

Visuelles Wissensmanagement mit adaptierbaren Dokumentenlandkarten

Andreas Becks

Fraunhofer-Institut Angewandte Informationstechnik
Schloss Birlinghoven, 53754 Sankt Augustin
andreas.becks@fit.fraunhofer.de

Dokumentiertes Wissen zu analysieren, zu strukturieren und zu organisieren ist ein wichtiger Bestandteil des Wissensmanagements. In der Literatur sind sogenannte Dokumentenlandkarten zur Visualisierung der inhaltlichen Ähnlichkeitsstruktur eines Textkorpus vorgeschlagen worden. Ausgehend von einem empirischen Aufgabenmodell wird in dieser Arbeit ein adaptierbares Rahmenmodell zur Generierung von Dokumentenlandkarten vorgestellt. Dieses bietet im Gegensatz zu anderen Ansätzen eine flexible Verknüpfung ähnlichkeits- und topologieerhaltender Visualisierungsmethoden, sichert die Austauschbarkeit der Komponente für die Bewertung der Textähnlichkeit und ermöglicht dem Benutzer, den Prozess der Kartengenerierung durch Regeln zu beeinflussen. Das interaktive Werkzeug DocMINER realisiert die vorgeschlagene Methodik und bietet zusätzliche, auf die Kartendarstellung abgestimmte Funktionen für die Analyse von Textsammlungen an. Durch Fallstudien in Industrie und Wissenschaft sowie eine vergleichende empirische Untersuchung in einer kontrollierten Laborumgebung wird das entwickelte Konzept ausführlich evaluiert.

1 Einleitung

Ein Problem im Wissensmanagement ist häufig nicht der Mangel an relevanten Informationen, sondern die geradezu überwältigende Vielfalt an Wissensquellen der unterschiedlichsten Art. Eine wichtige Quelle ist dabei die in der Regel umfangreiche, in der Praxis häufig in Form von unstrukturierten oder teilstrukturierten Textdokumenten vorkommende Dokumentation einer Organisation. Dokumentiertes Wissen zu analysieren, zu strukturieren und zu organisieren ist folglich ein wichtiger Bestandteil des betrieblichen Wissensmanagements.

Werkzeuge, die dem Benutzer einen aufgabenangemessenen Zugriff auf Textsammlungen bieten, sind dabei wichtige Schlüsseltechnologien. Von besonderer Bedeutung ist eine effektive Unterstützung explorativer Analysen von Textkorpora, bei denen es darum geht, einen Überblick über vorhandenes Material zu bekommen, Muster im Dokumentenraum zu entdecken oder inhaltliche Beziehungen zwischen Texten oder Textgruppen zu identifizieren und zu verstehen. Ein Beispiel hierfür ist die Patentanalyse: Für ein technologieorientiertes Unternehmen ist es wichtig, das Beziehungsgefüge zwischen dem eigenen Patentportfolio und dem der Konkurrenten zu erkennen. Hierzu werden häufig Kurzfas-

sungen von Patenten aus relevanten Technologiebereichen gesichtet und aufwändig im Hinblick auf Trends und Kompetenzschwerpunkte untersucht (vgl. [WD00]).

Eine vielversprechende Möglichkeit, derartige Korpusanalyseaufgaben im Wissensmanagement zu unterstützen, besteht darin, die Struktur von Textsammlungen bezüglich der inhaltlichen Verwandtschaft der Texte in Form grafischer Korpusübersichten explizit zu machen. Zur Visualisierung der inhaltlichen Ähnlichkeitsstruktur von Textsammlungen sind in der Literatur sogenannte Dokumentenlandkarten vorgeschlagen worden, welche die inhaltlichen Ähnlichkeitsbeziehungen zwischen einzelnen Textdokumenten und Dokumentengruppen durch eine kartografische Metapher veranschaulichen. Eine an den spezifischen Aufgaben der Dokumentenanalyse im Wissensmanagement orientierte Methodik und eine entsprechende applikationsspezifische Evaluierung ist in der Literatur allerdings nicht beschrieben. Die hier skizzierte Arbeit [Bec01] schließt diese Lücke: Ausgehend von einem empirischen Aufgabenmodell wird ein adaptierbares und parametrisierbares Rahmenmodell zur Generierung von Dokumentenlandkarten vorgestellt und systemtechnisch realisiert. In Fallstudien und einem empirischen Laborexperiment wird der Ansatz mit Rückbezug auf das Aufgabenmodell bezüglich seiner Angemessenheit im Anwendungskontext evaluiert.

