht eine Navigation zwischen verschiedenen, aber verbundenen Begriffen, wie in Abbildung 4 (rechts oben) zu sehen ist.

Die ContentMap und die ConceptMap bilden die zentralen Werkzeuge für den Nutzer, den gegebenen Informationsraum zu explorieren. Besonders bei großen Dokumentenmengen zeigt sich dabei die Notwendigkeit verschiedener Detaillierungsgrade. Entsprechend enthält sowohl die Content Map als auch die Concept Map eine Funktion des „semantischen Zooming“ mit mehreren (lokalen) Fokuspunkten. Die Idee des „sematischen Zooming“ im Gegensatz zum rein optischen Zooming ist, dass auf hoher Ebene nur ein repräsentativer Teil der Dokumente angezeigt wird. D.h. es wird nicht nur optisch verkleinert sondern auch semantisch gefiltert. Den lokalen Fokuspunkten liegt folgende Idee zugrunde: um Beziehungen zwischen verschiedenen Gruppen und Teilen der Karte erfassen zu können, muss der Benutzer detaillierte Informationen über gewisse Objekte erhalten und gleichzeitig den Überblick über den ganzen Informationsraum behalten. Die von uns eingesetzte Verzerrungstechnik ermöglicht eine solche Kombination von Detailansicht und Überblick (siehe Abbildung 2, links oben).

Basierend auf der Exploration der systemgenerierten Karte kann der Benutzer persönliche Karten erstellen. Dazu fügt er auf eine anfangs leere Karte, welche dieselbe Form hat wie die systemgenerierte ContentMap, beliebige Gruppen und Dokumente hinzu. Dabei kann er auch vollständige Gruppen aus der systemgenerierten Karte übernehmen. Neben ihrer direkten praktischen Bedeutung als Lesezeichen für den einzelnen Nutzer, bieten die persönlichen Wissenskarten die Grundlage für die Projektion, Explikation und den Austausch von Wissen. Zum einen können persönliche Wissenskarten genutzt werden, um sich eine personalisierte Karte einer Menge von Dokumenten erstellen zu lassen. Die Dokumente werden also gewissermaßen aus der Perspektive des Nutzers dargestellt, welcher die Wissenskarte erstellt hat (siehe Kap. 3.2). Zum anderen können Wissenskarten zwischen Nutzern ausgetauscht werden. Auf diese Weise kann ein Nutzer direkt von dem Wissen und der Arbeit eines anderen Nutzers profitieren (siehe Kap 3.3). Schließlich können die persönlichen Wissenskarten aller Benutzer zusammengenommen werden, um eine gemeinsame Ontologie (siehe Kap 3.4) oder eine personalisierte ConceptMap als einen persönlichen Begriffs-Blickpunkt bzw. eine Untermenge der Ontologie zu erstellen. Die personalisierte ConceptMap nimmt die Cluster-Namen, die seitens eines Nutzers vergeben wurden als persönliche Haupt-Konzepte und stellt sie in Bezug zu Begriffen aus der Textanalyse bzw. zu den Cluster-Namen und Begriffen aus den Karten anderer Nutzer. Die Auswahl der herangezogenen Begriffe und Cluster-Namen beruht auf einer Analyse der Ähnlichkeit von Konzepten, wie sie durch die von Nutzern erstellten und benannten Gruppen extrahiert werden kann. Ist eine kritische Masse von Nutzerinteraktionen vorhanden, kann die kollaborativ erstellte ConceptMap die auf Textanalyse beruhende ConceptMap komplett ersetzen (Novak & Wurst 2002).

Die Grundlage für alle diese Anwendungen ist die Annahme, dass Nutzer durch ihre persönlichen Wissenskarten, Wissen über einen gewissen Themenbereich oder einen Informationsraum zum Ausdruck bringen. Konkreter reflektieren Wissenskarten zum einen das Wissen des Nutzer über die gegenseitige Relevanz oder Ähnlichkeit von Dokumenten zum anderen das Wissen über Konzepte, d.h. das Wissen darüber, welche Dokumente zu welchen Begriffen gehören. Dies bildet den Ausgangspunkt für die Extraktion und Nutzung dieses Wissens.

