

MultiView-Systeme zur explorativen Analyse unstrukturierter Information

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften
der Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von
Diplom-Informatiker
Christian Seeling
aus Aachen

Berichter: Univ.-Prof. Dr. rer. pol. Matthias Jarke
Univ.-Prof. Dr. rer. nat. Thomas Seidl

Tag der mündlichen Prüfung: 28. Nov. 2007

**Fraunhofer Series in
Information and Communication Technology**

Band 3/2007

Christian Seeling

**MultiView-Systeme zur explorativen Analyse
unstrukturierter Information**

D 82 (Diss. RWTH Aachen)

**Shaker Verlag
Aachen 2007**

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Zugl.: Aachen, Techn. Hochsch., Diss., 2007

Copyright Shaker Verlag 2007

Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe, der Speicherung in Datenverarbeitungsanlagen und der Übersetzung, vorbehalten.

Printed in Germany.

ISBN 978-3-8322-6870-1

ISSN 1612-4863

Shaker Verlag GmbH • Postfach 101818 • 52018 Aachen

Telefon: 02407 / 95 96 - 0 • Telefax: 02407 / 95 96 - 9

Internet: www.shaker.de • E-Mail: info@shaker.de

Kurzfassung

Viele Aufgaben der Informationsrecherche erfordern die Verdichtung verschiedenartiger Informationen aus unterschiedlichen Quellen. Bei der Patentanalyse etwa müssen unstrukturierte Texte wie Patentkurzfassungen oder -ansprüche mit strukturierten Daten wie Patentklassen oder Firmendaten in Zusammenhang gebracht werden. Die schwierige Aufgabe der kombinierten Analyse strukturierter und unstrukturierter Information ist bei solchen Rechercheaufgaben mangels einheitlicher IT-Unterstützung meist mit aufwändiger Hand- und Kopfarbeit verbunden.

In der Dissertation wird ein neuartiges Systemkonzept zur integrativen Text- und Datenanalyse motiviert, entwickelt und bewertet, dessen Ziel es ist, die aufgabenrelevanten Informationen verknüpft und übersichtlich in einem System explorierbar zu machen. Kern des Konzepts sind grafische Übersichtsdarstellungen ('Views'). Zwischen den Views kann über Querbezüge in den Daten navigiert werden. Im System SWAPit wurde das Systemkonzept umgesetzt und empirisch untersucht. Eine konzeptionell-technische Kernherausforderung bei der Entwicklung war die aufgabenorientierte Anpassbarkeit mit dem Ziel, unterschiedliche Anwendungsdomänen und Rechercheaufgaben zu unterstützen. Daher stellen auch die entwickelten Methoden für Aufgabenbeschreibung, Systemdesign und Nutzenbewertung wichtige Beiträge dar. System und Konzept wurden systematisch in einem breiten Spektrum industrieller Fallstudien aus den Bereichen Business Intelligence (BI), Customer Relationship Management (CRM) und Kooperationssysteme (CSCW) evaluiert und optimiert.

Abstract

Information research tasks often require condensed information from different information sources. Patent analysis, for instance, requires the combination of structured data like patent classes or company names with unstructured text documents like patent abstracts or claims. Due to missing IT support, the combined analysis of structured and unstructured information often remains a manual and intellectual task.

This work motivates, develops and evaluates an approach for interactive text and data analytics. The goal of this approach is to offer concise exploration of the task-relevant information and relationships in a single system. Information from text documents, categories and relational data is presented in graphical views which are interactively coupled so that information relationships might be used for navigation among views. The system SWAPit implements the system concept. One conceptional-technical challenge has been the adaptability of the reference technology for different application domains, business processes, and tasks. Therefore, the methods for tailoring the system and for measuring usefulness are considered important contributions by themselves. SWAPit has been evaluated and optimized in a broad spectrum of industrial case studies from the fields of Business Intelligence (BI), Customer Relationship Management (CRM), and Cooperative Work (CSCW).

Danksagung

Ohne meinen Mentor und „alten“ Freund Andreas Becks wäre diese Arbeit nicht Realität geworden. Sein Ehrgeiz, Erkenntnisdrang und seine pragmatische aber gründliche Denk- und Arbeitsweise haben mich stets inspiriert und angespornt. Unsere gemeinsame Arbeitszeit war von fruchtbarem Dialog und Spaß geprägt und ich habe sehr viel gelernt. Danke dafür!

Ganz besonderer Dank gilt meiner geliebten Frau Anja, die auf ganz andere Weise den Nährboden für die Arbeit schaffte. So wenig sie mich inhaltlich unterstützen konnte, so sehr schaffte sie mir unser Zuhause und darin Ausgleich, Glück und Gelassenheit als Basis für die kreativen und anstrengenden Tage und Abende. Unsere beiden Kinder Torben und Mona sind mein ganzes Glück!

- Meiner kleinen Familie widme ich dieses Werk -

Professor Jarke verdanke ich den großzügigen Spielraum bei Forschung und Entwicklung. In den gemeinsamen Projekten habe ich das wissenschaftliche Arbeiten gelernt und geübt. Auch Professor Seidl danke ich herzlich für die auf den Punkt gebrachten Anregungen zur Arbeit.

Für die Unterstützung bei der Forschung und Entwicklung gilt mein Dank außerdem den studentischen Mitarbeitern Oliver Bresser, Omer Mushaid Khan und Ali Abbas, sowie den Diplomanden Patrick Eggert, Jinghui Xiang, David Kensche, Dino Frese, André Brunner und Stefan Hirschmeier.

Jessica Huster danke ich für die schöne gemeinsame Zeit im Büro sowie ihre Hilfsbereitschaft und ihr Interesse für mein Thema. Bei allen Freunden und Kollegen an der RWTH, bei FIT, bei Partnern und Kunden bedanke ich mich für die gute Zusammenarbeit und für ihre Unterstützung.

Aachen, im Dezember 2007

A handwritten signature in blue ink, consisting of several fluid, connected strokes that are difficult to decipher as specific letters.

Inhaltsverzeichnis

| | | |
|----------|--|------------------|
| 1 | <u>EINLEITUNG</u> | <u>1</u> |
| 1.1 | Strukturierte Daten für die Textrecherche..... | 1 |
| 1.2 | Ein MultiView-System für die Textanalyse | 3 |
| 1.3 | Forschungsfragen und Ziele..... | 4 |
| 1.4 | Bezugsrahmen | 5 |
| 1.5 | Forschungsmethodik..... | 7 |
| 1.6 | Übersicht und Lesehilfe | 8 |
| 2 | <u>MODELLE UND METHODEN.....</u> | <u>9</u> |
| 2.1 | Nutzen explorativer Suchsysteme | 9 |
| 2.2 | Modell zur Aufgabenbeschreibung..... | 11 |
| 2.3 | Portfolio zur Aufgabenbewertung | 14 |
| 2.4 | Verfahren zur Nutzenbewertung..... | 15 |
| 2.4.1 | BESONDERHEITEN EINER REFERENZTECHNOLOGIE | 15 |
| 2.4.2 | MODELLE ZUR SYSTEMANPASSUNG UND –BEWERTUNG..... | 16 |
| 3 | <u>STAND DER FORSCHUNG</u> | <u>19</u> |
| 3.1 | Auswahl und Klassifizierung relevanter Arbeiten..... | 19 |
| 3.2 | Integrierte Darstellung von strukturierten Daten und Textdokumenten..... | 20 |
| 3.2.1 | BEGRIFFE UND DOKUMENTE | 20 |
| 3.2.2 | KATALOGE UND DOKUMENTE..... | 25 |
| 3.2.3 | RELATIONALE DATEN UND DOKUMENTE | 29 |
| 3.2.4 | ZUSAMMENFASSUNG | 30 |
| 3.3 | Verknüpfende Darstellung von strukturierten Daten und Textdokumenten (MV)..... | 31 |
| 3.3.1 | AUTOFOCUS™..... | 32 |
| 3.3.2 | KNOWLEDGEEXPLORER..... | 34 |
| 3.3.3 | WEBTHEME™ (IN-SPIRE™) | 36 |
| 3.3.4 | CONVERSATION MAP..... | 37 |
| 3.4 | Frameworks zur Sichtenkoordination für die Text- und Datenanalyse..... | 38 |
| 3.4.1 | SNAP-TOGETHER..... | 38 |
| 3.4.2 | INVISION..... | 40 |
| 3.4.3 | DELAUNAY-VIEW..... | 41 |
| 3.5 | Überblick..... | 43 |
| 4 | <u>PROBLEMANALYSE AN FALLBEISPIELEN</u> | <u>45</u> |
| 4.1 | Fallstudie 1: Beschwerdemanagement in Versicherungen..... | 45 |
| 4.1.1 | HINTERGRUNDINFORMATION | 45 |
| 4.1.2 | IST-ANALYSE | 46 |
| 4.1.3 | POTENTIAL- UND AUFGABENANALYSE..... | 49 |
| 4.2 | Fallstudie 2: Produktsupport in Systemhäusern..... | 51 |
| 4.2.1 | HINTERGRUNDINFORMATION | 51 |
| 4.2.2 | IST-ANALYSE | 52 |
| 4.2.3 | POTENTIAL- UND AUFGABENANALYSE..... | 56 |
| 4.3 | Fallstudie 3: Aktivitätsanalyse in Kooperationsumgebungen..... | 59 |
| 4.3.1 | HINTERGRUNDINFORMATION | 59 |

| | | |
|------------|--|------------|
| 4.3.2 | IST-ANALYSE | 59 |
| 4.3.3 | POTENTIAL- UND AUFGABENANALYSE..... | 62 |
| 4.4 | Informationsarten und Verknüpfungen..... | 65 |
| 4.4.1 | ANALYSE DER INFORMATIONSTYPEN UND VERKNÜPFUNGEN | 65 |
| 4.4.2 | TEXTE UND TEXTÄHNLICHKEITEN..... | 66 |
| 4.4.3 | BESCHREIBUNGSKATEGORIEN..... | 67 |
| 4.4.4 | RELATIONALE DATEN..... | 69 |
| 4.4.5 | INFORMATIONSTYPEN U. VERKNÜPFUNGEN | 69 |
| 5 | <u>SYSTEMKONZEPT: INTEGRATIVE TEXTRECHERCHE MIT MULTIPLE VIEWS</u> | 73 |
| 5.1 | Rationale: Komplementäre Sichtenarten..... | 73 |
| 5.2 | Sichtenarten und Verknüpfungen | 74 |
| 5.3 | Sichtenkoordination und Systemverhalten | 75 |
| 5.4 | Relevanzmaße für Markierungen von Informationsobjekten | 78 |
| 5.5 | Informationsobjekte zur Beschreibung von Dokumenten..... | 80 |
| 5.6 | Assoziative Exploration | 81 |
| 5.6.1 | SICHTENÜBERGREIFENDE EXPLORATION – AUSWAHLEN ALS INTERAKTIVE ANFRAGEN | 82 |
| 5.6.2 | SICHTENINTERNE EXPLORATION – STRUKTUREN UND WERKZEUGE..... | 84 |
| 5.7 | Diskussion | 90 |
| 5.7.1 | CHARAKTERISIERUNG DES MULTIPLE VIEW-ANSATZES | 90 |
| 5.7.2 | SEMANTIK VON PRÄSENTATION UND INTERAKTION | 90 |
| 5.7.3 | ABGRENZUNG..... | 91 |
| 6 | <u>REFERENZTECHNOLOGIE SWAPIT</u> | 93 |
| 6.1 | Entstehung | 93 |
| 6.2 | Architektur | 93 |
| 6.3 | Administration mit DocMINER und SWAPitUp..... | 95 |
| 6.4 | Koordination der Sichten | 96 |
| 6.5 | Repräsentation der Informationsarten in Sichten | 97 |
| 6.5.1 | TEXTÄHNLICHKEITEN | 97 |
| 6.5.2 | BESCHREIBUNGSKATEGORIEN..... | 100 |
| 6.5.3 | RELATIONALE DATEN..... | 101 |
| 6.6 | Interaktive Werkzeuge | 101 |
| 6.7 | Sichtenspezifikation, Inhaltsextraktion und Informationsverknüpfung | 104 |
| 6.8 | Stellschrauben zur Systemanpassung..... | 105 |
| 6.8.1 | INFORMATIONSZUGRIFF | 105 |
| 6.8.2 | DOKUMENTENVERGLEICH..... | 106 |
| 7 | <u>EVALUATION IN FALLSTUDIEN.....</u> | 107 |
| 7.1 | Allgemeine Bewertungsziele | 107 |
| 7.2 | Systematik der Fallstudien | 108 |
| 7.3 | Fallstudie 1: Beschwerdemanagement in Versicherungen..... | 110 |
| 7.3.1 | BEWERTUNGSZIELE | 110 |
| 7.3.2 | SYSTEMDESIGN | 111 |
| 7.3.3 | INTEGRATION | 113 |
| 7.3.4 | VERSUCHSDESIGN UND DURCHFÜHRUNG | 113 |
| 7.3.5 | ERGEBNISSE | 113 |
| 7.4 | Fallstudie 2: Produktsupport in Systemhäusern..... | 116 |
| 7.4.1 | BEWERTUNGSZIELE | 116 |

| | | |
|------------|--|-------------------|
| 7.4.2 | SYSTEMDESIGN | 116 |
| 7.4.3 | INTEGRATION | 118 |
| 7.4.4 | VERSUCHSDSIGN UND DURCHFÜHRUNG | 118 |
| 7.4.5 | ERGEBNISSE | 120 |
| 7.5 | Fallstudie 3: Aktivitätsanalyse in Kooperationsumgebungen..... | 124 |
| 7.5.1 | BEWERTUNGSZIELE | 124 |
| 7.5.2 | SYSTEMDESIGN | 124 |
| 7.5.3 | INTEGRATION | 126 |
| 7.5.4 | VERSUCHSDSIGN UND DURCHFÜHRUNG | 128 |
| 7.5.5 | ERGEBNISSE | 129 |
| 7.6 | Systemverwendung | 133 |
| 8 | <u>FAZIT</u> | <u>135</u> |
| 8.1 | Zusammenfassung und Beiträge..... | 135 |
| 8.2 | Ergebnisse und Schlussfolgerungen | 135 |
| 8.2.1 | SWAP-SYSTEMKONZEPT UND SWAPIT-REFERENZTECHNOLOGIE | 135 |
| 8.2.2 | ERGEBNISSE FÜR DIE EXPLORATIVE TEXTRECHERCHE..... | 137 |
| 8.2.3 | INTERAKTIVE WERKZEUGE FÜR MULTIVIEW-SYSTEME | 137 |
| 8.3 | Ausblick..... | 138 |
| 9 | <u>ANHANG.....</u> | <u>141</u> |
| 9.1 | Notationen..... | 141 |
| 9.1.1 | GQM: GOAL, QUESTION, METRIC | 141 |
| 9.1.2 | K3: KOORDINATION, KOOPERATION, KOMMUNIKATION..... | 141 |
| 9.1.3 | EPK: EREIGNISGESTEUERTE PROZESSKETTE | 141 |
| 9.1.4 | SWAPIT-SYSTEMSKIZZE | 142 |
| 9.2 | Detaillierte Beschreibung des Aufgabenmodells | 143 |
| 9.2.1 | FUNKTIONALE UND INFORMATORISCHE MÄCHTIGKEIT | 144 |
| 9.2.2 | KONTROLLE DER INFORMATIONSRECHERCHE | 148 |
| 9.2.3 | QUALITÄT UND WERT DES SERVICE | 150 |
| 9.2.4 | AUFWAND DER ARBEITSPROZESSE | 151 |
| 9.3 | Weitere Anwendungsbeispiele | 154 |
| 9.4 | Veröffentlichen zur Arbeit | 166 |
| 9.4.1 | WISSENSCHAFTLICHE PUBLIKATIONEN | 166 |
| 9.4.2 | PRESSESPIEGEL..... | 166 |
| 9.5 | Fragebogen zur Darstellung von Matching-Werten | 167 |
| 9.6 | Positivlisten zur Fallstudie 1 (Beschwerdemgmt.)..... | 169 |
| 9.7 | Fragebögen zur Fallstudie 2 (Produktsupport)..... | 170 |
| 9.8 | Aufgaben- und Fragebögen zur Fallstudie 3 (Kooperationsumgebungen)..... | 176 |
| 10 | <u>LITERATURVERZEICHNIS.....</u> | <u>187</u> |

1 Einleitung

1.1 Strukturierte Daten für die Textrecherche

Unternehmen sind in hohem Maße auf die Erschließung und Nutzung ihrer Informationsressourcen angewiesen. Dabei liegt ein *Großteil der für Entscheidungen relevanten Information aus unternehmensinternen und -externen Informationsquellen in Textform* vor¹, darunter beispielsweise Branchennachrichten, Firmenprofile oder Produktbeschreibungen. Weniger als die Hälfte der für strategische Entscheidungen benötigten Informationen stammen aus dem eigenen Unternehmen [Ba96].

In vielen betriebswirtschaftlichen Entscheidungssituationen müssen nicht nur Informationen aus verschiedenen Quellen berücksichtigt und kombiniert werden, sondern auch *verschiedene Arten von Informationen*. Bei der Patentanalyse, Produktprogrammplanung, Umsatzbewertung und Kundenanalyse etwa wird der Zusammenhang zwischen strukturierten Daten und Informationen aus unstrukturierten Textdokumenten benötigt. Die Texte sind meist nach hierarchischen Ablage- oder Zugriffsstrukturen organisiert [BS01]. Dazu ein Beispiel: Online-Versandhäuser arbeiten kontinuierlich daran, die Präsentation der Produktpalette und die Servicequalität bei der Kaufabwicklung zu verbessern. Dabei spielen die Kundenrezensionen eine herausragende Rolle, weil sie direktes Kundenfeedback repräsentieren. Zusätzlich zu der quantitativen Produktbewertung erläutern die Kunden ihre Einschätzung in Texten. Das Marketing ist daran interessiert, welche Eigenschaften von Produkt, Präsentation und Service die Kundenbewertung und das Kaufverhalten maßgeblich beeinflussen. Dafür müssen die Zusammenhänge zwischen Kundenrezensionen, Produktkatalogen, Verkaufsrängen / Verkaufszahlen, Zugriffszahlen der Webseiten u.v.m. untersucht werden.

Am Beispiel wird klar, dass die Texte (Kundenrezensionen) zwar die entscheidende Informationsquelle darstellen um Gründe für Bewertung und Kaufverhalten zu ermitteln, aber nur *ein* Baustein des gesamten Analyseszenarios sind. Auch die strukturierten Daten spielen eine wichtige Rolle. Sie können einerseits dazu verwendet werden, Texte zu *spezifizieren* (betrachte z.B. Kundenrezensionen zu Top-Produkten mit hohem Verkaufsrang). Andererseits können sie helfen, Texte zu *beschreiben* (betrachte z.B. Produktgruppen und Web-Zugriffe zu den besonders negativen Kundenrezensionen)².

Solche komplexen Rechercheaufgaben dienen der Informationsversorgung zur Entscheidungsunterstützung – im skizzierten Szenario etwa entscheidet das Marketing über die Veränderung der Produktpräsentation. Bis zur Entscheidungsfindung erfahren die Rohdaten dabei mehrere Stufen der Veredelung. Lohmann beschreibt diesen Prozess in einem Pyramidenmodell mit den Stufen Rohdaten → Metadaten → Zusammenstellung → optische Aufbereitung → Auswertung → Bewertung → Entscheidungshilfe → Entscheidung ([Lo07], S.22). Keen et al. fordern von Systemen zur Entscheidungsunterstützung die Hilfestellung für Manager bei der Entscheidung in Bezug auf semi-strukturierte Aufgaben, die Unterstützung der Urteilsfällung durch Manager, sowie die Steigerung der Effektivität der Entscheidungsfällung [KM78]. Von Recherchesystemen wurde und wird zur Entscheidungsunterstützung gerade keine vollständige Automatisierung erwartet, sondern die geeignete Bereitstellung von Informationen und Funktionalitäten für den Experten. Besonders wichtig ist diese bei *unstrukturierten Entscheidungsproblemen*, bei denen der Lösungsweg a priori allenfalls vage bekannt ist. Holten empfiehlt für solche Probleme interaktive Recherchesysteme mit flexibler Benutzernavigation: *„Browsing ermöglicht die Datenquellen-übergreifende, nicht-sequentielle Nutzung eines Informationssys-*

¹ Eine 1995 durchgeführte Untersuchung öffentlich zugänglicher Datenbanken ergab einen Anteil an Information in Textform vom gesamten Informationsangebot von >70% [HH95]. Nach neueren Angaben liegen sogar 80-90% der Informationen eines Unternehmens in unstrukturierten Textdokumenten vor ([Ka03], S.189).

² Zur Unterscheidung siehe auch Kap.5.6.

tems und ist daher zur Unterstützung assoziativen und deduktiven Problemlösens besser geeignet als Operationen wie der Drill-Down, die fest vorgegebene Zugriffspfade unterstützen“ [Ho99].

Die *Informationsqualität*, die die Entscheidungsqualität maßgeblich beeinflusst, hängt nach Königer et al. [KR98] nicht nur von der „inneren Qualität“ der Informationen (z.B. Genauigkeit, Vertrauenswürdigkeit) ab, sondern auch von Kontext, Strukturierung, Darstellung und Zugangsqualität. Daher sollten die entscheidungsrelevanten Informationen in *einem* Analysekontext zusammen gebracht und auch die Struktur, also die logischen Verknüpfungen dazwischen, für den Zugriff auf Information nutzbar gemacht werden. Beschreibungskategorien und Metadaten zu Texten werden bereits von vielen Dokumentenmanagementsystemen bereitgestellt, allerdings eher zum Spezifizieren als zum Beschreiben von Texten und Textgruppen [FSH02]. Relevante strukturierte Daten müssen dann meist über separate Datenbankabfragen oder Reportingsysteme recherchiert werden. Diese Systeme bieten andere Zugriffsmechanismen. Daher müssen die Erkenntnisse aus der Textrecherche ohne Systemunterstützung in Beziehung zu den anderen Informationen gebracht werden, mit denen die Texte logisch, manchmal auch physisch, oft aber nicht systemisch verknüpft sind.

Während in operativen Prozessen elementare Informationen betrachtet werden, benötigen Entscheidungsprozesse verdichtete Informationen, beispielsweise aggregiert über Abteilungsgrenzen und Informationsarten hinweg ([BS06], S.65). Informationen zur Entscheidungsunterstützung können in vielerlei Hinsicht heterogen sein, also der Verdichtung bedürfen: Nach der *Quelle* unterscheidet Lohmann (Unternehmens-)externe von -interner Information und diskutiert die integrierte Verwendung beider Typen [Lo07]. Mertens unterscheidet qualitative und quantitative Information [Me02] in Bezug auf die Verwendung in Führungsinformationssystemen. Valente unterscheidet nach dem Grad der Strukturiertheit von Information [Va01].

Systeme für den Zugriff auf Datenbanken blicken auf eine lange Entwicklung zurück. Zielgerichtete Anfragen lassen sich in Anfragesprachen für relationale Datenbanken formulieren. Die meisten Datenbankhersteller haben Operatoren zur Suche in Textinformation in ihre Produkte aufgenommen [Su01]. Für die entscheidungsbezogene flexible (ad-hoc) Analyse der Information aus multidimensional organisierten Data Warehouses hat sich das OLAP-Paradigma etabliert [CCS93][KB+05].

In zahlreichen Arbeiten wurde entweder vorgeschlagen, Textressourcen in die OLAP-Kennzahlenanalyse einzubeziehen [CK+02][Ba04], oder das OLAP-Paradigma auf die Navigation in Textsammlungen zu übertragen [KHA98][Su01][AK+02][MC+03]. Andere Arbeiten verlangen wenigstens die gleiche Systematik und Kontinuität der Beschaffung und Bereitstellung strukturierter wie unstrukturierter Informationen [Ha99][He99]. Gerade der Analyse von unstrukturierten Datenressourcen wird dabei eine hohe Bedeutung beigemessen: „*UIM Applications [Unstructured Information Management] will change the world like database technology did“* [CK+02]. Solche Überlegungen führten zu einem erweiterten Verständnis von „Business Intelligence“, in dem auch die unstrukturierten Daten eine gleichberechtigte Rolle spielen, weil Entscheidungen aufgrund strukturierter *und* unstrukturierter Informationen getroffen werden [BR07]. Die traditionelle Spezialisierung von Systemen zur Bereitstellung strukturierter Information einerseits und unstrukturierter Information andererseits wurde in den letzten Jahren zunehmend relativiert. So zielen moderne Lösungen des „Enterprise Content Management“ (ECM) auf die ganzheitliche Informationsbereitstellung ab [CZ03].

In dieser Arbeit wird ein Ansatz zur explorativen Recherche in unstrukturierten *und* strukturierten Daten vorgestellt, der auf dem Einsatz multipler grafischer Sichten (Multiple Views) basiert und die entscheidungsrelevanten Informationen übersichtlich darstellt.

1.2 Ein MultiView-System für die Textanalyse

In Vorarbeiten wurde bereits deutlich, dass Textdokumente nicht isoliert existieren, sondern in Zusammenhang mit strukturierten Daten aus Datenbanken und Ablagestrukturen stehen. So belegten Interviews mit Unternehmen aus wissensintensiven Branchen, dass Texte meist in *hierarchische Ordnungsstrukturen* eingebettet sind [BS01]. In einer vergleichenden Studie stuften die Benutzer die *Tabelle von Metadaten* zu Textdokumenten (als Ergänzung einer grafischen Ähnlichkeitsdarstellung der Inhalte) als für die Recherche unerlässlich ein [BSM02]. Zahlreiche Fallstudien zeigten, dass bei der Recherche in unstrukturierten Datenbeständen solche strukturierten Daten berücksichtigt werden müssen (Kap.3). Daher bietet sich an, die strukturierten Daten unmittelbar für die Textrecherche nutzbar zu machen, indem die Textdokumente im Kontext aller relevanten strukturierten Daten „auf einen Blick“ dargestellt werden und die strukturierten Daten zum Navigieren (Beschreiben und Spezifizieren der Texte) bereitgestellt werden. Das MultiView-Systemparadigma adressiert den parallelen Zugriff auf verschiedene Informationsobjekte oder auf verschiedene Aspekte derselben Informationsobjekte und wird im Folgenden vorgestellt.

MultiView-Systeme stellen Informationen in mehreren gleichzeitig sichtbaren Fenstern (Sichten) dar. Die Fenster sind interaktiv gekoppelt, so dass Benutzerinteraktionen in einer Sicht in der Regel Veränderungen der verknüpften Sichten bewirken. Baldonado et al. definieren eine Sicht als Kombination von Daten und Darstellung: “[...] a single view of a conceptual entity as a set of data plus a specification of how to display that data visually” und “views are distinct if they allow the user to learn about different aspects of the conceptual entity, e.g. by presenting different information, or by emphasizing different aspects of the same information. A multiple view system uses two or more such distinct views to support the investigation of a given conceptual entity” [BWK00]. MultiView-Systeme kommen bereits seit langem zum Einsatz, zum Beispiel Werkzeuge zur Dateiverwaltung, wo in verschiedenen Sichten die Laufwerke, Ordner und Dateien dargestellt werden. Für die Analyse komplexer Daten eingesetzt wurden MultiView-Systeme u.a. zur Exploration von Multimedia-Datenbanken [CS+97], von großen Inhaltsverzeichnissen [CS94], von Weblog-Daten [NS00], von Simulationsdaten in der Medizin [GR+00], oder zur Analyse verschiedener struktureller Aspekte von Dokumentensammlungen [HW+98]. Auch die verbreitete facettenbasierte Navigation kann als MultiView-System aufgefasst werden [Po98]. Aufgrund der hohen Komplexität der Interaktion mit solchen Systemen wurde das MultiView-Konzept vor einigen Jahren explizit als Forschungsgegenstand adressiert³. Infolgedessen wurden Darstellungs- und Koordinationsarten systematisiert [NS00], Empfehlungs-Richtlinien zur Entwicklung solcher Systeme vorgeschlagen [BWK00] und Frameworks für die Konstruktion solcher Systeme entwickelt (Kap.3.4). Bei MultiView-Systemen steht die Komplexitätsreduktion durch Dekomposition⁴ dem potentiell hohen kognitiven Aufwand der Bedienung gegenüber: Aufgrund der geringen Erfahrung mit solchen komplexen Systemen ist die Informations- und Funktions-Vielfalt potentiell kritisch für Benutzbarkeit. Die Vorzüge und Nachteile von MultiView-Systemen werden unter Bezugnahme auf relevante Studien detaillierter in Kap.3.4 besprochen.

In dieser Arbeit wird ein MultiView-Systemkonzept vorgestellt, das Informationen aus Textdokumenten und damit verknüpfte strukturierte Daten, die für den Aufgabenkontext relevant sind, in separaten grafischen Sichten darstellt. Informationen und Querbezüge werden übersichtlich dargestellt und die strukturierten Daten werden zur Exploration der Texte nutzbar gemacht. Das Konzept stellt den Benutzer in den Vordergrund: Die menschlichen Kernfähigkeiten Kognition, Erfahrung und Intuition sollen durch grafische Übersichtsdarstellungen von Dokumentenräumen und durch flexible Interaktionspfade aktiviert werden. Die Referenz-

³ Seit 2003 findet jährlich eine internationale Konferenz zu diesen Thema statt (International Conference on Coordinated & Multiple Views in Exploratory Visualization, CMV).

⁴ “Multiple views can help the user to ‘divide and conquer’, complex data can be partitioned into multiple views to create manageable chunks.” [Sa03]

technologie SWAPit realisiert das Systemkonzept und unterstützt sowohl datengetriebene (explorative) als auch hypothesengetriebene (anfrageorientierte) Analysen. SWAPit lässt sich als Referenztechnologie für verschiedene Szenarien der Informationsrecherche anpassen und ermöglicht die praxisnahe Evaluation in realen Anwendungsszenarien.

1.3 Forschungsfragen und Ziele

Für die kombinierte Analyse von Texten und strukturierten Daten ist, wie oben angerissen und in Kap.5.1 vertieft wird, die Dekomposition in verknüpfte grafische Sichten vielversprechend. In der Arbeit wird untersucht, wie gut ein MultiView-Konzept für die explorative Text- und Datenanalyse geeignet ist, das auf die *Anforderungen der Textrecherche* spezialisiert ist, jedoch zugleich für ganz *unterschiedliche Anwendungsdomänen* eingesetzt werden kann. Zwar existieren einerseits zahlreiche MultiView-Spezialsysteme zur explorativen Textanalyse (bspw. Patentrecherche) und andererseits einige generische Frameworks zur technischen Realisierung von MultiView-Systemen, jedoch kein auf Textanalyse spezialisiertes MultiView-Framework.

Tabelle 1-1: Forschungsfragen und Ziele der Arbeit

| Ziele | Forschungsfragen |
|---|--|
| Entwicklung eines <i>MultiView-Systemkonzepts</i> zur Textrecherche basierend auf der Problem- und Anforderungsanalyse in realen Fallbeispielen. Das Konzept muss folgende Aspekte beschreiben: <ul style="list-style-type: none"> ▪ Darstellung von Texten und strukturierten Daten in Sichten, ▪ interaktive Koordination der Sichten mittels der Verknüpfungen der Informationen, ▪ interaktive Werkzeuge zur Spezifikation und Beschreibung von Texten und strukturierten Daten. | Welche <i>strukturierten Daten</i> spielen bei der Textrecherche eine Rolle, und in welchem Zusammenhang stehen sie zu den Textdokumenten? |
| | Wie können die relevanten strukturierten Daten und ihre Zusammenhänge mit den Texten in grafischen Sichten <i>dargestellt und interaktiv zur Textrecherche genutzt</i> werden? |
| Beschleunigung, Vereinfachung und Verbesserung des Entwurfs solcher Systeme für konkrete Szenarien der Textrecherche durch Konstruktion einer wieder verwendbaren <i>Referenztechnologie</i> und durch Spezifikation eines standardisierten <i>Prozesses zur Systemanpassung</i> . | Wie kann das Systemkonzept in einer <i>Referenztechnologie</i> realisiert werden, die ohne Entwicklungsaufwand für ein großes Spektrum von Szenarien der Textrecherche eingesetzt werden kann? |
| | Welche Schritte sind notwendig, um die Referenztechnologie an konkrete Szenarien der Textrecherche <i>anzupassen (Vorgehensmodell)</i> ? |
| Entwicklung eines <i>Bewertungsframeworks</i> , das betriebswirtschaftliche sowie informationstechnische Nutzenkriterien berücksichtigt. Empirische <i>Bewertung</i> von Systemkonzept und Referenztechnologie in realen Fallstudien basierend auf dem entwickelten Framework. | Wie lässt sich der <i>Nutzen</i> einer solchen Referenztechnologie als Ergänzung bestehender Anwendungssysteme zur Textanalyse nach betriebswirtschaftlichen und informationstechnischen Kriterien messen und wie hoch ist er in ausgewählten Szenarien? |

Eine *anwendungsübergreifende* Betrachtung macht einen technisch-methodischen Rahmen erforderlich, der zum einen eine *vielseitig einsetzbare Referenztechnologie*⁵ umfasst und zum anderen die *domänenübergreifende Modellierung von Analyseaufgaben* im Aufgabenmodell beinhaltet. Außerdem wird ein *einheitliches Maßsystem* benötigt, das den *Nutzen der Systemunterstützung* in Anwendungsdomänen erfasst. Bei der Nutzenbetrachtung finden sowohl informationstechnische Kriterien (z.B. Anpassbarkeit, Integrierbarkeit, Benutzbarkeit) als auch betriebswirtschaftliche Kriterien (z.B. Ergebnisqualität, Betriebskosten) Berücksichtigung. Dabei geht es um die *Bestimmung des Mehrwerts* der MultiView-Technologie als Ergänzung und im Vergleich zur bestehenden Infrastruktur zur Informationsrecherche.

Somit adressiert die Arbeit die in Tabelle 1-1 formulierten Forschungsfragen. Die Unterscheidung in die drei Abstraktionsebenen *Systemkonzept*, *Referenztechnologie*, *konkretes System* (Kap.1.5) spiegelt sich nicht nur in den Zielen der Arbeit wieder, die den Forschungsfragen zugeordnet sind, sondern auch in der gesamten Struktur der Arbeit.

1.4 Bezugsrahmen

Diese Arbeit adressiert die interaktive Informationsrecherche in Unternehmen. *Recherche* bezeichnet hierbei alle Aufgaben im Spektrum der Suche nach bekannter Information bis hin zur freien Exploration unbekannter Informationsräume ([Me92]⁶, [Be01] S.11). Solche Aufgaben der Informationsrecherche im betrieblichen Umfeld können nach folgenden drei Dimensionen unterschieden werden (Abbildung 1.1):

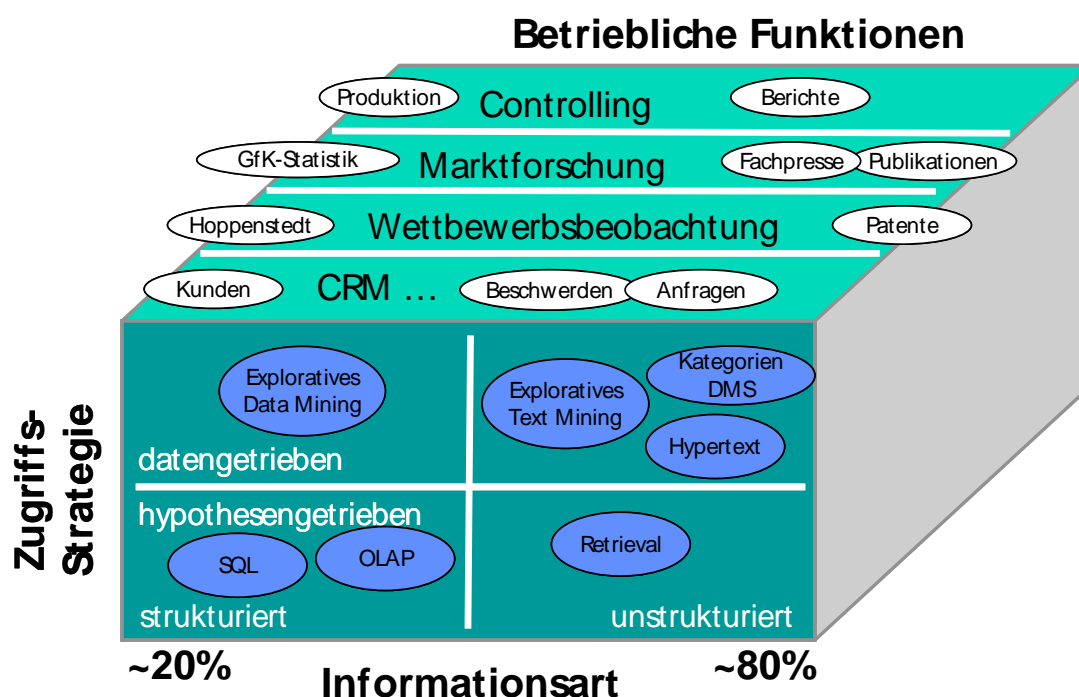


Abbildung 1.1: Bezugsrahmen⁷ der Arbeit – Betriebliche Funktionen, Informationsarten, Zugriffsstrategien

⁵ Der Begriff *Referenztechnologie* wird in Kap.1.5 definiert. Die Besonderheiten von Referenztechnologien werden in Kap.2.4 diskutiert.

⁶ Meadow unterscheidet Aufgaben danach, ob der Informationsbedarf formulierbar und die Daten bekannt sind.

⁷ Ein Bezugsrahmen ist „ein theoretisches Konzept, in dem der Forschungshintergrund ausführlich erläutert wird und sämtliche relevanten Einflussfaktoren gegenüber der Umwelt abgegrenzt werden“ [LN05]. Dabei geht es um die vorläufige

In allen **betrieblichen Funktionsbereichen** werden für operative oder dispositive Aufgaben Informationen benötigt. Diese lassen sich nach **Informationsarten** in strukturierte und unstrukturierte Informationen unterscheiden ([Me02], S.5). Die Methoden und Werkzeuge zur Informationsrecherche unterstützen meist entweder eine daten- oder eine hypothesengetriebene **Zugriffsstrategie**.

Viele analytische Aufgaben benötigen jedoch *verschiedenartige* Information aus *verschiedenen Funktionsbereichen* und erfordern die Kombination *verschiedener Zugriffsstrategien*. Dabei ergeben sich drei Barrieren, die ein Analyst überwinden muss:

Die **Funktionsbarriere** („gedanklicher Zaun“ [Me99]) besteht durch die unterschiedlichen betriebswirtschaftlich motivierten Sichten auf Information. So bewertet ein Controller die Umsatzentwicklung des Unternehmens anders als ein Mitarbeiter des Marketings, für den die Außenperspektive maßgeblich ist und von dem ein Umsatzzuwachs folglich nach der Entwicklung des Marktanteils bewertet wird. Weniger Kundenbeschwerden sind aus Sicht des Vertriebs im Allgemeinen positiv, das Controlling wird seine Einschätzung allerdings davon abhängig machen, ob auch der Absatz gesunken ist.

Der weitaus größere Teil der Informationen jedes Unternehmens liegt in Textform vor ([Ka03], S.189), die anderen Informationen in strukturierter Form. Die **Datenbarriere** besteht durch die gewachsene Unterscheidung von Systemen zur Recherche in strukturierter und unstrukturierter Information. So verwenden Werkzeuge zur Datenbankrecherche andere Algorithmen und Benutzeroberflächen als Werkzeuge des Information Retrieval für Textsammlungen. Die Datenintegration hilft mit semantischen Technologien verschiedene Informationsarten zu verknüpfen. Allerdings bleibt die Frage nach dem integrierten Zugriff offen.

Die **Zugriffsbarriere** besteht durch die Unterschiede zwischen daten- und hypothesenorientiertem Vorgehen bei der Recherche. Während Werkzeuge für exploratives Text und Data Mining die interaktive Analyse großer Informationsbestände unterstützen, helfen Suchsysteme bei der Spezifikation und dem Finden von bekannten Informationsobjekten.

Geeignete Zugriffsstrategien und einzubeziehende Informationsarten sind abhängig von der Funktion und Aufgabe eines Analysten. So sucht der Controller eine Rechnung, das Marketing will sich einen Überblick über die Konkurrenzsituation machen, und die Produktentwicklung prüft die Patente in einem Technologiebereich. Die Überwindung von Funktionsbarrieren erfordert die Veränderung von Geschäftsprozessen und möglicherweise der Organisationsstruktur. Solche Veränderungen werden in dieser Arbeit nicht adressiert (wenngleich es bei der ersten Fallstudie auch um die Überbrückung von Abteilungsgrenzen ging, vgl. Kap.4.1). In der Arbeit soll untersucht werden, wie Daten- und Zugriffsbarrieren durch neuartige Darstellungs- und Interaktionskonzepte überwunden werden können. Dabei spielen Text Mining-Verfahren, Informationsvisualisierung und verknüpfte grafische Sichten eine Rolle, die für die integrative Text- und Datenanalyse eingesetzt werden.

Die Verknüpfung der Information ist *Voraussetzung* für den integrativen Zugriff, nicht Gegenstand der Forschung (Kap.5.7.3). Vielmehr geht es um die interaktive *Verwendung* aller verknüpften Informationen in einem Aufgabenkontext. Allerdings wird in Kap.6.7 diskutiert, wie sich Informationen (semi-)automatisch verknüpfen lassen. Außerdem werden folgende Unterscheidungsmerkmale für Verknüpfungen besprochen:

- berechnete vs. explizit definierte Verknüpfungen (Kap.4.4)
- Verknüpfungen von Informationsobjekten der gleichen Informationsart bzw. verschiedener Informationsarten (Kap.4.4)
- Zustand bzw. Existenz der Verknüpfungen⁸ (Anhang 11.2.1)

fige und plausible Erläuterung der für die Arbeit wichtigen Objekte und Begriffe, ohne allerdings Lösungswege oder Algorithmen vorweg zu nehmen.

⁸ mögliche Ausprägungen des Aufgabenmerkmals sind: physisch, organisatorisch, logisch

Die Arbeit beschäftigt sich mit der Systemunterstützung für Rechercheaufgaben, die sich nicht vollständig automatisieren lassen, sondern Expertise und Intuition eines Experten erfordern. Daher liegt der Schwerpunkt auf Darstellungs- und Interaktionskonzepten zur Unterstützung der *explorativen Recherche*. Allerdings werden Verfahren der automatischen Text- und Datenanalyse eingesetzt, um Informationen für den interaktiven Zugriff zu berechnen, darunter

- Statistische Analyse von Texten und strukturierten Daten (Kap.6.6)
- Analyse der inhaltlichen Ähnlichkeiten von Textdokumenten zur Berechnung einer Clusterdarstellung (Kap.6.5.1 und 6.8.2)
- Relevanzmaße zur Gewichtung interaktiver Auswahlen (Kap.5.4)

Weitere Kriterien zur Einordnung der Arbeit werden im Zusammenhang mit der Auswahl verwandter Arbeiten besprochen (Kap.3.1).

1.5 Forschungsmethodik

Das Systemkonzept als Ganzes soll in Anwendungsszenarien studiert werden. Dabei soll der betriebswirtschaftliche Mehrwert als Erweiterung der bisher eingesetzten Technologien zur Textrecherche bewertet werden (Kap.7.1). Für eine solche Bewertung explorativer Suchsysteme werden empirische Feldstudien vorgeschlagen: *„If possible, experiments should be longitudinal, and take place in a naturalistic setting. The task domain should contain a mixture of task types: Some that relate closely to subjects regular activities, and some that are completely new“* [WMM06] und [St95], S.77. Gerade bei der Einführung explorativer Suchsysteme ist eine solche Nutzenbewertung typisch [SP04]⁹. Die Benutzereinschätzung und –zufriedenheit ist dabei der wichtigste Maßstab für den Nutzen eines Systems, weil die Interaktion zwischen Benutzer und System den Erfolg des Rechercheprozesses maßgeblich beeinflusst. Zusätzlich wird vorgeschlagen, das Benutzerverhalten zu beobachten, um Rückschlüsse auf die kognitive Belastung und das Lernverhalten mit dem System zu ermöglichen [WMM06]. In dieser Arbeit wird bei der Bewertung auch der betriebswirtschaftliche Einfluss der neuen Technologie zur Textrecherche explizit berücksichtigt (Kap.2.1).