2 Verwandte Arbeiten

Das allen Ansätzen von Dokumentenlandkarten gemeinsame Prinzip ist, dass die Ähnlichkeit von Dokumenten oder Dokumentengruppen durch räumliche Distanz ausgedrückt wird: Je ähnlicher zwei Texte oder Kategorien sind, desto näher liegen ihre Repräsentationen auf der Karte zusammen. Auf diese Weise soll es zum Beispiel möglich sein, die Themendichte oder Dokumentenverwandtschaft von Texten innerhalb eines Korpus zu studieren. Verschiedene Metaphern und Granularitäten der Darstellung sowie Methoden der Generierung von Dokumentenlandkarten finden sich in der Literatur: So lassen sich Dokumentenähnlichkeiten durch zwei- oder dreidimensionale Punktwolken [CC92], [WTP⁺95] oder 3D-Landschaften [DHJ⁺98] visualisieren, wobei Verfahren des mehrdimensionalen Skalierens (oder verwandte Ansätze) zur Berechnung der Karten auf Grundlage der Ähnlichkeitsinformation verwendet werden. Letztere wird in der Regel mit Standardmethoden des Information Retrieval aus der Textsammlung gewonnen. Die Abbildungsmethodik versucht dann, die Ähnlichkeit einzelner Dokumente möglichst maßgetreu als Distanz wiederzugeben.

Die Gruppenstruktur von Dokumenten eines Textkorpus kann durch Kategorienkarten [CSO96] dargestellt werden, wobei Textkategorien als farbige Felder repräsentiert sind, deren Größe die Anzahl der zum Thema vorhandenen Dokumente reflektiert. Methodisch werden hier selbstorganisierende Merkmalskarten (ein Modell neuronaler Netze, das auch unter dem Begriff "Kohonenkarten" bekannt ist) direkt mit einer Repräsentation der Dokumente als Schlagwortvektoren trainiert und mittels eines speziellen Verfahrens visualisiert. Im Ergebnis ist dann die Topologie des durch die Schlagwortvektoren definierten Eingaberaums zweidimensional abgebildet. Eine andere Anwendung von Kohonenkarten stellt die Dichte einer Textkollektion durch unterschiedliche Helligkeiten eines Farbtönen

dar [LHKK96], wobei Standardverfahren der Visualisierung dieses neuronalen Netzmodells benutzt werden.

Obgleich derartige grafische Darstellungen von Textkorpora sicherlich intuitiv einen gewissen Charme besitzen, stellt sich dennoch aus wissenschaftlicher Sicht die Frage nach ihrer Nützlichkeit. Hier fällt dann leider auf, dass nur wenige Resultate existieren, die bei der Eignungsbewertung kartografischer Korpusübersichten hilfreich sind. Und die wenigen vorhandenen Ergebnisse ernüchtern auf den ersten Blick, wie am Beispiel der bislang wohl am besten untersuchten Kategorienkarten deutlich wird: Zwar hat Lin [Lin95] festgestellt, dass die Visualisierung generell dem Benutzer helfen, Strukturen in der betrachteten Textsammlung zu lernen und zu erinnern – was insbesondere bei späteren zielgerichteten Suchvorgängen nützlich ist. In einem Vergleich von Kategorienkarten und dem hierarchisch organisierten Internetkatalog "Yahoo!" zeigte sich allerdings, dass Kategorienkarten beim ungerichteten Stöbern nicht besser geeignet sind als das textorientierte Katalogsystem, bei eher klaren Interessenslagen der Suchenden diesem aber sogar deutlich unterlegen sind [CHSS98].

Zusammenfassend lässt sich festhalten, dass viele technologische Ansätze für die Visualisierung von Textsammlungen vorgeschlagen worden sind. Allerdings sind die Vorteile der Karten, die in der Verdeutlichung von Strukturen und Mustern liegen – was bei der explorativen Analyse von Textkorpora potenziell hilfreich ist – nicht konsequent ausgenutzt worden. Ziel der in diesem Artikel skizzierten Arbeit [Bec01] war daher, das Potential der Kartenansätze systematisch für explorative *Analyseaufgaben im Wissensmanagement* einzubringen und zu bewerten.