3.2 Erlernen und Anwenden einer persönlichen Wissenskarte auf eine Menge von Dokumenten

Durch das Editieren einer persönlichen Karte definiert der Nutzer eine Menge von Klassen. Um eine persönliche Karte zu erlernen, muss eine Entscheidungsfunktion gefunden werden, die eine

automatische Zuordnung von neuen Dokumenten zu diesen Klassen ermöglicht. Es gibt mehrere Möglichkeiten, eine solche Funktion zu finden. Dabei hat sich die Nearest-Neighbour[21] Methode als besonders geeignet erwiesen. Ein Dokument wird klassifiziert indem es dieselbe Klasse erhält, wie das ihm ähnlichste, vom Nutzer angeordnete.

Diese Vorgehensweise bietet verschiedene Vorteile. Zum einen liefert sie zusätzlich zur Klassenzugehörigkeit eine Position auf der zweidimensionalen Ebene, welche das Dokument in den entsprechenden semantischen Kontext stellt. Zweitens ist das Verfahren sehr effizient, da die Dokumentenähnlichkeiten vorberechnet und in entsprechende Datenstrukturen gespeichert werden können. Effizienz spielt eine besondere Rolle, da der Nutzer normalerweise auf das Anwenden einer persönlichen Wissenskarte wartet. Die Antwortzeit muss dementsprechend so niedrig wie möglich sein. Mit unserer derzeitigen Implementierung liegt sie im Bereich unter einer Sekunde. Drittens, und das ist im Zusammenhang mit kollaborativen Systemen vielleicht am interessantesten, ermöglicht die Verwendung eines Ähnlichkeitsmaßes zur Klassifikation nicht nur die Repräsentation inhaltlicher Ähnlichkeit sondern auch von Kontextzusammenhängen. Die Idee ist folgende: Zwei Dokumente die häufig gemeinsam vorkommen, sehr selten aber einzeln, können normalerweise als ähnlich angesehen werden (im Sinne von „im selben Kontext relevant“). Wenn man die persönlichen Wissenskarten von allen Nutzern zusammennimmt, kann durch eine statistische Analyse der gemeinsamen und einzelnen vorkommenden Dokumenten ein „kollaboratives“ Ähnlichkeitsmaß abgeleitet werden. Ausgehend von der Annahme, dass sich in den persönlichen Wissenskarten das Wissen der Nutzer widerspiegelt, die diese Karten erstellt haben, repräsentiert ein solches Ähnlichkeitsmaß das gemeinsame Wissen aller Nutzer. Es kann auch ohne dass sich die Dokumente inhaltlich charakterisieren lassen (wie es bei Multi-Media Dokumenten beispielsweise schwierig ist) gewonnen werden.

Andererseits muss genügend Nutzerinteraktion vorhanden sein, um diese Art von Ähnlichkeit tatsächlich zu berechnen. Daher kombinieren wir Kontextähnlichkeit mit Textähnlichkeit (auf Grundlage des Kosinusmaßes für Wortvektoren). Wenn für ein Paar von Dokumenten ihr gemeinsames Vorkommen nicht statistisch signifikant ist, wird stattdessen Textähnlichkeit verwendet. Auf diese Weise vermeiden wir das aus dem „kollaborativen Filtern“ (Resnick et al. 1994) bekannte Problem, dass am Anfang keine Anfragen beantwortet werden können, da es keine Nutzerangaben gibt.