Für empirische Untersuchungen in realen Arbeitsumgebungen müssen spezialisierte Prototypen entwickelt werden. Dazu wird die Referenztechnologie SWAPit entworfen, deren Informationssichten und interaktive Koordination dazwischen flexibel konfiguriert werden können. Die Realisierung von SWAPit zeigt exemplarisch, wie das Systemkonzept von SWAPit umgesetzt werden kann. Die Erkenntnisse jeder Fallstudie werden zunächst an SWAPit, dann am Systemkonzept. Ergeben sich daraus Optimierungen für Konzept oder System, etwa aufgrund zusätzlicher Anforderungen oder aufgrund konzeptionell-technischer Innovationen oder beidem, so werden diese vor der Durchführung weiterer Studien realisiert. Insofern baut jede Fallstudie auf die vorangehende auf und die Chronologie der Fallstudien ist bedeutsam. Auch die Methodiken zur Systemanpassung und zur Systembewertung (Kap.2) werden auf diese Weise sukzessive verbessert und erweitert.

Analog zu Referenzmodellen im Bereich der Informationsmodellierung für Unternehmen, die von unternehmensspezifischen Eigenschaften zu abstrahieren helfen [TS06], dient eine Referenztechnologie der Abstraktion von konkreten Systemen. Eine *Referenztechnologie* bezeichne ein generisches Anwendungssystem, das Anpassungsmechanismen für wichtige Systemfunktionalitäten bietet, um konkrete Systeme daraus abzuleiten – eine Musterimplementierung für konkrete Systeme also, die eine Menge von Systemen, eine Systemklasse, realisieren hilft. Die Anpassungsmechanismen von SWAPit werden in Kap.6.8 erläutert. Der Einsatz von Referenzmodellen verspricht die Einsparung von Zeit und Kosten sowie die Erhöhung der Qualität des

⁹ *“Often, evaluations do not intend to compare different technologies, but rather intend to assess the usefulness of a given ESS in an application domain, as compared to the current situation and tool support”* [SP04]

zu konstruierenden Modells, da im Referenzmodell allgemeine Empfehlungen des Gegenstandsbereichs repräsentiert sind [FL04][TS06]. Durch das Vorhandensein einer Referenztechnologie können die Entwicklungsphasen von Prototyping, Implementierung und Einführung der Software stark verkürzt werden. Analyse- und Entwurfsphase werden beeinflusst durch die Rahmenbedingungen, welche die Referenztechnologie impliziert. Ausführlicher wird die Besonderheit der Softwareentwicklung auf Basis einer Referenztechnologie diskutiert in Kap.2.4.1.

1.6 Übersicht und Lesehilfe

Abbildung 1.2 stellt die Struktur der Arbeit übersichtlich dar. In Kap.1 wurde in einem Bezugsrahmen die Textrecherche in Unternehmen charakterisiert, wobei Arten von Systemen und Informationen unterschieden werden. Außerdem wurde die Forschungsmethodik erläutert, die im Kern in der induktiven Anwendung, Bewertung und Optimierung der Referenztechnologie besteht. In Kap.2 werden grundlegende Methoden für andere Kapitel entwickelt. Zum einen wird ein Modell (Scorecard) zur umfassenden empirischen Nutzenbewertung explorativer Suchsysteme vorgestellt. Zum anderen wird ein Vorgehensmodell zur Ableitung konkreter Systeme aus der in Kap.6 entwickelten Referenztechnologie vorgestellt. Einen wichtigen Baustein von Nutzenbewertung und Systemanpassung bildet das Aufgabenmodell zur Beschreibung der Eigenschaften von Rechercheaufgaben. Kap.3 beginnt mit einem Modell zur Auswahl und Klassifikation verwandter Arbeiten im Bereich der explorativen Textrecherche. Danach ausgewählte Systeme und MultiView-Frameworks werden im Rest des Kapitels besprochen. In Kap.4 werden drei aufeinander aufbauende Fallstudien analysiert mit dem Ziel, typische Arten von Informationen und Querbezügen dazwischen zu identifizieren und bezüglich der Bedeutung für die Textanalyse gegenüber zu stellen. Kap.5 stellt ein MultiView-Systemkonzept für die Textanalyse vor. In Kap.6 wird die erweiterbare Referenztechnologie SWAPit beschrieben, die das Systemkonzept umsetzt und für jede der identifizierten Informationsarten eine grafische (Über-)Sicht umfasst. Außerdem werden einheitliche Interaktionsmechanismen zur Auswahl, Spezifikation und Beschreibung von Informationsobjekten realisiert. Kap.7 beginnt mit den allgemeinen Bewertungszielen aller Fallstudien. Für jede der bereits in der Problemanalyse (Kap.3) untersuchten Anwendungsszenarien wird eine empirische Studie und ihre Ergebnisse beschrieben. Kap.8 fasst die Kernbeiträge und Ergebnisse zusammen und zeigt Potenzial für zukünftige Forschungsarbeiten auf. Der Anhang umfasst u.a. weitere Fallbeispiele sowie eine detaillierte Beschreibung des Aufgabenmodells vorgestellt.

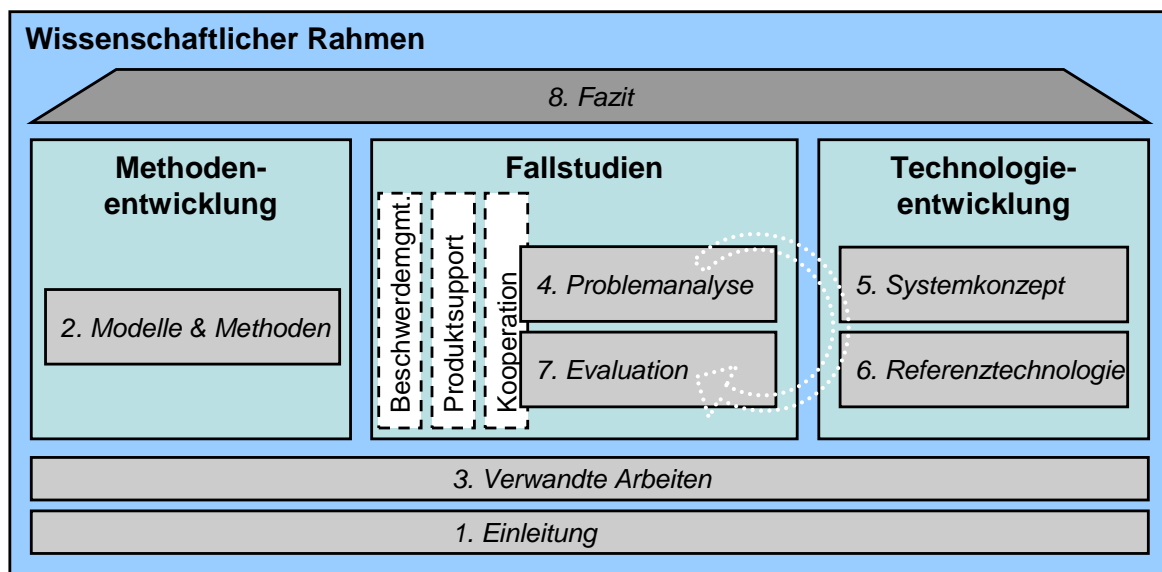


Abbildung 1.2: Struktur der Arbeit - eine Lesehilfe

2 Modelle und Methoden

Der Entwurf und die Bewertung spezialisierter Systemlösungen für die Textrecherche sind potenziell aufwändig und fehleranfällig. Im Folgenden wird ein „Best Practice“ in Form eines Beschreibungs- und Bewertungs-, sowie eines Vorgehensmodells vorgestellt. Ziel ist die systematische aufgabenorientierte Anpassung der zu entwickelnden Referenztechnologie an Informations- und Funktionalitätsbedarfe. Das Verfahren wird in den Fallstudien (Kap.7) angewendet und diskutiert.

2.1 Nutzen explorativer Suchsysteme

Für die Bewertung des Gesamtnutzens einer Technologie als Bestandteil von Geschäftsprozessen in Unternehmen spielen neben informationstechnischen auch betriebswirtschaftliche Kriterien eine Rolle. Schließlich müssen knappe Ressourcen und die Restriktionen der technischen Infrastruktur berücksichtigt werden. Außerdem steht bei jeder Produkteinführung die Frage nach dem wirtschaftlichen Netto-Nutzen an erster Stelle: „[...] *standard accounting techniques, drawing attention to design and engineering economic activity, to be judged in the end by its contribution to business value.*“ [FF01].

Beide Arten von Kriterien stehen in Zusammenhang: Obgleich die Optimierung der Kosten und des Nutzens der IT in der Regel nicht primäres Unternehmensziel ist, trägt sie zur Zielerreichung bei. Umgekehrt beeinflussen Geschäftsprozesse die IT-unterstützten Rechercheprozesse.

Auch Normen zur Bewertung der Softwarequalität (z.B. ISO/IEC 9126) beschäftigen sich mit Faktoren wie Funktionalität, Bedienbarkeit und Prozesseffizienz. Bietet ein Werkzeug die richtigen Funktionalitäten und hilft, gute Ergebnisse zu produzieren, so wird es als nützlich eingestuft. Wenn gleichzeitig der Aufwand durch Bedienung des Werkzeugs gering ist, dessen Eingliederung in den Arbeitsprozess gelingt und die Akzeptanz hoch ist, so ist dies ebenfalls der Nützlichkeit zuträglich.

Komplementär zur Unterscheidung der informationstechnischen und betriebswirtschaftlichen Kriterien lassen sich Bewertungskriterien in Kosten und Nutzen differenzieren. Ein geeignetes Informationssystem erzeugt hohen Nutzen und verursacht geringe Kosten.

Farbey & Finkelstein stellen zahlreiche Bewertungsmethoden für den Entwurf und die Einführung von Software gegenüber, die insbesondere ökonomischen Aspekte einbeziehen [FF01]. Eine Methode aus der Kategorie „qualitative & explorative Methoden“ ist die „Value Analysis“, die mit der gemeinsamen Bestimmung der Nutzenkriterien beginnt und dann auf Basis prototypischer Systeme und Abschätzungen die Bewertung von Kosten und Nutzen vornimmt: „*Value Analysis appears to be an attractive way of evaluating requirements of software systems and, because of its incremental nature, new features in product families.*“ [FF01].

Die *Akzeptanzforschung* beschäftigt sich mit der Annahmeentscheidung neuer Technologien durch den Benutzer ([Si01], S.87) und untersucht die Einflussgrößen auf diese Entscheidung. Die Akzeptanzforschung unterscheidet ex-ante Betrachtung (Prognose der Akzeptanz) und ex-post Betrachtung (Bewertung bereits eingesetzter Technologien).

Ein frühes Modell der Akzeptanzforschung (TAM¹⁰-Modell, [DBW89]) unterscheidet Nutzen und Aufwand (Bedienbarkeit) eines Systems. Im späteren TTFM¹¹-Modell [GT95] kommt eine Modellierung des Aufgabenkontexts als Einflussfaktor hinzu. Das DART-Modell [AHS03] strebt eine Kombination von systembezogenen Faktoren und Kontextfaktoren an und fokussiert dabei auf eine ausgewogene und übersichtliche

¹⁰ Technology Acceptance Model

¹¹ Task Technology Fit Model

Visualisierung aller Einflussfaktoren. Die Definition und Darstellung eines allgemeinen Bewertungsschemas im DART-Modell dient dem Ziel, verschiedene Untersuchungen, die individuelle Detail-Kriterien verwenden können, vergleichbar zu machen.

Balanced Scorecards ([KN98], S.566 ff.) dienen der umfassenden Beschreibung und Quantifizierung der Unternehmensperformanz und des Bezugs zur Unternehmensstrategie. Typische berücksichtigte Faktoren sind Mitarbeiter, Kunden und Prozesse. Analog lassen sich Einflussfaktoren auf die Nützlichkeit eines Recherchewerkzeugs für ein Unternehmen betrachten. Das DART-Modell wird im Folgenden als Grundlage einer *Scorecard zur Systembewertung* verwendet [Hi06]. Allerdings ist für die umfassende Bewertung von Recherchewerkzeugen die Benutzerakzeptanz meist nur *ein* Kriterium unter vielen. Qualität der Ergebnisse und Effizienz der Recherche gehören dazu. Hinzu kommen wirtschaftliche Faktoren wie Ressourcenverbrauch und Integrierbarkeit in Arbeitsprozesse. Die allgemeine Bewertungs-Scorecard der Arbeit ist in Abbildung 2.1 dargestellt.

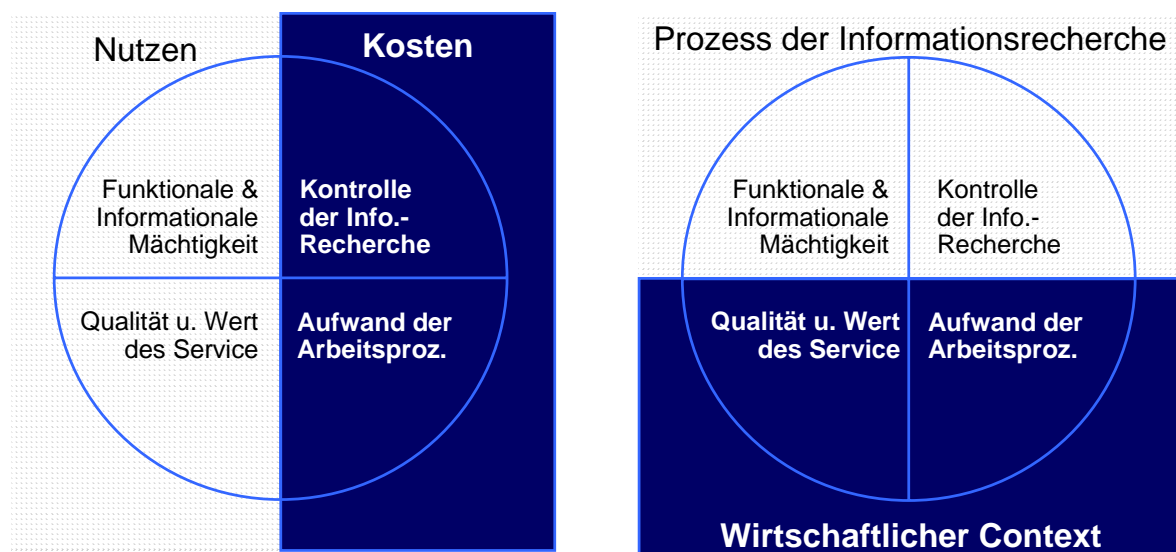


Abbildung 2.1: Ein Schema zur Bewertung explorativer Suchsysteme

Die 4 *Perspektiven* der Bewertungs-Scorecard sind Kosten, Nutzen, Prozess der Informationsrecherche und Wirtschaftlicher Kontext. Aus der Kombination der Perspektiven ergeben sich 4 *Bewertungsaspekte*:

- Funktionale und Informatorische Mächtigkeit des Systems
- Kontrolle der Informationsrecherche mit dem System
- Qualität und Wert des Service, der durch Verwendung des Systems realisiert wird
- Aufwand der Arbeitsprozesse, die nötig sind, um den Service zu realisieren

Diese Unterscheidung der *Perspektiven* und *Aspekte* dient als Denkmodell und zur Problemdekomposition von der Aufgabenbeschreibung bis zur Systembewertung. Diese Schritte werden in den nächsten Kapiteln beschrieben. Auch das in dieser Arbeit entwickelte Aufgabenmodell (Kap.2.2) orientiert sich an der Scorecard und beschreibt neben den informationstechnischen auch die ökonomisch relevanten Eigenschaften der Rechercheaufgaben.

Zur Bewertung eines konkreten Systems müssen die allgemeinen Perspektiven unter Berücksichtigung von System, Unternehmen und Aufgabe individuell spezifiziert bzw. herunter gebrochen werden, damit ein angepasstes und differenziertes Bild entsteht. Gleichzeitig bieten die einheitlichen Perspektiven die Chance, die Ergebnisse unterschiedlicher nach diesem Schema durchgeführter Systembewertungen qualitativ zu vergleichen.

2.2 Modell zur Aufgabenbeschreibung

Ein Aufgabenmodell dient als Vokabular zur systematischen Beschreibung und zum Vergleich von Aufgaben der Informationsrecherche. Eine Aufgabenbeschreibung lässt sich nutzen um Aufgabenklassen zu erkennen, Anforderungen für die Systemunterstützung zu identifizieren und Maße für die Systembewertung abzuleiten. Aufgabenmodelle dienen zudem bereits bei der Aufgabenerhebung als Leitfaden, die richtigen Fragen zu stellen. Die Aufgabenbeschreibung ist prinzipiell unabhängig von der Systemunterstützung. Aufgabenmodelle lassen sich unterscheiden nach der Granularität der beschriebenen Aufgaben und der Spezialisierung auf eine Domäne [Mo99]. Die Definition von Aufgaben reicht dabei von atomaren Benutzeraktivitäten bis hin zu komplexen Zielen der Benutzer. Aufgabenmodelle zählen meist die zur Beschreibung und Abgrenzung wesentlichen Dimensionen und Ausprägungen auf.

In [BSM02]¹² haben wir die bekanntesten Aufgabenmodelle aus den Bereichen Information Retrieval / Benutzermodellierung verglichen, darunter die Modelle von Shneiderman [Sh96], Wehrend & Lewis [WL90] und Belkin [BST94]. In Interviews mit Unternehmen aus wissensintensiven Branchen wurden die wichtigsten Dokumentenressourcen, die Arbeit damit, sowie eingesetzte Technologien zur Informationsrecherche erhoben [Se01]. Aus Literaturanalyse und Interviews als theoretischer Basis wurde ein verdichtetes Aufgabenmodell für Aufgaben der explorativen Informationsrecherche abgeleitet. Das Aufgabenmodell kam in zahlreichen Fallstudien und Projekten zum Einsatz (Anhang 9.3) und wurde kontinuierlich optimiert und erweitert, um die wesentlichen Aufgabenmerkmale möglichst vollständig und begriffsscharf abzubilden.

Während das ursprüngliche Aufgabenmodell für die Analyse von Textdokumenten konzipiert war, wurde es in dieser Arbeit ergänzt um Kriterien für die Einbeziehung relevanter strukturierter Daten in die Recherche. Außerdem wurde das Aufgabenmodell neu strukturiert gemäß den Perspektiven der Scorecard zur Systembewertung (Kap.2.1). Dabei kamen zahlreiche betriebswirtschaftliche Kriterien hinzu mit dem Ziel, den betriebswirtschaftlichen Aufgabenkontext mit zu erfassen. Ein einfaches Beispiel für ein technologierelevantes betriebswirtschaftliches Kriterium ist der Zeitdruck, unter dem ein Benutzer steht. Geht es dem Benutzer um schnelle Rechercheergebnisse, so wird ein Unterstützungssystem nur wenige Top-Ergebnisse präsentieren. Geht es aber um eine genaue Analyse wie die Identifikation kritischer Patente für eine neue Technologie, so müssen alle relevanten Informationen aufbereitet werden. Ein weiteres Beispiel ist die fachliche bzw. technologische Expertise des Benutzers. Als Grundlage dieser Erweiterungen diente das Modell von Mertens & Griese zur Beschreibung von Führungsinformationssystemen ([MG02] S.3, Abbildung 2.3).

Das neue Aufgabenmodell ist in Abbildung 2.4 dargestellt. Gemäß den 4 Aspekten der Scorecard zur Systembewertung (Kap.2.1), nach denen das Modell gegliedert ist, werden 4 Quadranten unterschieden. Zur Beschreibung einer Aufgabe werden nach dem Prinzip eines Morphologischen Kastens¹³ jedem Merkmal (z.B. „Ziel der Recherche“) Ausprägungen zugewiesen, die die Aufgabe gut beschreiben (z.B. „Lernen“, „Information verdichten“, usw.). So entsteht ein *Aufgabensteckbrief*. Zwei Besonderheiten gegenüber herkömmlichen Morphologischen Kästen gelten für das Aufgabenmodell:

- Die Merkmale des Modells sind nicht zwangsläufig unabhängig voneinander (z.B. hängen „Interaktionsmethode“ und „Interaktionsstrategie“ oft zusammen)
- Einer Aufgabe können beliebig viele der Ausprägungen eines Merkmals zugeordnet werden (nicht zwangsläufig genau eine)

¹² Der Beitrag liefert einen Überblick über die Entstehung des Aufgabenmodells. Für eine detailliertere Beschreibung siehe [Be01], S.9-27.

¹³ Kreativitätstechnik nach dem Schweizer Astrophysiker Fritz Zwicky (1898–1974). Der Kasten beschreibt kompakt einen Lösungsraum (Merkmale und Ausprägungen), in welchem sich Lösungen durch Auswahl einer Ausprägung zu jedem Merkmal darstellen und vergleichen lassen.

| checking the single source criterion in product documentation | | | | |
|---|-------------------------|----------------|---|---|
| goal of interaction | making use of documents | | learn | |
| | | | condense documents | |
| | | | select documents | |
| | maintaining documents | | control quality | • |
| | | assure quality | | |
| dynamics of focus of interest | | | fixed | • |
| | | | adaptive | |
| resource considered | | | information from documents | • |
| | meta-information | | document attributes | |
| | | | structural information | • |
| method of interaction | | | explorative | • |
| | | | goal-directed | |
| granularity | | | overview | |
| | | | details | • |
| categories | | | not relevant | • |
| | relevant | use existing | check classes | |
| | | | use classes | |
| | | | categorize | |
| focused relationship | | | external: document – specification | |
| | inherent | | document – document | • |
| | | | document – topic | |
| | | | topic – topic | |
| mode of communication | | | recognize (reflect, physically passive) | • |
| | | | specify (physically active) | |

Abbildung 2.2: Modell zur Beschreibung von Aufgaben der Explorativen Recherche in Dokumentensammlungen (Quelle: Becks, Seeling u. Minkenberg [BSM02]). Als Beispiel wird eine Aufgabe der Qualitätssicherung technischer Dokumentation charakterisiert: Das Ziel der Aufgabe ist inhaltliche Redundanz zu entdecken, also Dokumente zu identifizieren, welche die gleiche oder sehr ähnliche Information enthalten.

Beispiele für Aufgabenbeschreibungen finden sich bei der Problemanalyse jeder Fallstudie (Kap.3). Alle Merkmale und Ausprägungen des Aufgabenmodells werden detaillierter beschrieben in Anhang 9.2. Die Aufgabenbeschreibung sollte möglichst gemeinsam von Personen, die am Rechercheprozess beteiligt sind (Anwender), und technischen Experten durchgeführt werden, damit sie ein gemeinsames Aufgabenverständnis dokumentiert und Missverständnisse oder Unklarheiten bereits in dieser Phase aufgedeckt werden. Alternativ können mehrere Personen „konkurrierend“ Aufgabenbeschreibungen verfassen und die Ergebnisse werden anschließend diskutiert und konsolidiert.

| | | | | |
|--------------------------|------------------------------|--|--|---|
| Auslöser | Signale/Datenkonstellationen | Kalendertermine | Benutzerwünsche | Entscheidungsbedarfe |
| Adressatenzahl | Einzelpersonen | | Gruppen | |
| Adressatenhierarchie | Untere Führungsebene | Mittlere Führungsebene | Obere Führungsebene | |
| Informationsherkunft | Interne Quellen | | Externe Quellen | |
| Informationsart | Quantitative Informationen | | Qualitative Informationen | |
| Präsentationsform | Meldungen | Tabellen | Grafiken | Verbale Berichte Expertisen |
| Abfragemodus | Standardabfragen | | Freie Abfragen | |
| Informationsdistribution | Pull-Verfahren | | Push-Verfahren | |
| Entscheidungsmodell | Nicht vorhanden | Entscheidungsmodell mit statistischen Methoden | Entscheidungsmodell mit Operations-research-Methoden | Entscheidungsmodell mit Methoden der KI |
| Simulation | Nicht simulative | | What-if-Rechnungen | How-to-achieve-Rechnungen |

Abbildung 2.3: Morphologischer Kasten zur Beschreibung und Unterscheidung von Führungsinformationssystemen (Quelle: Mertens u. Griese [MG02]).

| Funktionale und Informatorische Mächtigkeit | | | | | | Kontrolle der Informationsrecherche | | | |
|---|---------------------------------|------------------|-------------------------|-------------------------|------------------|-------------------------------------|---------------------------|------------------------------|-----------------------------|
| Ziel der Info.-Recherche | lernen | Info. verdichten | Info. auswählen | Qualität prüfen | Qualität sichern | Interaktionsmethode | zielgerichtete Suche | explorative Recherche | Stöbern u. Entdecken |
| Betrachtete Info.-Arten | Text-Dok. | Dok.-Attribute | Beschreibungskategorien | Relationale Information | | Interaktionsstrategie | filtern | ausschließen | |
| Info.-Quellen | intern | | extern | | | Granularität | Übersicht | Details | |
| Relat. Größe d. Info.-Teilmenge | gesamtes Repository | große Teilmenge | | kleine Teilmenge | | Dynamik des Interesses | statisch | | wechselnd |
| Dynamik der Info.-Teilmenge | konstant | | unterschiedlich | | ad-hoc | Flexibilität Interaktion | frei | angeleitet | Workflow |
| Betrachtete Verknüpfungen | Dok.-Spezifikation | Dok.-Dok. | Dok.-Thema | Dok.-Kategorie | Dok.-Tupel | Kommunikationsmodus | wahrnehmen (phys. passiv) | | spezifizieren (phys. aktiv) |
| Existenz der Verknüpfungen | logisch | | organisatorisch | | physisch | Erwartetes Antwortverhalten | ohne Verzögerung | kurze Verzögerung akzeptiert | lange Berechnung akzeptiert |
| Verwendung der strukt. Info. | irrelevant | prüfen | verwenden | Dok. zuweisen | | | | | |
| Vokabular | allgemein | | Fachsprache | | Korrespondenz | | | | |
| Spektrum der Inhalte | allgemein | | domänenspezifisch | | | | | | |
| Relevanz im Unternehmen | Kernprozess | | Unterstützungsprozess | | | Vorhersagbarkeit | gelegentlich | wiederholt | regelmäßig |
| Absolute Bedeutung | „nice to have“ | | wichtig | kritisch | | Durchlaufzeit | Zeit ist irrelevant | Zeit ist wichtig | Zeit ist kritisch |
| Intendierte Wertsteigerung | Steigerung der Effizienz | | neue Aufgabe | | | Expertise in Domäne | Laie | | Domänen-Experte |
| Primäres Qualitätskrit. | Genauigkeit und Vollständigkeit | | grober Überblick | | | Expertise in Korpus | Information unbekannt | | Information bekannt |
| Qualität und Wert des Service | | | | | | Expertise in Recherche | gelegentl. Nutzer | Experte der Recherche | System-Experte |
| | | | | | | Aufwand der Arbeitsprozesse | | | |

Abbildung 2.4: Ein Aufgabenmodell zur ganzheitlichen Beschreibung von Rechercheaufgaben in Unternehmen

2.3 Portfolio zur Aufgabenbewertung

Aufgaben-Portfolios dienen der Aufgabenbewertung und -auswahl nach multiplen Kriterien. Sie können zur Moderation und zur Systematisierung eingesetzt werden. Aufgaben-Portfolios sind inspiriert durch „Kernprozessportfolios“, einer Methode des Strategischen Managements ([Sc06], S.31). Die Nützlichkeit einer Aufgabenunterstützung durch SWAPit lässt sich an drei Kriterien bewerten:

- *Technologiedruck*: Wie gut ist SWAPit für die Aufgabe geeignet? Der Technologiedruck betrifft die intendierte Systemunterstützung. Eine weniger gute Aufgabenunterstützung ist weniger hilfreich für das Unternehmen. Die Aufgabenangemessenheit von SWAPit ist für solche Aufgaben tendenziell gering.
- *Bedarfsog*: Wie gut ist die Aufgabe bereits unterstützt oder könnte durch andere bekannte Werkzeuge unterstützt werden? Der Bedarfsog betrifft die Ist-Unterstützung. Bereits gut lösbare Aufgaben sind nicht so optimierungsbedürftig. Das Alleinstellungsmerkmal von SWAPit ist für solche Aufgaben tendenziell gering.
- *Nutzenpotential*: Wie wichtig ist die Aufgabe für das Unternehmen? Das Nutzenpotential betrifft den Geschäftsprozess. In die Unterstützung nicht so wichtiger Aufgaben wird nicht so viel investiert. Die Wertsteigerung durch SWAPit ist für solche Aufgaben tendenziell gering.

Zur Aufgabenbewertung werden in Workshops mit den Anwendern nacheinander für jede Aufgabe die drei genannten Aspekte eingeschätzt und ein Aufgabensymbol entsprechend im Portfolio platziert und in der Größe festgelegt. Wurden alle Aufgaben eingeordnet, so dient das Portfolio als Grundlage zur Auswahl besonders wichtiger Aufgaben. Zum Vergleich der Aufgaben können und sollen im Spannungsfeld der Dimensionen Gewichtungen vorgenommen werden. Beispielsweise kann eine Aufgabe, die zwar durch die bislang verwendete Systemlandschaft schon recht gut unterstützt wird, jedoch mit einer besonders hohen Werthaltigkeit verbunden ist (z.B. aufgrund eines großen Multiplikators wie der Anzahl zu unterstützender Arbeitsplätze), durchaus mit der neuen Technologie adressiert werden, da auch eine kleine Optimierung einen hohen ökonomischen Mehrwert verspricht.

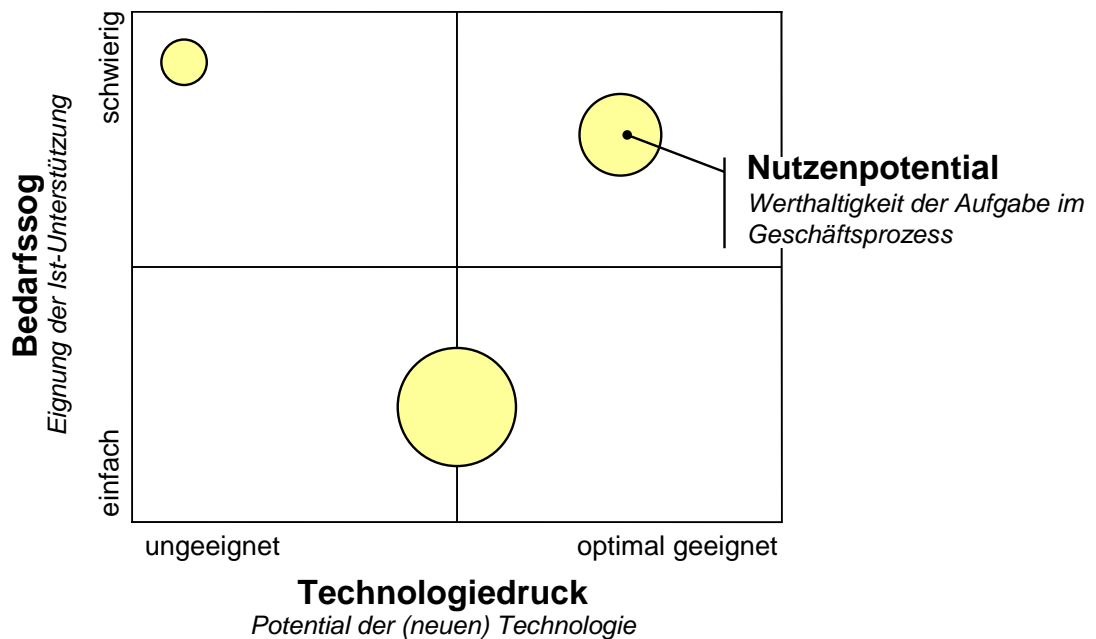


Abbildung 2.5: Portfolio-Methode zur Aufgabenbewertung

2.4 Verfahren zur Nutzenbewertung

Das wieder verwendbare *Vorgehensmodell* vereinheitlicht und vereinfacht den Anpassungsprozess der Referenztechnologie an individuelle Anforderungen. Dabei wird das Spannungsfeld von Technologie-/Daten-Druck und Bedarfs-/Funktions-Sog in zwei Schritten aufgelöst: Die Bottom-up Designphase dient dem schnellen Prototyping mit dem Ziel der Operationalisierung der Aufgaben und der Identifizierung von Anforderungen. Die Top-down Designphase dient der Optimierung der Aufgabenangemessenheit des Prototyps. Beide Phasen stellen Entwicklungszyklen dar. Das Vorgehensmodell wurde in zahlreichen Fallstudien entwickelt, angewendet und verbessert. Es repräsentiert ein „best practice“ des Systementwurfs von Rechtersystemen auf Basis einer Referenztechnologie wie SWAPit und versteht sich damit als systemunabhängiger Beitrag dieser Arbeit. Das Vorgehensmodell kommt in den Fallstudien 2 und 3 zur Anwendung. Dabei werden die ersten Phasen (Analyse) jeder Fallstudie in Kap.3 beschrieben und die letzten Phasen (Systemdesign und Bewertung) in Kap.7. Die verwendeten Notationen und Modelle werden in Anhang 9.1 erläutert.

2.4.1 Besonderheiten einer Referenztechnologie

Ein Systemkonzept und dessen Realisierung als Referenztechnologie verringert die Entwicklungszeit für spezialisierte Anwendungen beträchtlich. An Stelle der Neuentwicklung muss die Basistechnologie auf die Anwendung zugeschnitten werden. Dies stellt allerdings eine besondere Herausforderung dar, weil das *Spannungsfeld* zwischen den Defiziten und Chancen in der Ist-Situation (Bedarfssog) einerseits und den technischen Möglichkeiten und Grenzen der Basistechnologie (Technologiedruck) andererseits berücksichtigt werden muss. Eine *Potentialanalyse* ist nötig, die sich weder allein auf die Analyse der Informations- und Funktionalitätsbedarfe (Gefahr der Nicht-Machbarkeit mit der Referenztechnologie), noch allein auf die Analyse der verfügbaren Informationen (Gefahr der unmöglichen oder ineffizienten Unterstützung bestimmter Aufgaben) stützen kann.

Beim Entwurf von Anwendungen besteht die Besonderheit, dass die Referenztechnologie bereits Paradigmen für den Informationszugriff vorgibt. Daher kommen für den Anwendungsentwurf zwei Vorgehensweisen in Betracht: Wird *top-down (aufgabenorientiert)* vorgegangen, so werden wichtige Analyseaufgaben erhoben und nach einer adäquaten Systemunterstützung gesucht. Wird *bottom-up (datenorientiert)* vorgegangen, so wird angestrebt, alle vorhandenen Daten flexibel zugreifbar zu machen. Diese Ansätze werden auch für den Entwurf von Data-Warehouse-Lösungen diskutiert [BG05]. In dieser Arbeit sind beide Aspekte von Bedeutung. Die Kernidee ist, zunächst die verfügbaren Informationen und ihre Verknüpfungen mit Hilfe der Referenztechnologie zugreifbar zu machen (bottom-up), die ein schnelles Prototyping ermöglicht (z.B. Systemskizze, Anhang 9.1.4). Anschließend werden die Parameter der Referenztechnologie an die zu unterstützenden Benutzeraufgaben angepasst (top-down). Dazu sind Prozess- und Datenanalysen notwendig und die technischen Möglichkeiten und Grenzen der Referenztechnologie müssen berücksichtigt werden. Mit anderen Worten werden Anforderungsanalyse und Entwurf von zwei Kräften bestimmt: Bedarfs-/Funktions-Sog sowie Technologie-/Daten-Druck.

Neben den funktionalen Anforderungen steht bei Informationssystemen die Bereitstellung nützlicher Information im Vordergrund. Für Data Warehouses schlagen Burmester et al. [BG05] vor, die funktionalen Anforderungen durch *informativische Anforderungen* zu ersetzen. Ein bezüglich der informativischen Anforderungen gutes System zeichnet sich durch drei Aspekte aus: Es stellt nützliche Inhalte bereit („content“), es liefert angemessene Qualität („quality“) und präsentiert Information in angemessener Darstellung („visualization“). Für das auf Benutzerinteraktion ausgerichtete MultiView-Paradigma ist Benutzerinteraktion essentiell, weshalb hier neben den informativischen auch funktionale Anforderungen berücksichtigt werden. Zusätzlich finden sonstige Anforderungen Beachtung (z.B. Performanz, Ergonomie).

2.4.2 Modelle zur Systemanpassung und -bewertung

Alle Schritte des Vorgehensmodells folgen dem Denkmodell der Scorecard zur Systembewertung (Kap.2.1). Das Vorgehensmodell unterscheidet 5 Phasen:

Tabelle 2-1: Phasen des Vorgehensmodells zur Systemanpassung und -bewertung

| | | |
|------------------|------------|---|
| Analyse | Prozesse | Prozesse der Informationsrecherche und ihre Einbettung in Arbeitsprozesse des Unternehmens werden erfasst. |
| | Aufgaben | Aufgaben und ihre ökonomisch relevanten Rahmenbedingungen werden erfasst. Repräsentative Aufgaben für die Systembewertung werden definiert. |
| Design | Entwurf | In mehreren Iterationen wird ein Systemdesign entwickelt. |
| Bewertung | Maße | Ziele und Maße für die Systembewertung werden definiert. |
| | Ergebnisse | Die Bewertung wird durchgeführt und Ergebnisse beschrieben, dargestellt und interpretiert. |

Für jede Phase werden Modelle benötigt. Je später eine Phase im Vorgehensmodell angesiedelt ist, desto stärker hängt sie von individuellen Parametern einer Studie wie System, Bewertungszielen und dem technisch-ökonomischen Kontext ab. Abbildung 2.6 stellt das Vorgehensmodell übersichtlich dar.

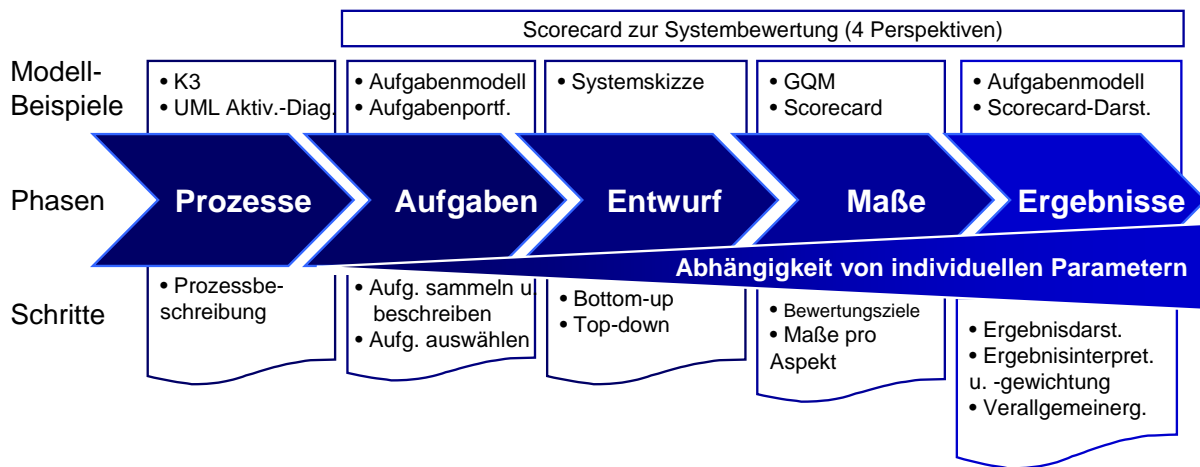


Abbildung 2.6: Das Vorgehensmodell für Systemanpassung und -bewertung im Überblick

Prozesse

Zunächst muss im Dialog mit den Benutzern ein Grundverständnis des Szenarios erarbeitet werden mit dem Ziel, die für die Recherche wichtigen Unternehmensprozesse zu erfassen. Dafür bieten sich Workshops oder strukturierte Interviews an. Bei der Modellierung der Unternehmensprozesse müssen alle für die Recherche relevanten Informationen und Verknüpfungen erfasst werden. Existiert bereits eine Werkzeugunterstützung, so werden die eingesetzten Systeme ebenfalls erfasst. Sofern vorhanden, können die Unterlagen einer bestehenden Zertifizierung bei der Prozessanalyse helfen. In den Unternehmensprozessen (z.B. Bearbeitung von Kundenmeldungen) werden Prozesse der Informationsrecherche identifiziert und in größerem Detail beschrieben (z.B. Analyse der Kundenmeinung). Schließlich kann eine Auflistung der relevanten Informationen erstellt werden, aus der die Informationsarten und die Beziehungen zu den Texten hervorgehen.

Ein bewährter Formalismus zur Beschreibung der Unternehmensprozesse ist das K3-Modell (Anhang 9.1.2), mit dem sich Kommunikation, Koordination und Kooperation aller Prozessteilnehmer erfassen lassen. K3 ist einfach genug um auch von den Anwendern leicht verstanden zu werden und bildet Informationen und Systeme ab. Für die detailliertere Beschreibung der Rechercheprozesse kann die EPK-Notation (Anhang 9.1) verwendet werden, mit der sich Handlungsauslöser und Aktivitäten modellieren lassen.

Aufgaben

In Kap.2.4.1 wurde die Herausforderung einer Potentialanalyse erläutert, wenn eine bestehende Technologie in einem neuen Anwendungsfeld eingesetzt und untersucht werden soll. Dabei kommt es auf eine gute Abstimmung von Technologie und Aufgabe an („task-technology fit“ [GT95]). In dieser Phase macht sich die Ambivalenz zwischen den Aufgabenorientierung einerseits und Orientierung an der Referenztechnologie andererseits besonders bemerkbar: Sind machbare Lösungen interessant und wichtig? Sind interessante und wichtige Lösungen mit der Referenztechnologie machbar?

In einem Brainstorming werden zunächst Ideen für viel versprechende Rechercheaufgaben gesammelt, die sich möglicherweise mit dem neuen Recherchesystem unterstützen oder optimieren lassen. Bei Aufgaben der Textanalyse ist die Frage zu beantworten, welches wertvolle Wissen in den Texten verborgen ist und welche wichtigen Fragestellungen sich damit adressieren lassen.

In einem zweiten Schritt werden diese Aufgaben bewertet und ausgewählt. Dazu können Defizite der bestehenden Systemunterstützung sowie Stärken und Schwächen des Einsatzes der neuen Technologie im Anwendungsfeld als Liste von Chancen und Defiziten erfasst werden. Zu unterstützende Aufgaben sollen außerdem werthaltig sein. Diese Faktoren werden im Aufgabenportfolio zur Unterstützung der multidimensionalen Aufgabenbewertung kombiniert (Kap.2.3). Schließlich werden die ausgewählten Aufgaben mit Hilfe des Aufgabenmodells (Kap.2.2) genauer untersucht: Für jede Aufgabe wird ein Aufgabensteckbrief erstellt, der die zentralen Eigenschaften der Rechercheaufgabe erfasst. Der Aufgabensteckbrief ist nach den Aspekten der Bewertungs-Scorecard (Kap.2.1) gegliedert.

Entwurf

Sind Prozess- und Aufgabenanalyse abgeschlossen, so folgt das Systemdesign in mehreren Schritten. Beim *Ersten Systemdesign (bottom-up)* wird eine Systemskizze (z.B. Anhang 9.1.4) des Zielsystems entwickelt, die eine erste plausible Lösung repräsentiert, weil sie alle verfügbaren und potentiell wichtigen Daten und Zusammenhänge auf Sichten und Sichten-Koordination abbildet und mit den Standard-Möglichkeiten der Referenztechnologie darstellt (z.B. Standardmaße für den Dokumentenvergleich). An dieser Systemskizze wird die Bearbeitung der ausgewählten repräsentativen Analyseaufgaben mit der Methode des *Cognitive Walkthrough*¹⁴ erprobt. Die sich für jede Aufgabe ergebenden Anforderungen (funktional, informatorisch und sonstige) werden schließlich konsolidiert und aufgelistet.

Beim *Zweiten Systemdesign (top-down)* wird eine revidierte Systemskizze entworfen, die nun auch als Software-Prototyp umgesetzt wird. Prototyping ist die einzige bewährte Methode um Erfahrungen mit den technischen Möglichkeiten zu sammeln [Go05]. Der Prototyp wird erneut einem Aufgabendurchlauf unterzogen, an dem nun auch die Endbenutzer teilnehmen sollen. Daraus lassen sich noch unberücksichtigte Anforderungen ableiten, und Parameter können optimiert werden (z.B. spezialisiertes Dokumenten-Vergleichsmaß).

Für das *Finale Systemdesign (Integration)* müssen alle erhobenen Anforderungen umgesetzt werden. Zudem ist ein Update-Prozess bei neuen oder geänderten Daten zu realisieren: Wann werden Projekte angelegt und

¹⁴ Der *Cognitive Walkthrough* ist eine Analytische Evaluationsmethode die, im Gegensatz zu empirischen Tests, vom Experten alleine durchgeführt wird.

Daten berechnet? Schließlich ist die technische Integration vorzunehmen, wobei vor allem Datenformate und Schnittstellen zu operativen Systemen eine Rolle spielen.