3 Korpusanalyse im Wissensmanagement

Für eine anwendungsspezifische Entwicklung und Evaluierung eines eigenen Dokumentenlandkartenansatzes bedarf es eines guten Verständnisses von Aufgaben, wie sie in der Praxis bei der Analyse von Textkorpora vorkommen. Doch wie lassen sich derartige Aufgaben angemessen beschreiben? Aufgabenmodelle repräsentieren die Ziele und Vorgehensweisen bei der Lösung komplexer Aufgabenstellungen. Im Bereich des Information Access gibt es eine Vielzahl verschiedenartiger Modelle, die jedoch häufig system- statt benutzerzentriert, zu allgemein und insbesondere nicht in der Lage sind, praktisch relevante Korpusanalyseaufgaben im Wissensmanagement hinreichend präzise und trennscharf zu beschreiben. Für den betrachteten Anwendungsbereich war daher die Entwicklung eines spezialisierten Modells notwendig.

Das entworfene Aufgabenmodell, welches eine Taxonomie zur Charakterisierung, zum Vergleich und zur Klassifikation von Korpusanalyseaufgaben zur Verfügung stellt, resultiert aus der Analyse und Verdichtung theoretischer Modellierungsansätze und aus Ergebnissen empirischer Untersuchungen: Zunächst wurden Aufgabenmodelle des Information Access aus der Literatur studiert. Kandidaten, die als mögliche Basis für ein spezialisiertes Modell geeignet waren, wurden zu einem Ausgangsmodell kombiniert. Durch das intensive Betrachten praktischer Korpusanalyseaufgaben im Rahmen von Fallstudien konnten weitere Facetten für die Modellbildung abgeleitet werden. Der Pool von analysierten und

in das Modell eingegangenen Aufgabenaspekten speist sich noch aus einer weiteren Quelle: Mittels teilstrukturierter Interviews mit Wissensmanagern aus der Industrie konnte das Modell auf einen empirisch solideren Sockel gestellt werden.

Mit Hilfe des so entstandenen Aufgabenmodells sind bekannte Ansätze der Visualisierung von Textkorpora – insbesondere die skizzierten Dokumentenlandkarten als vielversprechende Kandidaten – kritisch untersucht und Anforderungen an eine zu entwickelnde, dem Problemfeld angepasste Methodik abgeleitet worden. Zudem war das Modell Grundlage für den Entwurf eines umfangreichen Interaktionskonzepts für das Landkartensystem und Basis für die Festlegung von Evaluationskriterien.

4 Adaptierbare Dokumentenlandkarten für die Korpusanalyse

4.1 Ein flexibles Rahmenmodell für adaptierbare Dokumentenlandkarten

Aus der Charakterisierung relevanter Analyseaufgaben im Aufgabenmodell ergibt sich die Anforderung, die Ähnlichkeit von Dokumenten sowie Zusammenhänge von Textgruppen auf Detailebene darzustellen. Zudem sind natürlich nicht absolute Ähnlichkeitsinformationen bei der Analyse von Interesse, sondern *Ähnlichkeitsstrukturen*. Weitere Anforderungen ergeben sich aus den betrachteten Dokumententypen: Wesentlich für die Arbeit mit thematisch hochspezialisierten Texten ist, dass die Realisierung des Dokumentenvergleichs problemangepasst und methodisch und technisch austauschbar ist, so dass bei Bedarf nicht nur die typischen statistischen sondern auch spezialisierte wissensbasierte Verfahren eingesetzt werden können. Neben dieser methodischen Adaptierbarkeit spielt aber auch eine inhaltliche Adaptierbarkeit eine Rolle: So sollten für den individuellen Analytischen interessante inhaltliche Zusammenhänge der Anwendungsdomäne bei der automatischen Strukturierung der Textsammlung durch die Dokumentenlandkarten berücksichtigt und somit eine Personalisierung ermöglicht werden.