3.3 Austausch von Wissenskarten zwischen Nutzern

Um einen sinnvollen Austausch von Wissenskarten zu ermöglichen wird durch das System eine spezielle Suchfunktion zur Verfügung gestellt. Die Idee dabei ist: Nutzer haben normalerweise neben einem akuten auch einen langfristigen Informationsbedarf. Das System sollte, wenn möglich, solche Karten zurückliefern, die beiden entsprechen. Beispielsweise kann „Zeitreihenanalyse“ ein akuter Informationsbedarf eines Nutzers sein und „Maschinelles Lernen“ ein langfristiger. Falls es nun persönliche Wissenskarten gibt, welche Zeitreihenanalyse eher unter dem Aspekt von Lernverfahren darstellen, dann sollten diese eine höhere Bewertung erhalten als Karten, welche Zeitreihenanalyse unter anderen Aspekten enthalten. Um dies zu realisieren bietet das System zum einen die Möglichkeit Stichwörter für die Suche einzugeben, welche eher den akuten Informationsbedarf beschreiben. Der langfristige Informationsbedarf wird automatisch aus den bereits von dem entsprechenden Nutzer erstellten persönlichen Karten abgeleitet, indem charakteristische Dokumente aus diesen Karten extrahiert werden. Der Nutzer kann aber interaktiv die Gewichtung der beiden Aspekte bestimmen.

3.4 Generierung einer Ontologie durch Strukturvergleich persönlicher Karten

Komplementär und in Verbindung zu den maschinellen Verfahren und der flachen Strukturierung der Wissenskarten ist die systematische Erstellung von hierarchischen Klassifikationssystemen bzw. Ontologien erforderlich, um auf Metaebene explizite Zusammenhänge der Informationsressourcen zu definieren und somit eine inhaltsbasierte Suche nach einzelnen und speziellen Dokumenten zu ermöglichen. Eine solche, sogenannte semantische Suche erzeugt qualitativ hochwertige Suchergebnisse und dadurch kürzere Suchzeiten mit verringerter Redundanz. Problem hierbei ist jedoch, dass solche Ontologien sehr aufwendig manuell und dadurch fehleranfällig von Spezialisten, sogenannten Wissensingenieuren, erstellt werden müssen. Zusätzlich zum Aufwand der Erstellung einer Ontologie ist auch der Pflegeaufwand sehr hoch, was einen wesentlichen Kritikpunkt dieser Technologien darstellt. Aus dem Grund der aufwendigen Erstellung werden auch -relativ erfolgreiche - Versuche durchgeführt, Begriffsnetze aus dem zugrundeliegenden Textkorpus zu extrahieren. Beispiele hierzu sind das Wortschatz Projekt der Universität Leipzig (Heyer et al. 2001), TextToOnto (Maedche & Staab 2000) der Universität Karlsruhe und das LTG Projekt (Mikheev & Finch 1997) der Universität Edinburgh. TextToOnto extrahiert Begriffspaare und deren Relationen mittels linguistischer Abhängigkeiten, benötigt aber Hintergrundwissen über Begriffshierarchien für die Einbeziehung der extrahierten Begriffspaare. Eine Extraktion von Schlagwörtern, also inhaltsbeschreibende Wörter, ist aber grundsätzlich nicht möglich, wenn sie gar nicht erst in den Dokumenten auftauchen.

Da die nutzererstellten Wissenskarten Resultat einer sehr intentionalen Handlung sind, welche einen gewissen Grad an Reflexionsleistung bedingen und hierarchisch organisiert sein können, werden sie in dem hier verfolgten Ansatz als persönliche Klassifikationssysteme eines Teilausschnitts des Dokumentenpools betrachtet. Im Konkreten bedeutet dies, dass Gruppenbezeichner als thematischer Indikator für die darunter angeordneten Dokumente stehen. Durch Strukturvergleich werden die einzelnen Karten überlagert und zu einem gemeinsamen Begriffsnetz zusammengeführt. Abbildung 3 verdeutlicht dies. Ergebnis ist hierbei ein Informationsnetz, welches über hierarchisch gegliederte Begriffe verfügt, die mittels einem „hat zu tun mit“ Beziehungstyp vernetzt sind. In diesem Stadium ist sie schon für Such- und Explorationszwecke einsatzfähig.