Maße

Während für die Erfassung von Aufgabenmerkmalen für die Informationsrecherche ein allgemeines Modell angegeben werden kann, variieren Bewertungsziele und –Maße von Experiment zu Experiment stark. Daher bleibt die Festlegung von Bewertungszielen und die Dekomposition in Maße eine intelligible Aufgabe. Hilfestellung bieten können Qualitätsmodelle wie GQM (Anhang 9.1.1), die eine systematische Vorgehensweise in mehreren Schritten vorschlagen. Für eine umfassende Bewertung des Nutzens einer Recherche-Technologie im Unternehmenskontext sind für jeden Aspekt der Bewertungs-Scorecard (Kap.2.1) geeignete Maße zu definieren.

Auch die Aufgabenbeschreibung im analog aufgebauten Aufgabenmodell hilft bei der Spezifikation der Maße, weil die dort beschriebenen Eigenschaften der Rechercheaufgaben Anforderungen an Information und Funktionalität implizieren, deren Erfüllung bewertet werden soll. So könnten geeignete Maße für eine stark explorative Aufgabe (Aufgabenmerkmal *Interaktionsflexibilität: Ausprägung frei*) messen, inwiefern ein Recherchesystem den flexiblen Zugriff auf Information unterstützt, beispielsweise durch Brush&Link-Verknüpfung von Information oder durch assoziative Navigation.

Ergebnisse

Die Maße werden den Aspekten der Scorecard zugeordnet und verfeinern somit das allgemeine Schema. Für eine kompakte Darstellung der Ergebnisse der Studie bietet sich an, die Bewertungs-Scorecard als Spinnendiagramm zu visualisieren. Diese Darstellungsart unterstützt sowohl den groben Überblick durch Betrachtung der Perspektiven und Aspekte der Scorecard, als auch die detaillierte Betrachtung einzelner Maße. Außerdem können mehrere Messungen als Kennlinien im Spinnendiagramm gegenübergestellt werden. Zur Auswertung der Ergebnisse können nun Gewichtungen der Maße, Bewertungen und Interpretationen vorgenommen und diskutiert werden.

3 Stand der Forschung

3.1 Auswahl und Klassifizierung relevanter Arbeiten

Für diese Arbeit sind wissenschaftliche Beiträge potentiell relevant, die folgende notwendige Bedingungen erfüllen:

- Es geht um die Unterstützung der explorativen Recherche in Textsammlungen
- Der Forschungsschwerpunkt liegt auf der Gestaltung von Benutzerschnittstellen und Interaktionskonzepten
- Der Kontext relevanter strukturierter Daten zu den Texten wird für die explorative Recherche nutzbar gemacht

Nach dem Grad der Spezialisierung bzw. Abstraktion lassen sich solche Arbeiten in den folgenden Dimensionen klassifizieren (s. auch Kap.7.1):

| | | |
|---------------------|--|---|
| System: | Einzelssystem | Referenztechnologie (konfig.- u. erweiterbar) |
| Domäne: | Textanalyse | Datenanalyse allgemein |
| Daten: | spezialisiert auf eine Informationsart | mehrere Arten strukturierter Information |
| Darstellung: | eine Sicht | koordinierte Sichten |
| Navigation: | festes Paradigma | mehrere Paradigmen (konfigurierbar) |

Die ausgewählten Ansätze werden nach diesen Unterscheidungskriterien charakterisiert und in einer Pentagon-förmigen Darstellung mit konzentrischen Keilen für alle Dimensionen visualisiert (Abbildung 3.1): Die breite Außenseite jedes Keils symbolisiert Abstraktion bzw. Verallgemeinerung, während die Keilspitze im Zentrum für Spezialisierung steht. Zur Einordnung der Bedeutung verwandter Arbeiten für *diese* Arbeit ist der zu entwickelnde Ansatz in Abbildung 3.1 (rechts) charakterisiert: In *dieser* Arbeit wird eine Referenztechnologie konzipiert und untersucht, die auf Textanalyse spezialisiert ist, dabei mehrere Arten strukturierter Information berücksichtigt, sich koordinierter Sichten bedient und ein spezielles Navigationsparadigma verwendet. Anders ausgedrückt ist der Ansatz allgemein in den Dimension *System*, *Darstellung*, *Daten* und spezialisiert in den Dimensionen *Domäne* und *Navigation*. Das Pentagon stellt ein vereinfachtes einheitliches Schema für den Vergleich der Ansätze untereinander und den Bezug zum Ansatz der Arbeit dar.

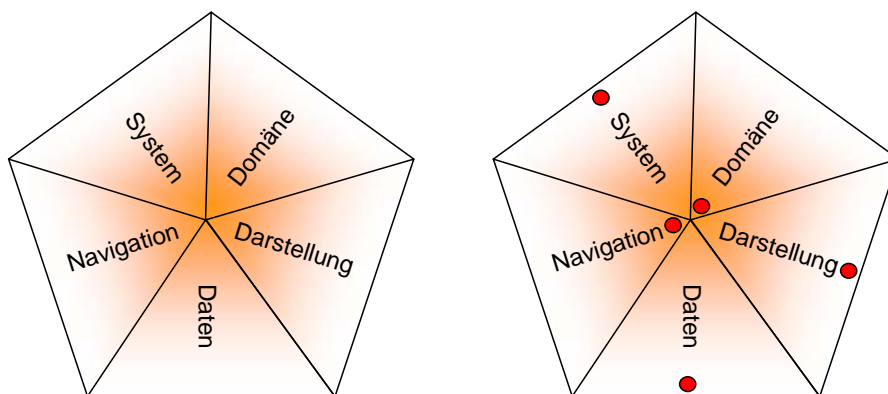


Abbildung 3.1: Links: Pentagon-Darstellung der Unterscheidungskriterien für relevante Forschungsarbeiten (innen~spezialisiert; aussen~allgemein). Rechts: Charakterisierung des zu untersuchenden Ansatzes

In den vergangenen Jahren sind zahlreiche Systeme zur Informationsvisualisierung und -analyse entstanden [WMM06]. Einige Online-Verzeichnisse haben sich zur Aufgabe gemacht, solche Arbeiten zu sammeln und zu klassifizieren, darunter *Atlas of Cyberspaces*¹⁵, *Visualcomplexity*¹⁶, *infovisWIKI*¹⁷ und *KDNuggets*¹⁸. Aus der Vielzahl der Werkzeuge und Metaphern wurden Ansätze ausgewählt, die im Zusammenhang mit dieser Arbeit besonders interessant sind.

Die Abschnitte zu den vorgestellten Ansätzen sind kanonisch aufgebaut. Nach einer Skizze von Idee und Anwendung werden jeweils die berücksichtigten Informationsarten und ihre Darstellung skizziert. Dann folgen Funktionalitäten und Interaktion sowie ggf. die Ergebnisse empirischer Studien. Schließlich wird, unterstützt durch das Pentagon-Schema, die Bedeutung des Ansatzes für diese Arbeit diskutiert. Im Zusammenhang mit vielen Ansätzen ist von *Brush&Link*-Metaphern zum interaktiven Auswählen von Daten die Rede. *Brushing* („anstreichen“) bezeichnet die Auswahl einer Teilmenge von Datenobjekten mit einem Eingabegerät. *Linking* bezeichnet die Hervorhebung damit verknüpfter Datenobjekte - üblicherweise in anderen grafischen Informationsdarstellungen.

3.2 Integrierte Darstellung von strukturierten Daten und Textdokumenten

Mit integrierter Darstellung von strukturierten Daten und Textdokumenten sind gemischte Darstellungen mehrere Informationsarten gemeint. Auch bei integrierten Darstellungen werden häufig Informationen in weiteren Fenstern (Sichten) angezeigt. Diese sind jedoch nicht interaktiv mit der zentralen Darstellung gekoppelt (koordinierte Sichten, s. Kap.3.3). Zunächst werden Darstellungen für Texte und *Kategorien* untersucht. Dann geht es um den Einsatz von *Katalogen* (z.B. Klassifikationshierarchien, Taxonomien) für die Navigation in Textsammlungen¹⁹. Schließlich werden *Relationale Daten* zur Navigation in Textsammlungen untersucht – insbesondere Metadaten zu Texten.

3.2.1 Begriffe und Dokumente

Begriffe (Kategorien) sind die einfachste Form strukturierter Daten. Beispiele sind Tags zur Annotation von Dokumenten oder Suchbegriffe. Immer mehr Suchmaschinen im Internet stellen Suchergebnisse auch grafisch dar, meist in Cluster-Darstellungen. Bei vielen solchen Ansätzen werden entweder die Suchbegriffe oder wichtige Begriffe, die aus den Suchergebnis-Dokumenten extrahiert wurden, zur Navigation nutzbar gemacht. Einige dieser Werkzeuge stellen lediglich die Begriffe und ihre Zusammenhänge grafisch dar (z.B. *Clusty*²⁰, *Quintura*²¹). Im Folgenden werden hingegen Werkzeuge vorgestellt, die sowohl Dokumente als auch Begriffe *in der gleichen Darstellung* visualisieren.

¹⁵ <http://www.cybergeography.org/atlas/>

¹⁶ <http://www.visualcomplexity.com>

¹⁷ <http://www.infovis-wiki.net>

¹⁸ <http://www.kdnuggets.com/software/visualization.html>

¹⁹ Eine alternative Navigationsmetapher zur katalogbasierten Navigation sind Facettierte Klassifikationshierarchien, wo durch Auswahl in komplementären Dimensionen die Spezifikation von Texten erfolgt (Kap.4.4.3)

²⁰ <http://www.clusty.com>

²¹ <http://www.quintura.com>

Sinnzeug

Für die flexible inhaltsbasierte Organisation von Webseiten haben Huber & Ammer²² ein Werkzeug entwickelt, das die Zusammenhänge zwischen Begriffen und Dokumenten bzw. Dokumentenähnlichkeiten darstellt. Sinnzeug wurde im Auftrag von DaimlerChrysler weiter entwickelt zu einem Werkzeug für Fonds-Manager (Abbildung 3.2, links).

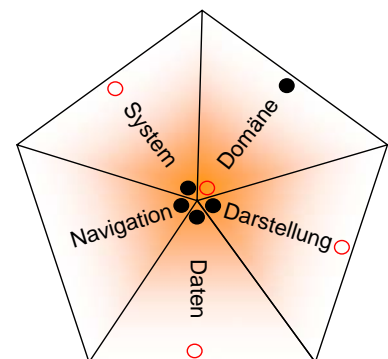


Abbildung 3.2: Das Werkzeug Sinnzeug (links für Fonds; rechts für Webseiten)

Sinnzeug verwendet eine Kräfte- bzw. Gravitations-Metapher („Spring Embedder“) zur Darstellung der multidimensionalen Zusammenhänge von Informationsobjekten wie Textdokumenten oder Fonds. Vordefinierte sowie freie Suchbegriffe können vom Benutzer hinzugefügt, verschoben oder entfernt werden, woraufhin sich die Anordnung der Dokumentenpunkte dynamisch anpasst. Wird der Zeiger über einen Punkt bewegt, so werden die Attribute des Informationsobjekts eingeblendet.

Diskussion

Das Interaktionskonzept von Sinnzeug zeichnet sich durch intuitive Bedienbarkeit aus. Sinnzeug zeigt eindrucksvoll, wie ein einfaches Werkzeug die flexible Exploration eines Informationsraumes unterstützen kann: Der Benutzer ist, ähnlich wie beim KnowledgeExplorer (Kap. 3.3.2), sehr frei in der Organisation und dynamischen Veränderung seiner individuellen „Sicht“ auf die Sammlung. Durch die Animation wird nachvollziehbar, welche Dokumente von den Begriffen angezogen bzw. abgestoßen werden. Allerdings lässt sich mit in der gemischten Darstellung sinnvoll nur mit wenigen Suchbegriffen gleichzeitig arbeiten, damit nicht zu viele „Kräfte“ gleichzeitig auf die Lage der Punkte wirken. Werkzeuge zur Inhaltsanalyse und die Navigation in existierenden strukturierten Daten sind nicht vorhanden. Sinnzeug ist ein Übersichtswerkzeug und kann bei einfachen Fragestellungen oder bei der groben Orientierung helfen.



²² <http://www.sinnzeug.de/>

KartOO

Die Meta-Suchmaschine KartOO²³ der Firma *KartOO S.A. Worldwide* unterstützt die Darstellung von Suchergebnissen im Internet und die interaktive Verfeinerung von Suchanfragen durch extrahierte Begriffe (Abbildung 3.3).

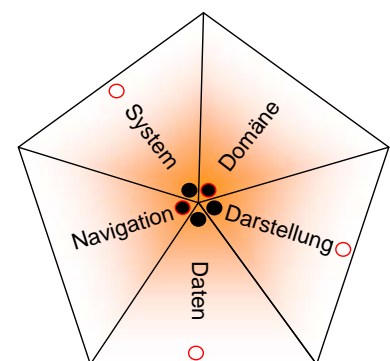


Abbildung 3.3: Meta-Suchmaschine KartOO zur Ergebnisdarstellung und Anfrageverfeinerung

Dazu verwendet KartOO eine interaktive Themenkarte, auf der Websites und extrahierte Begriffe angeordnet sind. Während beim Werkzeug „Sinnzeug“ explizit definierte Begriffe als organisierendes Prinzip für die Darstellung dienen, wird das Kartenlayout bei KartOO durch den multidimensionalen Vergleich der Dokumente bestimmt. Begriffe dienen der Kartenbeschriftung. Fährt der Benutzer mit dem Zeiger über ein Dokument, so wird eine Vorschau angezeigt, die einen visuellen Eindruck des Web-Dokuments und einen Text-Extrakt umfasst. Außerdem werden relevante Begriffe auf der Karte hervorgehoben. Fährt der Benutzer über einen Begriff, so werden die damit assoziierten Dokumente hervorgehoben. Einige Dokumentenattribute werden grafisch dargestellt (z.B. Aktualität, Dokumentenformat). Ein Klick auf einen Begriff übernimmt diesen in die Suchanfrage und berechnet eine neue Darstellung für die veränderten Suchergebnisse.

Diskussion

KartOO visualisiert die inhaltlichen Zusammenhänge von Websites und verwendet extrahierte Begriffe zur Beschriftung und Erklärung der Karte. KartOO erlaubt die flexible Navigation über die Begriffe. Da das System für die anfrageorientierte Suche konzipiert ist, lassen sich nur einzelne Dokumente auswählen und beschreiben. Die aufwändige Darstellung erlaubt keine gute Skalierung für große Dokumentensammlungen. So werden für große Suchergebnisse mehrere unzusammenhängende Karten berechnet, in die sich der Benutzer einzeln mit hohem kognitivem Aufwand einarbeiten muss. Die Suchbegriffe selbst sind nicht immer Bestandteil der Karte, obwohl diese für den Suchenden eine wichtige Rolle spielen und die Orientierung erleichtern könnten. KartOO hilft bei der Exploration von Suchergebnissen, ist aber für aufwändige Recherchen in größeren Textsammlungen ungeeignet.



²³ <http://www.kartoo.net>

VISUOS

Das Werkzeug VISUOS²⁴ [Bü04] versteht sich als „Operating Software for Knowledge Workers“ und adressiert verschiedene Aufgaben der Arbeit mit Dokumenten (Abbildung 3.4).

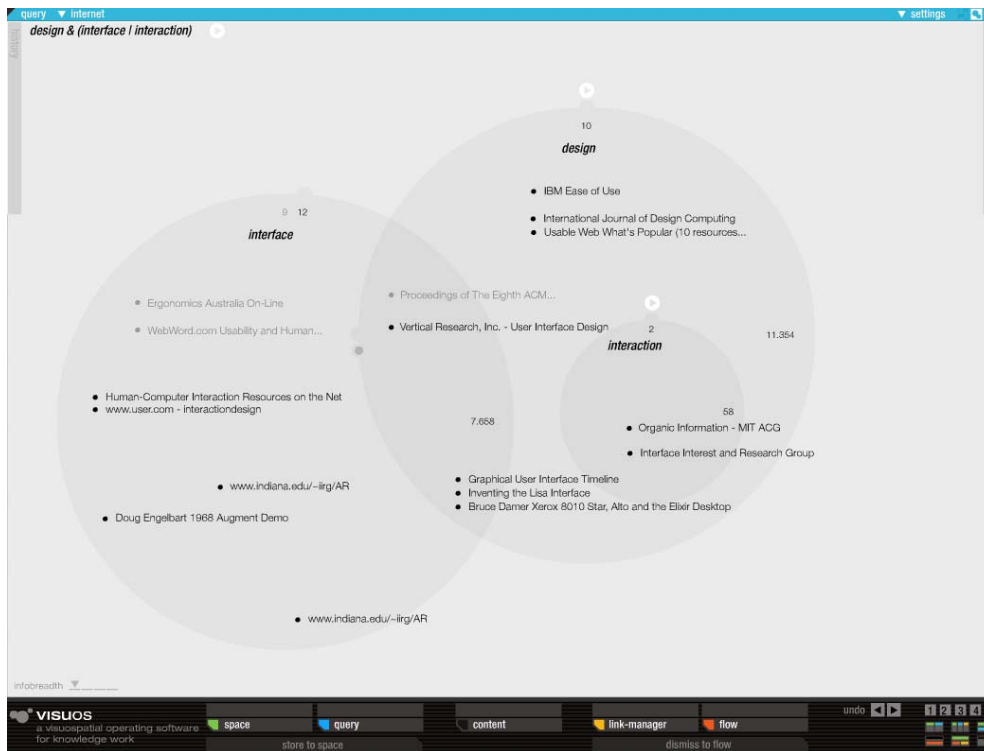
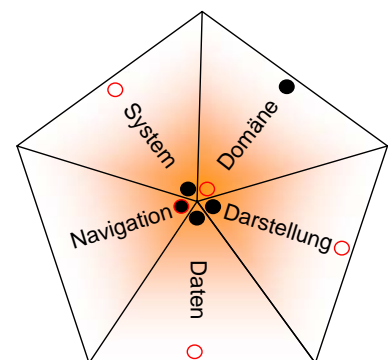


Abbildung 3.4: Das Werkzeug VISUOS verwendet Venn-Diagramme zur Organisation von Dokumenten und Anfragen

In dieser Arbeit interessieren die Funktionalitäten zur interaktiven Spezifikation von Suchanfragen mit entsprechender grafischer Darstellung der Ergebnisse, sowie die Möglichkeiten der interaktiven Organisation der Information. VISUOS arbeitet wie *Aduna AutoFocus* (Kap.3.3.1) mit Venn-Diagrammen. Der Benutzer kann Boolesche Anfragen interaktiv formulieren, indem er Begriffe eingibt und durch Platzierung miteinander kombiniert. Die Ergebnisdokumente werden im Anfrage-Diagramm angeordnet. Bewegt der Benutzer den Zeiger über einen Dokumentenpunkt, so wird eine Dokumentenvorschau eingeblendet – ein Klick öffnet das Dokument. Per Drag&Drop lassen sich Ergebnisdokumente oder –mengen in den privaten Arbeitsbereich ziehen, der ebenfalls als Venn-Diagramm von Begriffen organisiert ist, die vom Benutzer festgelegt wurden. Auf diese Weise lassen sich Dokumente interaktiv klassifizieren.

Diskussion

VISUOS zeichnet sich durch die einheitliche Metapher zur Darstellung von Anfragen, Suchergebnissen und persönlichen Informationsräumen aus und verfolgt das Ziel, die Grenzen unterschiedlicher Informationsquellen, unterschiedlicher Zugriffsmetaphern und unterschiedlicher Arbeitsvorgänge aufzuweichen. Wie bei *Sinnzeug* (s.o.) konstruiert der Anwender den Informationsraum dynamisch durch Spezifikation und Anordnung von Begriffen. Anders als bei *Sinnzeug* können Begriffe auch miteinander verknüpft werden. Allerdings unterliegt VISUOS den bekannten Limitationen von Venn-Diagrammen wie geringer Skalierbarkeit und potentieller Unübersichtlichkeit.



²⁴ <http://www.visuos.com>

newsmap

Weskamp & Albritton [Va05] haben newsmap mit dem Ziel entworfen, die gesammelten Nachrichten des Google News Aggregators aktuell und übersichtlich darzustellen (Abbildung 3.5).

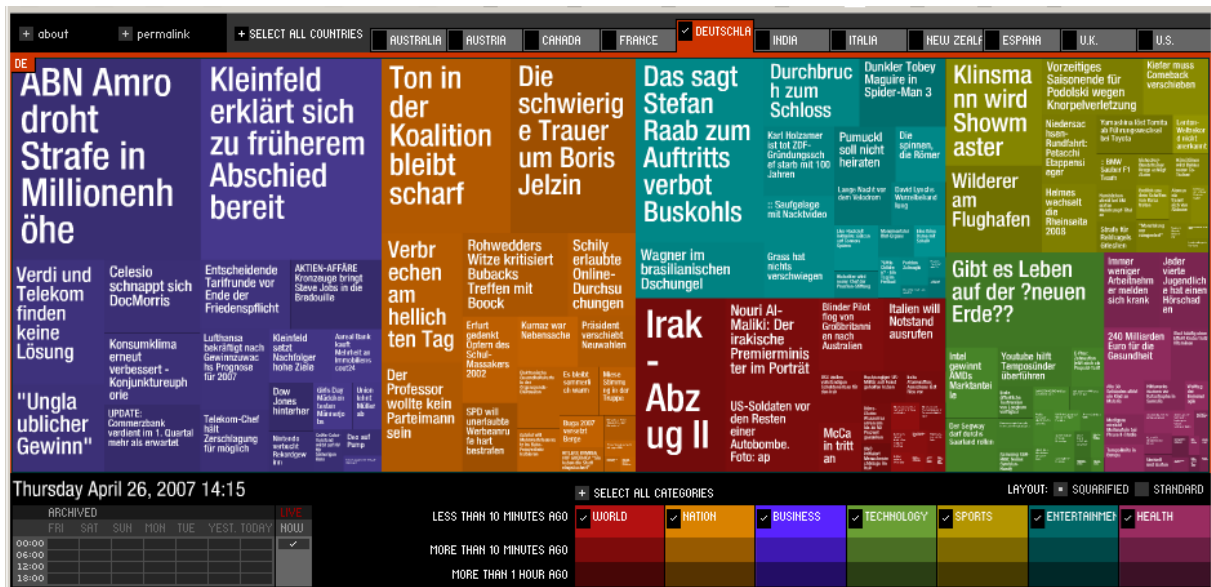
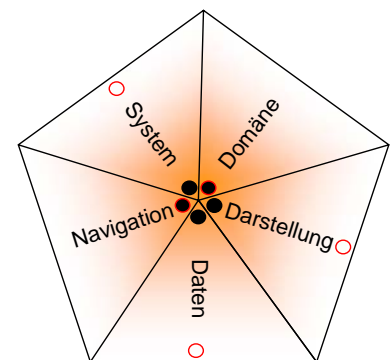


Abbildung 3.5: Das Werkzeug newsmap gewichtet Gruppen von Nachrichten nach der Anzahl ähnlicher Artikel.

Wichtige Themen werden gewichtet mit der Anzahl inhaltlich verwandter Artikel. In der Treemap-Darstellung ist die Größe der Darstellung einer Gruppe von Nachrichten proportional zu der Anzahl der Nachrichten. Beschriftet werden die Themenrechtecke mit den Titeln der wichtigsten Artikel. Die Einfärbung gibt Auskunft über den Themenbereich, die Helligkeit über die Aktualität einer Nachricht. Länder und Themenbereiche lassen sich auch als Nachrichtenfilter verwenden. Wird der Pfeil über ein Rechteck bewegt, so werden der ungekürzte Titel sowie die Anzahl der verwandten Artikel eingeblendet. Ein Klick führt zur Website der entsprechenden Online-Zeitung.

Diskussion

newsmap besteht durch die schnelle Berechnung und die einfache Bedienung. Statt einzelner Begriffe werden repräsentative Titel zur Beschriftung verwendet. Struktur und Einfärbung wurden mit einem Designpreis²⁵ ausgezeichnet. Die Darstellung von newsmap erinnert an Tag Clouds („Folksonomies“), wo die Schriftgröße zur Häufigkeit eines Begriffs korrespondiert. Allerdings kann dieser Vergleich irreführend sein: Die Schriftgröße eines Titels wird bei newsmap so angepasst, dass es das umgebende Rechteck ausfüllt. Die Größe eines Rechtecks ist proportional zur Relevanz eines Themas, gemessen an der Anzahl sehr ähnlicher Nachrichten. So hat die Länge einer Schlagzeile Einfluss auf die Schriftgröße, aber die Intuition mag daran die Bedeutung ablesen.



Die Navigationsmöglichkeiten von newsmap sind wenig ausgeprägt. Insbesondere lässt sich ein Themenrechteck nicht weiter explorieren: Weder die Titel einer Artikelgruppe, noch eine verfeinerte Darstellung von Unterthemen wird bereitgestellt. Der Benutzer kann lediglich zu einzelnen Artikeln navigieren. Insofern kann newsmap einen groben Überblick über die Nachrichtensituation liefern und beim

²⁵ ARS ELECTRONICA, Linz, <http://www.aec.at>, Designpreis 2004.

Lesen und Querlesen einzelner Nachrichten helfen, aber es fehlen Expertenwerkzeuge zur feingranularen Analyse von Textsammlungen.

Zusammenfassung

Ein weiteres interessantes Werkzeug mit integrierter Darstellung ist *AutoFocus*, das Kategorien, Verknüpfungen von Kategorien, Dokumente und Verknüpfungen von Kategorien und Dokumenten in einer Spring-Embedder Visualisierung darstellt. Weil AutoFocus zusätzlich multiple koordinierte Sichten einsetzt, wurde es Kap.3.3.1 zugeordnet. Ansätze zur integrierten Darstellung von Dokumenten und Begriffen lassen sich danach unterscheiden, ob sie Begriffe als organisierendes Prinzip der Darstellung verwenden (z.B. Sinnzeug, VISUOS), oder ob die Begriffe lediglich der Beschreibung und Erklärung der Darstellung dienen (z.B. KartOO, newsmap).

Die grundsätzliche Nützlichkeit der Navigation über Begriffe für die Dokumentenrecherche wird belegt durch die zunehmende Anzahl von Werkzeugen (z.B. „Folksonomies“). Eine Stärke solcher Ansätze ist die freie Verwendung von Begriffen zur Beschreibung und Navigation in Dokumentenräumen. *Integrierte* Darstellungen von Begriffen und Dokumenten sind allerdings oft schwierig zu interpretieren und skalieren meist schlecht.

3.2.2 Kataloge und Dokumente

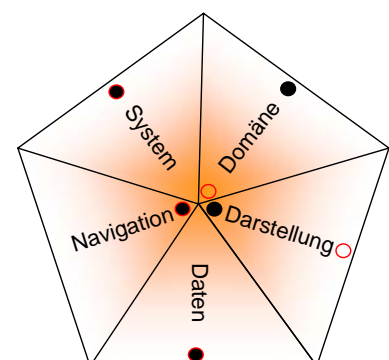
Hierarchisch strukturierte Beschreibungskategorien (Kataloge) werden traditionell zur Navigation in Textsammlungen verwendet (z.B. Inhaltsverzeichnisse, Web-Kataloge, Email-Ordner). Integrierte grafische Darstellungen von Dokumenten und Katalogen lassen sich, wie die gemischten Darstellungen von Dokumenten und Begriffen (Kap.3.2.1), danach unterscheiden, ob die strukturierten Daten den Aufbau der Darstellung bestimmen (z.B. Achsen, Ähnlichkeitsberechnung von Dokumenten aufgrund gemeinsamer Klassifikation, Distanz zwischen Kategorien im Katalog), oder ob die strukturierten Daten in der Darstellung lediglich symbolisch oder grafisch repräsentiert sind und so zur Erklärung und Navigation dienen (z.B. Symbole, Verknüpfungen).

Dotfire

Das System Dotfire²⁶ [SF+99] stellt Textdokumente als farbige Punkte („dots“) in einer Matrix dar. Attribute der Texte lassen sich konfigurierbar auf zwei räumliche (Achsen der Matrix) und eine grafische Dimension (Einfärbung) abbilden. Große Mengen gleichfarbiger Punkte in einer Matrixzelle werden durch Balken repräsentiert. Im Beispiel (Abbildung 3.6) werden juristische Fallbeschreibungen über ihre Metadaten explorativ zugreifbar gemacht.

Diskussion

Die Achsen können flach oder hierarchisch strukturierte Daten („hieraxes“) repräsentieren. Zu jedem Zeitpunkt wird genau ein Knoten einer hierarchischen Achse expandiert. Überblick und Skalierbarkeit sind zwei wichtige Designmerkmale von Dotfire. Die Navigation im Textbestand erfolgt durch Blättern entlang der Achsen bzw. durch Auswahl von Matrix-Zellen und Lesen der rechts aufgelisteten Titel und Kurztexte. In der Matrixdarstellung von Dotfire benachbarte Textpunkte sind bezüglich der damit assoziierten strukturierten Daten ähnlich. Allerdings fehlt eine Darstellung der *inhaltlichen* Querbezüge der Texte.



²⁶ Ein interaktiver Prototyp steht unter <http://www.cs.umd.edu/hcil/west-legal/dotfire/> zur Verfügung.



Abbildung 3.6: Dotfire [SF+99] stellt Textattribute räumlich und grafisch übersichtlich dar.

KnowledgeSea

Brusilovski & Rizzo adressieren mit ihrem System KnowledgeSea die Navigation in digitalen Textbüchern oder Lernmaterialien. Das System wurde motiviert durch frühere Arbeiten an der School of Computer Science der Carnegie Mellon University, Pittsburgh, USA, und durch die Beobachtung bei einem Programmiersprachen-Kurs, dass aufgrund der nicht vorhandenen Querverlinkung zwischen Kurseinheiten und Kapiteln verschiedener Programmier-Tutorials das angebotene Tutorial-Material kaum genutzt wurde [BR02]. Komplementär zur Navigation entlang Kategorienhierarchien, die beispielsweise Inhaltsverzeichnisse von Textbüchern repräsentieren, sollten inhaltliche Querbezüge *zwischen* Abschnitten („horizontale Navigation“) zur ähnlichkeitsbasierten Navigation genutzt werden können.

Das System KnowledgeSea stellt Textähnlichkeiten in einer Cluster-Darstellung dar (Abbildung 3.7). Mit einem neuronalen Kohonen-Netz [KK+00] wird jeder Text auf eine Zelle abgebildet. Je ähnlicher sich zwei Texte sind, desto stärker benachbart sind ihre Zellen auf der Karte. Besonders ähnliche Texte werden gruppiert, d.h. auf dieselbe Zelle abgebildet. Jede Zelle ist mit wichtigen Begriffen aus den Texten annotiert. Die Zellfärbung symbolisiert die Anzahl der Texte einer Zelle, wodurch sich Kernthemen identifizieren lassen, die in der Sammlung eine wichtige Rolle spielen. In der Metapher eines Wissens-Sees zeigt die Färbung also die Tiefe des Sees an.

Aus der Kategorienhierarchie kann in die Themenkarte verzweigt werden und umgekehrt, da die Zellen der Themenkarte mit relevanten Kategorien (z.B. Lerneinheiten eines Kurses) annotiert sind. Von der Kategorienhierarchie kommend, lassen sich so in der Themenkarte Einstiegspunkte für die Recherche identifizieren. Umgekehrt können die Verknüpfungen als Ausstiegspunkte aus der Themenkarte in die Hierarchie genutzt werden. Die Autoren wollen in Zukunft die kartenbasierte Navigation durch konzeptbasierte Navigation ergänzen, bei der genau ein Indexierungskonzept vom Benutzer ausgewählt wird, über das in verwandte Texte verzweigt wird. Nach einer Studie mit 21 Studenten bescheinigten die Teilnehmer der Karte eine sinnvolle räumliche Organisation der geclusterten Programmier-Tutorials und eine gelegentliche bis häufige Nützlichkeit für die Arbeit an den Kursinhalten.

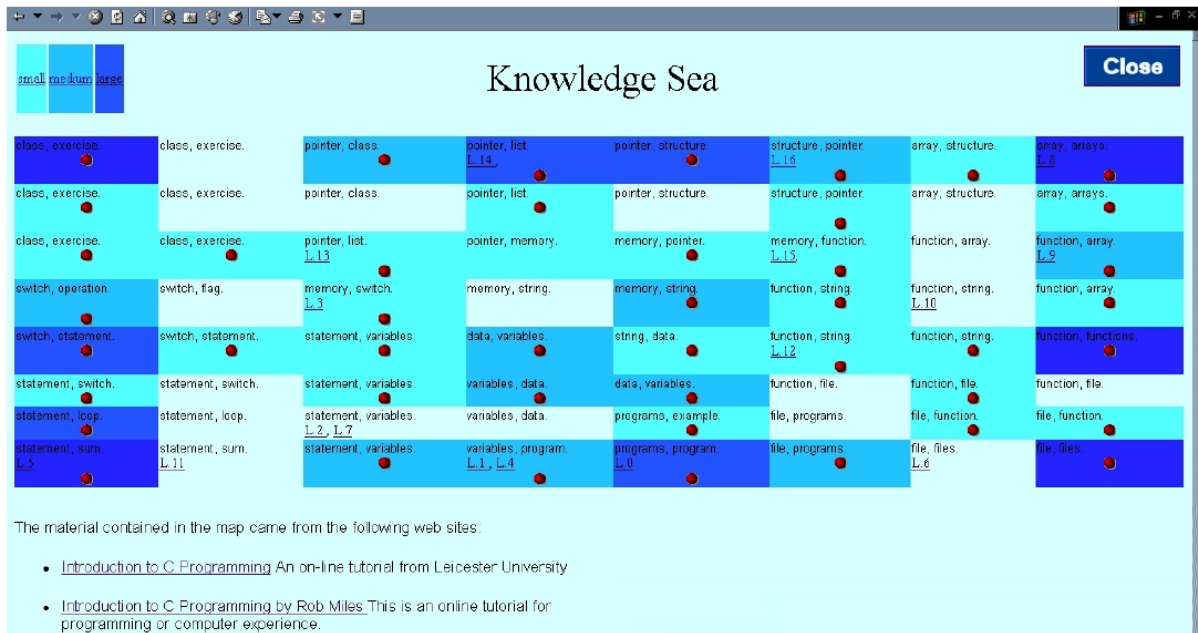
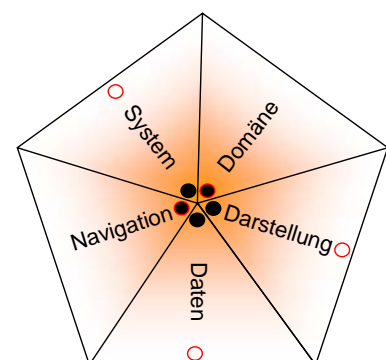


Abbildung 3.7: Das Werkzeug KnowledgeSea zum ähnlichkeitsbasierten Browsen in hierarchisch strukturiertem Hypertext [BR02]

Diskussion

KnowledgeSea realisiert Navigation basierend auf Katalogen und Textähnlichkeiten. Kategorien werden explizit in der Themenkarte symbolisiert. Die Karte bildet inhaltliche Verwandtschaft von Textdokumenten auf räumliche Distanz von Dokumentenpunkten ab. Die Achsen der zweidimensionalen Darstellung haben keine vordefinierte Semantik.

Ziel der Autoren ist die Ergänzung von herkömmlicher Navigation in hierarchischem Hypertext (z.B. Inhaltsverzeichnisse, Gliederung von Lerninhalten) durch inhaltsbasierte Navigation aufgrund automatisch berechneter Querbezüge. Auch in dieser Arbeit wird argumentiert, dass die Kombination beider Navigationsmöglichkeiten große Potentiale für die Textrecherche hat (z.B. Kap.4.4.5). KnowledgeSea besticht durch seine intuitive und übersichtliche Darstellung. Die Interaktionsmöglichkeiten beschränken sich auf Hypertext-artige Navigation. Der Übergang zwischen beiden Informationsarten, also die Navigation zwischen Kategorien und Texten, wird in beide Richtungen unterstützt. Obwohl die Darstellung Kategorien und Texte integriert („mixed visualization“), ist bei der Navigation innerhalb der Themenkarte die hierarchische Kategorienstruktur nicht erkennbar und nicht nutzbar, da die Kategorien unzusammenhängend auf der Karte dargestellt werden und lediglich als Ein- bzw. Ausstiegspunkte dienen. Der Benutzer muss zwischen den Darstellungen von Hierarchie und Clusterkarte wechseln, wodurch die Übersichtlichkeit gefährdet ist. KnowledgeSea ist zwar skalierbar, indem mehr Dokumente pro Zelle abgebildet werden, doch werden nur auf Zellebene Textähnlichkeiten durch räumliche Nähe dargestellt. Innerhalb einer Zelle werden Listen von Dokumententiteln eingesetzt, ohne Anzeige inhaltlicher Querbezüge. Insofern wird von KnowledgeSea nur die Navigation auf grobgranularer Ebene adäquat unterstützt. Auf Detailebene wird der Benutzer mit gleichen Problemen wie bei linearen Suchergebnislisten konfrontiert.



Grokker

Für die Visualisierung von Suchergebnissen im Internet realisiert die Meta-Suchmaschine Grokker²⁷ eine verschachtelte hierarchische Clusterdarstellung (Abbildung 3.8). Jedes Cluster ist mit einem extrahierten Begriff beschriftet. Ein Cluster kann durch Click vergrößert werden, wodurch zusammengefasste Dokumentenmengen der nächsten Hierarchiestufe expandiert werden. Ein Klick auf die Umgebung des gezoomten Clusters führt eine Hierarchiestufe zurück. So können Web-Dokumente über extrahierte Begriffe navigiert werden. In einer Werkzeugleiste lassen sich die Suchergebnisse interaktiv filtern.

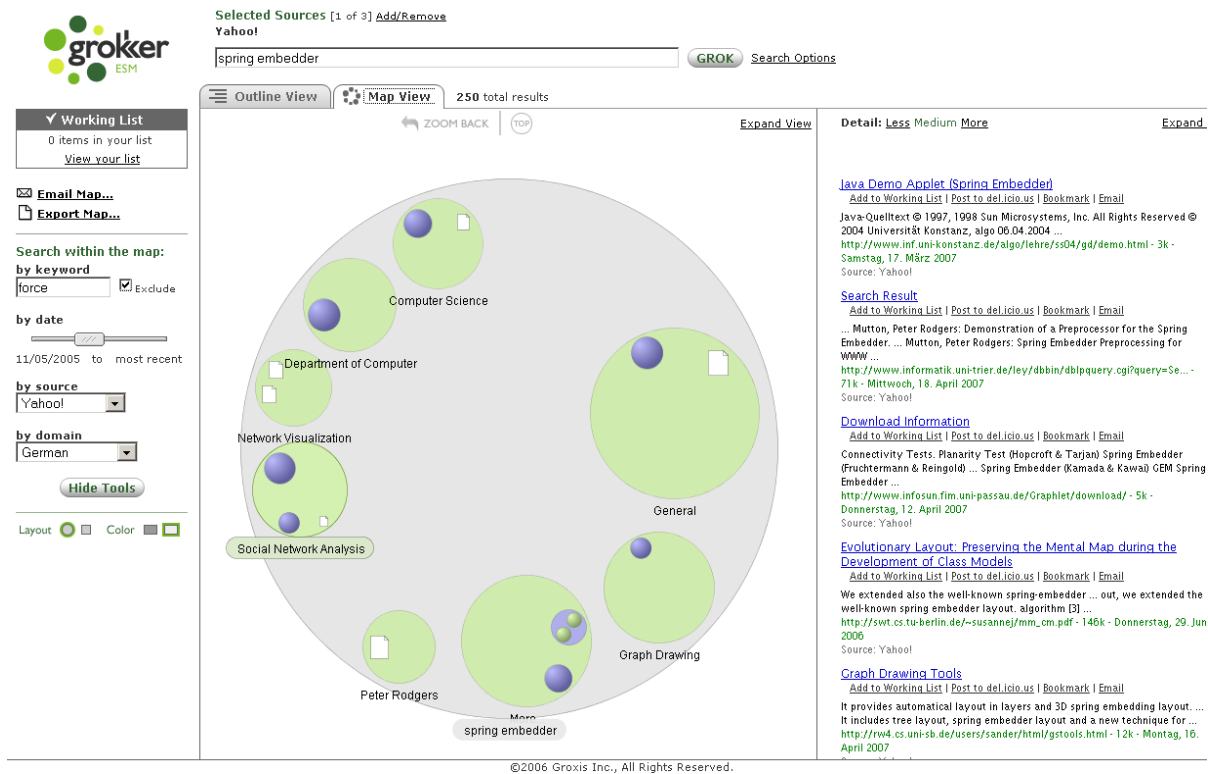
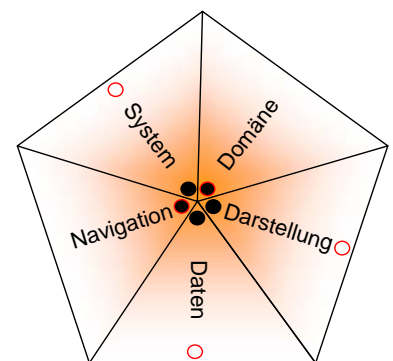


Abbildung 3.8: Das Werkzeug Grokker visualisiert Suchergebnismengen mittels einer verschachtelten hierarchischen Clusterdarstellung

Diskussion

Grokker ist ein skalierbares Werkzeug zur Suche im Web. Bei der Arbeit mit Grokker kann der Benutzer leicht den Überblick über den Kontext des gerade vergrößerten Clusters verlieren. Die Überlappung von Clustern (wie bei VISUOS, Kap.3.2.1) ist nicht vorgesehen. Die Beschränkung auf hierarchisches Clustering vernachlässigt Zusammenhänge zwischen Kategorien und inhaltliche Zusammenhänge zwischen Dokumenten zugunsten einer einfachen Darstellung. So ist beispielsweise die horizontale Navigation (wie bei KnowledgeSea, s.o.) nicht möglich. Die Navigationsmöglichkeiten von Grokker sind eingeschränkt und es stehen keine interaktiven Werkzeuge zur Auswahl und Inhaltsanalyse von Dokumentengruppen zur Verfügung. Grokker ist daher für die professionelle Analyse komplexer Zusammenhänge wenig geeignet.



²⁷ <http://www.grokker.com>

Zusammenfassung

In den vorgestellten Ansätzen werden inhaltliche Querbezüge zwischen Dokumenten und hierarchische Navigationsstrukturen gemischt dargestellt. Ein weiteres interessantes Werkzeug ist Aduna AutoFocus (Kap 3.3.1), wo die hierarchische Struktur sowohl als linearer Baum in einem separaten Fenster, als auch in Form gerichteter Verbindungslinien in der gemischten grafischen Darstellung abgebildet wird. In [SSS04] stellen die Autoren eine flexibel konfigurierbare Treemap-Repräsentation von Dokumentenstrukturen vor, in welcher grafisch und symbolisch das Matching von Suchbegriffen notiert wird.

Hierarchische Navigationsstrukturen können einerseits aus der Inhaltsanalyse der Dokumente hervorgehen (bottom-up, Clustering, z.B. Grotker), oder unabhängig von den Dokumenten definiert sein (top-down, z.B. KnowledgeSea). Während Clusterdarstellungen datenorientiert berechnet werden und die wesentlichen thematischen Merkmale und Strukturen einer Dokumentensammlung aufdecken sollen, werden Kategorien zur Klassifikation vorgegeben [GHK01], reflektieren also das konzeptionelle Modell, das einer Domäne, einem Arbeitsprozess, oder dem mentalen Modell eines Anwenders zu Grunde liegt. In diesem Sinne können die Kernmethoden des Data und Text Mining - Klassifikation und Clustering - als komplementär angesehen werden. Während Clustering datenzentriert arbeitet, funktioniert die Klassifikation modellzentriert [He06].

3.2.3 Relationale Daten und Dokumente

Metadaten werden als einfacher Spezialfall relationaler Daten in zahlreichen Systemen zur Textrecherche dargestellt und zur Navigation bereitgestellt. *AutoFocus* zeigt beispielsweise in einer interaktiven Sicht Metadaten der Dokumente an (Kap.3.3.1). Für die Recherche in relationalen Daten sind ebenfalls zahlreiche Werkzeuge vorhanden. Die Unterstützung der *kombinierten* Recherche von Texten und komplexeren relationalen Daten, die mit den Texten mit unterschiedlicher Kardinalität verknüpft sind, wird jedoch kaum durch integrierte Darstellungen unterstützt. Dabei stehen die Texte oft in einem natürlichen Zusammenhang zu relationalen Daten, die für die Recherche eine wichtige Rolle spielen, beispielsweise Unternehmensdaten zu Patenten oder Kundendaten zu Kundenbeschwerden. Im Folgenden wird das Recherchesystem *SuperTable* vorgestellt, das relationale Daten zu Texten tabellarisch darstellt. Systeme, die Texte und relationale Daten in mehreren koordinierten Sichten darstellen, sind häufiger (Kap.3.3) - möglicherweise ein Indiz dafür, dass sich bei großer Komplexität der Daten die Dekomposition in Sichten anbietet.