Das entwickelte Rahmenmodell, welches die Voraussetzung für diese Adaptierbarkeit im zweifachen Sinne schafft, geht zunächst davon aus, dass ein generisch spezifiziertes Dokumentenanalysemodul die inhaltliche Verwandtschaft je zweier Dokumente bestimmt. Die interne formale Dokumentenrepräsentation dieses austauschbaren Moduls ist mit keiner anderen Komponente des Rahmenmodells verknüpft. Die vom Dokumentenanalysemodul gelieferten Rohdaten werden als Distanzwerte aufbereitet, wobei gilt: Je ähnlicher zwei Dokumente sind, desto geringer ist ihr Distanzwert. Geeignete Transformationen legt das Rahmenmodell in Abhängigkeit von den Eigenschaften des Vergleichsmoduls fest. Mit der Methodik des mehrdimensionalen Skalieren [Dav83] (bzw. verwandten Verfahren [FL95]) wird dann ein sogenannter semantischer Dokumentenraum erzeugt, in den Repräsentanten der Dokumente (in Form von Ortsvektoren) möglichst distanzerhaltend – bezüglich der "gemessenen" Dokumentendistanz – eingebettet werden.

Dieser hochdimensionale semantische Dokumentenraum ist Ausgangspunkt für die Personalisierung und Visualisierung der Korpusstruktur. Für erstere wurde eine Regelsprache entwickelt, die dem Analytischen die Definition von Zusammenhängen wichtiger Konzepte der Anwendungsdomäne über gewichtete Wortfelder, Attribut-Wert-Paare und einfache

Regeln über diese Konzepte ermöglicht. Der Mechanismus zur Regelinterpretation wertet die Regeln mit Methoden der schwachen Sprachverarbeitung und der Fuzzy-Logik aus und erzeugt "Kraftfelder" im semantischen Dokumentenraum, die durch ein speziell entwickeltes Vektortranslationsverfahren interpretiert werden und die Anordnung der Dokumentenrepräsentanten beeinflussen.

Für die Visualisierung des resultierenden semantischen Dokumentenraums wurde ein Verfahren gewählt, das dessen Topologie, also die vom Dokumentenanalysemodul und gegebenenfalls von der Regelanwendung definierten Verwandtschaftsbeziehungen der Dokumente, sehr leistungsfähig auf eine zweidimensionale Karte abbildet. Hierbei kommen die bereits oben erwähnten Kohonenkarten [Koh95] zum Einsatz, welche die Vektoren des Dokumentenraums so auf eine zweidimensionale Gitterstruktur abbilden, dass auf dem Gitter benachbarte Objekte auch im Eingaberaum benachbart sind. Ein spezielles Visualisierungsverfahren [Sk196] extrahiert die topologischen Informationen aus dem neuronalen Netz und stellt sie – wie im nächsten Abschnitt erläutert – als Grauwertbild dar.

4.2 Das interaktive Dokumentenlandkartensystem DocMINER

Das interaktive Dokumentenlandkartensystem DocMINER realisiert beispielhaft die vorgeschlagene Methodik und bietet zusätzliche, auf die Kartendarstellung abgestimmte Funktionen für die Analyse von Textsammlungen an. Dabei wird die grafische Darstellung des Dokumentenraums eng mit explorativen und zielgerichteten Suchmethoden verknüpft. In der verwendeten Visualisierungsmetapher stellen benachbarte Dokumente, dargestellt durch grafische Symbole, Texte mit verwandtem Inhalte dar. Je stärker sich die Texte thematisch unterscheiden, desto weiter liegen sie auch auf der Karte entfernt. Dieser Eindruck wird zusätzlich – ähnlich einer Höhenkarte – durch eine Farbverteilung gestärkt: Gruppen verwandter Dokumente liegen in gemeinsamen, hellen Bereichen und werden durch dunkle, unterschiedlich hohe "Berge" getrennt. Je dunkler die Hintergrundfarbe, desto stärker liegen die Gruppen inhaltlich auseinander (Abb. 1).

Eine Reihe von Interaktionsmöglichkeiten hilft dabei, eine gegebene Textsammlung schrittweise zu erkunden und Informationen über die Sammlung auf unterschiedlichen Granularitätsgraden zu erhalten. Dokumente können durch Klicken auf die Dokumentensymbole geöffnet werden, wobei wahlweise jedes Dokumentensymbol durch den Titel des Dokuments beschriftet ist. Für die Dokumentenpunkte können verschiedene grafische Symbole festgelegt werden. Auf diese Weise ist es möglich, vorgegebene Kategorien oder auch während der Arbeit gebildete Klassen optisch hervorzuheben und ihre Zusammenhänge gezielt zu studieren. Auch eine Schlagwortsuche in der Textsammlung ist möglich, wobei das Ergebnis der Suche nicht nur eine Liste der gefundenen Texte ist: Vielmehr werden die entsprechenden Punkte auf der Karte hervorgehoben, so dass auch der Zusammenhang des Ergebnisses mit der Kartenstruktur untersucht werden kann (Abb. 1).