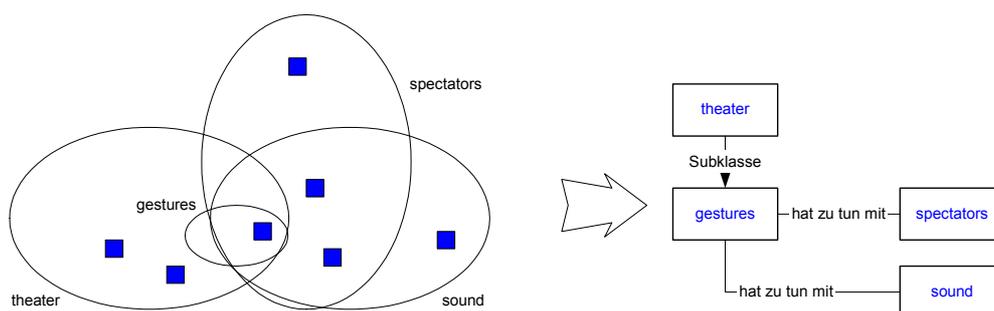


Abbildung 3: Überlagerung von einzelnen Kategoriensystemen

In den weiteren Stadien der Formalisierung kann dieses Informationsnetz durch Anreicherung von weiteren Begriffen und Beziehungstypen zu einer vollständigen Ontologie erweitert werden.

3.5 Themenbasierte Exploration und Suche

Die dem Informationsraum zugrundeliegende Ontologie kann nun mit Hilfe der OntologyMap, welche durch den MatrixBrowser (Ziegler et al. 2002) realisiert ist, exploriert und abgefragt werden. Nachdem eine solche Ontologie relativ groß werden kann, bietet die OntologyMap neben der explorativen Erschließung des Netzes einen Rechercheprozess in drei Schritten an: Um die gerade relevanten Ausschnitte des Netzes zu finden, beginnt der Nutzer mit der Eingabe eines Suchbegriffs. Hierauf vollzieht das System eine standardmäßige Volltextsuche, wobei nicht nur die reine Ergebnismenge als Liste dargestellt wird (siehe Abbildung 4, rechter Ausschnitt), sondern die gefundenen Dokumente den Konzepten in der Ontologie zugeordnet. Diese Teilbereiche der Ontologie werden visualisiert und die Treffermenge damit thematisch kontextualisiert (Kunz & Botsch 2002). Es werden zum einen alle Begriffshierarchien in denen Treffer erzielt wurden aufgelistet (Abbildung 4, links) zum anderen werden die beiden Begriffshierarchien mit den meisten Treffern und möglichen Relationen zwischen diesen auf der Matrix dargestellt. (Abbildung 4, Mitte). Gibt es gleiche Treffer in den auf den Achsen dargestellten Kategoriebegriffen, so wird in den zugehörigen Matrixzellen ein Kreis, dessen Größe mit der Anzahl der Treffer korrespondiert, dargestellt. Sämtliche auf diese Art gefundenen Begriffshierarchien können beliebig via Drag&Drop auf die Matrixachsen gezogen und damit exploriert werden.