SuperTable

| Relevance | Information | Time | Design | Title | URL | Size_in_bytes | Language | Server_Type |
|-----------|-------------|------|--------|-------|--|---------------|----------|-------------|
| 100 | 51 | 100 | 41 | 58 | Wisely's OS Yellow Pages | 123454 | en | sg |
| 88 | 100 | 88 | 41 | 35 | Geographical Information Systems (GIS) WWW Resource List | 4 | en | uk |
| 36 | 16 | 40 | 0 | 5 | Geosciences Information Society | 37124 | en | org |
| 23 | 51 | 18 | 16 | 11 | GLCEIS / Faculteit Ruimtelijke Wetenschappen, Universiteit Utrecht | 35120 | en | nl |
| 21 | 0 | 20 | 68 | 0 | RFC 2046 - Multipurpose Internet Mail Extensions (MIME) Part Two | 115722 | en | dk |
| 15 | 50 | 13 | 8 | 0 | JAPAN GIS MAPPING SCIENCES RESOURCE GUIDE: Table of Contents | 4 | en | edu |
| 15 | 0 | 15 | 16 | 0 | GISLink - What is a GIS? | 11369 | en | com |
| 13 | 0 | 11 | 56 | 0 | IDRISI FAQ 2 | 63520 | en | at |
| 12 | 16 | 7 | 16 | 31 | MA Portfolio Option - GIS | 27520 | en | edu |
| 11 | 0 | 12 | 8 | 0 | Important Notices - Natural Resources Canada | 12383 | en | ca |
| 11 | 0 | 12 | 8 | 0 | Useful and interesting GIS web sites | 22226 | en | uk |
| 10 | 0 | 4 | 100 | 0 | Information Technology - Department - From GITA | 4 | en | com |
| 10 | 0 | 0 | 0 | 100 | ceu certification / ceu accreditation / ceu course and path course | 127058 | en | com |
| 10 | 0 | 0 | 0 | 100 | ceu certification / ceu accreditation / ceu course and path course | 127058 | en | com |

Abbildung 3.9: SuperTable [KM+02] zeigt Dokumenteninhalte und -attribute (Granularitätsstufe 2)

Die Kernidee der Retrievalsystems SuperTable [KM+02] besteht darin, verschiedene Analysephasen durch eine gemischte Visualisierung zu unterstützen, die sich mittels unterschiedlicher Darstellungsebenen an den veränderten Granularitätsbedarf anpasst (grobe Übersicht bis hohe Detailstufe). SuperTable intendiert die aufgabenadäquate Unterstützung der Suche im Web.

In höheren Detailstufen werden grafische Darstellungen relationaler Textattribute (z.B. Balkenlänge und – position) durch Zahlen bzw. Texte ersetzt (erst Titel, später extrahierte Terme, schließlich Textabschnitte). In der höchsten Detailstufe werden die Volltexte zur Darstellung der Textattribute farbige annotiert.

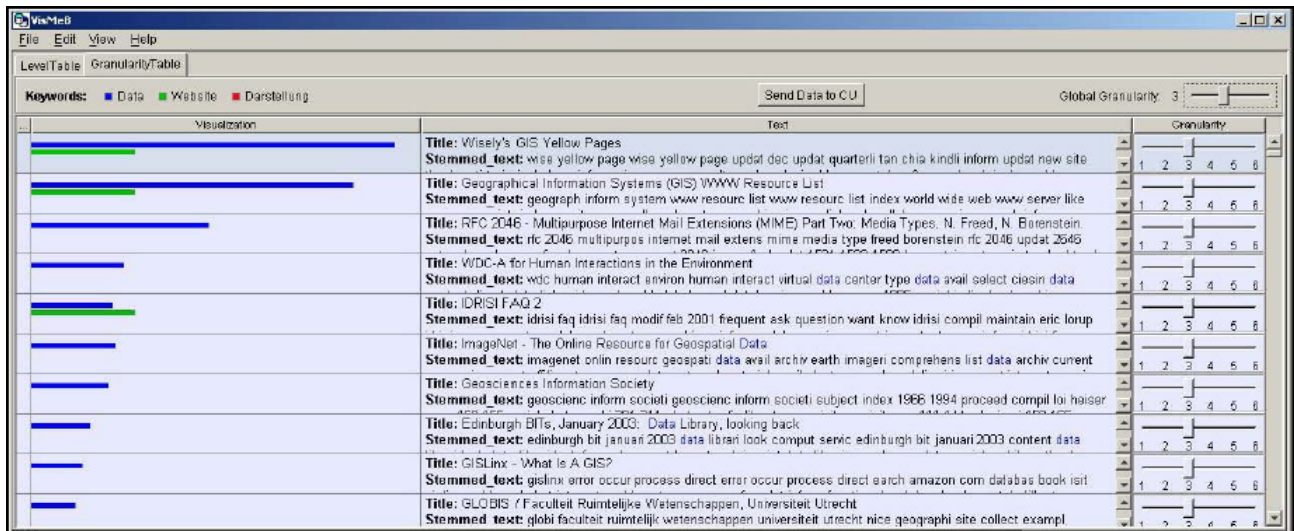
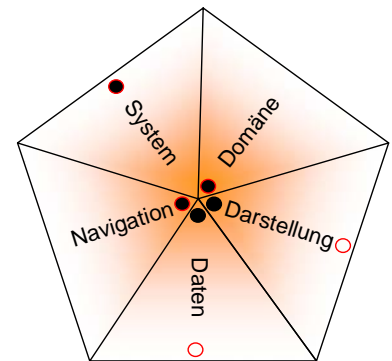


Abbildung 3.10: SuperTable [KM+02] zeigt Dokumenteninhalte und -attribute (Granularitätsstufe 3)

Diskussion

Mit zunehmender Detaillierung wandelt sich SuperTable von einer grafischen Darstellung hin zu einer vorwiegend textuellen Darstellung. Die dynamische Anpassung ist ein Alleinstellungsmerkmal von SuperTable. Wie Dotfire (Kap.3.2.2) verwendet das System eine tabellarische Darstellung. Allerdings werden keine hierarchischen, sondern relationale Attribute dargestellt. Das System versteht sich als Rahmenwerk, das verschiedene Anwendungsdomänen unterstützt (z.B. digitale Bibliotheken, Suche im WWW). Texte können mit SuperTable visuell über grafisch oder textuell dargestellte Attribute verglichen werden, die extern definiert oder aus Textinhalten extrahiert sein können. So repräsentieren eingefärbte Balken die kumulierte Relevanz eines Konzepts im Text und sogenannte TileBars zeigen im nächsten Schritt für ausgewählte Dokumente zusätzlich die Position des Vorkommens im Text an. SuperTable kann sowohl als integrierte Darstellung als auch als kombinierte Darstellung aufgefasst werden – schließlich verwenden die Tabellenzellen unterschiedliche Darstellungsmetaphern. In der Weiterentwicklung VisMeB wird die SuperTable interaktiv mit verschiedenen grafischen Darstellungen verknüpft [KR+03].



3.2.4 Zusammenfassung

Die Vorzüge der klaren und einfachen Ordnung von Kategorien-Hierarchien zur Navigation beschreibt Stürmer so: „Da für die Erkundung hunderter von Daten buchstäblich Millionen unterschiedlicher Wege zur Verfügung stehen, empfiehlt sich die Erstellung einer Art Klassifikation für die interessierenden Aspekte einer bestimmten Route“ ([St95], S.40). In der Literatur wird darauf hingewiesen, dass die Kombination von Kategorien-Hierarchien und Assoziativen Dokumentenkarten viel versprechend für die Textanalyse ist: „Two kinds of structures can be envisioned at this point, hierarchical and associative. The difference between the two is simply that associative structures involve no indications of subordination between categories. Hybrid arrangements are conceivable: one can have either hierarchies with associational cross-linkages, or association maps with arrows pointing toward subcategories. [...] we probably need more pathways than are provided by a pure hierarchy, but many fewer than are provided by a coordination system.“ [Do61]. Das

System KnowledgeSea realisiert gerade diese Kombination von Informationsarten in einer integrierten Darstellung (Kap.3.2.2). In ihrer Gegenüberstellung von Clustering und Klassifikation betont auch M. Hearst die komplementären Stärken beider Informationsarten: *“Clustering can be useful for clarifying and sharpening a vague query, by showing users the dominant themes of the returned results. [...] An underappreciated aspect of clusters is their utility for eliminating groups of documents from consideration. The disadvantages of clustering include their lack of predictability, their conflation of many dimensions simultaneously, the counterintuitiveness of cluster subhierarchies. [...] Usability results show that users do not like disorderly groupings like these, preferring understandable hierarchies in which categories are presented at uniform levels of granularity [...]. A category system is a set of meaningful labels organized in such a way as to reflect the concepts relevant to a domain. They are usually created manually [...]. Good category systems have the characteristics of being coherent and (relatively) complete.”* [He06].

Integrierte Darstellungen repräsentieren Querbezüge zwischen Texten und strukturierten Daten explizit symbolisch oder räumlich. Solche Darstellungen unterstützen die intuitive Wahrnehmung der Zusammengehörigkeit von Informationsobjekten unterschiedlicher Informationsarten (z.B. Texte zu Kategorien). Allerdings entsteht durch Vermischung der verschiedenen Informationsarten der potenzielle Nachteil der schwierigen Interpretierbarkeit der Querbezüge (Inwiefern sind die benachbarten Dokumente auch inhaltlich verwandt?). Verknüpfende Darstellungen gehen den entgegengesetzten Weg der Dekomposition.

3.3 Verknüpfende Darstellung von strukturierten Daten und Textdokumenten (MV)

Im vorangehenden Abschnitt wurden integrierte grafische Darstellungen von Textdokumenten und strukturierten Daten vorgestellt. In Kap.1.2 wurde bereits das Potenzial der Verwendung multipler koordinierter Sichten für Aufgaben skizziert, bei denen es um den Vergleich oder die Kombination unterschiedlicher Informationsarten geht.

Der visuelle menschliche Arbeitsspeicher kann drei bis fünf Informationseinheiten gleichzeitig erfassen ([Yo07], S.8). Insbesondere für Aufgaben des Erkennens von Mustern oder des Vergleichens und Kombinierens von Information, stellt diese Kapazitätsgrenze eine wichtige Beschränkung dar, die beim Entwurf unterstützender grafischer Darstellungen berücksichtigt werden muss. Die Verwendung Multipler Sichten verspricht für solche Aufgaben eine geringere kognitive Belastung des menschlichen Arbeitsspeichers [NS97][BWK00], weil die einzelnen Darstellungen weniger komplex sind und die Querbezüge zwischen zu kombinierenden Informationsobjekten nicht symbolisch oder grafisch repräsentiert werden, sondern interaktiv: *„Just as recognition is easier to perform than recall, visual comparison is easier to accomplish than memory-based comparison. Multiple views leverage the perceptual capabilities to improve understanding of relations among views...”* [Sa03]. Allerdings wurde befürchtet, dass die zusätzliche kognitive Belastung für den Wechsel zwischen Sichten („context switch“) das Nutzenpotenzial aufwiegen könnte²⁸ [BWK00]. Yost beschreibt das Spannungsfeld so: *„The complexity tradeoff is between the number of visual features of glyphs and the number of views. Presumably, both should be minimized to reduce perceptual and cognitive complexity.”* [Yo07].

Yost führt in ihrer Dissertation drei vergleichende Laborstudien mit koordinierten Sichten und integrierten Darstellungen durch, die im Folgenden skizziert werden sollen, weil sie die Eigenschaften von MultiView-Darstellungen differenziert untersuchen. Bei allen Studien wird eine multidimensionale Datenbasis verwen-

²⁸ Weitere Kosten-/Nutzen-Faktoren sind Einarbeitungszeit, Berechnungsaufwand und Platzbedarf [BWK00]

det. Jeder Datensatz umfasst vier abstrakte Attribute, damit Vorkenntnisse der Studienteilnehmer die Ergebnisse nicht beeinflussen. Die Datensätze sind einer geografischen Karte der USA zugeordnet.

In der *ersten Benutzerstudie* mit 57 Studenten ([Yo07], S.23-29) wird eine Darstellung mit 2 (bzw. 4) einfachen verknüpften Sichten mit einer komplexen integrierten Darstellung verglichen. Die Analyseaufgaben beschäftigen sich mit dem Finden einzelner Informationsobjekte und dem Erkennen von Trends in den Daten. Das Ergebnis belegt, dass bei der Suche nach zwei Attributen die multiplen Sichten der komplexen integrierten Darstellung bezüglich der Ausführungsgeschwindigkeit signifikant überlegen sind. Yost vermutet, dass die Möglichkeit vorteilhaft ist, in jeder der Sichten die optimale Codierung für ein Attribut zu verwenden ([CM84], s.u.).

Ein zweites Experiment adressiert kognitive Aspekte bei der Benutzung von MultiView-Systemen ([Yo07], S.30-42). Insbesondere die Fragen, welchen Einfluss der kognitive Wechsel zwischen Sichten („context switch“) auf die Benutzer-Performanz hat und auf welche Weise Benutzer die Informationen aus verschiedenen Sichten kombinieren, werden in der Benutzerstudie untersucht. Dabei werden psychologische Testverfahren eingesetzt und sogar Eye-Tracking-Equipment verwendet. Die 16 Teilnehmer sollen Aufgaben bearbeiten, bei denen entweder um die Suche nach Informationsobjekten oder um das Erkennen von Mustern in den Daten geht. Das System, extra für die Studie konstruiert, umfasst zwei grafische Sichten, die unterschiedliche Darstellungsmetaphern verwenden. Jede Sicht stellt die vier Datenattribute entweder durch einen Scatter Plot, Parallelkoordinaten oder eine geografische Karte dar. Das Ergebnis zeigt, dass der Wechsel zwischen den Sichten keinen Einfluss auf die Benutzer-Performanz hat. Außerdem war im Experiment das Sichten-Design problematisch, wo beide Sichten die gleiche Darstellungsmetapher verwendeten und dem Benutzer dadurch offenbar die Unterscheidung schwierig machen.

In der *dritten Studie* mit 30 Teilnehmern wird untersucht, für welche Aufgaben die Visualisierung von Verknüpfungen gegenüber der interaktiven Verknüpfung vorteilhaft (oder nachteilig) für Bearbeitungszeit und Ergebnisqualität einer Recherche sind ([Yo07], S.42-50). Dazu wurden die gleichen Daten in drei unterschiedlichen Systemen zugreifbar gemacht: Eine integrierte 3D-Darstellung, eine Darstellung mit zwei 2D-Sichten, sowie eine Darstellung mit vier 1D-Sichten. Bei den meisten Aufgaben performierte das System mit den vier einfachen Sichten am besten. Daraus lässt sich schließen, dass die Brush&Link-Interaktion gut dazu geeignet ist, Daten kognitiv zu verknüpfen – insbesondere besser als die visuell integrierte Darstellung. Die Studie bestätigt außerdem die bekannten Probleme mit 3D-Darstellungen: „*Rotation in the 3D scatter plot adds complexity to the user interface [as opposed to two 2D projections], and hence adds time without an increase in accuracy.*“ ([Yo07], S.50).

Bei komplexen integrierten Darstellungen führt die limitierte Anzahl möglicher Informationscodierungen dazu, dass für einige Attribute weniger effektive Codierungen [CM84] verwendet werden müssen als für andere. Bei multiplen Sichten hingegen können aufgrund der wenigen zu codierenden Attribute pro Sicht, die effektivsten Codierungen eingesetzt werden. Die Ergebnisse der ersten und dritten Studie bestätigen die Vorteilhaftigkeit dieses Ansatzes bei multidimensionalen Daten und komplexen Fragestellungen.

3.3.1 AutoFocus™

AutoFocus™ stellt, ähnlich wie KnowledgeSea (Kap.3.2), sowohl Kategorien als auch Texte in derselben Visualisierung dar. AutoFocus [SW+04] wurde 2001 entwickelt an der Universität Amsterdam und wird vertrieben von der Niederländischen Firma Aduna, die auf semantische Suchtechnologien spezialisiert ist. Ursprünglich eingesetzt wurde das Werkzeug im EU-Projekt SWAP²⁹ (2002-2004), wo es zur metadatenbasierten Exploration RDF-beschriebener Dokumente diente.

²⁹ EU-IST Project, Semantic Web and Peer-to-Peer, IST-2001-34103

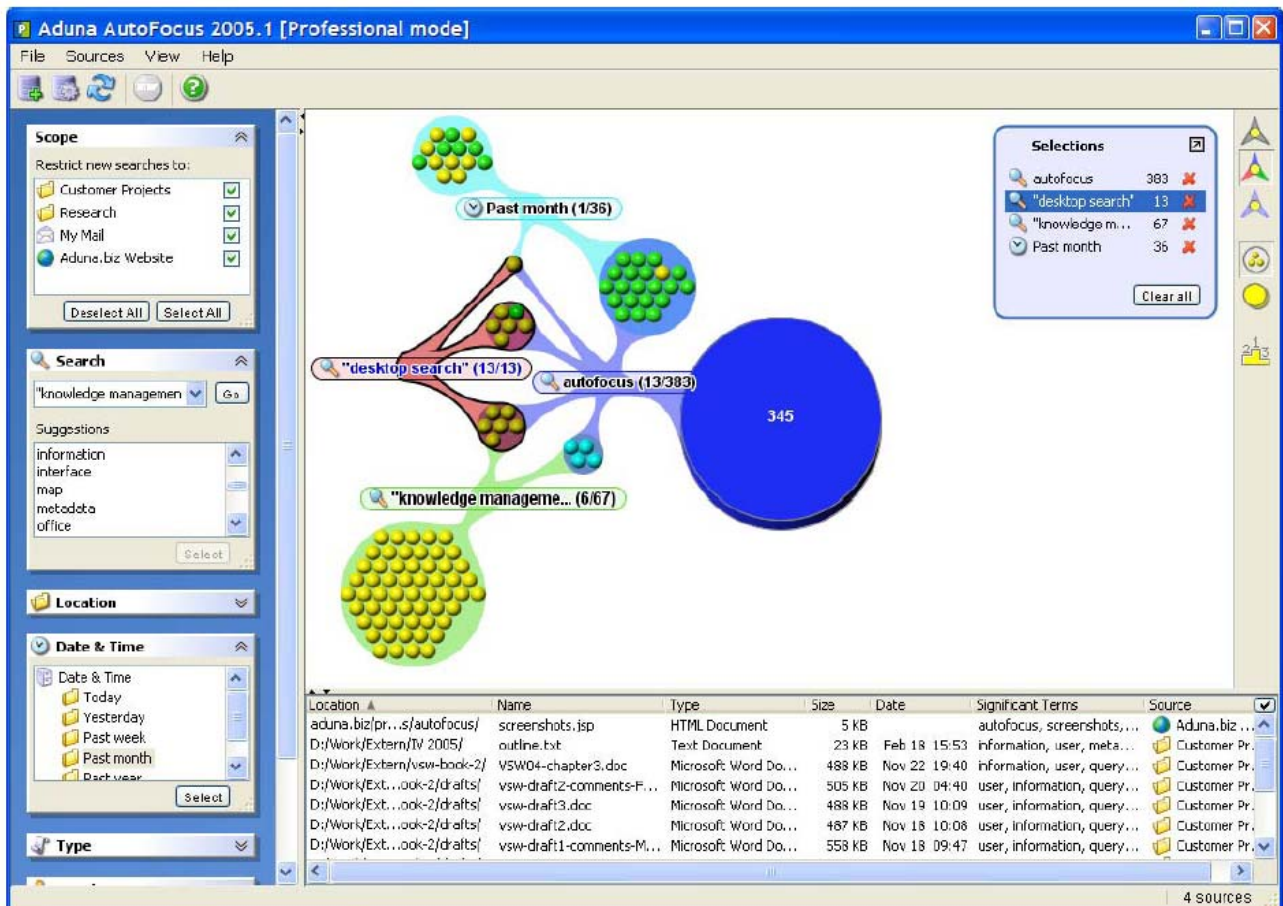


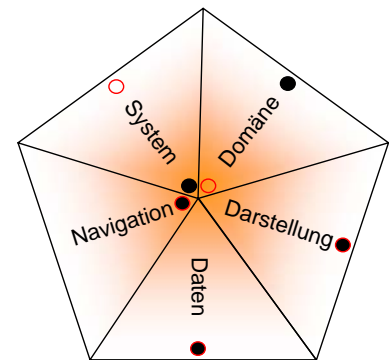
Abbildung 3.11: Das Werkzeug AutoFocus zum Browsen klassifizierter Dokumente [SW+04]

AutoFocus zeigt große Mengendiagramme (Venn-Diagramme) als Graphen an und basiert auf einer Variante des etablierten Spring-Embedder-Algorithmus (Abbildung 3.11). Klassifizierte Objekte, z.B. Textdokumente, werden als Punkte dargestellt. Gleich klassifizierte Dokumente werden zu Punktmengen zusammengefasst. An der Größe einer Punktmenge lässt sich die Anzahl darin enthaltener Dokumente ablesen. Der Spring-Embedder-Algorithmus sorgt dafür, dass Mengen ähnlich klassifizierter Dokumente nach Möglichkeit benachbart dargestellt werden. Zusätzlich zu den Dokumenten lassen sich Kategorien als dunkle Punkte anzeigen. Jede Kategorie ist mit jeder Dokumentenmenge verbunden, in der sich mit der Kategorie klassifizierte Dokumente befinden. Jeder Kategorie ist eine Farbe zugeordnet. Mengen, die mit mehreren Kategorien verbunden sind, werden mit entsprechenden Mischfarben eingefärbt. In der Kategorienhierarchie benachbarte Kategorien werden in der Darstellung durch Pfeile verbunden. Wenn das Graph-Layout es zulässt, werden Kategorien, die viele Dokumente gemeinsam haben, benachbart dargestellt.

Texte und Kategorien lassen sich durch interaktive Spezifikation in mehreren Sichten hervorheben (Facettierte Navigation). Realisierte Sichten sind Listen (z.B. Suchergebnisse, Dokumentengröße, Sprache), Baumdarstellungen (z.B. Dokumententyp, Ablageordner, Gruppen und Personen) und Metadaten der Dokumente. Ein Doppelklick auf eine Information (z.B. Suchterm, Person) führt dazu, dass diese Information als Kategorie in die zentrale Ansicht aufgenommen wird. Dann wird ein neues Graph-Layout berechnet, in dem gegebenenfalls Dokumentenmengen aufgespaltet werden. So lassen sich die dargestellten Dokumente und Kategorien dynamisch ein- und ausblenden.

Diskussion

AutoFocus hat sich für Suchen und Analysen im Bereich SemanticWeb bewährt. Es stellt Zusammenhänge zwischen Kategorien und Objekten, beispielsweise Dokumenten, in einer flexibel anpassbaren und nachvollziehbar animierten Ansicht dar. Die Analyse von Textsammlungen mit AutoFocus beschränkt sich allerdings auf die in der Textklassifikation erfassten Inhalte (z.B. Beschreibung eines Textes in einer Ontologie). Eine Darstellung der natürlichsprachlichen Textinhalte und Inhaltszusammenhänge ist nicht möglich – lediglich die Querbezüge zwischen einzelnen ausgewählten Termen und den Dokumenten lassen sich darstellen. Die Sichten bieten vielfältige Möglichkeiten zur Navigation und bestätigen die Bedeutung der in dieser Arbeit untersuchten Informationsarten. Über die Informationen in den Sichten lassen sich Dokumente und Kategorien in der zentralen Graph-Darstellung auswählen (Brush&Link). Umgekehrt werden für eine Auswahl von Dokumentenmengen in der zentralen Ansicht die damit verknüpften Informationen in den Sichten hervorgehoben. Für eine große Anzahl von Dokumenten skaliert das System gut: Ab einem Schwellwert werden einzelne Dokumentenpunkte nicht mehr angezeigt. Die Anzahl der gleichzeitig dargestellten Kategorien allerdings ist sehr kritisch in Bezug auf die Übersichtlichkeit. Schließlich ist es nicht ungewöhnlich, wenn sich für jede hinzugefügte Kategorie die Anzahl der darzustellenden Dokumentenmengen verdoppelt! Die platzaufwändige Separierung der Dokumente in disjunkte Mengen erschwert oft das Erkennen der Klassifikationsähnlichkeit, da nicht einmal garantiert werden kann, dass Dokumentenmengen, die sich in nur einer Kategorie unterscheiden, unmittelbar benachbart dargestellt werden. Das System ist sehr nützlich zur differenzierten Darstellung von Dokumenten (z.B. Suchergebnisse) nach (wenigen) Suchbegriffen und Metadaten. Zur explorativen Inhaltsanalyse fehlt jedoch eine Darstellung der inhaltlichen Querbezüge zwischen den Dokumenten. Als relationale Daten lassen sich lediglich Metadaten darstellen. Die Realisierung der Sichten als Facetten impliziert eine Kombination der in den Sichten abgebildeten Anfragekriterien zu einer einzigen interaktiven Anfrage.



3.3.2 KnowledgeExplorer

Novak & Wurst [NW05] stellen ihr System KnowledgeExplorer zur Unterstützung des Wissenstransfers innerhalb und zwischen Communities vor, das am ehemaligen Fraunhofer-Institut für Medienkommunikation (IMK) im Rahmen des AWAKE-Projekts³⁰ entwickelt wurde.

Das System verfügt über mehrere interaktiv koordinierte Sichten (Abbildung 3.12). Die Dokumenten-Ähnlichkeitssicht („ClusterMap“) stellt inhaltliche Ähnlichkeiten von Textdokumenten als räumliche Distanz von Dokumentenpunkten dar und setzt dazu ein neuronales Kohonen-Netz [KK+00] ein. Die Konzeptkarte („WordMap“) zeigt aus den Dokumenten extrahierte Terme und deren Zusammenhänge an. Dabei werden Terme als verwandt angesehen, wenn sie häufig im gleichen Dokumentenkontext vorkommen oder den gleichen Dokumenten zugeordnet wurden. Terme und Dokumente sind interaktiv bidirektional verknüpft (Brush&Link). Mit Drag&Drop-Aktivitäten lassen sich Teilmengen der Dokumente oder Terme als Detail-Sichten anzeigen.

³⁰ BMBF-Projekt AWAKE (Networked Awareness for Knowledge Discovery), 2001-2003 (http://awake.imk.fraunhofer.de/knowledge_explorer.html), <http://netzspannung.org/cat/servlet/CatServlet?cmd=netzkollektor&subCommand=showEntry&entryId=144011>

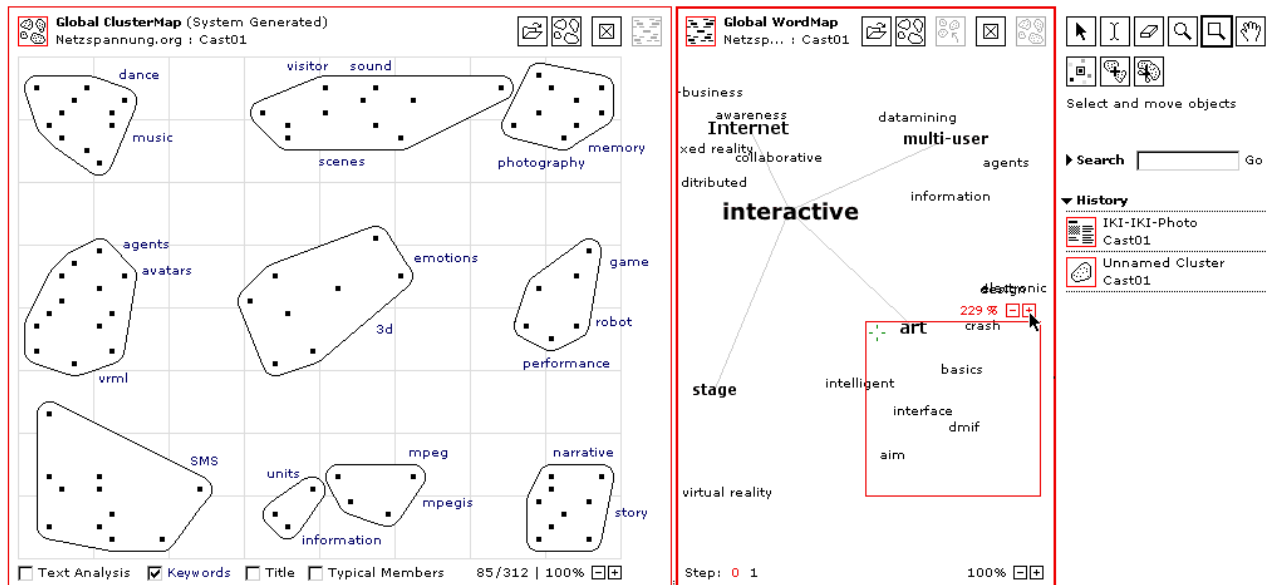


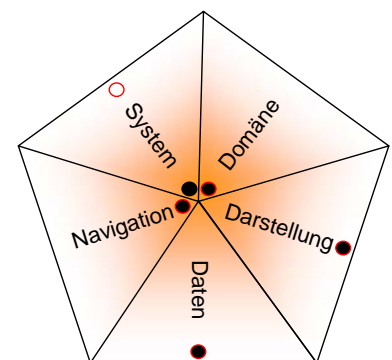
Abbildung 3.12: Das Werkzeug KnowledgeExplorer zum interaktiven Entwickeln und Darstellen individueller Perspektiven auf Dokumentensammlungen [NW05]

Der KnowledgeExplorer bietet reichhaltige Funktionen zur Personalisierung von Dokumenten- und Konzeptkarten. So lassen sich Dokumente und Terme verschieben und Gruppierungen beeinflussen. Außerdem können Beschriftungen hinzugefügt werden. Die individuellen Veränderungen („Interaktionsspuren“), welche die Sichtweise eines Nutzers oder einer Community repräsentieren, können als „Codebook“ die Strukturierung der jeweils anderen Sicht beeinflussen. Sie werden als nutzerspezifische Muster vom System gelernt („personal concept map“) und können auf neue Informationsmengen übertragen, also zu deren Strukturierung eingesetzt werden. Die Suchergebnisse lassen sich nicht nur in der systemgenerierten Dokumentenkarte anzeigen, sondern auch im Kontext beliebiger individuell arrangierter Karten, beispielsweise durch einen Domänenexperten.

Diskussion

Der KnowledgeExplorer ist ein vielfältiges Werkzeug zur Nutzbar-Machung impliziten Wissens, das aus der Benutzerinteraktion mit Dokumenten- und Termstrukturen abgeleitet wird. Das Sichten- und Koordinationskonzept baut auf unseren Arbeiten auf³¹. Die Sichten sind nicht nur interaktiv gekoppelt, sondern können auch wechselseitig zur dynamischen Restrukturierung der Sichten eingesetzt werden. Dadurch wird die Transparenz der dargestellten Strukturen jeder Sicht gefördert.

Der KnowledgeExplorer wurde entworfen für die Arbeit in (kulturwissenschaftlichen) Communities, wo eine hohe Diskursbereitschaft besteht. Komplexe Wissensstrukturen sinnvoll interaktiv zu verändern bedeutet hohen intellektuellen und editorialen Aufwand. Doch gerade bei der Textanalyse im Unternehmensumfeld ist Zeit eine kritische Ressource. Bei Textanalysen, die nicht primär auf Individualisierung und Wissenstransfer abzielen, steht dem Vorteil der Individualisierbarkeit der Nachteil des Funktionsoverheads gegenüber. Neben den Termen lassen sich im KnowledgeExplorer keine anderen strukturierten Daten einbinden. Somit können ggf. vorhandene und wichtige Daten mit Bezug zur Textsammlung nicht für die Recherche genutzt werden.



³¹ Novak et al. geben in [NW05] als Kern-Referenz ein SWAPit-Konzeptpapier an.

3.3.3 WebTheme™ (IN-SPIRE™)

Das Analysewerkzeug IN-SPIRE™³² ist das Ergebnis einer langen Entwicklungshistorie am PNNL³³, einem maßgeblich von den US-Energiebehörden getragenen Forschungsinstitut. 1996 gründete sich die Firma Cartia mit dem Werkzeug SPIRE [WT+95] aus, das zum Produkt WebTheme™ weiter entwickelt wurde. Das aktuelle System beinhaltet die Sichten GalaxyView und ThemeView [HT04].

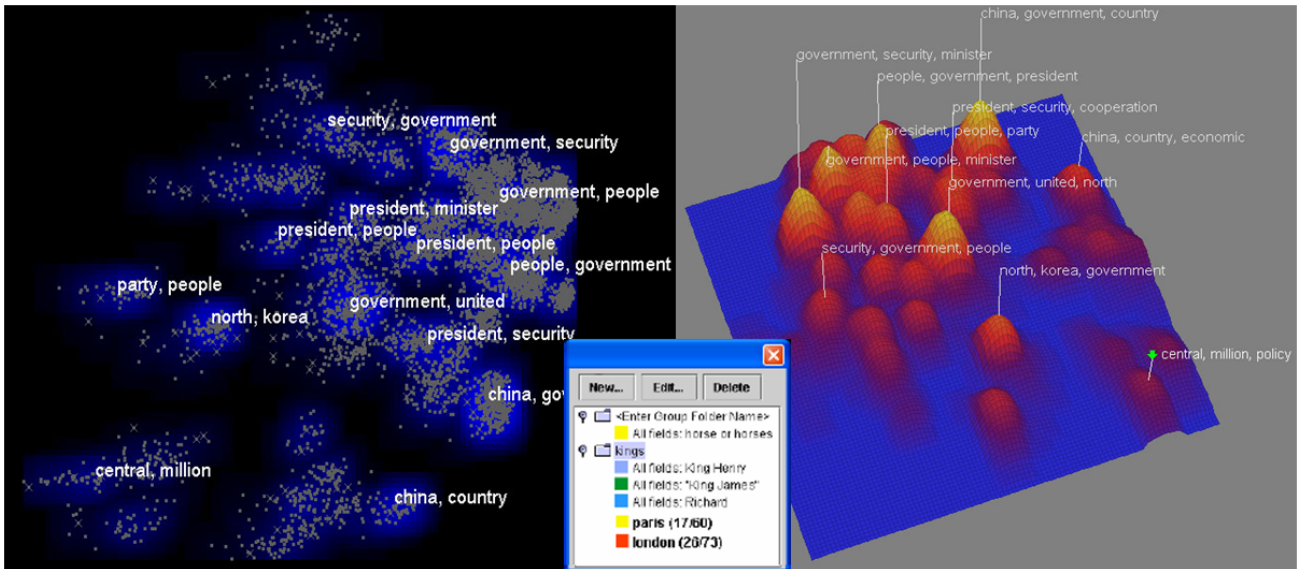
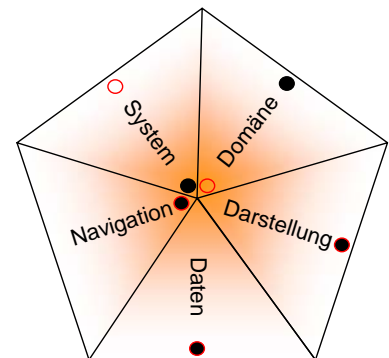


Abbildung 3.13: Sichten des Werkzeugs IN-SPIRE [HT04]: GalaxyView (links) und ThemeView (rechts)

IN-SPIRE verfolgt das Ziel, Text- und Themenräume von Web-Ressourcen übersichtlich darzustellen um lange Suchzeiten sowie das Risiko zu verringern, wichtige Dokumente bei der Suche unbewusst auszuschließen. Dazu stellt das Werkzeug, ähnlich wie der KnowledgeExplorer (Kap.3.3.2), eine dokumenten- und eine themenzentrierte Sicht bereit. GalaxyView stellt Dokumente als Punkte im 2D-Raum dar und verwendet eine Sternkarten-Metapher. Je benachbarter zwei Dokumentenpunkte sind, desto stärker ähneln sich zwei Texte inhaltlich. Ballungen repräsentieren Themen. Die Themendichte, also die Anzahl der Dokumente in einer Ballungsregion, wird durch Blauschattierung angedeutet. Genauer lassen sich Themenraum und Themendichte in der topographischen 3D-Themenkarte (ThemeView) identifizieren. Wie aus Abbildung 3.13 ersichtlich, beziehen sich beide Darstellungen auf das gleiche Layout (Informationsebene). Zur Berechnung beider Darstellung wird die Methode des Mehrdimensionalen Skalierens eingesetzt.

Dokumenten- und Themensicht können durch Mausklicks erkundet werden. Neben diversen Anfragemöglichkeiten (z.B. „query by example“) stehen Möglichkeiten zur *Gruppierung* von Dokumenten zur Verfügung. Dokumente lassen sich bezüglich Metadaten oder Suchanfragen gruppieren. Jeder Gruppe wird eine Farbe zugewiesen. Gruppen können hierarchisch organisiert werden. Die Gruppensicht ist mit den anderen Sichten interaktiv verknüpft (Brush&Link). Weitere Interaktionsmöglichkeiten von IN-SPIRE unterstützen u.a. die zeitbezogene Analyse und die Korrelationsanalyse von Themen (Gruppen), das Testen von Hypothesen und die Wiederverwendung von Suchstrategien (Workflows).



³² Spatial Paradigm for Information Retrieval and Exploration

³³ Pacific Northwest National Laboratory

Diskussion

IN-SPIRE unterstützt die Erkundung großer Textsammlungen durch ausgefeilte Darstellungen und zahlreiche Analysewerkzeuge. Extrahierte Terme oder Wertausprägungen der Metadaten können in einer Baum-Sicht eingefärbter Gruppen dargestellt und zur Exploration genutzt werden. Relationale Daten zu den Texten lassen sich nicht zur Navigation einsetzen. Das Konzept multipler Sichten wird in IN-SPIRE durch parallel geöffnete Dialoge unterstützt.

3.3.4 Conversation Map

Das System Conversation Map [Sa00] ist in Projekten an mehreren Universitäten entstanden (MIT, Berkeley, University of California). Conversation Map ist ein Newsgroup-Browser, der es den Teilnehmern erleichtert, große, umfangreiche Foren wie Mailinglisten oder Usenet-Newsgroups zu überblicken und inhaltlich zu verstehen. Die Conversation Map kann wie ein übliches elektronisches News- oder Mailprogramm benutzt werden. Zusätzlich analysiert und visualisiert das System die Inhalte und inhaltlichen Beziehungen der Beiträge. Dadurch werden die sozialen und semantischen Bezüge erkennbar, die sich im Laufe einer Debatte ergeben haben. Das System zeigt, wer mit wem kommuniziert, worüber kommuniziert wird, und welche zentralen Begriffe die Konversation dominieren. In vier koordinierten Sichten werden soziale und semantische Netze, Themen und Konversationsstränge („message threads“) dargestellt.

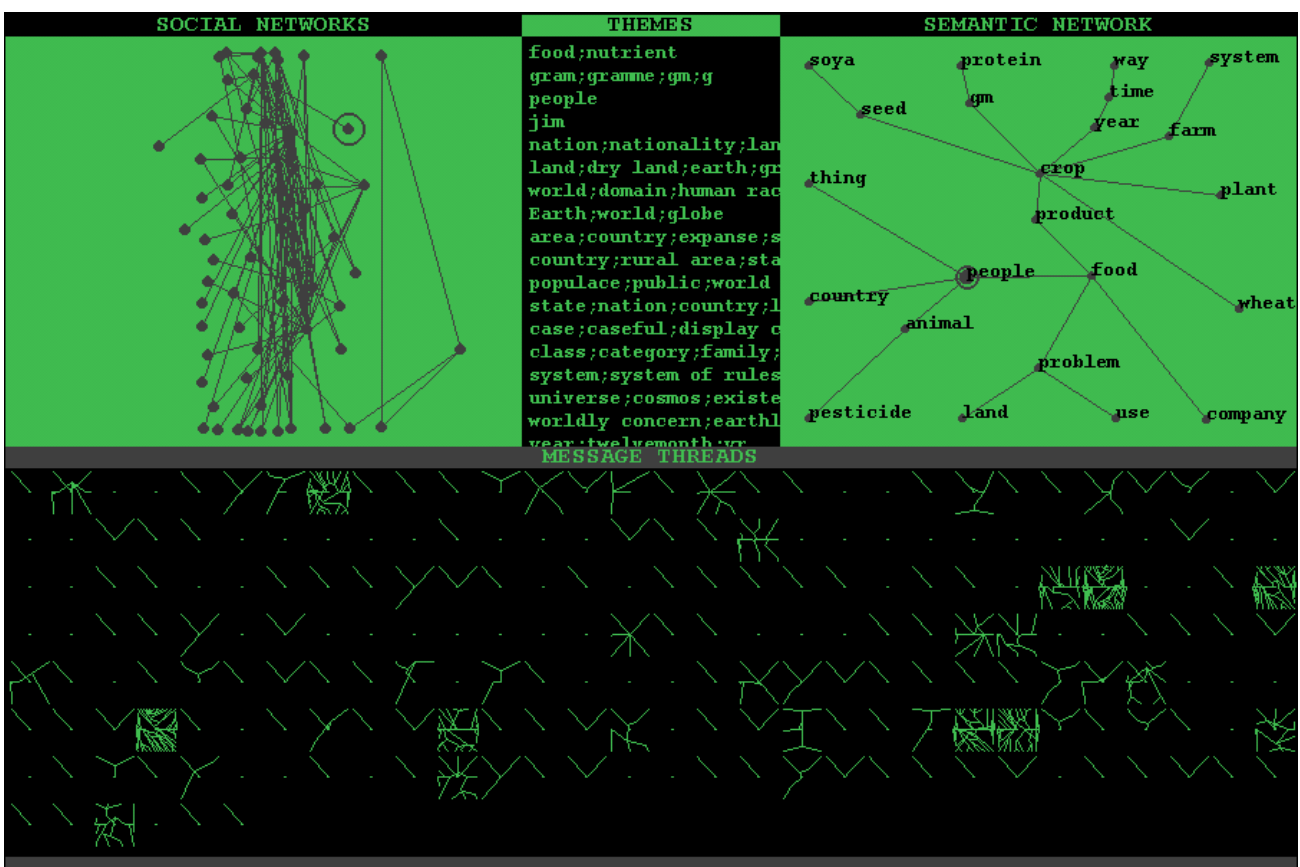
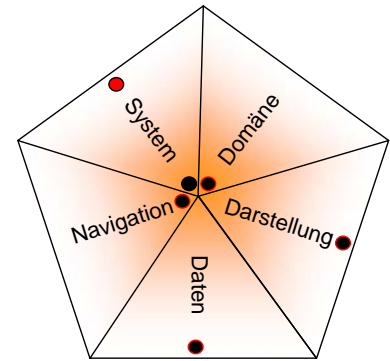


Abbildung 3.14: Conversation Map [Sa00] zur Analyse von Konversation

Die Interaktion erfolgt durch Auswählen mit einer zustandslosen Brush & Link-Metapher. Werden beispielsweise Themenkategorien ausgewählt, so zeigt Conversation Map damit verknüpfte Konversationsstränge, extrahierte Schlagworte und Phrasen (Themen), sowie die kommunizierenden Personen an. Dadurch wird eine umfangreiche Analyse unter thematischen, informationstechnischen und sozialen Gesichtspunkten möglich.

Diskussion

Wie der Knowledge Explorer (Kap. 3.3.2) adressiert auch die Conversation Map in erster Linie die Analyse von Sozialen Netzen. Sack [Sa00] stellt dafür prototypisch ein spezialisiertes MultiView-Konzept vor, das mit sparsamer Interaktion zeigen will, wie sich die unterschiedlichen Sichten und Sichtweisen interaktiv koppeln lassen. Infolgedessen werden keine spezialisierten Werkzeuge zur Analyse strukturierter Daten oder Texte präsentiert. Neben den Threads und Personen werden keine strukturierten Daten zu den Meldungen berücksichtigt. Eine für die Inhaltsanalyse zentrale Darstellung der inhaltlichen Ähnlichkeiten *einzelner Beiträge* ist nicht vorhanden.



3.4 Frameworks zur Sichtenkoordination für die Text- und Datenanalyse

Einige Forschungsarbeiten konzipieren und untersuchen wieder verwendbare Referenztechnologien, mit denen sich hochwertige MultiView-Systeme zur Datenanalyse einfach und schnell konstruieren lassen. Diese unterstützen u.a. die Konfiguration von Sichten und der Koordination dazwischen. Beim Design solcher Systeme gibt es viele Freiheitsgrade, darunter die Auswahl der relevanten Informationen und Verknüpfungen, der Darstellungsmetaphern, der Sichtenkoordination und der interaktiven Funktionalitäten.