Um den Inhalt einer beliebigen Dokumentengruppe zu "charakterisieren", kann die Verteilung der statistisch relevanten Schlagwörter (gemäß verschiedener Termgewichtungen) für ausgewählte Gruppen berechnet und angezeigt werden. Neben dieser wortbasierten Methode können auch solche Sätze aus den Dokumenten extrahiert und angezeigt werden,

welche Hinweise auf die Thematik der gewählten Textgruppe geben, wobei eine Kombination von Heuristiken der automatischen Textzusammenfassung zum Einsatz kommt. Das Spektrum der Interaktionsmöglichkeiten ist sicherlich nicht vollständig, realisiert aber einige wichtige Aspekte der Benutzerführung, die sich bei einer entsprechenden Betrachtung des Aufgabenmodells ergeben.

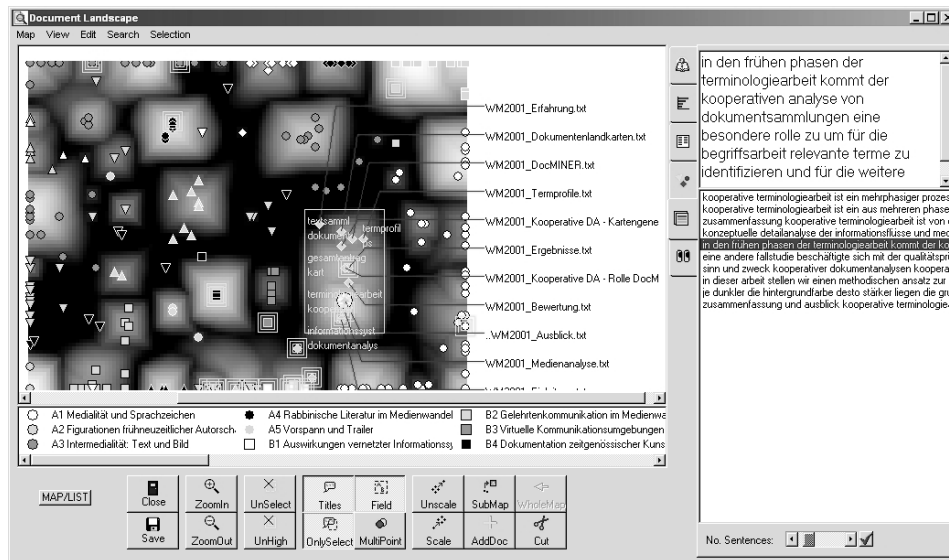


Abbildung 1: Benutzerschnittstelle des interaktiven Dokumentenlandkartensystems. Die Ergebnisdokumente zur Anfrage "Informationssysteme" sind in der Karte durch helle, konzentrische Rechtecke hervorgehoben. Die Dokumente des ausgewählten Feldes (helles Auswahlrechteck) sind mit Schlagworten beschriftet. Die Registerkarte zeigt statistisch wichtige Sätze der markierten Dokumentengruppe. Unter der Karte: Legende der verwendeten Dokumentensymbole.

5 Evaluation des Ansatzes im Anwendungskontext

Der vorgestellte Ansatz interaktiver Dokumentenlandkarten ist aufgabenorientiert für die Analyse spezialisierter Textkollektionen entwickelt worden. Eine wichtige Frage ist daher, ob und inwiefern Dokumentenlandkarten tatsächlich einen Mehrwert für die Analyse und Exploration dokumentierten Wissens in realen Anwendungsfeldern generieren können. Die immanente Vagheit von Korpusanalyseaufgaben hat zur Folge, dass formale Benchmarks kaum geeignet sind, einer Antwort näher zu kommen. Dieser Frage lässt sich daher nur mit qualitativen und empirischen Forschungsmethoden nachgehen. Im Folgenden werden Design und Ergebnisse von Fallstudien in unterschiedlichen Anwendungsfeldern und einem Benutzerexperiment in einer Laborumgebung skizziert.