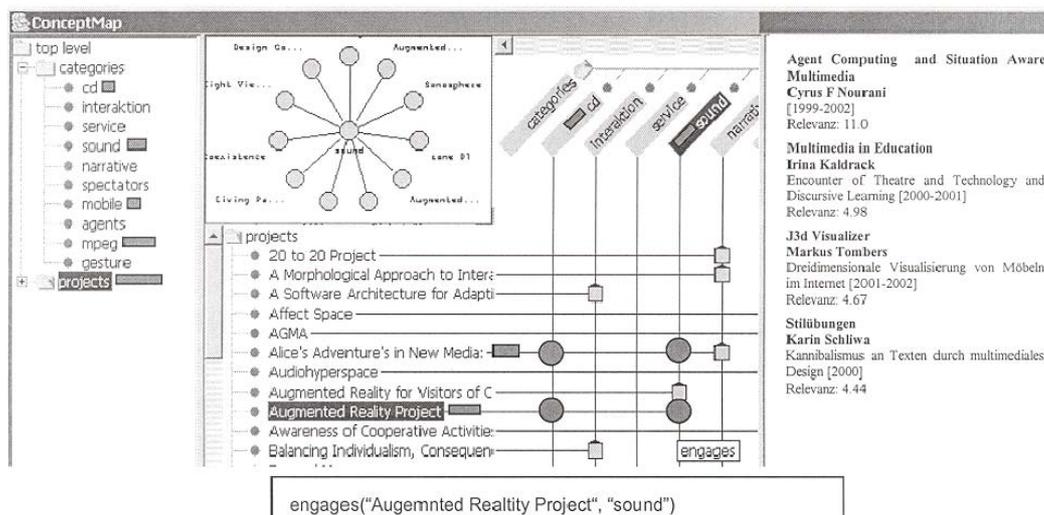


Abbildung 5: OntologyMap: Semantische Suche und Kontextualisierung von Treffern

Im zweiten Schritt kann die Ergebnismenge weiter eingeschränkt werden indem einzelne Begriffe auf den Matrixachsen selektiert werden. Es ergibt sich die Schnittmenge aus allen Trefferdokumenten und denen, die unter einem Kategoriebegriff als Instanzen aufgehängt sind. Im obigen Beispiel zeigt die Ergebnismenge alle Dokumente zum Thema „Sound“ bezogen auf das „Augmented Reality Project“, da diese Begriffe selektiert wurden, sowie zum Suchbegriff „Multimedia“. Im dritten Schritt können die expliziten Relationen dazu verwendet werden, tatsächliche semantische Abfragen zu generieren. Hierzu existiert in der OntologyMap ein Ausschnitt auf das wieder mittels Drag&Drop gewünschte Kategoriebegriffe und Relationen zwischen diesen gezogen werden können. Es wird möglich die Ontologie im Sinne von „gib mir alle Dokumente des

Augmented Reality Projects die sich mit dem Thema Sound beschäftigen“ abgefragt werden. Mit diesem Suchwunsch ergibt sich die in Abb. 4 unten dargestellte semantisch Abfrage.

4 Kollaborative Konstruktion und Anwendung des Wissens

Wird das System in der beschriebenen Weise angewandt (z.B. in einer Gemeinschaft von Experten aus verschiedenen Bereichen), kann eine Nutzergruppe eine Menge persönlicher Wissenskarten erstellen und diese gemeinsam nutzen. Ein Nutzer kann die Karte eines anderen Nutzers anwenden, um einen Informationspool mit dem „Wissen“ jenes Nutzers automatisch zu klassifizieren. Ein typischer Anwendungsfall wäre hier die Anwendung von Expertenwissen, um in Informationspools oder Themenfeldern zu navigieren, die einem selbst nicht so vertraut sind. Bei einer Nutzer-Gemeinschaft mit einem offenen Ansatz zur gemeinsamen Nutzung von Wissens, können Nutzer ihre eigenen Karten als Erweiterung der Karten anderer erstellen. So können sie aktiv die Wissensstrukturen anderer Nutzer erweitern und sie mit ihren eigenen verbinden. Die personalisierten und gemeinsamen Begriffstrukturen der ConceptMaps (siehe Kap. 3.1.) und der OntologyMap (siehe Kap. 3.4) ermöglichen es ferner, die nutzerdefinierten Karten zu einem großen Wissensnetz zusammenzufassen welches als übergreifende Navigationsstruktur dient und die relevanten Konzepte der jeweiligen Expertengruppe repräsentiert und strukturiert. Diese Begriffsstrukturen und -netzwerke bilden das implizite Wissen einer Gemeinschaft ab und machen es erkennbar und nutzbar.