3.4.1 SNAP-together

In seiner Dissertation stellt North [No00] unter der Bezeichnung *SNAP³⁴-Together Visualization* ein Modell und ein interaktives System für die Koordination verschiedener grafischer Sichten auf Informationen vor. Primäres Ziel ist die Vereinheitlichung und Systematisierung von Sichten- und Koordinationstypen zur schnellen Kombination und einfachen Nutzung verschiedener Darstellungsmetaphern in MultiView-Systemen.

Das Werkzeug Fusion [No03] ist eine Weiterentwicklung für das Internet, die zusätzlich DataMining-Algorithmen zur Datentransformation einbindet. Das SNAP-Framework (Abbildung 3.15) ist motiviert durch die Beobachtung, dass zahlreiche Systeme zur explorativen Text- und Datenanalyse mit koordinierten Sichten arbeiten, aber die adäquate Gestaltung solcher Systeme stark von Daten, Benutzern und Aufgaben abhängt. SNAP realisiert die Idee, die Entscheidung über zu betrachtende Verknüpfungen zwischen Sichten dem Benutzer selbst zur Laufzeit zu überlassen und somit die Parameter Benutzer und Aufgabe flexibel zu gestalten.

Den Entwicklungsprozess SNAP-basierter MultiView-Systeme unterstützt das Framework durch vordefinierte grafische Metaphern (z.B. Scatterplot, Tabelle), mit denen Daten in Sichten dargestellt werden können. Die Sichten lassen sich über vordefinierte Koordinationsmöglichkeiten koppeln (z.B. Brush & Link, Overview & Detail). Die so definierten Verknüpfungen zwischen den Daten verschiedener Sichten kann der Benutzer nach Bedarf aktivieren. Daten und Koordination basieren auf dem relationalen Datenmodell. Angezeigte Daten können aus materialisierten Datentabellen stammen oder durch SQL-Anfragen spezifiziert sein. Sichtenkoordination wird durch Datenbank-Operationen erreicht. SNAP bietet eine API, mit der sich neue Visualisierungen einbinden lassen. Als wichtiger Vorteil des komponentenbasierten Ansatzes (technisch unabhängige koordinierte Sichten) gegenüber einem System mit programmatisch fest integrierten Sichten wird die konzeptionelle Klarheit und daraus resultierende Konsistenz der Koordination hervorgehoben

³⁴ to snap bedeutet *einrasten*

([No06], S.83). SNAP's Koordinationsmodell basiert auf der Spezifikation und Kombination von „action invocations“. Das sind Tupel der Form <Visualisierung, Aktionstyp, Datensatz-ID>.

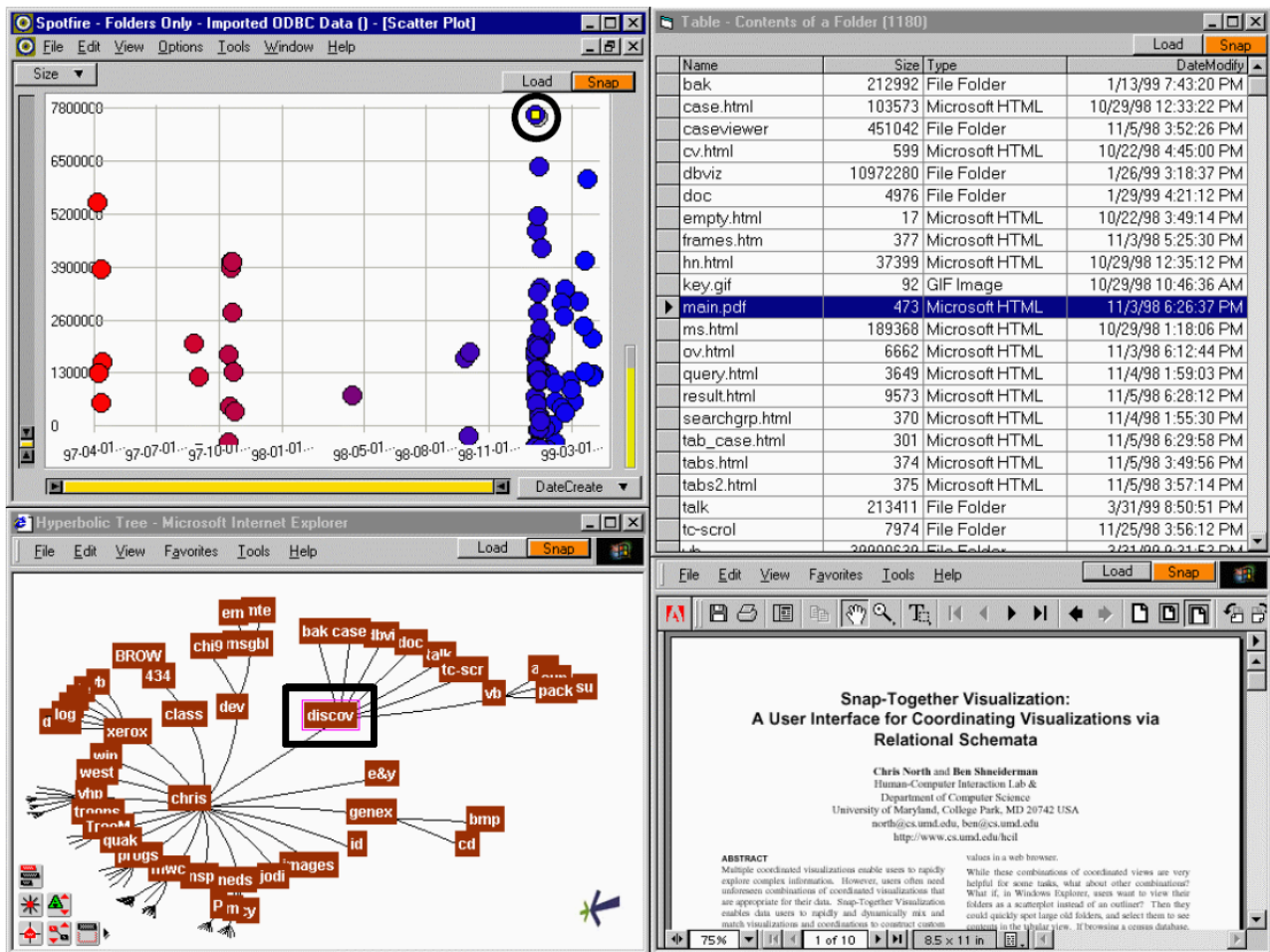
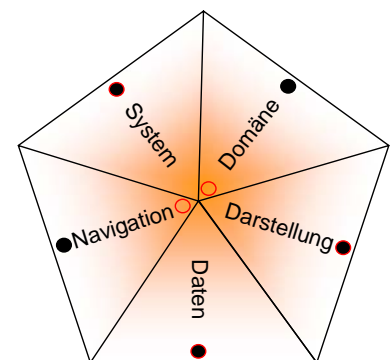


Abbildung 3.15: Das SNAP-Together Framework [No00] ermöglicht die interaktive Koppelung von Datensichten durch den Benutzer.

Diskussion

SNAP ist ein vielseitiges *technisches* Framework für die Wiederverwendung visueller Komponenten in verschiedenen Anwendungskontexten. SNAP adressiert die explorative Datenanalyse im Allgemeinen. Dies bedeutet einerseits vielfältige Einsatzmöglichkeiten, aber andererseits die Vernachlässigung von Besonderheiten spezieller Anwendungsdomänen. Insbesondere die Frage der *Aufgabenangemessenheit* des Systemdesigns wird kaum adressiert: *Wie* soll mit Hilfe des Frameworks ein angemessenes MultiView-System für bestimmte Fragestellungen im Bereich Informationszugriff gestaltet werden, damit auf intuitive Weise gerade die benötigten Informationen und Zusammenhänge dargestellt werden. Der Gestaltungsprozess ist schwierig und bleibt intelligible Aufgabe, auch wenn SNAP ein einmal entwickeltes Design schnell umzusetzen hilft: „*Implementing a coordinated-visualization interface is very tricky. An interface with two coordinated visualizations may be quite straightforward, but complexity quickly increases with the number of visualizations and coordinations.*“ ([No00], S.85). Besonders kritisch für die Benutzbarkeit ist die Nachvollziehbarkeit durch den Anwender – eine Herausforderung, die wie die Entwicklungskomplexität mit der Anzahl der Sichten zunimmt. Der Ansatz dieser Arbeit stellt, anders als SNAP-together, die *Textanalyse* in den Vordergrund: Alle anderen verfügbaren Daten dienen primär der Spezifikation und Beschreibung von Texten. Aus dieser Spezialisierung ergeben sich ein vereinfachtes



Koordinationsdesign und dadurch eine einfacher nachvollziehbare Navigation zwischen Sichten. Außerdem werden spezialisierte Werkzeuge zur Suche und Beschreibung von Texten benötigt die beispielsweise herauszufinden helfen, *aus welchem Grund* Texte zu Themen zusammen gefasst sind oder *welche Texte* zu einer Menge von Schlagworten passen.

3.4.2 InVision

Pattison & Phillips schlagen für die Realisierung visueller Informationsdarstellungen in explorativen Suchsystemen ein an das Architekturmuster Model-View-Controller (MVC) der Softwaretechnik angelehntes Modell vor [PP01]. Die *Präsentationsschicht* ist für das Rendern der Grafik zuständig und basiert auf der Datenrepräsentation in der *Modellschicht*, wo die Daten auf grafische Attribute abgebildet werden – ggf. mit optionalen Transformations- bzw. Berechnungsschritten. In der darunter liegenden *Spezifikationsschicht* werden die Abbildungsmöglichkeiten determiniert. Koordination kann nach Pattison & Phillips auf diesen Ebenen stattfinden. Für den Informationsaustausch zur wechselseitigen Benachrichtigung über Änderungen werden dafür Koordinationsobjekte definiert. Durch die gesonderte Berücksichtigung der Darstellungsaspekte lassen sich beispielsweise grafische Elemente unmittelbar in der Präsentationsschicht koppeln – etwa durch gleiche Einfärbung, räumliche Gruppierung oder Verbindungslinien zwischen Objekten unterschiedlicher Sichten.

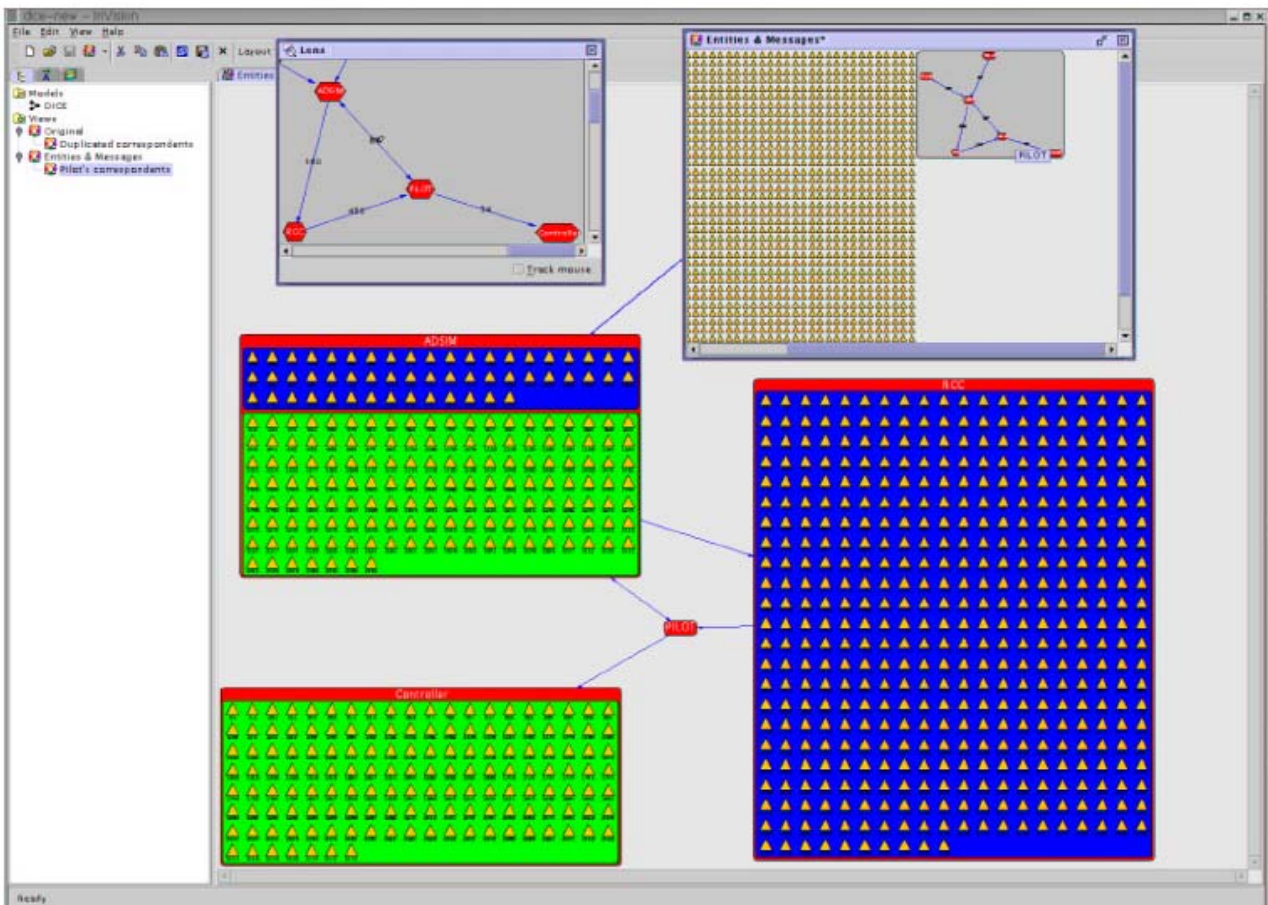


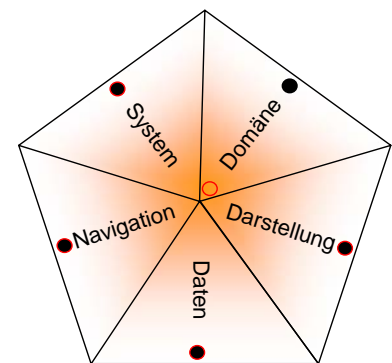
Abbildung 3.16: InVision [PP01] unterscheidet bei der Sichtenkoordination die Ebenen Präsentation, Modell, Transformation und Spezifikation.

Diskussion

Insbesondere durch die Koordination auf Darstellungs- und Transformationsebene lassen sich MultiView-Systeme spezifizieren, die mit dem SNAP-Framework (Kap.3.4.1) nicht realisierbar wären. Wie SNAP-together zielt InVision darauf ab, viele Arten der Sichtenkoordination technisch abbilden zu können. Arbeitsabläufe der explorativen Textanalyse im Besonderen und spezialisierte Interaktionskonzepte werden dagegen nicht behandelt.

Boukhelifa & Roberts stellen ein Architekturmodell zur Sichtenkoordination vor, das dem von InVision ähnelt [BR03]. Allerdings werden mehr Ebenen unterschieden, auf denen Koordination spezifiziert werden kann. Roberts & Wright haben kürzlich das allgemeine Prinzip des „Ubiquitous Brushing“ (UB) eingeführt [RW06]. UB bezeichnet die Möglichkeit, in jeder Sicht eines MultiView-Systems Informationsobjekte auszuwählen, ohne die anderen Auswahlen aufzuheben – eine Trennung von Auswahl und Hervorhebung also. Die Unterscheidung unterschiedlicher Auswahlen wird durch benutzerdefinierte farbige Codierungen des Auswahl-Werkzeugs („Brush“) erreicht. UB stellt eine Verallgemeinerung³⁵ des von uns publizierten System- und Farbkonzepts von SWAPit dar [BS04]. Obwohl die willkürliche Festlegung der Farbcodierung durch den Benutzer eine große kognitive Herausforderung darstellt, diskutieren Roberts & Wright nicht die Frage, ob und wie Benutzer mit der Vielzahl von Farben umgehen können.

In einem Prototyp [RW06] werden Informationsobjekte, die mit ausgewählten Informationsobjekten verschiedener Farben verknüpft sind, in der Farbe der zuerst ausgewählten Informationsobjekte dargestellt. Somit können bei verschiedener Auswahlreihenfolge verschiedene Farb-Hervorhebungen (Auswahlzustände) entstehen. Um die Hervorhebung unabhängig von der Auswahlreihenfolge zu machen und zudem die Überlagerung von Hervorhebungen zu visualisieren, wurde bereits in [BS04] ein Darstellungskonzept vorgestellt, das in Kap.6.4 erläutert und in Kap.6.5 durch DOI³⁶-Gewichtungen verfeinert wird.



3.4.3 Delaunay-View

Mit ihrem Werkzeug *Delaunay^{View}* stellen Cruz & Huang [CH04] ein Framework für die Visualisierung und explorative Analyse von Daten aus heterogenen objektorientierten Datenbanken vor. *Delaunay^{View}* ist eine Weiterentwicklung des Werkzeugs *Delaunay* zur Erzeugung einzelner Visualisierungen, das auf einer proprietären Beschreibungssprache für die Übersetzung von Datenobjekten und ihren Attributen auf grafische Attribute basiert. Ein Compiler („query engine“) verarbeitet solch eine Beschreibungsvorlage und erzeugt durch Anfragen an die darin referenzierten Datenbanken dynamisch eine Visualisierung. Auch Delaunay-View zeichnet sich durch die dynamische, also benutzer- und aufgabenspezifische Generierung von Visualisierungen zu Daten aus und adressiert die integrierte Exploration von Suchergebnissen aus heterogenen Datenquellen.

Im Beispiel zur Produktionsplanung (Abbildung 3.17) werden Anlagen für Flugzeugmotoren, deren Standorte, Produktionskapazitäten und Lagerbestände, sowie relevante Flugzeugmodelle in interaktiv koordinierten Sichten dargestellt.

³⁵ Bei UB kann der Benutzer innerhalb einer Sicht Auswahlen vornehmen, die durch *verschiedene* Farben codiert sind

³⁶ Degree-of-Interest [CN02], eingeführt durch Furnas [Fu81]

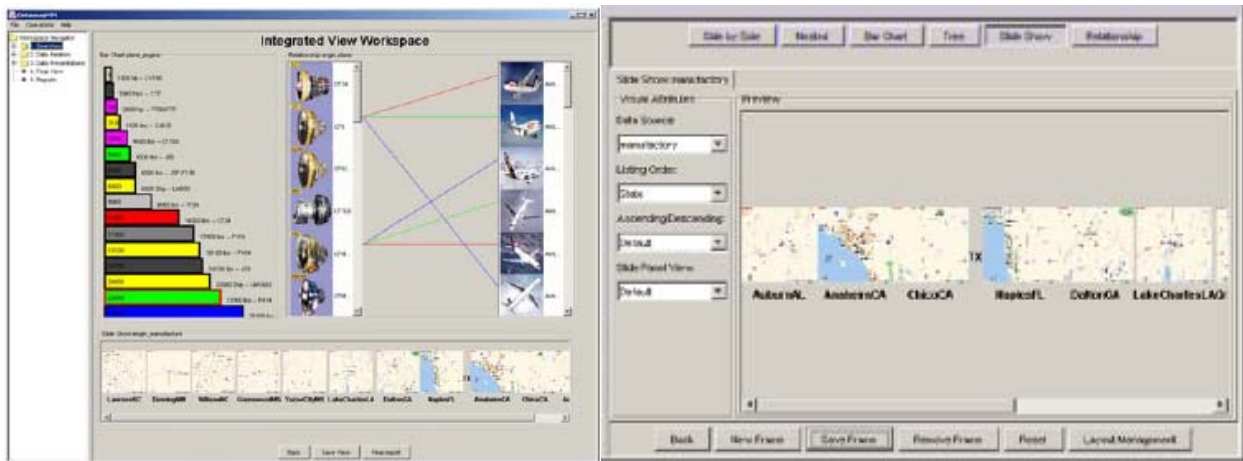
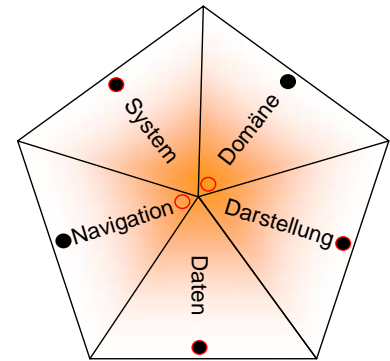


Abbildung 3.17: Beispielsystem zur Produktionsplanung, das auf Delaunay-View [CH04] basiert.

Diskussion

Die Autoren diskutieren und erweitern die Arbeiten SNAP und InVision (siehe vorangehende Abschnitte), beispielsweise durch die Verwendung von XML-Beschreibungen zur Integration von Sichten, welche auf unterschiedlichen Plattformen ausgeführt werden. Ähnlich dem InVision-Ansatz werden drei Ebenen unterschieden: In der Schicht “data processing” kann der Experte ein globales Datenschema definieren. In der Schicht “view building” lassen sich die einzelnen Sichten spezifizieren. In der Schicht “view integration” werden Gesamlayout und Sichtenkoordination festgelegt. Zur dynamischen Koordination der Sichten kommt, wie in dieser Arbeit, ein zentraler Koordinationsmechanismus („mediated notification service“) zum Einsatz.



MultiView-Systeme sind vielversprechend für die explorative Datenanalyse. Empirische Studien belegen gute kognitive Eigenschaften gegenüber integrierten Darstellungen [Yo07]. Leitfäden [BWK00] und technische Frameworks [No00] helfen beim Entwurf der verknüpften Sichten.

Allerdings sind die Frameworks nicht spezialisiert auf die Analyse von Textdokumenten, die sich in wesentlichen Aspekten von der allgemeinen Datenanalyse unterscheidet. So spielen bei der Textanalyse interaktive Werkzeuge zur Beschriftung und Beschreibung eine wichtige Rolle, um Mengen natürlichsprachlicher Texte inhaltlich erfassen und quer lesen zu können.

Die technischen Frameworks systematisieren zwar Sichtendarstellung und –Koordination, liefern aber kaum Hilfestellung für den Entwurf *guter* Systeme für die Textanalyse. So hängt die Systemqualität stark ab von den Fähigkeiten des Systementwicklers. Leitfäden zählen allgemeine Empfehlungen für MultiView-Systeme auf und sind daher bedingt hilfreich.

Eine MultiView-Referenztechnologie für die Textanalyse sowie ein Vorgehensmodell für die Konfiguration von aufgabenangemessenen Systemen für spezielle Anwendungsszenarien liefern hier wertvolle Ergänzung. Außerdem wird ein umfassendes Bewertungsschema für den Nutzen solcher Systeme benötigt.

3.5 Überblick

| | | |
|----------------------------------|--------------------------------------|-------------------------------------|
| <p>Sinnzeug (Kap.3.2.1)</p> | <p>KartOO (Kap.3.2.1)</p> | <p>VISUOS (Kap.3.2.1)</p> |
| <p>newsmap (Kap.3.2.1)</p> | <p>KnowledgeSea (Kap.3.2.2)</p> | <p>Grokker (Kap.3.2.2)</p> |
| <p>Dotfire (Kap.3.2.2)</p> | <p>SuperTable (Kap.3.2.3)</p> | <p>Conversation Map (Kap.3.3.4)</p> |
| <p>AutoFocus (Kap.3.3.1)</p> | <p>KnowledgeExplorer (Kap.3.3.2)</p> | <p>IN-SPIRE (Kap.3.3.3)</p> |
| <p>SNAP-together (Kap.3.4.1)</p> | <p>InVision (Kap.3.4.2)</p> | <p>Delaunay-View (Kap.3.4.3)</p> |

4 Problemanalyse an Fallbeispielen

In drei industrienahen Fallstudien werden wissensintensive Fragestellungen untersucht, die den integrierten Zugriff eines Experten auf Texte und auf damit in Beziehung stehende relevante strukturierte Daten erfordern. Die Fallstudien werden von Problem- und Anforderungsanalyse in diesem Kapitel bis zur Evaluation in Kapitel 7 ausgeführt.

Beschwerdemanagement und Stornoprophylaxe in Versicherungen: In einem gemeinsamen Projekt mit einem Deutschen Versicherungskonzern wurde untersucht, wie sich Informationen aus Kundenbeschwerden für die Prozesse der operativen Beschwerdebearbeitung, des Beschwerdemanagements und der Kündigungsvermeidung nutzbar machen lassen.

Produktsupport in Systemhäusern: In einem gemeinsamen Projekt mit einem großen deutschen Systemhaus wurde untersucht, wie ein exploratives Recherchewerkzeug als Ergänzung zu herkömmlichen Suchtechnologien die Mitarbeiter des Produktsupports bei der Recherche nach Kundenmeldungen unterstützen kann.

Aktivitätsanalyse in Virtuellen Kooperationsumgebungen: Im Europäischen Forschungsprojekt EcoSpace³⁷ wurde untersucht, wie durch integrative Navigation über Akteure, Zeit, Aktivitäten und kooperative Artefakte die Wahrnehmung von Aktivitäten in professionellen virtuellen Arbeitsgruppen sowie der explorative Zugriff auf gemeinsame Dokumente verbessert werden können.

4.1 Fallstudie 1: Beschwerdemanagement in Versicherungen

4.1.1 Hintergrundinformation

Domäne: Beschwerdemanagement

Dienstleistungsunternehmen aus der Banken-, Versicherungs- oder Telekommunikationsbranche sind auf eine hohe Kundenbindung angewiesen. Bei Versicherungen dauert die Kundenbindung, je nach Versicherungstyp, bis zu mehrere Jahrzehnte. Versicherungsunternehmen setzen vieles daran, die Kündigungsbereitschaft von Kunden frühzeitig zu erkennen, um Kündigungen mit gezielten Maßnahmen entgegen zu wirken (Stornoprophylaxe). Allerdings werden unternehmensweite CRM-Maßnahmen erschwert durch den traditionell dezentralen Vertrieb über Niederlassungen und Ausschließlichkeitsvertreter. Es gibt jedoch Erfolgsbeispiele einer Umsetzung von CRM-Maßnahmen im Versicherungssektor (z.B. bei der AXA Colonia Versicherung [Ra98]). Um Probleme in Service und Leistungen aus Kundensicht kennen zu lernen, werden Beschwerdeabteilungen installiert und den Informationen aus Servicecentern ein hoher Stellenwert beigemessen.

Die Prozesse im Beschwerdemanagement werden in primäres und sekundäres Beschwerdemanagement unterschieden [SS98]. Das primäre Beschwerdemanagement umfasst alle Aufgaben der operativen Beschwerdebearbeitung. Daneben kennt das aktive Beschwerdemanagement den Sekundärprozess, der das Ziel verfolgt, aus den Kundenbeschwerden etwas über Defizite und Chancen in Produkt- und Servicebereich zu lernen.

Hintergrundinformation zum Unternehmen und Fallbeschreibung: Beschwerdemanagement bei einem Versicherungs-Konzern

³⁷ eProfessionals Collaboration Space, 6. Rahmenprogramm, IST-5 35208.

Wissen aus Kundenbeschwerden zu extrahieren ist keine leichte Aufgabe. Bei der Beschwerdebearbeitung durch die Versicherung werden Metadaten zu den Beschwerden angelegt, die zum Beispiel Sparte, Art der Beschwerde oder Beschwerdegrund erfassen. Zusätzlich erscheint jedes Halbjahr ein unternehmensinterner Beschwerdebericht. Dieser besteht aus zahlreichen Statistiken und berichtet über die Verteilung von Beschwerdegründen, über den Bearbeitungsprozess (z.B. Durchlaufzeiten), über Beschwerden je Organisationseinheit und Sparte, über Beschwerden pro Vertragsphase und über die Ergebnisse der Bearbeitung (z.B. Einschätzung der Beschwerde als begründet oder unbegründet). Pro Jahr fallen über 1.000 Kundenbeschwerden an. Für die Beschwerdebearbeitung gibt es Richtlinien, die beispielsweise maximale Durchlaufzeiten festlegen. So wird u.a. angestrebt, vom Zwischenbescheid bis zur Erledigung der Beschwerde nicht länger als zwei Wochen zu benötigen. Die Hauptbeschwerdegründe, nach denen Beschwerden klassifiziert werden, sind für die Branche bekannt³⁸ und werden bei der Versicherung in 8 Bereiche mit insgesamt 58 Gründen unterschieden.

Seit einiger Zeit interessiert sich das Unternehmen verstärkt für die Zusammenhänge zwischen Beschwerden und Kündigungen. Ein externer Dienstleister prognostiziert im Auftrag der Versicherung jedes Halbjahr für jeden Vertrag eine Kündigungswahrscheinlichkeit (Storno-Score) auf Basis der demografischen Daten, der Kundenstammdaten und der Vertragsdaten. Ziel ist es, zukünftig drohenden Kündigungen frühzeitig durch gezielte Rabatt- oder Werbeaktionen für die gefährdeten Kundengruppen entgegen wirken zu können (Aussagen über die Qualität der Stornoscores liegen nicht vor). Nun stellt sich die Frage, ob auch aus dem Beschwerdeverhalten auf das Kündigungsverhalten von Kunden geschlossen werden kann. Für die Mobilfunkbranche beschreibt Aschenbrenner [As04], wie der Zusammenhang zwischen Gesprächsdauer und Vertragsauflösung zur Kündigungsvermeidung genutzt werden kann. Er betont dabei die wichtige Rolle des Experten bei der Analyse der Daten.

Sowohl für die Meldungsbearbeitung als auch für die Stornoprophylaxe stellt die Tatsache ein Problem dar, dass kundenbezogene Daten auf verschiedene Abteilungen verteilt sind. Die Daten werden meist nur in den Organisationseinheiten gespeichert, in denen sie verarbeitet werden (z.B. Schadensabteilung, Call Center, Vertrieb, Beschwerdeabteilung). Große Versicherungen unternehmen daher erhebliche Anstrengungen, Serviceaufgaben zu bündeln und die Datenhaltung zu digitalisieren und zu zentralisieren.

4.1.2 Ist-Analyse

Herausforderung: Beschwerdeanalysen zum besseren Verständnis auslösender Faktoren von Beschwerden und Kündigungen

Die Versicherung verfolgt das Ziel, im Bestand der Kundenbeschwerden nach Erklärungen für das Kündigungsverhalten von Kunden zu suchen. Im Durchschnitt gehen nur etwa 10% der Kündigungen Beschwerden voraus. Dabei handelt es sich oft gerade um solche Kunden, die stark an das Unternehmen gebunden sind und deshalb vor der Kündigung versuchen, ihr Problem auf dem Beschwerdeweg zu lösen.

Für diese Fälle könnten anhand der Beschwerdevorgänge Kündigungs- oder Beschwerdegründe sowie Kundenmerkmale identifiziert werden mit dem Ziel der langfristigen Verbesserung der Kündigungsvorhersage (Stornoprognose). Aber auch die Optimierung von Service und Produkten macht die detaillierte Analyse von Kundenbeschwerden sinnvoll.

Die in der Schadensabteilung berechneten Kundenklassen (A bis E, maßgeblich abhängig von Beitragsvolumen, Bonität, Kauf- und Zahlverhalten) könnten dabei helfen die Beschwerde- oder Kündigungsgründe ent-

³⁸ Die Bundesanstalt für Finanzdienstleistungsaufsicht, bei der pro Quartal mehrere tausend Beschwerden eingehen, veröffentlicht eine Statistik der Hauptbeschwerdegründe, an der sich Unternehmen orientieren können.

sprechend ihrer Bedeutung für die Versicherung zu gewichten. Zudem könnten bereits die Beschwerdebearbeiter, beispielsweise durch kulante Reaktion, kündigungsvermeidend tätig werden, wenn bei der Beschwerdebearbeitung bekannt wäre, wie *kündigungskritisch* eine Beschwerde ist. Dazu müssen Merkmale von kritischen Beschwerden (gegenüber Merkmalen von Beschwerden, auf die keine Kündigung folgte) bestimmt werden.

Für eine zusammenhängende Untersuchung der Beweggründe für Kündigungen, der Merkmale kündigungsgefährdeter Kunden und der Merkmale kündigungskritischer Beschwerden reichen die vorhandenen Statistiken nicht aus. Vielmehr müssen die Beschwerdetexte selbst analysiert und in den Kontext aller verfügbaren Kunden- und Vertragsbezogenen Information gestellt werden.

Aus Unternehmenssicht besteht die Motivation zu dieser Fallstudie darin, das in Beschwerdeschreiben verborgene Wissen über Beweggründe und Merkmale von Kunden über die operative Beschwerdebearbeitung hinaus auch für die Stornoprophylaxe nutzbar zu machen. Die Sonderbehandlung von kündigungskritischen Beschwerden und die Berücksichtigung von „weichen Faktoren“ (z.B. der Tonfall des Kunden in der Beschwerde) bei der Stornoprognose sind neue Wege, das Kündigungsverhalten besser zu verstehen und dadurch die Chance der Kündigungsvermeidung zu erhöhen. Die automatisierte Ermittlung von Haupt- und Zweitbeschwerdegründen aus der Korrespondenz mit dem Kunden und die Bewertung von Beschwerdegründen mit Kundenklassen bieten das Potential, die Berichterstattung im Beschwerdewesen und damit mittelbar die Produkte und Leistungen zu optimieren.

Prozesse: Beschwerdebearbeitung, Beschwerdeberichterstattung und Stornoprognose

Die Beschwerdebearbeitung (Abbildung 4.1) beginnt damit, dass eine Tabelle zur Dokumentation des Vorgangs angelegt wird. Darin aufgenommen werden u.a. Sparte, Vertragsphase, Beschwerdetyp, Beschwerdegrund und Datum. Nachdem offene Fragen mit dem Kunden geklärt wurden, wird beim Abschluss der Meldung dem Kunden eine Mitteilung geschrieben und die Beschwerdeakte ergänzt. Halbjährlich werden die Beschwerdeakten in einem Beschwerdebericht für das Management zusammengefasst. Vierteljährlich werden aus diversen Kunden- und Vertragsdaten mit Data Mining-Verfahren Kündigungswahrscheinlichkeiten für jeden Vertrag ermittelt.

Systemunterstützung

Die Systemunterstützung der Beschwerdebearbeitung befindet sich zum Zeitpunkt der Studie im Umbruch: Beschwerden werden noch nicht digital erfasst und die Beschwerdevorgänge werden in Excel™ dokumentiert. Wenn das geplante unternehmensweite Intranet auch für Beschwerdevorgänge genutzt werden wird, können auch spezialisierte Systeme³⁹ zum Einsatz kommen.

Datenanalyse: Beschwerdevorgänge, Kunden- und Vertragsdaten

Die Beschwerdetexte liegen idealer Weise in einem Dokumentenverwaltungssystem vor. Bei der Versicherung wurden Papierdokumente zunächst digitalisiert. Diese Texte sind logisch, aber nicht physisch, mit zahlreichen wichtigen Hintergrundinformationen für eine umfassende Analyse von Beschwerde- und Kündigungsgründen verknüpft. Zu den Informationsquellen gehören die Stammdatenverwaltung, der Beschwerdebericht und die Vertriebsdatenbank. In Tabelle 4-1 sind die wichtigsten Daten aufgeführt und bezüglich ihrer Datenart sowie ihres Zusammenhangs mit den Texten (Spalte „Bezug“) charakterisiert.

³⁹ wie z.B. „Sorry!“, ein Produkt von Rödl & Partner (<http://www.roedl.de/crm>)

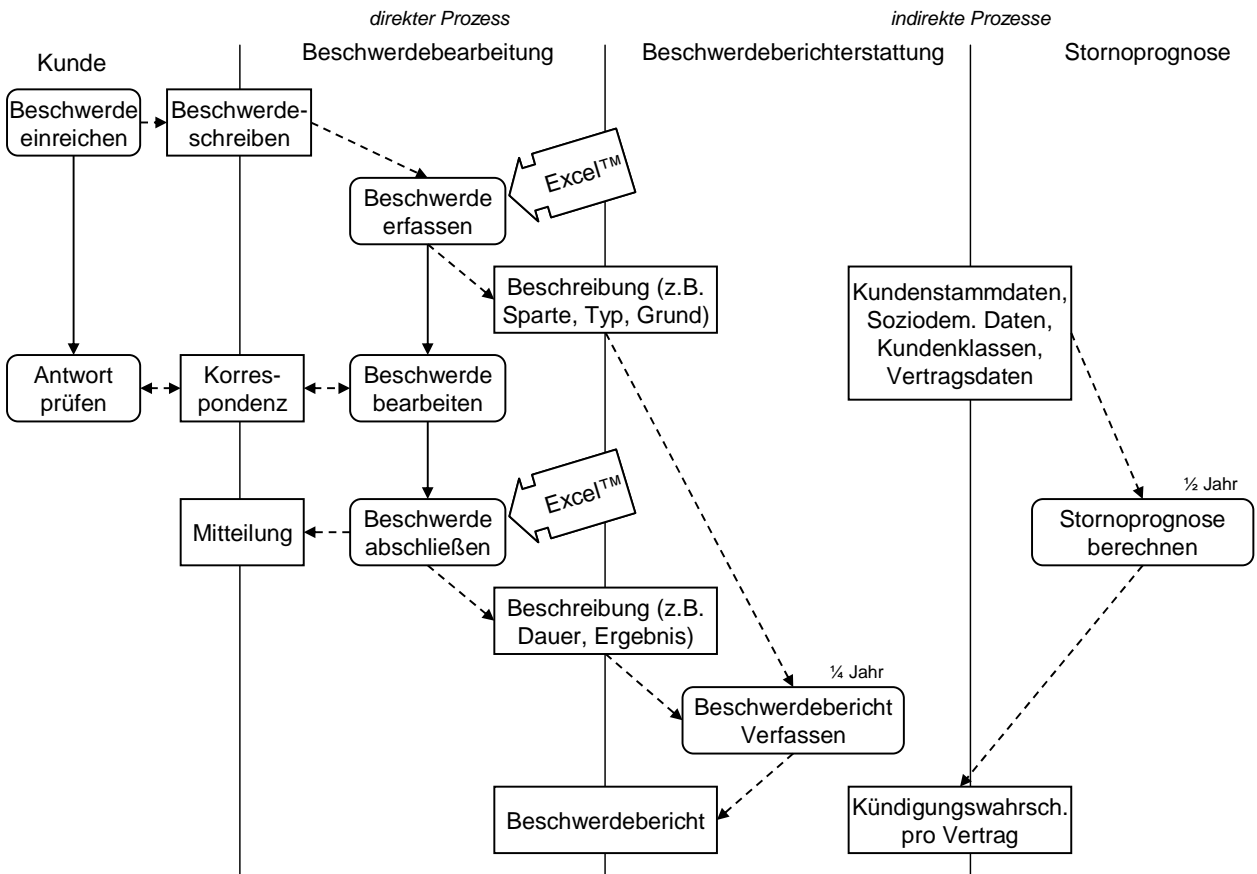


Abbildung 4.1: Beschwerdemanagement und Stornoprophylaxe bei der Versicherung (K3-Notation⁴⁰)

Für dieses Fallbeispiel wurden ca. 100 abgeschlossene Beschwerdevorgänge aus verschiedenen Versicherungssparten ausgewählt. In ungefähr 50% der Fälle erfolgte einige Zeit darauf eine Kündigung des Vertrags. Ein Beschwerdevorgang umfasst den gesamten Dialog zwischen Kunden und Versicherung, allerdings sind zur Analyse der Kundenmeinung die Anschreiben maßgeblich, die meist inhaltlich abgeschlossen und selbst erklärend sind.

Zunächst wurden die Anschreiben statistisch und inhaltlich analysiert. Tabelle 4-1 führt die wichtigsten Merkmale auf. Ein Anschreiben umfasst im Mittel knapp eine Seite Text. Als Nachdrucksmittel wird neben dem häufigen Verweis auf ein langjähriges Kundenverhältnis in fast jeder zehnten Beschwerde mit Kündigung gedroht.

⁴⁰ Koordination, Kooperation, Kommunikation (K3), vgl. Anhang 9.1.2.

Tabelle 4-1: Wichtige Daten bei der Analyse von Kundenbeschwerden im Hinblick auf das Kündigungsverhalten

| | Text | Kategorien | Tupel | Bezug zu den Texten |
|--|------|------------|-------|---------------------|
| Beschwerde: Datum, Dauer | | | • | 1:1 |
| Beschwerdeschreiben | • | | | 1:1 |
| Typ (z.B. BaFin⁴¹), Ergebnis | | • (flach) | | m:1 |
| Sparte, Vertragsphase | | • (flach) | | m:1 |
| Hauptbeschwerdegrund | | • (hier.) | | m:1 |
| Kunde: ID, Beruf, Alter | | | • | m:1 |
| Kundenklasse | | • (flach) | | m:1 |
| Vertrag: Sparte, Summe | | | • | m:1 |
| Storno-Score | | | • | m:1 |

Tabelle 4-2: Merkmale der untersuchten Kundenbeschwerden [Eg05]

| | | |
|--|----------------------------|----------------------------|
| Textmerkmale | Mittelwert (Zeichen) | 1.950 |
| | Min / Max (Zeichen) | 250 / 10.500 |
| | Wortstamm-Konzepte (Terme) | 1.895 |
| Nachdruck* (wichtigste Argumente) | Langjährige Mitgliedschaft | 18 % |
| | Drohung | Kündigung: 9 % |
| | | Rechtl. Schritte: 8 % |
| | | BaFin-Beschwerde: 6 % |
| | | Presse/Öffentlichkeit: 6 % |
| Stimmung | Besonders Wütend | 6 % |
| Sprache | Grammatikalisch fehlerhaft | etwa 10 % |
| | Sprachlich herausragend | etwa 10 % |

4.1.3 Potential- und Aufgabenanalyse

Defizite der Situation: Beschwerdemgmt. und Stornoprophylaxe basieren auf Informationsinseln

Die Dokumentation des Beschwerdeverfahrens wird in erster Linie zur operativen Beschwerdebearbeitung verwendet. Der vierteljährlich erscheinende Beschwerdebericht dient u.a. der Optimierung der Produkt- und Servicequalität. Allerdings beschränkt sich der Bericht auf Statistiken zu einigen Merkmalen von Kundenbeschwerden – die Textinhalte selbst bleiben weitgehend ungenutzt. Die halbjährlichen Stornoprognosen beruhen auf Kunden- und Vertragsdaten. Für die Kündigungsvermeidung wird nicht auf die Daten der Beschwerdebearbeitung zugegriffen, obwohl von einem Kausalzusammenhang zwischen Beschwerden und Kündigungen ausgegangen wird.

Chancen der Technologie: Texterschließung und Informationsverknüpfung

Die inhaltliche Analyse der Texte kann helfen, Muster in den Texten zu erschließen, die bei der operativen Beschwerdebearbeitung nicht strukturiert erfasst werden, wie z.B. Stimmung oder Nachdrucksmittel der Kunden. Die übersichtliche Darstellung aller relevanten Daten zu einer Beschwerde hilft bei der Identifikation von Zusammenhängen über Abteilungs-, Datenart- und Systemgrenzen hinweg. Eine stark explorative / interaktive Systemunterstützung bietet sich an, da Experten in die Lage versetzt werden sollen, Indikatoren

⁴¹ Die Bundesanstalt für Finanzdienstleistungsaufsicht übernimmt seit 2002 die Aufgaben des Bundesaufsichtsamtes für das Versicherungswesen

für Benutzerverhalten zu identifizieren und Hypothesen zu generieren und zu prüfen. „Fingerspitzengefühl“, nicht formalisiertes Expertenwissen, kann interaktiv in die Betrachtung eingebracht werden.

Potentiale: Umfassende Nutzung der Beschwerdeinhalte – auch zur Kündigungsvermeidung

Der verknüpfte Zugriff auf Kunden-, Vertrags- und Beschwerdeinformationen kann vielfältig nützlich sein. Vielleicht lässt sich die Kündigungsgefährdung einer Beschwerde bereits bei der Bearbeitung abschätzen. Vielleicht kann die turnusmäßige Kündigungsprognose verbessert werden. Vielleicht lassen sich über den Beschwerdebericht hinaus wertvolle Ansatzpunkte für die Verbesserung von Leistungen oder Kundenansprache entdecken.