5.1 Fallstudien in technischen und kulturwissenschaftlichen Disziplinen

Anhand von Fallstudien in industriellen und wissenschaftlichen Anwendungsdisziplinen wurde der vorgestellte Dokumentenlandkartenansatz für verschiedene Aufgaben der Analyse von Textsammlungen studiert. Dazu wurden jeweils die Analyseaufgaben im betrachteten Kontext untersucht und Methoden der Aufgabenunterstützung durch Dokumentenlandkarten entwickelt. An authentischen Textkollektionen sind dann – gemeinsam mit Fachleuten des Anwendungsgebietes – die ausgewählten Analyseaufgaben durchgeführt worden. Am Ende stand immer eine qualitative Bewertung des Unterstützungspotentials der Karten für den betrachteten Anwendungsbereich durch die Anwendungspartner.

In einer ersten Fallstudie im Rahmen des von der EU geförderten Projekts CAPE OPEN, in dem Schnittstellen für chemische Prozess-Simulatoren standardisiert wurden, ist das Landkartensystem zur Untersuchung von kollaborativ erstellten Use Cases (Anwendungsszenarien der Simulatoren für die Anforderungsanalyse) eingesetzt worden. Das Augenmerk lag hierbei auf der Frage, wie einzelne Anwendungsszenarien zusammenhängen und wie beschriebene Module miteinander interagieren. Mit Hilfe der Karte gelang es nicht nur, schnell interessante Zusammenhänge zwischen den beschriebenen Use Cases zu entdecken und eine hierarchische Gruppierung vorzunehmen, sondern auch, eine gegebene Struktur der Module sinnvoll zu verfeinern.

Eine andere Fallstudie beschäftigt sich mit der Qualitätsprüfung technischer Produktinformationen. Die zugrundeliegende Frage war dabei, ob die inhaltliche Kapitelstruktur komplexer Benutzerhandbücher konsistent aufbereitet war. In der Tat ließen sich visuell sehr gut Zusammenhänge zwischen den dargestellten Themenbereichen erkennen und überprüfen. Einige mit der Karte entdeckte Auffälligkeiten führten sogar zu einer partiellen Überarbeitung des Textmaterials. Im Kontext technischer Dokumentation spielt auch die Verdichtung von Informationen und die Schaffung von einheitlichen Informationsquellen (Single Sources) eine wichtige Rolle. Auch hier konnten Dokumentenlandkarten das Management von umfangreichen Sammlungen technischer Texte erfolgreich unterstützen. Schließlich wurde der Kartenansatz für die kooperative Terminologiearbeit im kulturwissenschaftlichen Forschungskolleg "Medien und kulturelle Kommunikation" (FK/SFB 427) eingesetzt. Ziel war dabei, die Verteilungen von Begriffen und die terminologischen Zusammenhänge in der Sprache der einzelnen Forschergruppen zu erkennen, wie sie sich in den Publikationen der beteiligten Wissenschaftler manifestieren. Hier half die Visualisierung der Korpusstruktur, relevante Begriffe für die Terminologiearbeit zu identifizieren und zu erkennen, in welchen unterschiedlichen Kontexten sie vorkommen.

In den skizzierten Anwendungsfeldern hat sich der Einsatz von Dokumentenlandkarten als fruchtbar herausgestellt. Als Mehrwert wurden insbesondere das einfache Zurechtfinden in komplexen Textsammlungen und die Möglichkeit genannt, schnell einen intuitiven Überblick über die Struktur des Korpus zu erlangen und Einsichten in thematische Zusammenhänge explorativ zu gewinnen. Ergebnisse der Fallstudien sind somit qualitative Hinweise auf die Nützlichkeit des Ansatzes, aber auch Methoden für die effektive Anwendung von Dokumentenlandkarten in praktischen Kontexten konnten studiert werden.