5 Anwendung auf netzspannung.org

Die praktische Testumgebung und erster Anwendungskontext der beschriebenen Arbeiten ist die Internetplattform netzspannung.org. netzspannung.org strebt an, ein Wissensportal einzurichten, das Einblick in die Schnittpunkte zwischen digitaler Kunst, Kultur und Informationstechnologie bietet (Fleischmann et al. 2001). Die grundlegenden Bedingungen zur Realisierung eines solchen Wissensportals, das eine Gemeinschaft von Experten aus unterschiedlichen Bereichen verbindet, sind: Ein sich ständig entwickelnder Informationspool muss strukturiert und zugänglich gemacht werden, entsprechend den vielen unterschiedlichen Klassifizierungssystemen, die den Bedarf von unterschiedlichen Nutzergruppen und individuellen Nutzerkontexten (z.B. Künstler und Forscher, erfahrene Nutzer und die interessierte Öffentlichkeit) berücksichtigen. Typische Nutzer von netzspannung.org sind Experten und Fachleute wie Künstler, Forscher, Designer, Kuratoren und Journalisten.

Durch die Benutzung des beschriebenen Systems wird es für diese heterogene Nutzergruppe möglich ein Informationspool nicht nur kollaborativ aufzubauen, sondern es auch interaktiv und kollaborativ zu strukturieren, mittels personalisierten Wissenskarten zu visualisieren und zu explorieren, sowie eine gemeinsame Navigationsstruktur zu konstruieren, welche die unterschiedliche persönliche Blickpunkte auf die Bedeutung der Informationen verbindet. Ein Szenario ist die Anwendung zur Exploration möglicher Beziehungen zwischen den Arbeiten zu verschiedenen Themen und in unterschiedlichen Berufsbereichen (z.B. zum Auffinden von Projekten aus unterschiedlichen Disziplinen, die für die eigene Arbeit von Wichtigkeit sind). Ein weiteres Szenario ist der Vergleich von Projekten mit der eigenen Perspektive und der Sichtweise von anderen Experten (z.B. zum Entdecken von Bezügen und verborgenen Annahmen). Und schließlich wird die breite Öffentlichkeit das Wissen der Experten-Gemeinschaft nutzen, indem Wissenskarten einzel-

ner Experten angewendet und die gemeinschaftliche Begriffsstruktur als Mittel zur Navigation und Wissensentdeckung im Informationsraum von netzspannung.org genutzt wird.

Der aktuelle System-Prototyp wurde bereits intern eingesetzt und angewandt: zum einen als Wissens-Browser für die Einreichungen der Konferenz cast01^[1] und des Wettbewerbs digital sparks, zum anderen zur internen Erprobung der Nutzung von personalisierten Wissenskarten. Dieses simuliert ein Nutzungsszenario, der es ermöglicht Beziehungen zwischen Informationen aus normalerweise separaten Archiven unterschiedlichen Nutzer-Gemeinschaften in den Bereichen der Medienkunst, -forschung und -technologie zu explorieren. Die Ergebnisse können derzeit öffentlich in Form von Guided Tours und Teilweise in Form von online verfügbaren interaktiven Demos erprobt werden^[2].

6 Zusammenfassung und Ausblick

In diesem Beitrag haben wir einen Ansatz vorgestellt, wie die Metapher einer Wissenskarte verwendet werden kann, um mehrere Methoden zur interaktiven Informationssuche zu integrieren und ein Modell zur kollaborativen Entdeckung und Austausch von Wissen zu realisieren. Darüber hinaus haben wir gezeigt, wie überwachtes und unüberwachtes Lernen zur Erstellung von Wissenskarten verwendet werden kann, die dem Nutzer verschiedene Sichten auf einen Informationsraum bieten und dessen semantische Struktur aufzeigen. Wir haben ein Modell aufgezeigt, welches auf Basis zweidimensionaler Wissenskarten auf unaufdringliche Art und Weise Profile für personalisierte Nutzeragenten erstellt. Diese bieten sowohl einen Kontext für die implizite Kommunikation zwischen menschlichen Nutzern und Agenten als auch eine Form der visuellen Repräsentation der daraus resultierenden Wissensstrukturen. Wir haben gezeigt wie solche Wissenskarten visualisiert und zur Entdeckung von Kontexten und Beziehungen in heterogenen Informationspools eingesetzt werden können. Die automatische Erstellung von personalisierten und kollaborativen Begriffsstrukturen in Form von ConceptMaps und Ontologien und die Möglichkeiten ihrer Nutzung zur begrifflichen und thematischer Navigation wurden auch behandelt. Des Weiteren zeigten wir Möglichkeiten auf, wie Wissenskarten als Medium für den Austausch von explizitem und implizitem Wissen zwischen verschiedenen Nutzern dienen können. Wie in diesem Beitrag aufgezeigt, weicht unser System stark von sogenannten „kollaborativen Filter“-Systemen ab, da einzelne Informationselemente nicht nur von der Nutzergemeinde bewertet werden, sondern auch in verschiedene Nutzungskontexte gesetzt werden. Dies geschieht auf eine unauffällige Art, weil es in die Hauptaktivität der Nutzer eingebettet ist. In diesem Sinne ermöglicht unser System mehr ein „kollaboratives Strukturieren“ als nur ein „kollaboratives Filtern“.