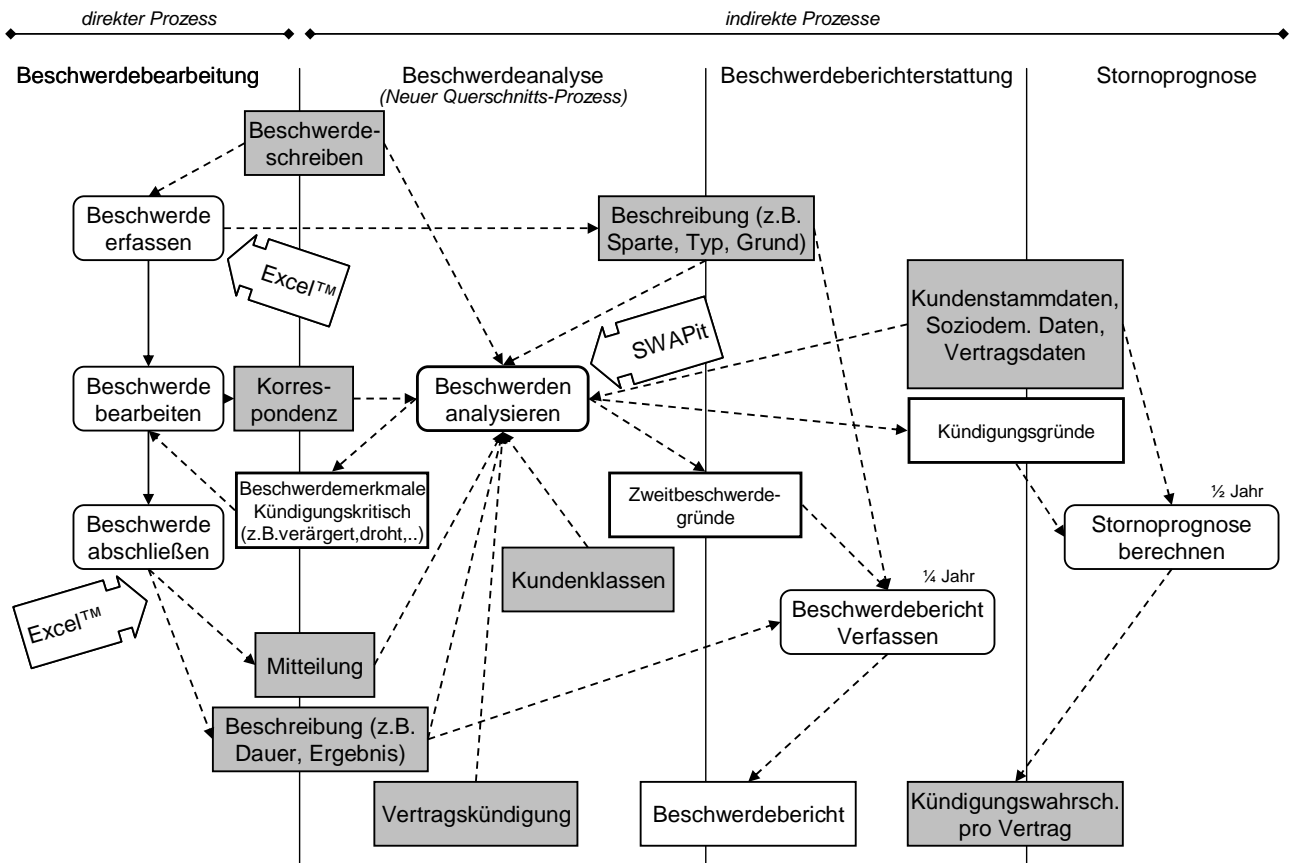


Abbildung 4.2: Neuer Querschnittsprozess der Beschwerdeanalyse (K3-Notation). Grau hinterlegt sind Daten, die in die Beschwerdeanalyse eingehen. Dick umrandet sind Daten, die daraus hervorgehen.

Um diese Fragen zu beantworten, Indikatoren zu identifizieren und Scores zu definieren, muss ein Experte aber in die Lage versetzt werden, mit den *integrierten* Daten zu experimentieren. Die avisierte explorative Systemunterstützung soll als Grundlage dafür dienen.

Dieser neue Querschnittsprozess ist ein neuer Baustein im analytischen Customer Relationship Management (CRM) und

- macht Beschwerdevorgänge der Vergangenheit nutzbar,
- macht Wissen aus Beschwerdevorgängen durch Verfahren der Textanalyse erschließbar,
- verknüpft Daten aus verschiedenen Organisationseinheiten,
- verbindet Text- und Datenanalyse.

Alle verfügbaren kunden- und vertragsbezogenen Informationen (Kundenklassen, Kunden- und Vertragsdaten, demografische Daten, Storno-Scores, Beschwerdetexte und -Metadaten) sollen experimentell aus den

unterschiedlichen Abteilungen zusammengetragen und in einem interaktiven Analysesystem zugreifbar gemacht werden. Der neue Querschnittsprozess ist in Abbildung 4.2 mit Bezug zu den Ist-Prozessen der verschiedenen Abteilungen skizziert. Die Beschwerdeanalyse verwendet und erzeugt Daten für die drei anderen Prozesse. Eine solche Analyseaufgabe lässt sich als Aufgabensteckbrief (Kap.2.2) beschreiben:

| Funktionale und Informatorische Mächtigkeit | | | | | | Kontrolle der Informationsrecherche | | | | | |
|---|---------------------------------|------------------|-------------------------|-------------------------|------------------|-------------------------------------|---------------------------------|------------------------------------|-----------------------------|------------------------------|-----------------------------|
| Ziel der Info.-Recherche | lernen | Info. verdichten | Info. auswählen | Qualität prüfen | Qualität sichern | Interaktionsmethode | zielgerichtete Suche | explorative Recherche | Stöbern u. Entdecken | | |
| Betrachtete Info.-Arten | Text-Dok. | Dok.-Attribute | Beschreibungskategorien | Relationale Information | | Interaktionsstrategie | filtern | | ausschließen | | |
| Info.-Quellen | intern | | | extern | | | Granularität | Übersicht | | | |
| Relat. Größe d. Info.-Teilmenge | gesamtes Repository | große Teilmenge | | kleine Teilmenge | | | Dynamik des Interesses | statisch | | | |
| Dynamik der Info.-Teilmenge | konstant | unterschiedlich | | ad-hoc | | | Flexibilität Interaktion | frei | angeleitet | | |
| Betrachtete Verknüpfungen | Dok.-Spezifikation | Dok.-Dok. | Dok.-Thema | Dok.-Kategorie | Dok.-Tupel | Kommunikationsmodus | wahrnehmen (phys. passiv) | | spezifizieren (phys. aktiv) | | |
| Existenz der Verknüpfungen | logisch | | organisatorisch | | physisch | | | Erwartetes Antwortverhalten | ohne Verzögerung | kurze Verzögerung akzeptiert | lange Berechnung akzeptiert |
| Verwendung der strukt. Info. | irrelevant | prüfen | verwenden | Dok. zuweisen | | | | | | | |
| Vokabular | allgemein | Fachsprache | | Korrespondenz | | | | | | | |
| Spektrum der Inhalte | allgemein | | | domänenspezifisch | | | Vorhersagbarkeit | gelegentlich | wiederholt | regelmäßig | |
| Relevanz im Unternehmen | Kernprozess | | | Unterstützungsprozess | | | Durchlaufzeit | Zeit ist irrelevant | Zeit ist wichtig | Zeit ist kritisch | |
| Absolute Bedeutung | „nice to have“ | | wichtig | kritisch | | | Expertise in Domäne | Laie | | | |
| Intendierte Wertsteigerung | Steigerung der Effizienz | | | neue Aufgabe | | | Expertise in Korpus | Information unbekannt | | Information bekannt | |
| Primäres Qualitätskrit. | Genauigkeit und Vollständigkeit | | | grober Überblick | | | Expertise in Recherche | gelegentl. Nutzer | Experte der Recherche | System-Experte | |
| Qualität und Wert des Service | | | | | | Aufwand der Arbeitsprozesse | | | | | |

4.2 Fallstudie 2: Produktsupport in Systemhäusern

4.2.1 Hintergrundinformation

Domäne: Produktsupport

Große Systemhäuser bieten für ihre IT-Lösungen ausgefeilte Support-Konzepte an. Finden Kunden in Lösungsdatenbanken selbst keine zufrieden stellenden Antworten, wird eine Meldung an das Systemhaus verfasst. Mitarbeiter des Systemhauses stehen vor der Herausforderung, die Kundenprobleme rasch, ressourcenschonend und nachhaltig zu lösen. Häufig ist der Produkt-Support gestuft: Zunächst versuchen Generalisten das Problem kostengünstig zu lösen (First Level Support). Im ungünstigen Fall müssen sich Spezialisten, meist Entwickler, mit dem Problem beschäftigen. Vielleicht ist ein zu korrigierender Produktfehler das Problem oder Funktionalitäten, die von mehreren Kunden vermisst werden, müssen nachentwickelt werden. Jedenfalls steigen die Kosten von Support-Stufe zu Support-Stufe erheblich, weshalb die frühzeitige Erkennung, auf welcher Stufe eine Meldung bearbeitet werden muss, so wichtig ist.

Hintergrundinformation zum Unternehmen und Fallbeschreibung: Meldungsbearbeitung im Product Support eines großen Systemhauses

Der Product Support des Systemhauses bearbeitet rund 800.000 Meldungen pro Jahr. Dabei erfolgt die Bearbeitung in bis zu drei Stufen. Die ersten beiden Stufen – der First und Second Level Support – werden in den

Global Support Centern (GSC) durchgeführt. Für den deutschsprachigen Raum ist das GSC Austria in Wien zuständig. Von den GSC werden ca. 70% aller Meldungen erfolgreich bearbeitet (geschlossen). Die verbleibenden 30% werden an den *Development Support* weitergeleitet. Diese aufwändigste Supportstufe wird durch die Unternehmenszentrale in Walldorf / St. Leon-Rot realisiert. Hier sind detaillierte Systemkenntnisse gefragt, denn oft sind Code-Korrekturen vorzunehmen. Im Development Support müssen Mitarbeiter mit größerer Produktkompetenz beschäftigt werden, und die Meldungsbearbeitung nimmt mehr Zeit in Anspruch als im Global Support.

Im Product Support des Systemhauses kann auf Problemlösungen der Vergangenheit zurückgegriffen werden. Dazu recherchieren Mitarbeiter in archivierten Lösungsbeschreibungen ihrer Kollegen (Hinweise) oder in abgeschlossenen Akten der Meldungsbearbeitung (Meldungen) nach ähnlichen Fällen. Hinweise und Meldungen liegen als schwach strukturierte Textdokumente vor.

4.2.2 Ist-Analyse

Herausforderung: Einsparungen durch Optimierung der Recherche

Ein Ziel des Systemhauses ist, den kostengünstigeren Primary Support zu optimieren, damit noch mehr Kundenanfragen bereits dort geschlossen werden können. Da die Recherche nach Hilfestellungen im Archiv eine wichtige Rolle bei der Meldungsbearbeitung spielt, kann dort angesetzt werden. Je effektiver und effizienter die Recherche in Hinweisen und Meldungen durch die IT unterstützt wird, desto größer die Ersparnis, zumal der Gesamtnutzen durch einen hohen Multiplikationsfaktor, der großen Anzahl von Bearbeitungsvorgängen pro Jahr, bestimmt wird.

Ein Optimierungsaspekt ist die *Verkürzung der Recherchedauer* (Effizienz). Dabei ist neben der Beschleunigung der Suchzeit bei erfolgreicher Recherche auch die Beschleunigung der Gewissheit nützlich, dass *keine* hilfreichen Hinweise oder Meldungen im Archiv vorhanden sind, damit unverzüglich andere Schritte zur Problemlösung unternommen werden können. Ein zweiter Optimierungsaspekt ist die *Verbesserung der Ergebnisqualität* der Recherche (Effektivität). Werden wertvolle Hilfestellungen übersehen, so kann dies erheblichen Mehraufwand aufgrund redundanter Schritte bei der Lösungsbearbeitung verursachen. Zusätzliche Optimierungskriterien sind aus Unternehmenssicht die Arbeitsbelastung der Mitarbeiter sowie die IT-Kosten der Recherche.

Das Systemhaus hat bereits in der Vergangenheit in die Verbesserung der Nutzung von Hinweisdokumenten investiert. So wurde mit Hilfe eines externen Beratungsunternehmens ein Case Based Reasoning–Ansatz konzipiert mit dem Ziel, Kunden in mehr Fällen automatisch gefundene Lösungsvorschläge machen zu können. Das Projekt scheiterte am zu hohen Wartungsaufwand.

Prozesskontext: Meldungsbearbeitung

Die Meldungsbearbeitung beginnt mit der Zuteilung einer Meldung an einen Mitarbeiter (vgl. Abbildung 4.3). Der zuständige Mitarbeiter übernimmt die Meldungsbearbeitung, korrespondiert bei Unklarheiten mit dem Kunden und recherchiert dann nach hilfreichen Informationen in der Hinweis-Datenbank und im Meldungsarchiv. Diese Recherche ist ressourcenkritisch, da sie einerseits einen großen Anteil der Bearbeitungszeit in Anspruch nimmt und ihr Ausgang andererseits entscheidend zur Vermeidung von Doppelarbeit in der Abteilung ist. Für die Recherche entscheidende Daten sind Produktversion, Systemkomponente, Problembeschreibung und Zeitpunkt.

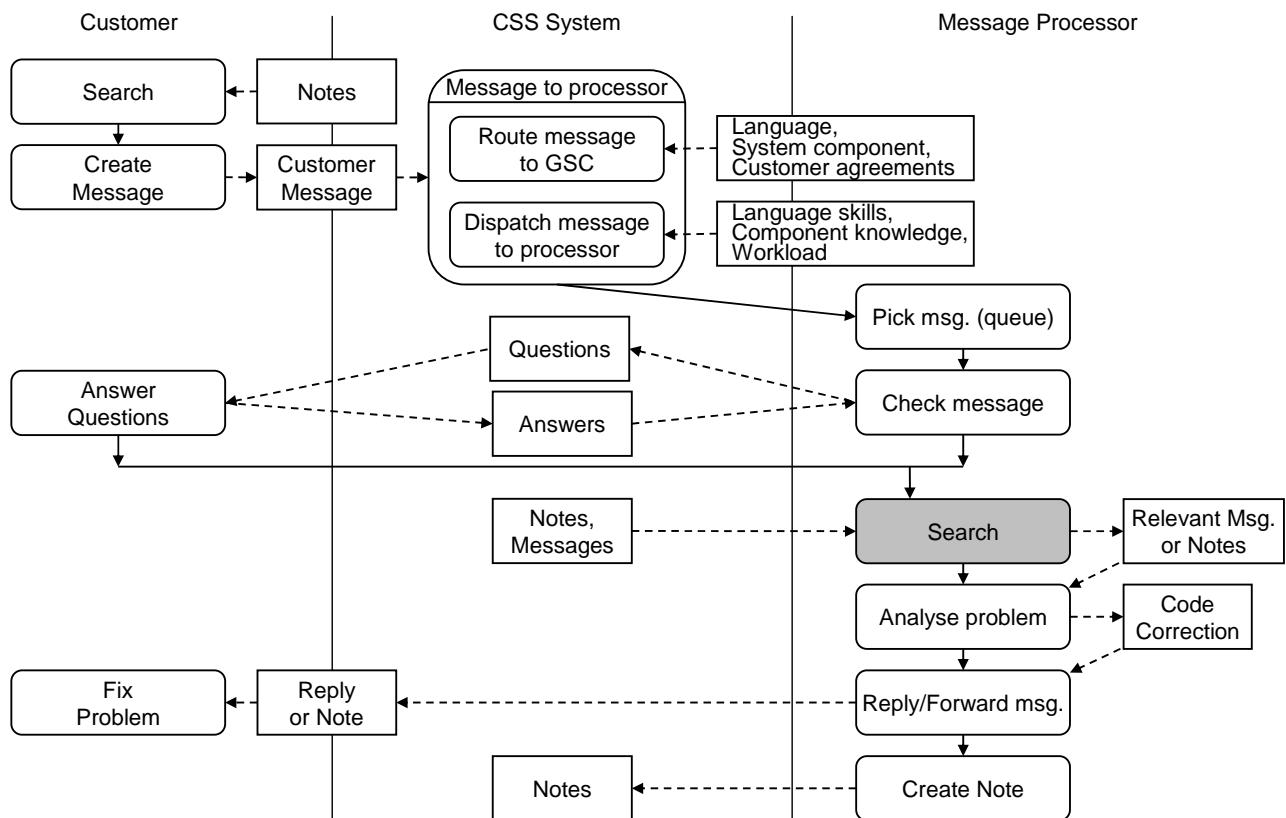


Abbildung 4.3: Bearbeitung von Kundenmeldungen beim Systemhaus (K3-Notation⁴²)

Schließlich wird eine Problemlösung vorgeschlagen und dem Kunden geht eine Antwort zu, die gefundene Hinweise referenziert oder Code-Korrekturen enthält. Meldungsbearbeiter sind angehalten, gewonnene Erkenntnisse zur Problemlösung in neuen Hinweisen festzuhalten. Die Bearbeitung einer Meldung kann mehrere Stunden in Anspruch nehmen. Die Prozesse der Meldungsbearbeitung des Systemhauses werden alle drei Jahre in einem externen Audit ISO-zertifiziert. Eine interne Prüfung erfolgt sogar jährlich.

Wie bereits erwähnt, unterscheidet der Meldungsbearbeiter beim Systemhaus zwei Dokumentenarten, *Hinweise* und *Meldungen*, die im Folgenden erläutert werden:

Hinweise enthalten gesichertes Wissen von Mitarbeitern für Mitarbeiter. Sie werden mit hohem redaktionellem Aufwand erstellt mit dem Ziel, Lösungswege zu dokumentieren und dadurch für ähnliche Problemfälle in der Zukunft nutzbar zu machen. Hinweise besitzen eine einheitliche Struktur und unterscheiden Verwaltungsdaten, Problem- und Lösungsbeschreibung, sowie Referenzen. In diesem Sinne sind Hinweise qualitativ hochwertige Dokumente, die sich sowohl maschinell als auch durch den Menschen gut finden und verarbeiten lassen. Mitarbeiter sind dazu angehalten Hinweise zulegen, wenn Meldungen erfolgreich geschlossen worden sind.

Meldungen dienen zunächst nur der operativen Meldungsbearbeitung. Im Meldungsarchiv werden ganze Bearbeitungsvorgänge gespeichert. Solche Vorgänge umfassen die Kommunikation zwischen dem Meldungsbearbeiter und dem Kunden, interne Notizen, Lösungsanleitung, ggf. Patches (Code-Korrekturen) sowie Referenzen auf wichtige Hinweise. Meldungen und Hinweise haben einen beschränkten Gültigkeitsbereich in Bezug auf Produktversionen und veralten aufgrund zahlreicher Produktmodifikationen schnell.

⁴² K3 (Koordination, Kooperation, Kommunikation) ist eine Notation für Zustände und Aktivitäten, die als Erweiterung von UML am Institut für Arbeitswissenschaft (iaw) der RWTH Aachen entwickelt wurde, s. Anhang 9.1.2

thoden umfassen Boolesche Verknüpfungen von Suchbegriffen, Wildcards, Phrasensuche und automatische Anfrage-Erweiterung. Gesucht werden kann u.a. in Volltext oder Kurzttext.

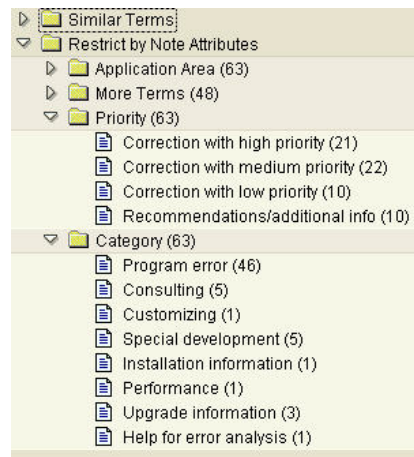


Abbildung 4.5: Attribute gefundener Meldungen werden in Baumdarstellung angezeigt.

Die Suchergebnisse werden absteigend nach Relevanz sortiert und tabellarisch dargestellt. In einem eingblendeten Kriterien-Baum werden alternative Suchbegriffe sowie Suchattribute (erweiterte Suchkriterien) zusammen mit der Anzahl der jeweils zugeordneten Ergebnisdokumente dargestellt (Abbildung 4.5) – eine wertvolle Unterstützungsfunktion für die Interpretation der Suchergebnisse und die Anfrageverfeinerung.

Datenanalyse: Meldungen und Metadaten

Die Meldungen sind mit reichhaltigen Metadaten annotiert. Das Systemhaus verwendet für die Beschreibung einer Kundenmeldung u.a. die Informationen aus Tabelle 4-3. Das Kriterium *Systemkomponente* ist hierarchisch aufgebaut und wird häufig durch den Suchenden variiert: In einer Oberkomponente zu suchen vergrößert die Treffermenge, allerdings möglicherweise zu Lasten der Genauigkeit.

Tabelle 4-3: Wichtige Daten bei der Suche von Meldungen

| | Text | Kategorien | Tupel | Bezug zu den Texten |
|---------------------------|------|------------|-------|---------------------|
| Kurzttext | • | | | 1:1 |
| Meldungsschreiben | • | | | 1:1 |
| Priorität | | • (flach) | | m:1 |
| Problemkategorie | | • (flach) | | m:1 |
| Bearbeitungsstatus | | • (flach) | | m:1 |
| Systemkomponente | | • (hier.) | | m:n |
| Bearbeitungsdatum | | | • | 1:1 |
| Bearbeiter | | | • | m:1 |

Tabelle 4-4: Statistische Charakterisierung der Dokumente (Bsp. der Meldungen zur Systemkomponente DBIF)

| | | |
|------------------|---|----------|
| Dokumente | Sprache | Englisch |
| | Anzahl Dokumente | 684 |
| | Index-Terme (Wortstamm) | 21878 |
| | Mittl. Anz. Index-Terme pro Dok. | 32 |

4.2.3 Potential- und Aufgabenanalyse

Defizite der Ist-Situation: Reine Schlagwortsuche für vage Fragestellungen inadäquat

Die Inhalte von Meldungen sind schwer zu erschließen, aber immer dann wichtig, wenn von Mitarbeitern nicht der Aufwand betrieben wird, ein Hinweisdokument anzulegen. Insbesondere in den Problemkategorien „Consulting“, „Customizing“, „Workarounds“ und „External Errors“ werden Meldungen den besser strukturierten Hinweisen vorgezogen [Hi06].

Bei vielen Recherchen ist das Problem klar eingrenzbar, oder es soll schlicht eine bekannte Meldung wieder gefunden werden. Die Suche funktioniert in diesen Fällen gut. Allerdings ist aus Sicht der Meldungsbearbeiter bei etwa jeder zweiten Meldung die reine Verwendung einer Stichwortsuche nicht angemessen [Hi06]. Die Formulierung einer „treffsicheren“ Anfrage setzt nicht nur ein gutes Verständnis des Kundenproblems voraus, sondern auch eine gute Kenntnis des Vokabulars im Meldungsbestand.

Lange lineare Ergebnislisten machen die Recherche oft mühsam, zumal das Übersehen eines Lösungsvorschlags Kosten verursacht wie Mehrfachbearbeitung oder Erhöhung der Supportstufe.

Vom bestehenden System werden die Ergebnisdokumente zwar im Kontext der erweiterten Suchkriterien (z.B. Priorität, Problemkategorie) dargestellt, jedoch nur eindimensional geordnet, z.B. nach Ranking-Wert. Es werden weder Informationen über die thematischen Zusammenhänge gefundener Meldungen bereitgestellt, noch sind Möglichkeiten zur direkten interaktiven Variation der Suchparameter vorgesehen. Somit fehlen die Anhaltspunkte zum systematischen Stöbern in Kundenmeldungen.

Chancen der Technologie: Ähnlichkeitssuche und Überblick

Über Ähnlichkeitssuche lassen sich auch Dokumente finden, die im Suchergebnis gar nicht oder sehr weit unten auftauchen. Zusätzlich verspricht die Verwendung von Clusterdarstellungen Einprägungseffekte (z.B. wie viele Meldungen es zu einem Thema gibt) und bessere Orientierung bei der Recherche. Die Strukturierung von Suchergebnissen kann helfen, schneller die richtigen Meldungen zu finden bzw. schnell auszuschließen, dass ähnliche Kundenprobleme in der Vergangenheit aufgetreten sind. Auch die rasche Identifikation wichtiger Rückfragen an den Kunden wird begünstigt durch einen guten Überblick über ähnliche Kundenprobleme. Die einfache interaktive Variation der Suchparameter (z.B. Modul, Bearbeiter, Problemkategorie) ermöglicht einen guten und effizienten Einblick in die Textsammlung.

Potentiale: Explorative Ergänzung der Suchunterstützung

Die Unterstützung von Themen- und Suchkriterien-orientierter Exploration *als Ergänzung* zum Ergebnislisten-orientierten Zugriff auf Meldungen ist vielversprechend. Damit beim explorativen Zugriff auch Meldungen gefunden werden können, die nicht zu den initialen Suchkriterien passen, wird eine Themendarstellung vorgeschlagen, die *alle* Dokumente einer Systemkomponente einschließt, zumal die Systemkomponenten ein „hartes“ Differenzierungskriterium darstellen, eine Meldung also immer nur für bestimmte Systemkomponenten Gültigkeit hat. Solch eine thematische Unterstützung bietet potentiell den Vorteil einer *Orientierungshilfe* bei der Suche: Ergebnisdokumente werden im Kontext der Suchkriterien dargestellt und thematische Zusammenhänge zwischen den Ergebnisdokumenten untereinander, aber auch zwischen Ergebnisdokumenten und anderen Meldungen zur betrachteten Komponente, werden aufgezeigt. Ausgehend von interessanten Meldungen der Treffermenge können über thematische Ähnlichkeiten oder über interaktiv variierbare Suchkriterien weitere wichtige Meldungen identifiziert werden.

Aufgrund dieser Überlegungen bietet sich ein stark interaktiver Ansatz mit grafischer Themenübersicht an, in dem alle für die Recherche relevanten Daten in parallel sichtbaren und navigierbaren Sichten abgebildet werden. Zu erwarten ist dadurch eine Optimierung der Recherchedauer, da die Variation der Suchkriterien einfacher und übersichtlicher wird. Außerdem ist mit einer Steigerung der Ergebnisqualität der Suche zu

rechnen. Schließlich können auch interessante Meldungen identifiziert werden, die auf die Suchanfrage *nicht* zurückgeliefert wurden. Könnten mehr Meldungen schneller im ersten Supportlevel geschlossen werden, so wäre die erhoffte Entlastung der teureren höheren Support-Stufe erreicht.

Zusammenfassend lassen sich folgende viel versprechende Systemergänzungen festhalten:

- Analyse *aller* Meldungen zu einer Systemkomponente statt der Beschränkung auf die Teilmenge der Suchergebnisse
- Kombination zielorientierter und explorativer Suchstrategien
- Gleichzeitige Darstellung und Navigierbarkeit aller Suchkriterien sowie der Ähnlichkeiten von Meldungen

Mit den Systemergänzungen könnte ein typischer Recherche-Prozess folgendermaßen ablaufen: Ein Meldungsbearbeiter sucht mit der Suchmaschine nach bestimmten Schlagworten. Einige hundert Meldungen werden zurückgeliefert. Die meisten Meldungen betreffen ein bestimmtes Systemmodul. Der Meldungsbearbeiter ruft das Explorationssystem für das betreffende Systemmodul auf. Auf der thematischen Karte, die alle Meldungen des letzten Jahres zu dem Systemmodul zeigt, werden die im Suchergebnis enthaltenen Meldungen hervorgehoben. Der Meldungsbearbeiter erkennt eine Ballung solcher Meldungen. Er markiert diese Gruppe ähnlicher Meldungen und sieht in der Liste der Meldungsbearbeiter, dass fast alle Meldungen der Gruppe vom gleichen Kollegen bearbeitet wurden, der offenbar über Expertise in dem Themenbereich verfügt. Er markiert diesen Meldungsbearbeiter und erkennt in der Themenkarte, dass dieser noch weitere Meldungen in der (thematischen) Nähe der Suchergebnisse bearbeitet hat. Das Öffnen einer dieser Meldungen liefert ihm Hinweise zur Lösung des Kundenproblems.

In einem Brainstorming konnten drei grundsätzlich viel versprechende Unterstützungsmöglichkeiten für Rechercheaufgaben identifiziert werden:

- Kunden recherchieren Hinweise im Web-basierten Kundenportal
- Meldungsbearbeiter recherchieren Hinweise
- Meldungsbearbeiter recherchieren Meldungen

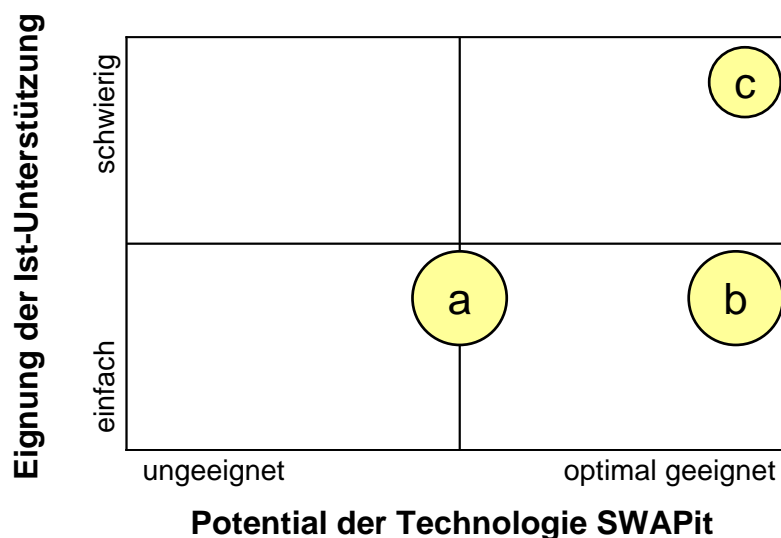


Abbildung 4.6: Aufgabenportfolio für die Recherche im Produktsupport (große Aufgabenblasen deuten auf werthaltige Aufgaben hin; Buchstaben beziehen sich auf Aufgabennummerierung des vorangehenden Abschnitts)

Das Aufgabenportfolio (Abbildung 4.6) basiert auf folgenden Überlegungen:

- Die Kundenrecherche nach Hinweisen ist werthaltig, weil gute Rechercheergebnisse einen direkten Nutzen für Kunden und Unternehmen haben: Die Kundenzufriedenheit steigt und weitere Support-Kosten werden vermieden. Die Recherche wird durch die bestehende Suchmaschine recht gut unterstützt, weil Hinweisdokumente redaktionell bearbeitete Dokumente sind und gute Retrieval-Ergebnisse versprechen. Das Expertenwerkzeug SWAPit ist für den gelegentlichen Nutzer nur bedingt geeignet.
- ähnlich wie (a.), aber Meldungsbearbeiter sind Recherche-Experten, weshalb ein Werkzeug wie SWAPit besser geeignet ist als bei (a.)
- Wird in Hinweisdokumenten keine relevante Information gefunden, so ist die Suche in Meldungsdokumenten obligatorisch. Mit dem bestehenden Suchsystem ist die Meldungsrecherche allerdings oft mühsam und der Erfolg hängt u.a. von der Qualität der Anfrage und der Dokumente ab. SWAPit's Inhaltsanalyse unstrukturierter Textdokumente kann einen großen Beitrag zur thematischen Übersicht leisten und durch erweiterte Explorationsmöglichkeiten die Probleme von Stichwortsuchen kompensieren.

Das Systemhaus hat sich dazu entschieden, die Recherche von Meldungsdokumenten (c) mit SWAPit zu adressieren. Diese Aufgabe lässt sich als Aufgabensteckbrief (Kap.2.2) beschreiben:

| Funktionale und Informatorische Mächtigkeit | | | | | |
|---|---------------------------------|------------------|-------------------------|-------------------------|------------------|
| Ziel der Info.-Recherche | lernen | Info. verdichten | Info. auswählen | Qualität prüfen | Qualität sichern |
| Betrachtete Info.-Arten | Text-Dok. | Dok.-Attribute | Beschreibungskategorien | Relationale Information | |
| Info.-Quellen | intern | | extern | | |
| Relat. Größe d. Info.-Teilmenge | gesamtes Repository | große Teilmenge | | kleine Teilmenge | |
| Dynamik der Info.-Teilmenge | konstant | unterschiedlich | | ad-hoc | |
| Betrachtete Verknüpfungen | Dok.-Spezifikation | Dok.-Dok. | Dok.-Thema | Dok.-Kategorie | Dok.-Tupel |
| Existenz der Verknüpfungen | logisch | organisatorisch | | physisch | |
| Verwendung der strukt. Info. | irrelevant | prüfen | verwenden | Dok. zuweisen | |
| Vokabular | allgemein | Fachsprache | | Korrespondenz | |
| Spektrum der Inhalte | allgemein | | domänenspezifisch | | |
| Relevanz im Unternehmen | Kernprozess | | Unterstützungsprozess | | |
| Absolute Bedeutung | „nice to have“ | wichtig | | kritisch | |
| Intendierte Wertsteigerung | Steigerung der Effizienz | | neue Aufgabe | | |
| Primäres Qualitätskrit. | Genauigkeit und Vollständigkeit | | grober Überblick | | |
| Qualität und Wert des Service | | | | | |

| Kontrolle der Informationsrecherche | | | |
|-------------------------------------|---------------------------|------------------------------|-----------------------------|
| Interaktionsmethode | zielgerichtete Suche | explorative Recherche | Stöbern u. Entdecken |
| Interaktionsstrategie | filtern | | ausschließen |
| Granularität | Übersicht | | Details |
| Dynamik des Interesses | statisch | | wechselnd |
| Flexibilität Interaktion | frei | angeleitet | Workflow |
| Kommunikationsmodus | wahrnehmen (phys. passiv) | | spezifizieren (phys. aktiv) |
| Erwartetes Antwortverhalten | ohne Verzögerung | kurze Verzögerung akzeptiert | lange Berechnung akzeptiert |
| Vorhersagbarkeit | gelegentlich | wiederholt | regelmäßig |
| Durchlaufzeit | Zeit ist irrelevant | Zeit ist wichtig | Zeit ist kritisch |
| Expertise in Domäne | Laie | | Domänen-Experte |
| Expertise in Korpus | Information unbekannt | | Information bekannt |
| Expertise in Recherche | gelegentl. Nutzer | Experte der Recherche | System-Experte |
| Aufwand der Arbeitsprozesse | | | |

4.3 Fallstudie 3: Aktivitätsanalyse in Kooperationsumgebungen

4.3.1 Hintergrundinformation

Domäne: Virtuelle Kooperationsumgebungen für eProfessionals

Zunehmende Globalisierung und die Mobilisierung von Arbeitsplätzen sind Treiber für die Entwicklung hochwertiger kooperativer Arbeitsumgebungen. Allerdings entwickeln sich professionelle virtuelle Kooperationsumgebungen schnell zu komplexen Netzwerken von Akteuren und Artefakten (z.B. Dokumente oder Blog-Beiträge).

Mit zunehmender Vernetzung steigt auch der Bedarf an innovativen Werkzeugen für den Überblick über Aktivitäten anderer Projektteilnehmer, wie die Veränderung von Wissensressourcen. Ebenso bedeutsam ist der kontextbezogene Zugriff auf die gemeinsamen Dokumente und Daten. So kann es wichtig sein, Experten für ein Thema zu identifizieren oder Änderungen an gemeinsam (asynchron) bearbeiteten Dokumenten zu erkennen.

Hintergrundinformation zum Projekt und Fallbeschreibung: Referenzarchitektur für Kooperationsumgebungen

Das EU-Projekt EcoSpace entwickelt eine produktunabhängige Architektur für kooperatives Arbeiten, die synchrone und asynchrone Kommunikation über Systemgrenzen hinaus zum Ziel hat, aber auch neue Technologien für eine eher Personen- und Aktivitäts- als Dokumenten- und Technikzentrierte Arbeitsunterstützung entwickelt und integriert. eProfessionals sollen sich weniger um Dokumentenformate, technische Kompatibilität und Zugriffsparadigmen von Einzellösungen kümmern müssen und auf eine technologisch durchgängige Systemunterstützung aufbauen können, welche auf die eigentlichen Arbeitsaufgaben abgestimmt ist.

Die web-basierte kooperative Arbeitsumgebung BSCW⁴⁴ wurde am Fraunhofer Institut für Angewandte Informationstechnik (FIT) entwickelt und wird vertrieben durch die Firma OrbiTeam GmbH. BSCW wurde und wird erfolgreich in zahlreichen Unternehmen und Projekten eingesetzt. BSCW verfügt über einige Mechanismen, mit denen sich Benutzer über Aktivitäten anderer Benutzer informieren können: So lassen sich Mail-Benachrichtigungen abonnieren und Symbole weisen auf Aktivitäten an Artefakten hin, z.B. Lese-Zugriffe oder die Modifikation von gemeinsamen Dokumenten. Zu jedem Artefakt lässt sich zudem die gesamte Aktivitätshistorie abfragen.

Die Koordination und Kooperation im Forschungsprojekt EcoSpace, an dem etwa 20 Partnerorganisationen aus Forschung und Industrie beteiligt sind, wird auf Basis der BSCW-Plattform durchgeführt. Bereits nach wenigen Monaten umfasste die Arbeitsumgebung über 100 gemeinsam bearbeitete Word-Dokumente sowie zahlreiche andere Dokumente. Diese Textsammlung, die kontinuierlicher Veränderung unterzogen ist, diente als Datenbasis für die Fallstudie.

4.3.2 Ist-Analyse

Herausforderung: Beschleunigung und Verbesserung asynchroner Kooperation

Kooperative Arbeitsprozesse, die so innovativ und komplex sind wie im Bereich Forschung & Entwicklung, erfordern eine gute Abstimmung der Beteiligten. Das gemeinsame Schreiben an Projektberichten oder Veröffentlichungen sind Beispiele. Durch die Beteiligung an mehreren Arbeitsgruppen und Projekten, oder durch

⁴⁴ Basic Support of Cooperative Work

das vorübergehende Ausscheiden aus einer virtuellen Arbeitsgemeinschaft, etwa bedingt durch Dienstreisen, wird der Überblick über das Projektgeschehen zur Herausforderung.

Aus Sicht eines Projektkoordinators ist ein guter Überblick über Aktivitäten des Konsortiums Pflicht. Aber auch jeder Projektteilnehmer hat die Aufgabe, im Rahmen seines Verantwortungs- und Aktivitätsbereichs informiert zu bleiben. Ziele einer guten Wahrnehmung von Kooperationsaktivitäten betreffen die Faktoren *Qualität* und *Zeit*, denn es geht u.a. darum,

- sich einen *schnellen und umfassenden Überblick* über den Stand des Projektes zu verschaffen, z.B. nach Reisetätigkeiten,
- *zeitliche Engpässe* oder *inhaltliche Defizite* bei der Bearbeitung von Projekteinreichungen frühzeitig zu erkennen,
- dem Auftraggeber *konsistente Einreichungen* abzuliefern und ein *konsistentes Gesamtbild* des Projektes zu vermitteln, also geeignete Querbezüge zwischen Kapiteln, Dokumenten, Arbeitspaketen etc. zu schaffen,
- bei überlappenden Tätigkeiten und Kompetenzen *Mehrfacharbeit* sowie *redundante Inhalte* zu vermeiden.

Eine stark auf Personen und Aktivitäten fokussierte Sichtweise, wie sie im EcoSpace-Projekt entwickelt wird, impliziert die Notwendigkeit einer guten Wahrnehmung des virtuellen Arbeitsumfelds. Über die konventionellen Benachrichtigungsmechanismen hinaus werden neue Werkzeuge zur Aktivitätsanalyse benötigt, die nicht nur auf letzte Änderungen hinweisen, sondern einen umfassenden Blick⁴⁵ auf die Wissensressourcen erlauben und flexible Zugriffspfade für verschiedenste, gerade auch inhaltsbezogene Fragestellungen bieten.

Prozessanalyse: Koordination von (Forschungs)projekten in virtuellen Kooperationsumgebungen

Die Projektbeteiligten müssen sich in regelmäßigen Abständen einen Überblick über den Stand des Projektes verschaffen. Werden relevante Dokumente hinzugefügt oder geändert, so müssen die Änderungen identifiziert und inhaltlich nachvollzogen werden. Dies kann Handlungsauslöser sein für einen Dialog mit Projektpartnern oder für editorielle Aktivitäten. Meist werden den Projektbeteiligten in der virtuellen Arbeitsumgebung individuelle Zugriffsrechte zugewiesen, die die Sicht jedes Partners auf die gemeinsame Arbeitsumgebung beeinflussen.

Das Projektmanagement koordiniert die Partner und sorgt für die Erfüllung des Vertrags gegenüber dem Auftraggeber, beispielsweise für die fristgerechte, inhaltlich vollständige, konsistente und qualitativ hochwertige Einreichung von Dokumenten. Daher ist ein guter Überblick über die Aktivitäten im Projekt Voraussetzung zur Projektsteuerung. Werden wichtige Dokumente nicht gelesen oder nicht bearbeitet, so kann dies zu Vertragsverletzungen führen. Das Projektmanagement orientiert sich am Projektplan und konzentriert sich daher jeweils auf bestimmte zu erledigende Aufgaben und Prozesse, auf Partner und Dokumente. Das Projektmanagement ist in besonderem Maße auf Werkzeuge zur umfangreichen Aktivitätsanalyse und zur flexiblen Recherche angewiesen. In einem zweistündigen *Workshop* mit 10 Mitarbeitern der Gruppe Kooperationsysteme (CSCW⁴⁶) um Prof. W. Prinz wurden wichtige Analyseaufgaben gesammelt und bewertet (nächstes Kap.).

⁴⁵ Eine erhöhte Transparenz der Aktivitäten aller Projektpartner ist in Bezug auf den Schutz von Daten und Privatsphäre potentiell kritisch. Abhilfe schaffen können Anonymisierungskonzepte (z.B. schrittweise dialogische De-Anonymisierung). Diese Thematik wird im Forschungsprojekt vertieft, nicht aber in dieser Arbeit.

⁴⁶ Computer Supported Cooperated Work

Systemunterstützung: BSCW und Werkzeuge zur Aktivitätsanalyse

Das Web-basierte BSCW-System [Ap99] bietet als kooperative Arbeitsumgebung vielfältige Funktionalitäten zum Informationsaustausch, darunter Dokumentenablage, Kalender, Adressbuch, Lesezeichen, Diskussionen, Umfragen, Blogs und Workflows. BSCW zeichnet alle Benutzeraktivitäten auf. Der Zugriff auf Artefakte durch den Benutzer kann auf zwei Arten erfolgen:

- Browsing in der Ordnerstruktur,
- Umfangreiche und flexible Suche in Inhalten und Metadaten.

Die Analyse von kooperativen Aktivitäten erfolgt in BSCW durch:

- Aktivitäts-Historie (Alle Aktivitäten zu einem ausgewählten Artefakt als Liste),
- Aktivitäts-Reports (Push-Service per e-Mail; alle Aktivitäten des letzten Tages),
- Awareness-Icons (Aktivitäten bis zur letzten Bestätigung durch den Benutzer).

Für das BSCW-System wurden zahlreiche Werkzeuge und Forschungsprototypen zur Aktivitätsanalyse entwickelt. In [PPS04] klassifizieren Pankoke-Babatz et al. die Beiträge zur Aktivitätsanalyse nach vier Zugriffsdimensionen. Sie unterscheiden Akteur, Artefakt, Aktivität und Zeit. Hier seien zwei der klassifizierten Spezialwerkzeuge (Abbildung 4.7) heraus gegriffen: Das System *Readers* [PR+06] verwendet einen Hypergraph zur verknüpften Anzeige von Akteuren und Artefakten. Mit *Readers* lässt sich beispielsweise leicht herausfinden, von welchen Personen ausgewählte Dokumente bearbeitet bzw. gelesen werden. Das Werkzeug *SmartMaps* [GWG03] verwendet eine Treemap⁴⁷-Metapher zur kompakten Anzeige von Aktivitäten in hierarchischen kooperativen Ablagestrukturen. *SmartMaps* unterstützen den Überblick über Aktivitäten bzgl. der Ablagestruktur.