5.2 Ein Laborexperiment

Die Fallstudien lieferten deutliche anekdotische Evidenz für die Hypothese, dass der vorgeschlagene Ansatz den traditionellen Weg der Aufgabenlösung verbessert. Aber erzeugt die Visualisierung der Kollektionsstruktur tatsächlich einen *messbaren* Mehrwert für die Lösung von praktisch relevanten Analyseaufgaben? Lassen sich die Ergebnisse der Fallstudien verallgemeinern? Wie sieht es mit der Benutzerakzeptanz einer solchen Visualisierung aus? Zur Beantwortung dieser Fragen wurde ein kontrolliertes Laborexperiment mit Versuchspersonen entworfen und durchgeführt, bei dem das entwickelte Dokumentenlandkartensystem DocMINER mit einem alternativen Textlistensystem verglichen wurde. Die beiden Werkzeuge unterschieden sich lediglich in der Darstellung der Textkollektion: Statt der Karte wurde im alternativen Fall eine Liste der Titel aller in der Dokumentensammlung vorkommenden Texte präsentiert. Alle anderen Funktionen von DocMINER waren auch im Vergleichswerkzeug verfügbar.

Für das Experiment wurden ca. 30 Studierende der Informatik gebeten, in einer kontrollierten Experimentalsitzung verschiedene, von ihrer Natur her generische Korpusanalyseaufgaben – wie das schnelle Einarbeiten in eine Textsammlung, das Identifizieren von Klassifikationsfehlern oder das Gruppieren und Charakterisieren von Dokumenten – mit den beiden Werkzeugen durchzuführen. Dazu wurden die Versuchspersonen zufällig auf zwei Gruppen verteilt, die für die Lösung der Aufgaben je eines der beiden Systeme zur Verfügung hatten. Als Dokumentenkollektion dienten 300 Texte eines deutschsprachigen Nachrichtenmagazins aus den Bereichen Internet, Kultur und Wissenschaft. Die Aufgaben wurden mit Hilfe des Aufgabenmodells entworfen, das auch schon den Entwurf der Methodik und des Systems geleitet hat.

Die Qualität der Aufgabenlösungen der beiden Gruppen unterschied sich zum Teil signifikant. Dabei waren es jeweils die Benutzer der Dokumentenlandkarte, die die besseren Ergebnisse vorzuweisen hatten. Insbesondere im Bereich der Identifikation struktureller Irregularitäten, z.B. dem Erkennen von Klassifikationsfehlern, zeigte die grafische Darstellung wie erwartet ihre Stärke. Die Versuchspersonen gaben an, für die meisten der gestellten Aufgaben die Dokumentenlandkarte als Werkzeug zu bevorzugen. Interessanterweise gab es aber auch Situationen, in denen Aufgaben zwar mit der Karte bezüglich der gemessenen Qualität besser gelöst werden konnten, aber dennoch die eher traditionelle Textdarstellung bevorzugt wurde.

6 Zusammenfassung und Ausblick

Bei der hier beschriebenen aufgabenorientierten Entwicklung einer Methodik zur Berechnung von Dokumentenlandkarten im Wissensmanagement und ihrer systemtechnischen Umsetzung wurden die Potenziale grafischer Korpusübersichten konsequent ausgenutzt. Eine intensive qualitative und quantitative Evaluation hat gezeigt, dass der vorgeschlagene Ansatz tatsächlich einen effektiven und nützlichen Einblick in die Strukturen spezialisierter Textsammlungen ermöglicht, die zu erschließen für das Wissensmanagement so wichtig ist. Insbesondere helfen die Dokumentenlandkarten, relevante und interessante in-

haltliche Zusammenhänge zwischen Einzeltexten, Textgruppen und Dokumentenklassen auf Detailebene zu entdecken und zu analysieren.

Aktuelle Arbeiten nutzen die erzielten Resultate in verschiedenen Kontexten: Die Kombination des Dokumentenlandkartensystems mit einem ontologiebasierten Werkzeug für das Beobachten, Strukturieren und personalisierte Filtern von Informationsquellen ermöglicht beispielsweise neuartige wissensgestützte Navigations- und Analysemechanismen für Dokumentenräume. Im Rahmen der Entwicklung eines Führungsinformationssystems für die Wettbewerberanalyse werden geeignete Werkzeuge für die Überwachung und Untersuchung interner und externer, quantitativer und qualitativer Daten integriert, darunter DocMINER als dimensionsreduzierende Visualisierung für numerische und textbasierte Daten. Eine medientheoretische Untersuchung von Dokumentenlandkarten wird u.a. im kulturwissenschaftlichen Forschungskolleg "Medien und kulturelle Kommunikation" (SFB/FK 427) betrieben. Die leitende Frage ist dabei, welche Übersetzungsmechanismen – sogenannte "Transkriptionen" [Jäg01] – zwischen Symbolsystemen (wie etwa zwischen textbasierten Daten und ihrer Visualisierung in Form grafischer Korpusübersichten) für die Wissensorganisation in Kulturen angemessen sind.