Aktuell befassen wir uns mit verschiedenen Methoden das System auszubauen und zu optimieren. Erstens wollen wir zusätzliche Ähnlichkeitsaspekte für das Lernen der persönlichen Karten hinzufügen. Zweitens wird ein zusätzlicher Editor, der das Erstellen von hierarchischen Strukturen und anderen Beziehungen zwischen Objekten ermöglicht in das System integriert. Dieser wird deswegen benötigt, da der Nutzer beim Editieren einer Wissenskarte Objekte nur in einer flachen Struktur anordnen kann, was zwar einfach und intuitiv ist, aber nicht immer ausreicht. Desweiteren werden Methoden zur Verarbeitung einzelner Strukturen untersucht, welche auch den gesamten Informationsgehalt ausnutzen, der in diesen verborgen ist. Letztlich ist ein Evaluationsworkshop

[1] Öffentlicher Demo-Einsatz eines frühen Prototyps eines Wissenskarten-Browsers: <http://netzspannung.org/cast01/semantic-map>

[2] Guided Tour und interaktive Demos des aktuellen Systems: <http://awake.imk.fraunhofer.de>

geplant, um die Nützlichkeit des Systems im Gesamten zu analysieren und die verschiedenen Beiträge der einzelnen Ansätze zu vergleichen.

7 Literatur

- Becks, A., Sklorz, S., Jarke, M.:A (2000): *Modular Approach for Exploring the Semantic Structure of Technical Document Collections*, ACM Proc. of AVI 2000, Palermo.
- Bederson, B.B. et al., (1996): *Pad++: A zoomable graphical sketchpad for exploring alternate interface physics*. J. Vis. Lang. Comput.,7:3.
- Chalmers, M. (2001): *Paths and Contextually Specific Recommendation*. DELOS Workshop: Personalisation and Recommender Systems in Digital Libraries.
- Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. (1996): *From data mining to knowledge discovery: An overview*. In Advances in Knowledge Discovery and Data Mining, AAAI Press and the MIT Press, Chapter 1, 1-34.
- Fleischmann, M., Strauss, W., Novak, J. et al. (2001): *netzspannung.org : an internet media lab for knowledge discovery in mixed realities*, in Proc. of cast01 // living in mixed realities, Fraunhofer IMK, ISSN 1618-1379.
- Herlocker, K., Konstan, J., Reidl, J., (2000): *Explaining Collaborative Filtering Recommendations*. In Proc. of CSCW 2000, December 2-6, Philadelphia, Pennsylvania, USA.
- Heyer, G., Quasthoff, U., Wittig, T., Wolf, C. (2001): *Learning Relations Using Collocations*, Proceedings of the Second Workshop on Ontology Learning.
- Joachims, T. et al. (1997): *Web Watcher: A Tour Guide for the World Wide Web*, IJCAI.
- Kohonen, T., Kaski S., et al. (2000): *Self Organization of a Massive Document Collection*, IEEE Transactions on Neural Networks, Vol. 11, No. 3.
- Kunz, C., Botsch, V. (2002): *Visual Representation and Contextualization of Search Results: List and Matrix Browser*, Proceedings of the DC-2002, Florence, Italy.
- Lin, X., Soergel, D., Marchionini, G., (1991): *A self-organizing semantic map for information retrieval*. In Proc. of 14th ACM/SIGIR Conf. Research and Development in Information Retrieval.
- Maedche, A., Staab, S. (2000): *A Generic Architecture for Text Knowledge Acquisition*.
- Mikheev, A., Finch, S. (1997): *A Workbench for Finding Structure in Texts*, Proceedings of ANLP-97, Seite 8.
- Nonaka, I., Takeuchi, H. (1995): *The Knowledge-Creating Company*. Oxford University Press, London.
- Novak, J., Wurst, M. et al. (2003): *Discovering, visualizing and sharing knowledge through personalized learning knowledge maps – an agent-based approach*, to appear in Proc. of the AAAI Spring Symposium on Agent-mediated Knowledge Management (AMKM03), March 24-26, Springer Lecture Notes on AI.