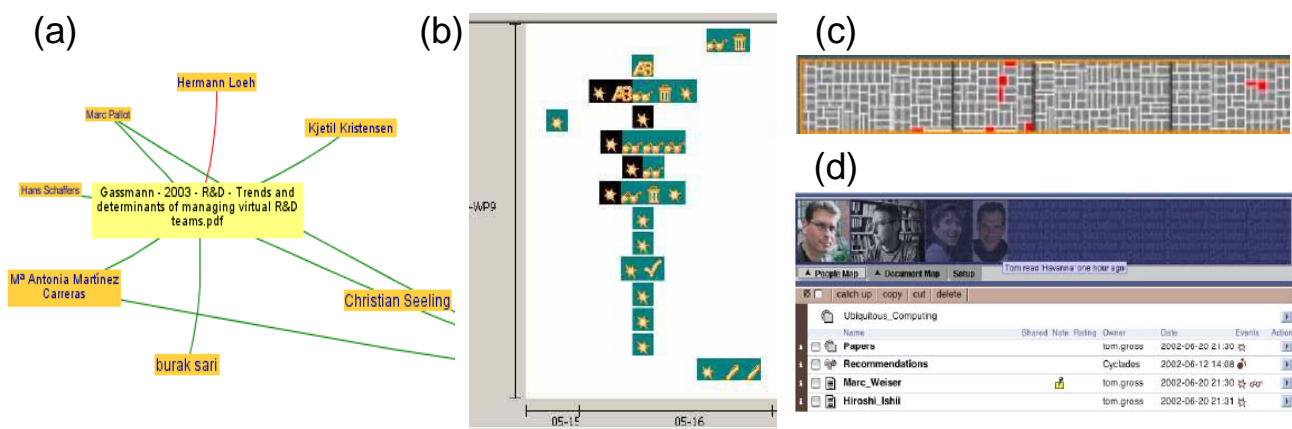


Abbildung 4.7: Einige Werkzeuge zur Aktivitätsanalyse in kooperativen Arbeitsumgebungen (u.a. aus [PPS04]): (a) *Readers*, (b) TOA (Time-Object-Activity) -Diagramme, (c) *SmartMaps*, (d) *AwarenessMaps*.

Datenanalyse: Artefakte, Personen, Aktivitäten, etc.

Der primäre Zweck des BSCW-Systems ist es, den gemeinsamen Zugriff auf Artefakte (z.B. Dokumente) zu ermöglichen. Die Artefakte sind in einer explizit definierten hierarchischen Struktur abgelegt und verfügen über umfangreiche Metadaten (z.B. Ersteller, MIME-Typ, Beschreibung). Mitglieder einer Arbeitsumgebung können zu Akteuren werden, wenn sie Aktivitäten an Artefakten ausführen, z.B. Lesen, Modifizieren, Erstellen, Sperren, Löschen. Solche Aktivitäten werden mit Typ und Zeit in einem Aktivitäts-Log abgelegt.

⁴⁷ <http://www.cs.umd.edu/hcil/treemap/>

Tabelle 4-5: Wichtige Daten bei der Analyse von Aktivitäten in virtuellen kooperativen Arbeitsumgebungen

| | Text | Kategorien | Tupel | Bezug zu den Texten |
|----------------------------|------|------------|-------|---------------------|
| Artefakt (z.B. Dok., Blog) | • | | | 1:1 |
| Metadatum (zu Artefakt) | | | • | 1:1 |
| Ablagestruktur | | • (hier.) | | m:1 |
| Akteur | | | • | m:n |
| Aktivität, Zeit | | | • | m:n |

Für die Benutzerstudie (Kap.7.5) wurden Dokumente mit rein administrativem Inhalt (z.B. Progress Reports) ausgeschlossen, weil es um die Analyse der wissenschaftlichen Themen des Projekts ging. Die verbleibenden 89 Word-Dokumente und die damit verknüpften strukturierten Daten sind in Tabelle 4-6 beschrieben.

Tabelle 4-6: Statistische Charakterisierung der Daten (Zeitraum: 5/2006 bis einschl. 1/2007)

| | | |
|----------------------------|---|----------|
| Dokumente | Sprache | Englisch |
| | Anzahl Dokumente | 89 |
| | Index-Terme (Wortstamm) | 3321 |
| | Mittl. Anz. Index-Terme pro Dok. | 37 |
| Strukturierte Daten | Aktivitäten an den Dokumenten | 1324 |
| | Lese-Aktivitäten | 1074 |
| | Schreib-Aktivitäten | 161 |
| | andere Aktivitäten | 89 |
| | Akteure | 65 |

4.3.3 Potential- und Aufgabenanalyse

Defizite der Ist-Situation: Eingesetzte Werkzeuge unterstützen nur Teilaspekte, insbesondere keinen *inhaltsbasierten* Überblick über Artefakte und Aktivitäten.

Beim Browsen im BSCW-System ist der Benutzer gezwungen, entlang der vordefinierten Ordnerstrukturen zu navigieren, die eine allgemeine (aufgabenunabhängige), statische Perspektive repräsentieren und für manche Fragestellungen inadäquat sind. Die anfrageorientierte Suchfunktion des BSCW-Systems setzt klar formulierbare Fragestellungen voraus. Aktivitätshistorie, Aktivitäts-Reports und Awareness-Icons stellen Aktivitäten immer in Bezug zu Artefakten dar. Somit werden nur Recherchen direkt unterstützt, deren Ausgangspunkt ein oder mehrere bekannte Artefakte sind. Deutlich aufwändiger ist hingegen beispielsweise die Recherche nach den Dokumenten oder Themen, mit denen sich ein Akteur beschäftigt hat. Neben der Ordnerstruktur gibt es keine übersichtliche Darstellung der Objekte und Mitglieder eines Workspace. Zeit- oder Personenbezogene Analysen, wie die Suche nach Experten oder Aktivitätsanalysen innerhalb eines Arbeitspaketes, sind unzureichend unterstützt. Neben der Suchfunktion gibt es keine Möglichkeit der Inhaltsanalyse von Artefakten. Die im letzten Abschnitt vorgestellten grafischen Werkzeuge zur Aktivitätsanalyse unterstützen diese Aufgaben teilweise, allerdings:

- Sie adressieren als Speziallösung jeweils eine bestimmte Aufgabe und unterstützen eine bestimmte Perspektive (meist durch Kombination von 2 bis 3 Zugriffsdimensionen)
- Sie bieten keinen *inhaltsbasierten* Zugriff auf Artefakte (komplementär zu den vorgegebenen Wissensstrukturen)

Chancen der Technologie: Inhaltsüberblick und multiperspektivischer Informationszugriff.

Novak et al. [NW05] fokussieren mit ihrem MultiView-System⁴⁸ *KnowledgeExplorer* (Kap. 3.3.2) die *inhaltliche Analyse* von Artefakten in Kooperationsumgebungen. Die wichtigsten koordinierten Sichten sind Dokumentenkarte und Konzeptkarte (Codebook), die sich wechselseitig beeinflussen. Der *KnowledgeExplorer* bietet interaktive Werkzeuge zur Modifikation der vordefinierten Strukturen von Dokumenten und Konzepten. Aus den Modifikationen werden individuelle wiederanwendbare Perspektiven auf Informationsbestände berechnet, welche die Sichtweisen von Individuen oder Communities repräsentieren.

Wie beim KnowledgeExplorer stellt die Inhaltsanalyse von Textdokumenten eine Stärke von SWAPit dar. Die übersichtliche und einprägsame Darstellung erleichtert die Arbeit mit großen Mengen komplexer und inhaltlich zusammenhängender Artefakte. Zusätzlich zum inhaltsbasierten Zugriff lassen sich alle Zugriffsdimensionen aus [PPS04] auf natürliche und intuitive Weise mit SWAPit's Sichtenkonzept zur Navigation nutzbar machen. Eine Aktivität besteht aus Akteur, Artefakt, Zeit und Aktivitätstyp. In SWAPit bietet sich an, diese Aktivitäten zur *interaktiven Verknüpfung* der Sichten zu nutzen. Dadurch unterstützt das System zugleich die Inhalts- und die Aktivitätsanalyse und wird zu einem mächtigen Recherchewerkzeug.

Potentiale: Ergänzung der ordnerbasierten durch inhaltsbasierte Exploration von Artefakten, Überblick über Artefakte, Akteure und Aktivitäten, sowie Unterstützung umfassender Aktivitäts- und Inhaltsanalysen

SWAPit's Berechnung und kompakte grafische Darstellung von inhaltlichen Querbezügen zwischen Textdokumenten eröffnet für eine Kooperationsumgebung wie den BSCW neue Recherchemöglichkeiten. Die kompakte Darstellung aller verfügbaren Informationen zeigt Projektmanagern, Projektadministratoren und Mitarbeitern „auf einen Blick“ die Situation des Projekts. Der Einsatz von SWAPit als Analysewerkzeug in kooperativen Arbeitsumgebungen bietet folgende Potentiale:

- Übersichtliche, intuitive und einprägsame Darstellung relevanter Informationen und Zusammenhänge,
- Inhaltserschließung und übersichtliche Darstellung inhaltlicher Querbezüge aller Artefakte,
- Ergänzung der Anzeige individueller Neuigkeiten um die Darstellung beliebiger Ausschnitte der gesamten Aktivitätshistorie,
- Ergänzung der ordnerbasierten Exploration um inhalts- u. themenbasierte Navigation,
- Kombination von explorativen inhalts- und aktivitätsbezogenen Analysen.

In einem zweistündigen *Workshop* mit 10 Mitarbeitern der Gruppe Kooperationssysteme (CSCW⁴⁹) um Prof. Prinz wurde am 23.11.2006 am Fraunhofer-Institut für Angewandte Informationssysteme (FIT) eine Aufgaben- und Potentialanalyse durchgeführt. Nach einer Präsentation der technischen Möglichkeiten konnten in einem Brainstorming zahlreiche Aufgaben in kooperativen Forschungsprojekten identifiziert werden.

Projektmanagement:

- a. Erkennen von inhaltlichen Communities (gemeinsame Themen, ggf. orthogonal zu den explizit vordefinierten Ablagestrukturen)
- b. Analyse der Expertise (welche Themen werden vom Projekt faktisch vertreten)
- c. Analyse von Defiziten in der Kooperation (z.B. warum und von wem werden Dokumente nicht gelesen?)
- d. Identifikation von Querschnittsthemen zwischen Forschungsprojekten
- e. Zeitbezogener Überblick/Navigation über die Aktivitäten von Partnern an Kerndokumenten

⁴⁸ Bei der Entwicklung des Sichtenkonzeptes referenzieren die Autoren das Konzept von SWAPit.

⁴⁹ Computer Supported Cooperated Work

Projektmitarbeit:

- f. Schnelles Aufspüren von Dokumenten über bekannte Metadaten, z.B. Schlagworte, Personen, Zeit (ohne Umweg über Ordnerhierarchie)
- g. Hilfe bei der Einordnung neuer Dokumente (Vorschlagskomponente um zu verhindern, dass Dokumente gar nicht oder falsch in die Ablagestruktur eingeordnet werden)
- h. Zeitbezogener Überblick/Navigation über Aktivitäten in bestimmten Ordnern

Die Bewertung der Aufgaben in einem Aufgabenportfolio (Kap.2.3) ist in Abbildung 4.8 dargestellt. Für die Benutzerstudie (Kap.7.5) wurde auf die Aufgaben (a), (b), (e), (f), (h) fokussiert. Diese Aufgaben kommen häufig vor, sind werthaltig, sind relevant für Projektmanagement und Projektmitarbeit, werden durch die vorhandenen Technologien noch nicht ausreichend unterstützt und lassen sich, geeignete Anpassungen vorausgesetzt, durch SWAPit gut unterstützen.

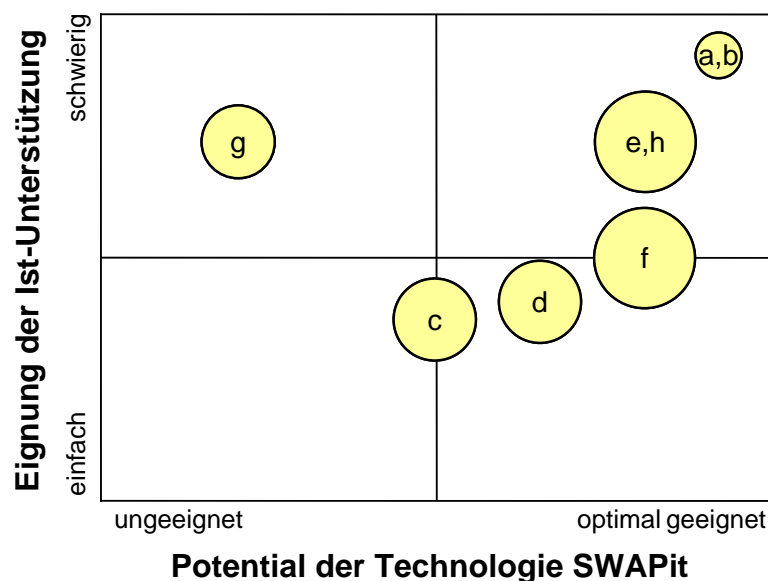


Abbildung 4.8: Aufgabenportfolio für Aufgaben in kooperativen Forschungsprojekten. Große Aufgabenblasen deuten auf werthaltige Aufgaben hin (Buchstaben beziehen sich auf die Aufgabennummerierung oben)

Die in der Benutzerstudie gestellten Aufgaben lassen sich folgendermaßen umreißen:

1. Thematischer Überblick
 - Kernthemen pro Arbeitspaket (Workpackage, WP)
 - Arbeitspakete-übergreifende Themen
2. Personen
 - Experten u. Interessenten zu Themen
 - Interessensgebiete einer Person
 - Personen mit gemeinsamen Interessen („Communities of Interest“)
3. Suchergebnisse und Aktivitätsanalyse
 - Dokumente und WPs zu Thema
 - Aktivitäten (Personen, Zeit) zu Thema

Wer hat viele Dokumente zum Thema gelesen? Wer hat aktiv am Thema gearbeitet?
Was ist neu? Wann gab es viele Aktivitäten an Dokumenten zum Thema?

Diese Gruppe von Aufgaben lässt sich als Aufgabensteckbrief (Kap.2.2) beschreiben:

| Funktionale und Informatorische Mächtigkeit | | | | | |
|---|---------------------|------------------|-------------------------|-------------------------|------------------|
| Ziel der Info.-Recherche | lernen | Info. verdichten | Info. auswählen | Qualität prüfen | Qualität sichern |
| Betrachtete Info.-Arten | Text-Dok. | Dok.-Attribute | Beschreibungskategorien | Relationale Information | |
| Info.-Quellen | intern | | extern | | |
| Relat. Größe d. Info.-Teilmenge | gesamtes Repository | große Teilmenge | | kleine Teilmenge | |
| Dynamik der Info.-Teilmenge | konstant | unterschiedlich | | ad-hoc | |
| Betrachtete Verknüpfungen | Dok.-Spezifikation | Dok.-Dok. | Dok.-Thema | Dok.-Kategorie | Dok.-Tupel |
| Existenz der Verknüpfungen | logisch | organisatorisch | | physisch | |
| Verwendung der strukt. Info. | irrelevant | prüfen | verwenden | Dok. zuweisen | |
| Vokabular | allgemein | Fachsprache | | Korrespondenz | |
| Spektrum der Inhalte | allgemein | | domänenspezifisch | | |

| Kontrolle der Informationsrecherche | | | |
|-------------------------------------|---------------------------|------------------------------|-----------------------------|
| Interaktionsmethode | zielgerichtete Suche | explorative Recherche | Stöbern u. Entdecken |
| Interaktionsstrategie | filtern | | ausschließen |
| Granularität | Übersicht | | Details |
| Dynamik des Interesses | statisch | | wechselnd |
| Flexibilität Interaktion | frei | angeleitet | Workflow |
| Kommunikationsmodus | wahrnehmen (phys. passiv) | | spezifizieren (phys. aktiv) |
| Erwartetes Antwortverhalten | ohne Verzögerung | kurze Verzögerung akzeptiert | lange Berechnung akzeptiert |

| | | | |
|-----------------------------------|---------------------------------|-----------------------|------------------|
| Relevanz im Unternehmen | Kernprozess | Unterstützungsprozess | |
| Absolute Bedeutung | „nice to have“ | wichtig | kritisch |
| Intendierte Wertsteigerung | Steigerung der Effizienz | | neue Aufgabe |
| Primäres Qualitätskrit. | Genauigkeit und Vollständigkeit | | grober Überblick |
| Qualität und Wert des Service | | | |

| | | | |
|-------------------------------|-----------------------|-----------------------|---------------------|
| Vorhersagbarkeit | gelegentlich | wiederholt | regelmäßig |
| Durchlaufzeit | Zeit ist irrelevant | Zeit ist wichtig | Zeit ist kritisch |
| Expertise in Domäne | Laie | | Domänen-Experte |
| Expertise in Korpus | Information unbekannt | | Information bekannt |
| Expertise in Recherche | gelegentl. Nutzer | Experte der Recherche | System-Experte |
| Aufwand der Arbeitsprozesse | | | |

4.4 Informationsarten und Verknüpfungen

Im Folgenden werden die Informationsarten und ihre Verknüpfungen systematisiert und verglichen, die in den drei Anwendungsbeispielen (und vielen weiteren, s. Anhang 9.3), von Bedeutung sind.

4.4.1 Analyse der Informationsarten und Verknüpfungen

Informationen lassen sich nach ihrem Strukturierungsgrad unterscheiden. Mertens et al. [MG02] unterscheiden quantitative von qualitativer Information aus Textdokumenten. Valente et al. [Va01] entwickeln eine feinere Unterscheidung von Informationsarten nach dem Grad der Strukturiertheit, welche der Bewertung und dem Vergleich von Wissensmanagement-Lösungen dient. Als völlig unstrukturiert wird das implizite Wissen in den Köpfen von Mitarbeitern bezeichnet, während als völlig strukturiert mathematisch-logisch formalisiertes Wissen gilt. Dazwischen angeordnet sind, von unstrukturiert nach strukturiert, Rohtext, gegliederter Text (z.B. HTML), strukturierter Text (z.B. XML), strukturierte Daten in Datenbanken und kategorisierte Information (Taxonomien). Diese Informationsarten – Textinformation, Datenbankdaten und Kategorien – sind auch in den Fallstudien zur Textrecherche von Bedeutung, wie aus Tabelle 4-7 ersichtlich wird.

Tabelle 4-7: Informationen und Informationsarten der Fallstudien.

| | Textdokumente | Beschreibungs-kategorien | Relationale Daten |
|-----------------------|--|---|--|
| Beschwerde-management | Kundenbeschwerden, Korrespondenz | Typ, Ergebnis, Sparte, Vertragsphase, Kundenklasse, Beschwerdegrund | Metadaten, Kundenstamm- u. Vertragsdaten |
| Produktsupport | Kundenanfragen, Korrespondenz | Priorität, Status, Problem, Systemkomponente | Datum, Bearbeiter |
| Aktivitätsanalyse | Kooperativ bearbeitete Dokumente (Artefakte) | Ablagestruktur | Metadaten, Akteur |

Die Informationen werden über explizit definierte oder automatisch berechnete Assoziationen verknüpft. Diese Verknüpfungen können zur Navigation in Informationsräumen genutzt werden (Kap.5.6), oder zur Gruppierung und kompakten Darstellung von Informationsobjekten. In dieser Arbeit werden Verknüpfungen von Informationsobjekten *der gleichen Informationsart* und Verknüpfungen von Objekten *verschiedener Informationsarten* (zwischen Texten und strukturierten Daten) unterschieden (Kap.5.6). Häufig vorkommende Arten von Verknüpfungen sind in Tabelle 4-8 dargestellt. *Zwischen* den Informationsarten werden nur strukturelle Verknüpfungen betrachtet, die untypisiert, ungerichtet und ungewichtet sind.

Tabelle 4-8: Verknüpfungen von Informationsobjekten gleicher bzw. verschiedener Informationsarten

| | Textdokumente | Beschreibungs-kategorien | Relationale Daten |
|--------------------------|---|---|--|
| Textdokumente | definierte Verlinkung oder berechnete Textähnlichkeit | definierte oder berechnete Klassifikation | definierte oder berechnete Relation |
| Beschreibungs-kategorien | | definierte Struktur oder berechnete Nachbarschaft | |
| Relationale Daten | | | definierte Relation oder berechnete multidimensionale Wertebeziehung |

Die Verknüpfungen der *gleichen Informationsart* können genutzt werden (Kap.5.6) zur

- *Assoziativen Navigation*, z.B. zwischen ähnlichen Texten, zwischen benachbarten Beschreibungskategorien, zwischen benachbarten sortierten relationalen Datentupeln,
- *Abstraktion*, z.B. Textgruppe als Thema, Oberkategorie zu Kategorien, Aggregierte Werte für Gruppen relationaler Tupel

Die Verknüpfungen *verschiedener Informationsarten* können genutzt werden (Kap.5.6) zur

- *Spezifikation von Texten*, z.B. durch Auswahl von Kategorien und relationalen Daten
- *Beschreibung von Texten*, z.B. durch Betrachtung verknüpfter Kategorien u. relationaler Daten

Liegen wichtige Verknüpfungen nicht vor, müssen sie entweder explizit definiert oder berechnet werden. In Kap.6.7 werden einige Werkzeuge für die automatische und semi-automatische Verknüpfung von Texten und strukturierten Daten mittels Regulärer Ausdrücke und Boolescher Suche vorgestellt.

4.4.2 Texte und Textähnlichkeiten

„Ausser unter speziellen Bedingungen ist es schwierig für einen Benutzer, ein Retrievalsystem genau das zu fragen, was er oder sie wissen möchte, denn normalerweise ist weder bekannt, was verfügbar ist, noch, wovon es differenziert werden muss“ ([St95], S.39). Ein Potential automatisch berechneter Ähnlichkeitsdarstellungen von Texten besteht darin, die wesentlichen Abgrenzungsmerkmale von Texten der gesamten Doku-

mentensammlung zu vermitteln. Von diesen Themenstrukturen kann sich ein Anwender bei der interaktiven Suche oder Erkundung leiten lassen. Dazu stellt der Literaturwissenschaftler und Vordenker L.B. Doyle in seinem frühen Artikel über Assoziative Karten fest: „*The decision as to what is ‚just the right amount of information‘ depends on the need for discrimination of a document from topically close documents, and this decision is perhaps best made by a computer, which is the only agency capable of having ‚full consciousness‘ of the contents of a library. [...] Proxies can be designed to conspicuously indicate the things about a document which make it different from its closest topical relatives*“ [Do61].

Textähnlichkeiten können zur assoziativen Suche eingesetzt werden, um ausgehend von einem Einstiegsdokument inhaltsbasiert durch die Dokumentensammlung zu navigieren. Dadurch lassen sich insbesondere dynamische oder vage Informationsbedürfnisse adressieren (vgl. Harter [Ha92]). Außerdem helfen Textähnlichkeiten, einen Überblick über den Themenraum der Textsammlung zu gewinnen, was u.a. für die thematische Einordnung identifizierter Texte hilfreich ist.

Allerdings hängt die Nützlichkeit solcher Ähnlichkeitsdarstellungen stark von der *Aufgabenangemessenheit* des Ähnlichkeitsmaßes ab ([Be01], S.67-70). Ein spezialisierter Ähnlichkeitsbegriff soll einerseits gerade die für eine Aufgabe *entscheidenden* inhaltlichen Konzepte der Texte miteinander vergleichen und ist daher bestrebt, nicht zu viele Merkmalsdimensionen in den Vergleich einzubeziehen – schließlich werden die Ähnlichkeiten meist auf wenige Abbildungsdimensionen reduziert. Andererseits sollen aber auch gerade bislang *unbekannte* Querbezüge aufgezeigt werden. Somit geht es um die Balance zwischen Unerwartetem und Erwartetem: „*Literature searchers find and value both the unexpected and the expected. When something unexpected is found, one thereby obtains information; when something expected is found, one obtains confirmation. However, when one formulates a [keyword] search, the unexpected is hardly ever involved. [...] chances that some of these [displayed associations] will ‚ring a bell‘. Note that, instead of depending on his imagination to think up a search request, he [the searcher] is depending on his recognition of semantic relationships. We have put him in close touch with the contents of the library, qualitatively and quantitatively.*“ [Do61]. Eine Stärke der Assoziativen Exploration mittels Ähnlichkeitsdarstellungen als Ergänzung der anfrageorientierten Suche besteht in der Unterstützung des Entdeckens nicht antizipierter Information. Das Generieren von Ideen oder das Anregen neuer Suchstrategien sind Beispiele. Diesen Effekt des gewünschten aber letztlich zufallsbedingten Entdeckens nützlicher Information bezeichnet man als das Serendipity-Prinzip⁵⁰.

4.4.3 Beschreibungskategorien

Beschreibungskategorien für Texte sind meist hierarchisch organisiert oder werden hierarchisch dargestellt⁵¹ (vgl. Klement et al. [KHA98]). Unbestritten sind Analogien zum Organisationsprinzip menschlichen Denkens einerseits und zur Organisation der Welt andererseits. Aristoteliker⁵² betrachten „*Hierarchien als etwas, das die Struktur der Welt selbst widerspiegeln kann*“ [Sm02]. Zahlreiche Experimente legen nahe, dass das Orientierungswissen im Gehirn hierarchisch abgelegt ist, vermutlich zum Zweck der effizienten Speicherung und des effizienten Abrufs der Information ([Mo97], S.6). Neben diesen Analogien mag ein wesentlicher Grund für die grundsätzliche Beliebtheit von Hierarchien auch der Grad der Formalisierung sein, der dem Prinzip „Formalisierung gleich einfache Strukturierung“ folgt und somit einen guten Kompromiss zwischen detaillierten mathematisch-logischen Beschreibungsmodellen wie Ontologien einerseits und flacher bzw.

⁵⁰ <http://de.wikipedia.org/wiki/Serendipity>: Das Serendipity-Prinzip ist auch zur Bezeichnung zufallsgesteuerter Erfindungen wie der des Schwarzpulvers oder der Röntgenstrahlung gebräuchlich.

⁵¹ Jeder Graph kann hierarchisch dargestellt werden, indem nur ein Ausschnitt betrachtet wird, einige Kategorien redundant aufgeführt werden, oder beides.

⁵² Von Aristoteles wird der Begriff *Kategorie* nur für die Wurzel eines hierarchischen Schemas verwendet.

Nicht-Organisation von Objekten andererseits darstellt. Klement et al. stellen dazu fest: „*The primary design goal of hierarchical catalogue categorizations is to create a comprehensive, memorable and convenient semantic organization of information relevant to the user’s task and information needs*“ [KHA98].

Das hierarchische Organisationsprinzip ist als vertraute Alltagsmetapher auch auf den Informationszugriff übertragen worden. So belegen Studien, dass Hierarchien für die Informationsrecherche bevorzugt eingesetzt werden [He06]. Berühmte Beispiele sind die Kategorien von Yahoo!⁵³ und der WindowsTM File Explorer. Bei Hierarchien zur Textklassifikation unterscheidet man einfache Klassifikation und Multi-Klassifikation, wo ein Text mehreren Beschreibungskategorien zugeordnet werden kann.

Robertson problematisiert die üblichen Darstellungen von Hierarchien [Ro00]: Oft sind die Visualisierungen nicht den Daten angepasst (z.B. lineare Baumdarstellungen bei großer Hierarchie-Tiefe). Außerdem erfordern viele Aufgaben den Vergleich oder die Spezifikation mehrerer Aspekte. Daher untersucht Robertson die simultane Navigation in mehrere komplementären Hierarchien bzw. die gleichzeitige Fokussierung mehrerer Informationsobjekte („multiple focus“). Diese Überlegungen führen zu Polyhierarchien⁵⁴ bzw. zur Facettenklassifikation (auch analytisch-synthetische Klassifikation genannt) als flexiblere Beschreibungs- und Navigationsstrukturen. Facettierte Kategorien-Hierarchien (HFC nach [He06]) sind ein Spezialfall der Multiklassifikation und werden erfolgreich zur Navigation in Dokumentenbeständen eingesetzt. Ein Sachverhalt wird dabei immer unter mehreren Gesichtspunkten betrachtet. Für jeden Gesichtspunkt, Facette genannt, steht eine eigene Teilklassifikation zur Verfügung (Beispiel-Facetten für Kochrezepte: Art der Speise, Form der Zubereitung). Die Zusammenführung von Kategorien verschiedener Facetten erzeugt aus mehreren semantisch unabhängigen (allgemeinen) Begriffen der einzelnen Facetten einen speziellen Begriff. Daher sind interaktive Auswahlen in Facettierten Hierarchien geeignet, Benutzeranfragen zu repräsentieren. Marchionini [Ma06] setzt facettierten Hierarchien zur explorativen Suche ein, u.a. für Video-Datenbanken und Statistiken der Amerikanischen Regierung. Klamma et al. [KSJ05] verwenden solche Hierarchien, um Nutzerspezifische Klassifikationssichten auf Multimediadaten zu kombinieren. Das Werkzeug iTunes der Firma Apple zum Verwalten und Abspielen von Musikdateien bietet beispielsweise Genre, Interpret und Album als Facetten zum Browsen an. Das Werkzeug HIBROWSE [Po98] unterstützt durch facettenbasiertes Browsen u.a. die Recherche in medizinischen Krankheitsdaten. In [Po98] ist von komplementären Sichten die Rede. Bei Cody et al. [CK+02] werden Facetten, angelehnt an Multidimensionale Datenmodelle für Business Intelligence-Anwendungen, als Dimensionen bezeichnet und in das Multidimensionale Kennzahlenmodell integriert. Dadurch wird dem Anwender das kombinierte Browsen in Texten und Kennzahlen ermöglicht. Die Parallelen der Navigation über Facetten und der Navigation in OLAP-Datenwürfeln beobachten auch Sullivan [Su03] sowie Klement et al. [KHA98]. Letztere schlagen als Werkzeug den ICE-Browser vor.

Empfehlungen für den Entwurf von Facetten und ihre Verwendung, insbesondere für Navigation und Suche, formulieren W. Denton⁵⁵ sowie M. Yin⁵⁶. Die Ebenen konventioneller Hierarchien, die als Entscheidungspunkte bei der Navigation aufgefasst werden können, lassen sich, wenn die Hierarchie einheitlich gestaltet ist, meist unmittelbar als Facetten repräsentieren. Während Hierarchien die Reihenfolge der Ebenen vorgeben, sind Facetten unabhängig voneinander. Daher kann die optimale Auswahlreihenfolge, bei der frühzeitig möglichst viele Alternativen ausgeschlossen werden, in Abhängigkeit von der Fragestellung mathematisch bestimmt werden. Die Unabhängigkeit der Facetten und die damit verbundene Flexibilität kann aber auch

⁵³ <http://www.yahoo.com>

⁵⁴ Bei Polyhierarchien kann jede Kategorie mehr als eine Oberkategorie haben.

⁵⁵ Miskatonic University Press, <http://www.miskatonic.org/library/facet-web-howto.html>

⁵⁶ Open Source Applications Foundation (OSAF),

<http://wiki.osafoundation.org/bin/view/Journal/HierarchyVersusFacetsVersusTags>

gerade ein Nachteil bei der Bedienung sein, weil der Anwender viel mehr Freiheitsgrade bei der Wahl der Vorgehensweise hat und die optimale Vorgehensweise selbst erkennen muss.

In der dritten Fallstudie dieser Arbeit wird die Anforderung identifiziert, umfangreiche Tabellen mit Benutzeraktivitäten in kooperativen Arbeitsumgebungen kompakt darzustellen. Dazu wird prototypisch eine grafische Sicht entwickelt, die jede Tabellenspalte als eigene Facette repräsentiert (FacetTable, Kap.7.5.5).

4.4.4 Relationale Daten

In vielen Bereichen werden Informationen tabellarisch dargestellt, darunter Publikationen, Kontakte oder Produkte. Listen sind einspaltige Tabellen. Von besonderem Interesse sind hier Tabellen zur Darstellung von Informationen zu Texten. Ist jedem Text höchstens ein Datensatz (Tabellenzeile) zugeordnet, sprechen wir von Meta-Informationen des Dokuments. Dazu gehören neben bibliografischen Informationen wie Publikationsjahr oder Autoren auch inhaltliche Deskriptoren, die meist mit Hilfe standardisierter Attributsätze wie dem der Dublin Core Metadata Initiative (DCMI⁵⁷) oder den ACM Categories⁵⁸ spezifiziert werden. Solche Meta-Informationen können bei der Texterschließung helfen, da ein Anwender kontextuelle oder inhaltliche Merkmale der Texte für den Zugriff nutzen kann, z.B. zur Suche, Filterung oder Inhaltsbeschreibung von Texten. Die inhaltsbezogenen Meta-Informationen ergänzen den Textinhalt, da sie modelliert sind und den Texten explizit zugewiesen wurden. Somit können in den Meta-Informationen auch andere Informationen transportiert werden als in den Texten selbst.

Über solche Meta-Informationen hinaus lassen sich in Tabellen Datensätze darstellen, die nicht ursprünglich und unmittelbar der Dokumentenbeschreibung dienen, sondern Objekte beschreiben, die mit den Dokumenten zu tun haben. Solche Informationen können im Allgemeinen mehreren Dokumenten zugeordnet sein oder umgekehrt. Die meisten Unternehmen speichern heute ihre wichtigen Geschäftsdaten (z.B. zu Kunden, Produkten, Verträgen, Konkurrenten) in relationalen Datenbanken. Weil die Daten ohnehin in logischem Zusammenhang zu den Texten der Unternehmen stehen (z.B. Kundenbeschwerden, Fachpresse, Projektdokumentation, Patente), liegt eine Nutzbarmachung dieser relationalen Daten für die Textanalyse nahe.

Die meisten Systeme verwenden zur Detaildarstellung von Daten-Tupeln die Tabellen-Metapher, während vielfältige Arten von Diagrammen zur Zusammenfassung verwendet werden. Solche Diagramme machen meist von statistischen Lage- oder Streumaßen Gebrauch, aggregieren also die Daten. Einige Ansätze zur Exploration relationaler Daten favorisieren *tabellarische Übersichtsdarstellungen* in Form komprimierter Tabellen. Dabei werden Werte in kompakter grafischer Form dargestellt, identische Werte verschmolzen, und die Darstellung dem verfügbaren Platz angepasst. Beispiele solcher Werkzeuge sind FOCUS (später InfoZoom[®] [SBB96]), MatrixBrowser [BK02] und VisMeB [KR+03].

4.4.5 Informationsarten u. Verknüpfungen

In den vorangehenden Abschnitten wurden die Eigenschaften von Textähnlichkeiten, Beschreibungskategorien und Relationalen Daten für die Textanalyse diskutiert. Eine Gegenüberstellung der drei adressierten Informationsarten, insbesondere im Hinblick auf ihre komplementären Stärken zur interaktiven Navigation, ist Gegenstand von Tabelle 4-9. Der Vergleich orientiert sich an den Kriterien Information, Darstellung und Nutzung.

Information. Unabhängig von Darstellung und Zugriff ist festzustellen, dass natürlichsprachige Texte unstrukturierte, schwer formalisierbare, und in diesem Sinne besonders komplexe Objekte sind. Demgegenüber

⁵⁷ <http://dublincore.org/>

⁵⁸ http://www.jucs.org/jucs_info/acm_categories

repräsentieren Kategorien homogene Begriffe. Fakten-Tupel setzen sind aus definierten Feldern zusammen (Datenmodell) und beschreiben ebenfalls einzelne Objekte. Somit können Kategorien und Fakten als vereinfachte Zugriffspfade auf komplexe Textobjekte genutzt werden. Die Beschreibungskategorien nehmen aufgrund ihrer einfachen hierarchischen Struktur eine Sonderrolle ein, zumal die Beziehungen zwischen Kategorien modelliert und damit Bestandteil der Daten sind – im Gegensatz zu der Ähnlichkeit von Texten oder Fakten-Tupeln, deren Inhalts- bzw. Werte-Ähnlichkeiten extra zum Zweck der Darstellung und Interaktion berechnet werden. Während die feste hierarchische Struktur eines Kategoriensystems den Vorteil besitzt gut einprägsam zu sein, bieten die berechneten Ähnlichkeitsbeziehungen zwischen Texten einerseits und Fakten andererseits die Möglichkeit der dynamischen Anpassung durch den Anwender (z.B. flexible Sortierung) und somit das Potential der Optimierung für eine Analyseaufgabe.

Darstellung. Metaphern zur Informationsdarstellung beschreibt L.A. Wozny folgendermaßen: „*a user interface metaphor is the process of representing the computer system with objects and events from a noncomputer domain*“ (Wozny 1989 in Barr et al. [BBN02]). Barr et al. unterscheiden *orientationale* von *strukturellen* Metaphern. Die Darstellung von Kategorien als Hierarchie ist eine strukturelle Metapher, denn Anwender kennen das Ordnungsprinzip der Verfeinerung und Vergrößerung nicht nur aus dem Dateisystem oder aus Internetkatalogen, sondern auch beispielsweise aus Inhaltsverzeichnissen von Büchern. Die Raum-Metapher für Textähnlichkeiten ist *orientational*, da sie eine direkte Übertragung von inhaltlicher Ähnlichkeit auf räumliche Nähe liefert nach dem Prinzip „nah ist ähnlich“. Auch die Tabellenmetapher für Tupel ist *orientational*, da sie Wertebeziehungen durch Reihenfolgen von Tupeln repräsentiert, z.B. nach dem Prinzip „oben ist groß“.

Während Textinhalte und –Eigenschaften symbolisch repräsentiert werden, können selbst in den meisten Übersichtsdarstellungen alle Details zu Kategorien und Fakten angezeigt werden. Alle Darstellungen bieten Möglichkeiten der Gruppierung von Informationsobjekten, die ausgedrückt werden durch Nachbarschaft bzw. grafische Verbindung.

Nutzung. Die Assoziative Navigation (z.B. mit dem Ziel der Suche nach ähnlichen Informationsobjekten) bedient sich der dargestellten Beziehungen. Einen Überblick gewinnt der Anwender, indem er mit Hilfe der Darstellungen von Textsammlungen und Kategorienhierarchien Themen, deren Größe und Diversifizierung untersucht. Bei Relationalen Daten geht es um den Überblick über statistische Muster und Abhängigkeiten von Tupeln.

Das Prinzip der Zusammenfassung einer Menge von Informationsobjekten ist bei Kategorienhierarchien zugleich Prinzip der Strukturierung selbst (Subsumption), denn zu einer Menge von Kategorien lassen sich gemeinsame Oberkategorien identifizieren. Textgruppen lassen sich thematisch gruppieren durch Nähe im Themenraum. Tupelmengen lassen sich über gemeinsame Attributwerte zusammenfassen.

Ein bewertender Vergleich der Informationsarten wird in Tabelle 4-10 vorgenommen. Zusammenfassend lässt sich feststellen, dass die Informationsarten, die jeweils für spezielle Informationsarten und Anwendungszwecke konzipiert wurden und in unterschiedlichen Forschungsbereichen entstanden sind, zahlreiche komplementäre Stärken und Schwächen aufweisen. Daher ist die Kombination solch komplementärer Navigations- und Zugriffspfade auf Textsammlungen viel versprechend.

Tabelle 4-9: Gegenüberstellung der Informationsarten (Textähnlichkeiten, Beschreibungskat., Relationale Daten)

| | | Textähnlichkeiten | Beschreibungs-kategorien | Relationale Daten |
|-------------|-------------------------------|---|---|---|
| Information | Atomare Objekte | Texte | Kategorien | Relationale Tupel |
| | Objektkomplexität | hoch (natürlichsprachliche Texte) | niedrig (Bezeichner; Schema-Charakter) | mittel (Werte numerisch oder alphanumerisch) |
| | Objektstruktur | unstrukturiert (unmodelliert) | strukturiert (modelliert) | strukturiert (modelliert) |
| | Objektgruppierung | bzgl. Ähnlichkeitsmaß auf Textinhalt | bzgl. expliziter Relation über Kategorien | bzgl. Ähnlichkeitsmaß auf Attributen |
| | Beziehungstyp | Ähnlichkeitsfunktion | Relation | Ähnlichkeitsfunktion |
| | Beziehungsherkunft | berechnet | modelliert (Bestandteil der Daten) | berechnet |
| | Dynamik der Bez. | für Darstellung u. Navigation berechnet; ggf. interaktive Anpassung | statisch (gegeben) | für Darstellung u. Navigation berechnet; ggf. interaktive Anpassung |
| Darstellung | Metapher (Typ ⁵⁹) | orientational | strukturell | orientational |
| | Metapher (Bsp.) | Karte, Raum | Katalog, Treemap | Tabelle, Liste |
| | Abbildungs-Dim. | geometrisch (2-D,3-D), grafisch (n-D) | geometrisch (2-D,3-D), grafisch (n-D) | geometrisch (1-D) |
| | Objektdetails | Einige (Texte symbolisiert) | Alle (Namen) | Alle (Werte) |
| | Objektgruppen | Benachbarte Texte | Verbundene Kategorien | Benachbarte Tupel |
| Nutzung | Assoz. Navigation | Benachbarte Texte | Benachbarte Kategorien (nur lokal: paarweise schwierig) | Auswahl Attributwerte für Reihenfolge (Sortierung) |
| | Überblick | Themen, Themengröße, Themendiversifizierung | Themen, Themengröße, Themendiversifizierung | Statistische Muster und Zusammenhänge |
| | Zusammenfassung | Themenraum | Oberkategorien | Werte-Aggregation |

Tabelle 4-10: Bewertung von Stärken und Schwächen der drei Sichtenarten (Stärke +, Schwäche -, neutral 0).

| | Textähnlichkeiten | Kategorien | Tupel |
|--------------------------------------|-------------------|------------|-------|
| Strukturen beeinflussen u. verstehen | - | + | + |
| Unbekannte Zusammenhänge entdecken | + | - | + |
| Assoziativ Suchen | + | 0 | + |
| Verfeinerung erkennen | 0 | + | - |
| Gruppen erkennen | 0 | + | + |
| Gruppen zusammenfassen (Labels) | 0 | + | + |
| Exakte Wertigkeit von Bez. erkennen | - | + | + |
| Alle paarweisen Beziehungen erkennen | + | 0 | - |
| Globale Topologie erkennen | + | - | - |
| Strukturen einprägen | + | + | - |

⁵⁹ nach Barr et al. [BBN02]

5 Systemkonzept: Integrative Textrecherche mit Multiple Views

Das Systemkonzept beschreibt Informations- und Interaktions-Design („HCI-Architektur“) einer neuartigen Systemunterstützung, die auf koordinierten grafischen Sichten (Kap.1.2) basiert und die integrierte Analyse von Textdokumenten und strukturierten Daten unterstützt. Das Konzept definiert eine Klasse explorativer Recherchesysteme (SWAP-Systeme). SWAP (engl. für „wechseln“) versteht sich als Aufforderung an den Benutzer, für die Recherche in Textsammlungen multiple Perspektiven (Sichten) einzunehmen bzw. zu kombinieren.

Das Konzept berücksichtigt drei Arten von Informationsobjekten und zeigt auf, inwiefern diese sich für die Textanalyse ergänzen können. Die Definition der SWAP-Systemklasse umfasst Systemkomponenten, Systemverhalten und interaktive Werkzeuge. So lassen sich beispielsweise Interaktionsmuster beschreiben oder die Konsistenz von Auswahlen in verschiedenen verknüpften Sichten überprüfen. Außerdem hilft das Konzept bei der systematischen Realisierung konkreter SWAP-Systeme. Die explorativen Möglichkeiten solcher Systeme, sowie die Abgrenzung gegen Analysekonzepte für strukturierte Daten, werden ebenfalls diskutiert.

5.1 Rationale: Komplementäre Sichtenarten

In Kap.4.4 wurden drei Arten von Informationsobjekten und Querbezüge dazwischen unterschieden, deren Berücksichtigung bei der Textanalyse in der Praxis wichtig und typisch ist: Neben den Textdokumenten selbst sollten die inhaltlichen Ähnlichkeiten der Texte, Beschreibungskategorien zu den Texten, sowie mit den Texten verknüpfte Relationale Daten (Metadaten als Spezialfall) für die Recherche verfügbar gemacht werden. Damit Experten in die Lage versetzt werden, interaktive Anfragen über diesen komplementären Informationsarten zu spezifizieren, müssen die vorhandenen logischen Verknüpfungen zwischen den verschiedenartigen Informationen für die Navigation nutzbar gemacht werden: Jede Informationsart soll als Zugriffspfad auf die Texte genutzt werden können. Solche interaktiven Anfragen verstehen sich als Ergänzung herkömmlicher schlagwortbasierter Suchmöglichkeiten.

Da bei dispositiven Aufgaben der Lösungsweg a priori meist allenfalls vage bekannt sind, schlägt Holten [Ho97] gerade hierfür explorativ orientierte Interaktionsstrategien vor, die sich durch eine hohe Flexibilität der Navigation sowie kurze Interaktionszyklen auszeichnen. Diese Faktoren sind wichtig, damit Experten ihre Erfahrung und Intuition bei der Textanalyse einbringen können und stets in der Lage sind, ihre Recherchestrategie anzupassen, welche von den untersuchten Informationen selbst und dem veränderlichen Erkenntnisstand des Experten abhängt. Auch OLAP-Systeme zur Kennzahlenanalyse zeichnen sich durch ein hohes Maß an Flexibilität bei der Navigation aus.