Literaturverzeichnis

- [Bec01] Andreas Becks. *Visual Knowledge Management with Adaptable Document Maps*, volume 15 of *GMD research series*. GMD, Sankt Augustin, 2001. Dissertation, RWTH Aachen.
- [CC92] Matthew Chalmers and Paul Chitson. Bead: Explorations in Information Visualization. In *Research and Development in Information Retrieval*, pages 330–337, 1992.
- [CHSS98] Hsinchun Chen, Andrea L. Houston, Robin R. Sewell, and Bruce R. Schatz. Internet Browsing and Searching: User Evaluations of Category Map and Concept Space Techniques. *Journal of the American Society for Information Science*, 49:582–603, 1998.
- [CSO96] Hsinchun Chen, C. Schuffels, and R. Orwig. Internet categorization and search: a self-organizing approach. *Journal of Visual Communication and Image Representation*, 7(1):88–102, 1996.
- [Dav83] Mark L. Davidson. *Multidimensional Scaling*. John Wiley & Sons, 1983.
- [DHJ⁺98] G. S. Davidson, B. Hendrickson, D. K. Johnson, Ch. E. Meyers, and B. N. Wylie. Knowledge Mining With VxInside: Discovery Through Interaction. *Journal of Intelligent Information Systems*, 11(3), 1998.
- [FL95] Christos Faloutsos and King-Ip Lin. FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. In Michael J. Carey and Donovan A. Schneider, editors, *Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data*, pages 163–174, San Jose, California, 22–25 1995.
- [Jäg01] Ludwig Jäger. Transkriptivität – Zur medialen Logik der kulturellen Semantik. In Georg Stanitzek Ludwig Jäger, editor, *Transkribieren – Medien/Lektüre*, 2001.
- [Koh95] Teuvo Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer, Berlin, Heidelberg, 1995. (Second Extended Edition 1997).

- [LHKK96] Krista Lagus, Timo Honkela, Samuel Kaski, and Teuvo Kohonen. Self-organizing maps of document collections: A new approach to interactive exploration. In Evangelios Simoudis, Jiawei Han, and Usama Fayyad, editors, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pages 238–243. AAAI Press, Menlo Park, California, 1996.
- [Lin95] Xia Lin. Searching and Browsing on Map Displays. In *Proceedings of the ASIS '95 Annual Conference, American Society for Information Science*, 9–12 October 1995.
- [Sk196] Stefan Sklorz. A Method for Data Analysis Based on Self-Organizing Feature Maps. In *Proceedings of the World Automation Congress*, 1996.
- [WD00] Paul Warren and Graham Davies. Knowledge Management at BT Labs. Research, Technology, Management. *Journal of the Industrial Research Institute*, May–June 2000.
- [WTP⁺95] James A. Wise, James J. Thomas, Kelly Pennock, David Lantrip, Marc Pottier, Anne Schur, and Vern Crow. Visualizing the Non-Visual: Spatial Analysis and Interaction with Information from Text Documents. In Nahum D. Gershon and Steve Eick, editors, *Proc. IEEE Symp. Information Visualization, InfoVis*, pages 51–58. IEEE Computer Society, 30–31 October 1995.

Andreas Becks, geboren im Dezember 1971 in Wattenscheid, studierte Informatik mit Nebenfach Psychologie an der Universität Dortmund. Schwerpunkt seines Studiums waren wissensbasierte Systeme in der Medizin. In diesem Kontext entstanden seine ersten wissenschaftlichen Arbeiten im Bereich der Repräsentation vagen Wissens. Nach Abschluss seines Diploms im Dezember 1997 war Becks Stipendiat der DFG im Graduiertenkolleg "Informatik und Technik" an der RWTH Aachen, wo er im Juli 2001 mit der Arbeit "Visual Knowledge Management with Adaptable Document Maps" bei Matthias Jarke promovierte. Seit Januar 2001 ist er am Fraunhofer-Institut für Angewandte Informationstechnik FIT (vormals GMD-FIT) verantwortlich für das Arbeitsgebiet Wissensmanagement und leitet Projekte unter anderem zum Themengebiet "prozessorientiertes Wissensmanagement".