- Novak, J., Wurst, M. (2003): *Supporting Communities of Practice through Personalisation and Collaborative Structuring based on Capturing Implicit Knowledge*, to appear in Proc. of I-KNOW '03, Graz, July 2-4.
- Resnick, P. et al. (1994): *GroupLens: An Open Architecture for Collaborative Filtering of Netnews*. In Proc. of ACM Conference on CSCW, Chapel Hill, North Carolina.
- Robertson, G.G., Mackinlay, J.D. (1993): *The document lens*. In *UIST '93*, Atlanta GA, USA, ACM Press.
- Sack, W. (2000): *Conversation Map: An Interface for Very Large-Scale Conversations*. Journal of Management Information Systems.
- Sarkar, M. et al. (1993): *Stretching the Rubber Sheet: A Metaphor for Viewing Large Layouts on Small Screens*. ACM Symposium on User Interface Software and Technology (UIST).
- Ziegler, J., Kunz, C., Botsch, V., Schneeberger, J. (2002): *Visualizing and Exploring Large Networked Information Spaces with Matrix Browser*. In Proc. of the 6th International Conference on Information Visualisation (IV'02), London.

Danksagung

Die in diesem Beitrag beschriebenen Arbeiten wurden in den Projekten *AWAKE – Networked Awareness for Knowledge Discovery*³ und *netzspannung.org – an Internet Media Lab*, die beide vom BMBF unterstützt werden, durchgeführt.

Dank an Martin Schneider und Kresimir Simunic für Ihre Mitarbeit an der Entwicklung des Kohonen-Clustering-Moduls und an Stefan Paal für seine Mitarbeit und Rat am software-technischen System-Design. Danijela Djokic und Hartmut Bohnacker haben am aktuellen Interface-Design mitgearbeitet, während Boris Müller an einer frühen Visualisierung für die Kohonen-basierte Wissenskarten maßgeblich beteiligt gewesen ist. Jens Wagner und Roger Sennert haben an dem MapManager-Modul für die Metadatenverwaltung und Reni Banov am Software-Engineering mitgearbeitet.

Kontaktinformationen

Jasminko Novak, Monika Fleischmann, Wolfgang Strauss
Fraunhofer Institute for Media Communication
MARS Exploratory Media Lab
Schloss Birlinghoven, 53754 Sankt Augustin
Tel: 02241-14-3437, Email: jasminko.novak@imk.fraunhofer.de
Michael Wurst, Katharina Morik
Universität Dortmund, Abteilung Künstliche Intelligenz,
44221 Dortmund
Tel: 0231-755-5101, Email: wurst@kimo.cs.uni-dortmund.de

³ <http://awake.imk.fraunhofer.de/>

Christoph Kunz, Jürgen Ziegler
Fraunhofer Institut für Arbeitswirtschaft und Organisation, Competence Center Human-Computer
Interaction
Nobelstraße 12, 70569 Stuttgart
Tel: 0711-970-2334, Email: christoph.kunz@iao.fraunhofer.de