Der für eine Aufgabe relevante Informationskontext der Texte sollte für den Analysten integriert dargestellt werden. Das SWAP-Konzept führt dazu drei simultan angezeigte Sichten ein, welche die drei Informationsarten in grafischen Übersichtsdarstellungen interaktiv zugreifbar machen. Jede Sicht ist auf eine Informationsart spezialisiert. Sichten können durch Querbezüge zwischen den Informationen interaktiv miteinander verknüpft (koordiniert) sein. Im Folgenden werden die Anforderungen an eine solche Systemunterstützung zusammengefasst:

Integrative Systemunterstützung. Aufgabenrelevante strukturierte Daten und deren aufgabenrelevante Verknüpfungen mit den betrachteten Texten sollen als Informationskontext in die Textanalyse einbezogen werden. So sollen Beschreibungskategorien und Relationale Daten als Metadaten der Texte aufgefasst werden. Alle Informationsobjekte sollen übersichtlich „auf einen Blick“ dargestellt werden.

Interaktive Systemunterstützung. Ein Benutzer soll im engen Dialog mit dem System (kurze Interaktionszyklen) die Texte und damit verknüpfte Information analysieren können. Anfrageorientierte und explorativer Aktivitäten sollen flexibel kombiniert werden können. Textähnlichkeiten, Kategorienstrukturen, relationale Daten sowie die Verknüpfungen zwischen den Texten und den strukturierten Daten sollen auf intuitive Weise navigiert werden können. Die strukturierten Daten sollen zur Spezifikation aufgabenrelevanter Texte (interaktive Anfragen) und zur Beschreibung ausgewählter Texte verwendet werden können. Während jede einzelne Sicht Darstellungs- und Explorationsmetaphern verwendet, die auf die repräsentierte Informationsart spezialisiert ist, soll zur intuitiven Navigation *zwischen* Sichten ein einheitliches Konzept eingeführt werden. Dabei soll eine Brush&Link-Metapher zur Navigation verwendet werden, bei der die Auswahl in einer Sicht die Markierung verknüpfter Informationsobjekte der anderen Sichten zur Folge hat, aber die vom Benutzer durchgeführten Auswahlen in anderen Sichten *nicht* beeinflusst.

Wiederverwendbarkeit und Erweiterbarkeit. Das Systemkonzept und seine Realisierung als Referenztechnologie sollen für ein breites Spektrum verschiedener Informationen und Aufgaben einsetzbar sein. Durch Konfiguration (und ggf. modulare Erweiterungen) sollen sich spezifische Systeme für konkrete Rechercheszenarien leicht ableiten lassen.

5.2 Sichtenarten und Verknüpfungen

Ein SWAP-System umfasst mehrere Sichten. Eine Sicht stellt Informationsobjekte gleicher Art dar. Informationsobjekte verschiedener Sichten können verknüpft sein. Auswahlen und Markierungen der Informationsobjekte in Sichten werden über Verknüpfungen zwischen Informationsobjekten koordiniert. Zunächst werden Sichtenarten und Verknüpfungen formal definiert.

Def. 5.1: Ähnlichkeitssicht

Eine *Ähnlichkeitssicht* $S_{\bar{A}}$ ist eine grafische Übersichtsdarstellung einer Menge D von Textdokumenten (Informationsobjekten) und deren paarweisen Ähnlichkeiten $\sigma: D \times D \rightarrow [0,1]$. Die paarweisen Ähnlichkeiten werden auf wenige Darstellungsdimensionen abgebildet, von denen eine die räumliche Nähe der Informationsobjekte ist. Eine *Nachbarschaftsfunktion* $\mu: D \times D \rightarrow [0,1]$ beschreibt die durch einen Benutzer aufgrund der Darstellung wahrgenommenen paarweisen Ähnlichkeiten. μ ist vollständig und symmetrisch.

Def. 5.2: Faktensicht

Eine *Faktensicht* S_F ist eine tabellarische (textuelle oder grafische) Übersichtsdarstellung einer Menge F von Relationalen Tupeln (Informationsobjekten) mit $F \subseteq F_1 \times F_2 \times \dots \times F_n$; F_i sind Grundmengen.

Def. 5.3: Kategoriensicht

Eine *Kategoriensicht* S_C ist eine mono-hierarchische⁶⁰ (textuelle oder grafische) Übersichtsdarstellung eines gerichteten Graphen $G = (C,E)$ mit Knotenmenge C (Informationsobjekte) und Kantenmenge $E \subseteq C \times C$. Die Knoten heißen Kategorien.

⁶⁰ Polyhierarchische Darstellungen, bei denen eine Kategorie mehrere übergeordnete Knoten haben darf, werden nicht betrachtet.

Def. 5.4: Verknüpfungen zwischen Informationsobjekten

Die Menge der *Verknüpfungen zwischen Informationsobjekten* der Faktensicht (Tupelmengende F) und Textdokumenten der Ähnlichkeitssicht (Dokumentenmenge D) ist $R_F \subseteq D \times F$.

Die Menge der Verknüpfungen zwischen Informationsobjekten der Kategoriensicht (Kategorienmenge C) und Textdokumenten der Ähnlichkeitssicht (Dokumentenmenge D) ist $R_C \subseteq D \times C$.

Erläuterungen:

- Die Verknüpfungen zwischen Dokumenten und Fakten/Kategorien repräsentieren strukturelle **Assoziationen**, also ungewichtete, untypisierte und ungerichtete (symmetrische) Beziehungen zwischen Informationsobjekten verschiedener Sichtenarten.
- Die Formalisierung der Verknüpfungen als Relation lässt beliebige Verknüpfungen (**m:n**) zwischen Textdokumenten und Informationsobjekten der anderen Sichtenarten zu. Somit kann die Kategoriensicht bspw. eine Multi-Klassifikation der Texte darstellen. Linkseindeutige Verknüpfungen (injektive Abbildung; **1:1**) stellen einen Spezialfall dar.
- Da Fakten/Kategorien zwar mit Dokumenten, nicht jedoch untereinander verknüpft sein können, ergibt sich eine **Stern-Topologie** (vgl. Abbildung 5.1) mit der Ähnlichkeitssicht im Zentrum. Aus den angegebenen direkten Verknüpfungen zu den Dokumenten können indirekte (transitive) Verknüpfungen zwischen Fakten und Kategorien berechnet und dargestellt werden. Das Ausschließen direkter Verknüpfungen zwischen Fakten und Kategorien vermeidet transitive Zyklen und verringert dadurch die Systemkomplexität, insbesondere aus der Benutzerperspektive. Außerdem liegt der Fokus der Systemunterstützung auf der Textanalyse, weshalb die Verknüpfungen zwischen *Dokumenten* und anderen Informationsobjekten von primärem Interesse sind.
- Die hier beschriebenen Verknüpfungen sind Verknüpfungen auf **Instanzen-Ebene**, nicht auf Schema-Ebene (s. auch Kap.5.7.3). Welches Datenbank-Schema zum Erzeugen der Faktensicht verwendet wurde, ist für strukturelle Betrachtungen unerheblich, mag aber für die Erzeugung der dargestellten Verknüpfungen zwischen den Fakten und anderen Informationsobjekten sowie für die Interpretation durch den Benutzer eine wichtige Rolle spielen.

Def. 5.5: SWAP-System

Ein *SWAP-System* S besteht aus einer Ähnlichkeitssicht, einer Faktensicht, einer Kategoriensicht und 2 Verknüpfungsmengen (nach Def. 5.4.): $S := (S_A, S_F, S_C, R_F, R_C)$.

Ohne Beschränkung der Allgemeinheit wird hier die Einschränkung auf *eine* Sicht pro Sichtenart vorgenommen. Umfasst ein System mehrere Sichten der gleichen Art, so kann es leicht durch mehrere SWAP-Systeme mathematisch modelliert werden.

5.3 Sichtenkoordination und Systemverhalten

Def. 5.6: Auswahl von Informationsobjekten

Jeder Sicht eines SWAP-Systems ist eine *Auswahlmenge* zugeordnet. Die Auswahlmenge der Ähnlichkeitssicht ist $D_{\text{sel}} \subseteq D$. Die Auswahlmenge der Faktensicht ist $F_{\text{sel}} \subseteq F$. Die Auswahlmenge der Kategoriensicht ist $C_{\text{sel}} \subseteq C$.

Def. 5.7: Markierung von Informationsobjekten

Jeder Sicht eines SWAP-Systems sind 2 Markierungsmengen zugeordnet.

Die Markierungsmengen der *Ähnlichkeitssicht* sind $D_{\text{mark}_C} \subseteq D, D_{\text{mark}_F} \subseteq D$.

Die Markierungsmengen der *Faktensicht* sind $F_{\text{mark}_D} \subseteq F, F_{\text{mark}_C} \subseteq F$.

Die Markierungsmengen der *Kategoriensicht* sind $C_{\text{mark}_D} \subseteq C, C_{\text{mark}_F} \subseteq C$.

In SWAP-Systemen werden Markierungen (im Gegensatz zu Auswahlen) nicht explizit durch den Benutzer festgelegt, sondern unter Berücksichtigung der Koordinationsgleichungen der Sichten (s. Def. 5.8) aus den Auswahlen in den jeweils anderen Sichten abgeleitet.

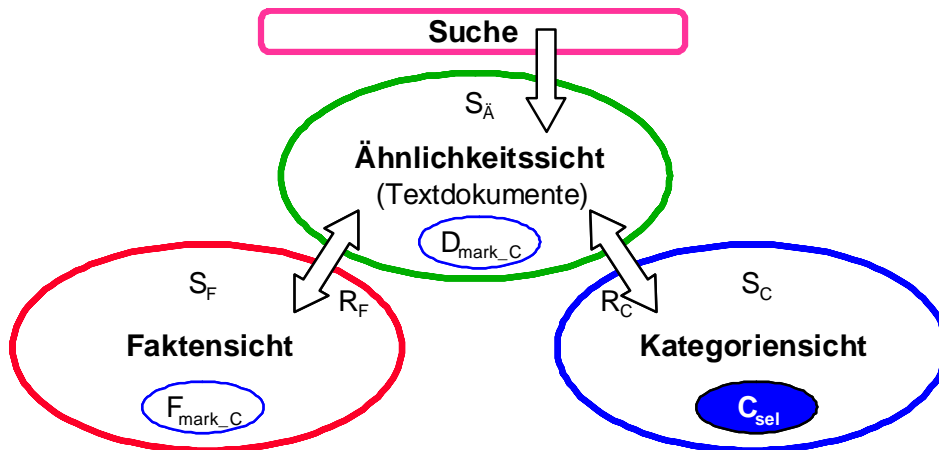


Abbildung 5.1: Beispiel einer Auswahl (von Kategorien) und der dadurch implizierten Markierungen (Texte und Fakten)

Def. 5.8: Koordination von Sichten

Die *Koordination der Sichten* wird so definiert, dass eine Auswahl in *einer* Sicht zu Markierungen von Informationsobjekten in den damit verknüpften Sichten führt. Außerdem führt eine Markierung in der Ähnlichkeitssicht zu Markierungen in der damit verknüpften Sicht der *anderen* Sichtenart (transitive Markierung über die zentrale Ähnlichkeitssicht). Dies wird ausgedrückt durch folgende *Koordinationsgleichungen*. Seien F und C die Informationsobjekt-Mengen der Sichten S_F und S_C :

$$F_{\text{mark}_D} := \{ f \in F \mid \exists d \in D_{\text{sel}} : (d, f) \in R_F \}$$

$$C_{\text{mark}_D} := \{ c \in C \mid \exists d \in D_{\text{sel}} : (d, c) \in R_C \}$$

$$D_{\text{mark}_F} := \{ d \in D \mid \exists f \in F_{\text{sel}} : (d, f) \in R_F \}$$

$$C_{\text{mark}_F} := \{ c \in C \mid \exists d \in D_{\text{mark}_F} : (d, f) \in R_F \}$$

$$D_{\text{mark}_C} := \{ d \in D \mid \exists c \in C_{\text{sel}} : (d, c) \in R_C \}$$

$$F_{\text{mark}_C} := \{ f \in F \mid \exists d \in D_{\text{mark}_C} : (d, c) \in R_C \}$$

Erläuterungen:

- Die in den Koordinationsgleichungen ausgedrückten **Abhängigkeiten** werden in Tabelle 5-1 durch Pfeile angezeigt: Während die Auswahl von Dokumenten direkt zu Markierungen in den anderen Sichten führt, werden die Auswahlen von Fakten oder Kategorien über die Markierung in der Ähnlichkeitssicht transitiv auf die Kategorien- bzw. Faktensicht übertragen. Abbildung 5.1 veranschaulicht die Effekte einer Auswahl in der Kategoriensicht.

- Die transitiven Markierungen C_{mark_F} und F_{mark_C} lassen sich in der Relationenalgebra ausdrücken als Join aus R_F und R_C über das gemeinsamen Attribut D (z.B. URL). Vorsicht ist geboten bei Relationen mit mehr als 2 Schlüsselattributen, deren Attribute in verschiedenen Sichten dargestellt werden, wie z.B. Benutzeraktivitäten in kooperativen Arbeitsumgebungen, die durch Tupel $\langle \text{Person}, \text{Dokument}, \text{Datum} \rangle$ beschrieben seien. Werden Personen in C und Zeitpunkte in F dargestellt und somit die Ursprungsrelation in R_C ($\langle \text{Person}, \text{Dokument} \rangle$) und R_F ($\langle \text{Datum}, \text{Dokument} \rangle$) aufgebrochen, so kann ein Join kontraintuitive Markierungen erzeugen. Beispiel: $\langle \text{Prinz}, d1, 2004 \rangle, \langle \text{Seeling}, d1, 2006 \rangle$. Wird nun vom Benutzer „Prinz“ ausgewählt, so werden die Jahre 2004 und 2006 markiert, obwohl Prinz nur in 2004 an d1 gearbeitet hat. Daher müssen große Relationen (zusammengehörende Attribute) *innerhalb* derselben Sicht abgebildet werden (z.B. Aktivitätstabelle $\langle \text{Person}, \text{Datum} \rangle$ in Sicht F).
- Zur Unterscheidung der Herkunft von Auswahlen und dadurch ausgelösten Markierungen wird bei der Darstellung der Sichten im Beispielsystem SWAPit (vgl. Kap.6) jeder Tabellenzeile in Tabelle 5-1 eine andere **Farbe** zugeordnet.

Tabelle 5-1: Sichtenkoordination - Auswahlen und Markierungen

| Auswahl in der ... | Ähnlichkeitssicht | Faktensicht | Kategoriensicht |
|--------------------|---------------------|---------------------|---------------------|
| Ähnlichkeitssicht | D_{sel} | F_{mark_D} | C_{mark_D} |
| Faktensicht | D_{mark_F} | F_{sel} | C_{mark_F} |
| Kategoriensicht | D_{mark_C} | F_{mark_C} | C_{sel} |

Def. 5.9: Auswahlzustand

Der *Auswahlzustand* eines SWAP-Systems $SEL(t)$ ist definiert durch die Auswahlen von Informationsobjekten zum Zeitpunkt t , also: $SEL(t) := (D_{\text{sel}}, F_{\text{sel}}, C_{\text{sel}})$.

Erläuterungen:

- Markierungen leiten sich nach Def.5.8 aus Auswahlen her. Deshalb genügt die Berücksichtigung aller Auswahlen bei der Definition des Auswahlzustands.
- Die Berücksichtigung eines Zeitparameters t kann zur Beschreibung des dynamischen Systemverhaltens durch Zustandsfolgen hilfreich sein.

Def. 5.10: Systemzustand

Der *Systemzustand* eines SWAP-Systems $STATE(t)$ ist definiert durch die Auswahlen und Markierungen von Informationsobjekten zum Zeitpunkt t , also:

$$STATE(t) := SEL(t) \times (D_{\text{mark}_C}, D_{\text{mark}_F}, F_{\text{mark}_D}, F_{\text{mark}_C}, C_{\text{mark}_D}, C_{\text{mark}_F})$$

Def. 5.11: Konsistenz (eines Systemzustands)

Ein Systemzustand $STATE(t)$ heißt *konsistent* \Leftrightarrow zum Zeitpunkt t gelten alle Koordinationsgleichungen (nach Def. 5.8).

5.4 Relevanzmaße für Markierungen von Informationsobjekten

In diesem Kapitel wird ein strukturelles Relevanzmaß eingeführt, das markierte Informationsobjekte einer Sicht bezüglich der Auswahl in einer anderen Sicht gewichtet. Außerdem wird diskutiert, inwiefern solche Gewichtungen Analogien zu Termgewichtungen im Information Retrieval aufweisen, und inwiefern solche Gewichtungen für MultiView-Systeme im Allgemeinen eine wertvolle Neuerung darstellen.

Def. 5.12: Matchingfunktion

Eine *Matchingfunktion* τ gewichtet die Markierungen eines SWAP-Systems S bezüglich der Auswahl, welche die Markierung ausgelöst hat. Gegeben zwei verschiedene Mengen von Informationsobjekten $I_1, I_2 \in \{D, F, C\}$, $I_1 \neq I_2$, $D \in \{I_1, I_2\}$, die Verknüpfungs-Relation dazwischen $R_{12} \in \{R_F, R_C\}$, sowie die Auswahl $I_{1\text{-sel}}$ und die Markierung $I_{2\text{-mark}}$. Dann ist eine Matchingfunktion definiert als $\tau_{I_1\text{-sel}}: I_{2\text{-mark}} \rightarrow [0, 1]$.

Erläuterungen:

- In der Systemrealisierung sollen Matchingfunktionen helfen, die Aufmerksamkeit des Benutzers auf wichtige Informationsobjekte in anderen Sichten zu lenken, die mit den interaktiv durch den Benutzer ausgewählten Informationsobjekten verknüpft sind.
- Weiter unten werden zwei spezielle Matchingfunktionen vorgeschlagen und diskutiert.
- Neben der Auswahl von Informationsobjekten könnten Matchingfunktionen auch die gesamten Grundmengen von Informationsobjekten in die Berechnung des Gewichts eines markierten Informationsobjektes einbeziehen, wie unten vorgeschlagen wird.
- Bei 1:1-Verknüpfungen zwischen Sichten ist die Bedeutung jedes markierten Informationsobjektes für die Auswahl von Informationsobjekten immer gleich groß. Da eine Matchingfunktion aber gerade dazu dient, Kriterien für eine Differenzierung (Ranking) zu liefern, scheint die Anwendung hier überflüssig.

(1) Signifikante Matchingfunktion

Unter der Prämisse, dass ein markiertes Informationsobjekt besonders wichtig ist, wenn es mit vielen ausgewählten Informationsobjekten verknüpft ist, liegt es nahe, ein markiertes Informationsobjekt B aus I_2 zu gewichten mit dem Anteil der mit B verknüpften ausgewählten Informationsobjekte an allen ausgewählten Informationsobjekten aus $I_{1\text{-sel}}$:

$$\tau(B) := |\{ b \in I_{1\text{-sel}} \mid (b, B) \in R_{12} \}| / |I_{1\text{-sel}}|$$

Diese Matchingfunktion heißt *signifikant* und kann verwendet werden, um – unabhängig vom Rest der Informationsobjekte aus I_1 – die ausgewählte Menge von Informationsobjekten näher zu untersuchen (bzgl. die damit verknüpften Informationsobjekte anderer Sichten). Damit signifikante Matchingfunktionen besonders für kleine Werte, die bei großen Auswahlen typisch sind, eine gute Differenzierung liefern, kann beispielsweise eine logarithmische Skalierungsfunktion eingesetzt werden.

(2) Charakteristische Matchingfunktion

Unter der Prämisse, dass ein markiertes Informationsobjekt besonders wichtig ist, wenn es mit mehr ausgewählten als nicht ausgewählten Informationsobjekten verknüpft ist, kann ein markiertes Informationsobjekt B aus I_2 gewichtet werden mit dem Anteil der mit B verknüpften ausgewählten Informationsobjekte an allen mit B verknüpften Informationsobjekten aus I_1 :

$$\tau(B) := |\{ b \in I_{1\text{-sel}} \mid (b, B) \in R_{12} \}| / |\{ b \in I_1 \mid (b, B) \in R_{12} \}|$$

Diese Matchingfunktion heißt *charakteristisch* und kann eingesetzt werden um charakteristische Merkmale der ausgewählten Menge von Informationsobjekten (also damit verknüpfte Informationsobjekte anderer Sichten) bezüglich der gesamten Menge von Informationsobjekten aus I_1 zu untersuchen.

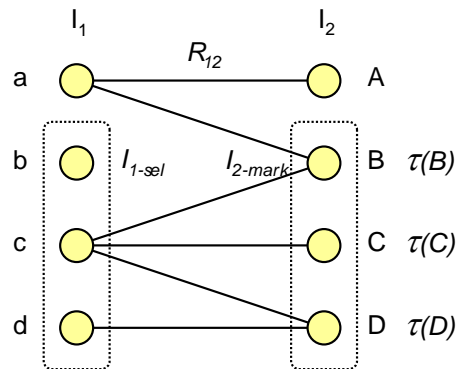


Abbildung 5.2: Die Matchingfunktion τ bewertet die Relevanz der markierten Informationsobjekte aus I_2

Zur Anschauung sind die Werte der vorgeschlagenen Matchingfunktionen für das Beispiel aus Abbildung 5.2 in Tabelle 5-2 angegeben. Dort führt eine Auswahl von 3 Informationsobjekten der linken Sicht (I_1) zur Markierung von 3 Informationsobjekten der rechten Sicht (I_2). Auch Kombinationen aus beiden vorgeschlagenen Matchingfunktionen sind denkbar, haben jedoch den potentiellen Nachteil der schwierigeren Interpretierbarkeit durch den Benutzer aufgrund einer Vermischung der Überlegungen, die in den Einzelmaßen realisiert sind.

Tabelle 5-2: Auswertung zweier Matchingfunktionen für das Beispiel aus Abbildung 5.2

| | $\tau(\mathbf{B})$ | $\tau(\mathbf{C})$ | $\tau(\mathbf{D})$ |
|-------------------------|--------------------|--------------------|--------------------|
| Signifikant | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{2}{3}$ |
| Charakteristisch | $\frac{1}{2}$ | 1 | 1 |

Diskussion

Analogien der Matchingfunktionen zu bewährten Maßen zur Termgewichtung im Information Retrieval [SWY75] lassen sich leicht aufzeigen:

Die *signifikante Matchingfunktion* erinnert an normierte Termhäufigkeit (TF), die den Anteil der Vorkommenshäufigkeit eines Terms an der Summe der Vorkommenshäufigkeiten aller Terme eines Dokumentes als Indikator der Wichtigkeit eines Terms für ein Dokument wertet. Die signifikante Matchingfunktion wertet analog den Anteil der mit einem markierten Informationsobjekt verknüpften ausgewählten Informationsobjekte an allen ausgewählten Informationsobjekten als Indikator für die Wichtigkeit eines verknüpften Informationsobjektes für die Auswahl.

Die *charakteristische Matchingfunktion* erinnert an das Maß TFxIDF ⁶¹, das die o.g. Termhäufigkeit normiert mit der Anzahl der Dokumente, in denen ein Term vorkommt. Die charakteristische Matchingfunktion wertet den Anteil der mit einem markierten Informationsobjekt verknüpften ausgewählten Informationsobjekte an allen mit dem markierten Informationsobjekt verknüpften Informationsobjekten der Grundmenge als Indikator für die Wichtigkeit eines verknüpften Informationsobjektes für die Auswahl. Auch hier werden (wie bei TFxIDF) nicht nur die Auswahl von Informationsobjekten (das Dokument), sondern alle Dokumente berücksichtigt.

⁶¹ TFxIDF steht für: *term frequency times inverse document frequency*

Die vorgestellten Berechnungsvorschriften für solche Matchingwerte repräsentieren ein DOI⁶²-Modell zur Leitung der Aufmerksamkeit von Benutzern in grafischen Informationssystemen. Sahling führt eine DOI-Gewichtung für interaktive Auswahlen ein mit dem Ziel, den Benutzer beim Auswahlvorgang zu unterstützen („smooth brushing“ [Sa03]). In dieser Arbeit geht es um die Gewichtung der *markierten* Informationsobjekte anderer Sichten unter Berücksichtigung der interaktiven Benutzerauswahl in einer Sicht. Dabei ist das Ziel, die Aufmerksamkeit des Benutzers *nach* einem Auswahlvorgang auf andere potentiell relevante Informationsobjekte zu lenken.

Eine mögliche Darstellung der Matchingwerte wird in Kap.6.5 vorgeschlagen und bewertet. Die Wichtigkeit solcher Maße für die Informationsexploration wird von den Benutzern in Fallstudie 3 bestätigt, allerdings bedarf die Nachvollziehbarkeit für den Benutzer der Verbesserung (Kap.7.5.5). In weiteren Benutzerstudien sollte gezielt überprüft werden, wie hoch der Nutzen einer durch Relevanzmaße unterstützten Benutzerführung gegenüber ungewichteten Markierungen ist, und ob der Nutzen der zusätzlichen Information tatsächlich die Nachteile zusätzlicher kognitiver Belastung aufwiegt (Kap.8.3).

5.5 Informationsobjekte zur Beschreibung von Dokumenten

Def. 5.13: MultiView-Annotation

Eine *MultiView-Annotation* für eine Menge von Informationsobjekten I_{1-sub} ist eine Menge von Informationsobjekten einer anderen Sicht, welche durch eine *Zählfunktion* $freq$ gewichtet werden.

Gegeben zwei verschiedene Mengen von Informationsobjekten $I_1, I_2 \in \{D, F, C\}$, $I_1 \neq I_2$, $D \in \{I_1, I_2\}$, die Verknüpfungs-Relation dazwischen $R_{12} \in \{R_F, R_C\}$, sowie $I_{1-sub} \subseteq I_1$. Sei I_{2-sub} die Menge der mit I_{1-sub} durch R_{12} verknüpften Informationsobjekte, d.h. $I_{2-sub} = \{ i_2 \mid i_2 \in I_2, \exists i_1 \in I_{1-sub} \text{ mit } (i_1, i_2) \in R_{12} \}$.

Gegeben eine *Beschriftungs-Funktion* $label$, die jedem Informationsobjekt aus I eine Zeichenkette aus einer Grundmenge L zuordnet, d.h. $label: I \rightarrow L$

Gegeben eine *Zählfunktion* $freq$ für die Anzahl der Informationsobjekte aus I mit der gleichen Beschriftung ℓ , d.h. $freq: L \rightarrow \mathbb{N}$ mit $freq(\ell) = | \{ i \mid i \in I, label(i) = \ell \} |$

Dann ist eine *MultiView-Annotation* definiert als

$$A_{I_{1-sub}} = \{ (\ell, w) \mid \exists i_2 \in I_{2-sub} \text{ mit } label(i_2) = \ell, w = freq_{I_{2-sub}}(\ell) \}$$

Erläuterungen:

- Eine exemplarische Realisierung von interaktiven Werkzeugen, die auf MultiView-Annotation basieren, ist in Kap.6.6 beschrieben. Dort wird das Konzept der MultiView-Annotation eingesetzt, um Dokumente mit strukturierten Daten aus anderen Sichten zu annotieren.
- Die Label-Funktion kann eine Projektion aus den Informationsobjekten sein: Bei Tabellen können Einträge einer Spalte als Beschriftung dienen. Bei Kategorien können Kategoriennamen oder Teile daraus verwendet werden. Die Label-Funktion sollte die für die Recherche wichtigen Eigenschaften der Informationsobjekte abbilden.
- $freq$ liefert ein Maß dafür, wie häufig eine relevante Beschriftung in einer anderen Sicht vorkommt. Somit kann $freq$ zur gewichteten Darstellung der annotierten Beschriftungen verwendet werden. Die Häufigkeiten können zusätzlich gewichtet werden, beispielsweise durch Normalisierung oder Durchschnittsbildung.

⁶² Degree-of-Interest [CN02], eingeführt 1981 durch Furnas [Fu81]

SWAP-Systeme unterstützen die Textanalyse durch Bereitstellung des Informationskontexts in Form von strukturierten Daten, die mit den Texten verknüpft sind. Dadurch entsteht die Chance, die Textdokumente nicht nur durch interaktive Werkzeuge zur Inhaltsanalyse (z.B. Termstatistiken, Zusammenfassungen, Termhervorhebung, Kap.6.6) zu beschreiben, sondern auch durch damit verknüpfte strukturierte Daten (z.B. Personen, Jahre, Produkte). Das Sichtendesign eines SWAP-Systems erlaubt, über Verknüpfungen zwischen Texten und strukturierten Daten zu navigieren. Die MultiView-Annotation kann zusätzlich eingesetzt werden, den räumlichen Abstand zwischen zusammengehörenden Informationsobjekten zu verringern, indem die relevanten strukturierten Daten gruppiert und in unmittelbarer Nähe zu den beschreibenden Texte notiert werden. Dadurch wird die semantische Nähe⁶³ von Informationsobjekten durch die räumliche Nähe ihrer grafischen Repräsentation illustriert, um die kognitive Belastung bei der Navigation zu verringern, denn „*visual comparison is easier to accomplish than memory-based comparison*“ [Sa03].

Die meisten der in MultiView-Systemen verwendeten Sichten bieten Darstellungen oder Funktionalitäten, um die semantischer Nähe der Information innerhalb einer Sicht durch die räumliche Nähe der Repräsentation zu reflektieren (z.B. Abstände in Clusterdarstellungen oder Bäumen, Sortierung von Tabellen). Durch MultiView-Annotation kann dieses Prinzip auch *zwischen* Sichten eingesetzt werden, also im Fall von SWAP-Systemen für die Beziehung zwischen Informationsobjekten *unterschiedlicher* Art.

In Kap.3.2 wurden die Nachteile integrierter grafischer Darstellungen, also der gemischten Darstellung verschiedener Informationsarten diskutiert. Die MultiView-Annotation bietet die Grundlage für interaktive Werkzeuge, die auf Benutzerwunsch temporär für eine Teilmenge der Informationsobjekte eine integrierte Darstellung erzeugen.

5.6 Assoziative Exploration

Nachdem das SWAP-Konzept definiert wurde, sollen nun in der Vielzahl interaktiver Möglichkeiten zielführende und effiziente Interaktionsmuster identifiziert werden, die sinnvolle Arbeitsschritte der Analyse repräsentieren. Im Folgenden wird unterschieden zwischen sichtenübergreifender und sichteninterner Exploration.

Der Begriff *Navigation* bezeichnet häufig, in Abgrenzung zum *Browsen*, den systematischen bzw. *zielorientierten* Übergang von einem Informationsobjekt zu einem anderen. So unterscheidet Furnas [Fu97] für Informations-Strukturen wie Netze, Bäume und Tabellen die Navigation vom schlichteren Übergang („*traversal*“): „[...] *in addition the selections try to be informed and reasonable in the pursuit of a desired target*“. Bei der Navigation durch eine Informations-Struktur ist zu unterscheiden, ob es sich um einen Detaildarstellung (lokale Sicht, „*view navigation*“ [Fu99], „*or-style-traversal*“ [Ro97]) oder eine Übersichtsdarstellung („*and-interactions*“ [Ro97]) handelt (vgl. Kap.5.6.2). So kann durch Hypertext detailliert mittels eines Dokumenten-Browsers oder übersichtlich mittels einer Netzdarstellung bzw. Themenkarte navigiert werden [SM99]. Ein Baum lässt sich beispielsweise in Detaildarstellung erkunden, indem Verknüpfungen auf Vater- und Kindknoten als lokale Verzweigungen angeboten werden, während eine hierarchische Übersichtsdarstellung große Ausschnitte der Baumstruktur gleichzeitig darstellt. Übersichtsdarstellungen erlauben die gleichzeitige Auswahl entweder genau eines oder mehrerer Informationsobjekte (Single vs. Multiple Focus).

Für SWAP-Systeme bietet sich aufgrund der Anforderungen aus Kap.5.1 der Einsatz flexibel navigierbarer *Übersichtsdarstellungen* an. Auch Craft et al. argumentieren, dass zur effizienten Bearbeitung komplexer Analyseaufgaben neben Detail- auch Übersichtswissen benötigt wird [CC05]. Ausserdem sollten die Übersichtsdarstellungen die *gleichzeitige Fokussierung mehrerer Informationsobjekte* („*multiple focus*“) erlau-

⁶³ Voraussetzung ist, dass die Funktion *label* so berechnet wird, dass semantisch zusammengehörende Informationsobjekte die gleiche Beschriftung erhalten.

ben, da für interaktive Anfragen im Allgemeinen zugleich mehrere Informationsobjekte einer Sicht ausgewählt werden müssen.

Navigation durch assoziative Strukturen ist in vielen Forschungsdisziplinen ein wichtiges Thema. Für die *Informationsrecherche* spielt die Navigation neben der anfrageorientierten Suche eine herausragende Rolle. Insbesondere die durch die Verbreitung des Internet angetriebene Entwicklung von *Hypertext-Systemen* hat zahlreiche grafische Übersichtsdarstellungen und Navigationsparadigmen hervorgebracht [SM99]. Oft bedienen sich die Darstellungen von Informationsräumen kartografischer Metaphern aus dem Bereich der Geoinformatik.

Neben der Navigation in Übersichtsdarstellungen spielen, neben den physischen Aktivitäten wie Auswahlen, auch rein kognitive Aktivitäten eine wichtige Rolle [BS01]. Ein Beispiel ist das „Querlesen“ einer Menge von Informationsobjekten oder die Untersuchung, zu welchen Themen Informationsobjekte gehören. Wenn in dieser Arbeit von *Exploration* die Rede ist, sind die physischen und kognitiven Aktivitäten eingeschlossen.

5.6.1 Sichtenübergreifende Exploration - Auswahlen als Interaktive Anfragen

In einem SWAP-System impliziert eine Auswahl von Informationsobjekten einer Sicht die Markierung von Informationsobjekten in damit verknüpften Sichten (vgl. Def. 5.8, Koordinationsgleichungen). Außerdem können Markierungen in der Ähnlichkeitssicht (transitive) Markierungen in den anderen Sichten bedingen. Somit lassen sich die Auswahlen D_{sel} , F_{sel} , C_{sel} und die Markierungen D_{mark_F} , D_{mark_C} als interaktive Benutzeranfragen auffassen⁶⁴. Das Anfrage-Ergebnis ist die resultierende Markierung, also die jeweils linke Seite der entsprechenden Koordinationsgleichung. Die Anfragen sind disjunktiv (s. Existenzquantor auf rechter Gleichungs-Seite).

Interessant ist die Kombination solcher einfachen Anfragen zu multidimensionalen Anfragen: Da Markierungen und Auswahlen jeder Sicht unterschieden werden, erlaubt ein SWAP-System die simultane Darstellung mehrerer unterschiedlicher Anfragen und Ergebnisse. Parallelen der multidimensionalen interaktiven Anfragen in SWAP-Systemen zur interaktiven Spezifikation von Kennzahlenberichten in OLAP-Systemen werden in Kap.5.7.3 diskutiert. Bei SWAP-Systemen lassen sich zwei Interaktionsmodi unterscheiden (vgl. Abbildung 5.3):

1. *Spezifikation von Textgruppen* durch Auswahl damit verknüpfter strukturierter Daten: Die Anfrage wird gebildet aus den Auswahlen von Informationsobjekten von Kategorien- und Faktensicht. Das Ergebnis bilden die Markierungen der Ähnlichkeitssicht.
 - Anfragen: $\{ F_{sel}, C_{sel} \}$
 - Ergebnisse: $\{ D_{mark_F}, D_{mark_C} \}$
2. *Beschreibung von Textgruppen* durch damit verknüpfte strukturierte Daten: Die Anfrage wird gebildet aus Auswahl und Markierungen der Ähnlichkeitssicht. Ergebnis sind die Markierungen der Kategorien- und Faktensicht.
 - Anfragen: $\{ D_{sel}, D_{mark_F}, D_{mark_C} \}$
 - Ergebnisse: $\{ F_{mark_D}, F_{mark_C}, C_{mark_D}, C_{mark_F} \}$

⁶⁴ Diese Informationsmengen tauchen in den Koordinationsgleichungen (Def. 5.7) auf der rechten Seite auf.

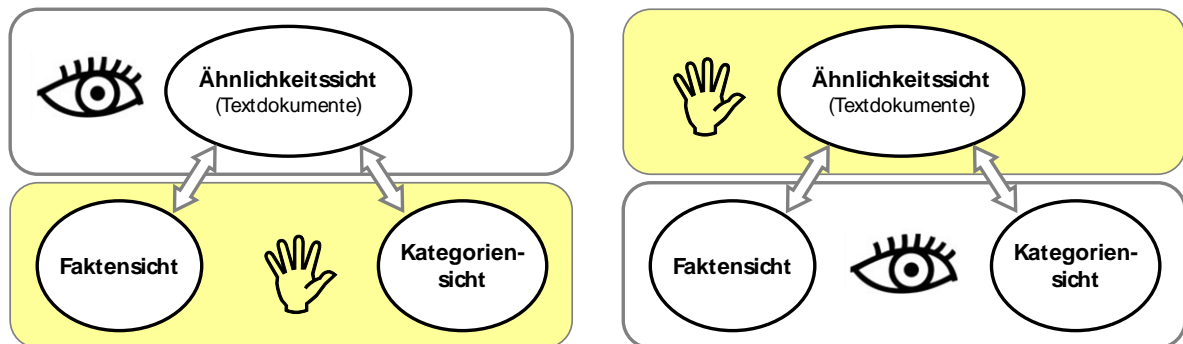


Abbildung 5.3: Unterscheidung zweier Interaktionsmodi: Spezifizieren (links) vs. Beschreiben von Textgruppen (rechts)

Aufgrund der simultanen Darstellung aller Anfragen und Ergebnisse in einem SWAP-System lassen sich bei multidimensionalen Anfragen neben den Einzelergebnissen der Anfragen auch alle Booleschen Kombinationen der Ergebnisse ablesen. Ein Benutzer legt also diese Kombinationen nicht mittels Operatoren bei der Anfrage fest, sondern kann sie nach Bedarf am Ergebnis erkennen. Ähnliche Ansätze im Bereich der Suchmaschinen zeigen in grafischen Darstellungen der Suchergebnisse die Kombinationen der einzelnen passenden Terme an. So zeigt das Werkzeug *TileBars* pro Suchbegriff und Trefferdokument einen Balken an [He95]. Daran lassen sich für ein Dokument zutreffende Termkombinationen bis auf Abschnittsebene ablesen.

Neben den dargelegten Möglichkeiten der Auswahl von Informationsobjekten in Sichten können dem Benutzer eines SWAP-Systems weitere („weiche“) Hilfsmittel für interaktive Anfragen zur Verfügung gestellt werden, insbesondere:

Sichten-interne Navigationsstrukturen (Kap.4.4)

- bei der (themenorientierten) Auswahl von Texten helfen Text-Ähnlichkeiten
- bei der Beschreibung von Texten helfen Text-Ähnlichkeiten (wie gliedert sich eine Textgruppe in die Themenlandschaft ein?)
- bei der (wertorientierten) Auswahl von relationalen Daten helfen Spalten-Sortierungen
- bei der Auswahl von Kategorien helfen die Verbindungs-Kanten dazwischen

Analyse-Werkzeuge

- bei der Auswahl von Texten, Fakten und Kategorien helfen Filter- und Suchmöglichkeiten
- bei der Beschreibung von Textgruppen helfen Statistiken und Text-Zusammenfassungen

Relevanzmaße

- bei der Beschreibung von Texten helfen Matchingfunktionen für Kategorien bzw. relationale Daten, die mit den Texten verknüpft sind

Matchingfunktionen gewichten die Informationsobjekte der Ergebnismenge danach, wie stark sie mit den Informationsobjekten der Anfragemenge verknüpft sind. Daher können an den gewichteten Ergebnissen potentiell auch Konjunktionen der Informationsobjekte der Anfragemenge abgelesen werden: Ein hoher Wert der Matchingfunktion deutet darauf hin, dass das betreffende Informationsobjekt der Ergebnismenge mit zahlreichen, vielleicht sogar allen Informationsobjekten der Anfragemenge verknüpft sind. Somit wird das Anfragekonzept um quasi-konjunktive Anfragen erweitert.

Das interaktive Wechseln zwischen Sichten ist Kernidee jedes MultiView-Konzeptes. In SWAP-Systemen wird dieses Interaktionsmuster ebenfalls unterstützt (Engl.: *to swap*). Insbesondere Phasen des Spezifizierens

und Beschreibens der Texte abzuwechseln, ist ein typisches Interaktionsmuster und eröffnet interessante Analysemöglichkeiten, wie z.B.:

- Texte finden, die ähnlich Begriffe enthalten (Vorgehen z.B.: Texte suchen, interessante Beschreibungsterme identifizieren, dann prüfen, welche Dok. diese Terme ebenfalls enthalten)
- Texte finden, die ähnlich klassifiziert sind (Vorgehen z.B.: Texte suchen, interessante Kategorien identifizieren, dann prüfen, welche Texte ähnlich klassifiziert sind)
- Kategorien finden, die ähnlich zur Klassifikation verwendet werden (Vorgehen z.B.: Kategorien auswählen, damit klassifizierte Texte identifizieren, dann prüfen, welche Kategorien diese Texte noch beschreiben)

Von einem *Roundtrip* wird beim Wechsel zwischen zwei Sichten gesprochen, wenn zwei SWAP-Schritte hintereinander ausgeführt werden, die Markierung also wieder zur Auswahl wird. Im Grunde geht es bei einem Roundtrip darum, weitere IO einer Sicht zu finden, die zu den ausgewählten IO bezüglich der Verknüpfungen zu IO einer anderen Sicht ähnlich sind.

Def. 5.14: Roundtrip

Eine Funktion $round: I_{sel} \times R \rightarrow I$, $I \in \{D, F, C\}$ ist definiert durch:

$$round(C_{sel}, R_C) := \{ c \in C \mid \exists d \in D_{mark_C} : (d, c) \in R_C \}$$

$$round(F_{sel}, R_F) := \{ f \in F \mid \exists d \in F_{mark_F} : (d, f) \in R_F \}$$

$$round(D_{sel}, R_C) := \{ d \in D \mid \exists c \in C_{mark_D} : (d, c) \in R_C \}$$

$$round(D_{sel}, R_F) := \{ d \in D \mid \exists f \in F_{mark_D} : (d, f) \in R_F \}$$

Ergebnis eines Roundtrips ist eine Markierung in Sicht A, die dadurch entsteht, dass eine Auswahl in Sicht A eine Markierung in Sicht B ergibt, die wiederum - als Auswahl in Sicht B aufgefasst - zur Markierung in Sicht A führt. Die Verknüpfungsrelation wird also ausgehend von einer Auswahl zweimal (je einmal „in jede Richtung“) angewendet. Der Roundtrip ist ein Beispiel für ein typisches Interaktionsmuster in SWAP-Systemen, für das eine direkte Systemunterstützung bereitgestellt werden kann.

5.6.2 Sichteninterne Exploration - Strukturen und Werkzeuge

Assoziative Strukturen zwischen gleichartigen Informationsobjekten einer Sicht können genutzt werden, um von fokussierten IO zu damit verwandten IO zu navigieren. Assoziative Strukturen bei SWAP-Systemen umfassen die Ähnlichkeit von Texten – abgebildet auf räumlich-grafische Nachbarschaft, strukturelle Beziehungen zwischen Konzepten – abgebildet auf Verbindungslinien, sowie Wertebeziehungen von relationalen Tupeln – abgebildet auf lineare Reihenfolge.

In den folgenden Abschnitten wird für jede Sichtenart das Prinzip der assoziativen Navigation auf Basis der Literatur diskutiert mit dem Ziel, Benutzerstrategien besser verstehen und beschreiben zu können.

Ähnlichkeitssicht

Die Ähnlichkeitsbasierte Navigation in Informationsräumen geht zurück auf Hypertext-Systeme wie das Internet. Dort bieten Links die Möglichkeit, zu Dokumenten zu gelangen, die mit dem betrachteten Dokument explizit verknüpft sind. Nach Shipman et al. [SM99] wurde das klassische Navigationsparadigma des Hypertext-Browsing in den 80er Jahren erweitert um *grafische Übersichtsdarstellungen* der Dokumente und Verknüpfungen. Diese Entwicklung wiederum war der Grundstein für Überlegungen zu Abstandsmaßen für

Dokumente, aus denen in den 90er Jahren räumliche Darstellungen hervor gingen: Studien belegen, dass der ‚Spatial Hypertext‘ häufig dem traditionellen Hypertext vorgezogen wird (vgl. Marshall et al. [Ma92]). Rosenberg [Ro97] nennt die Navigation im Dokumentenzentrierten (klassischen) Hypertext ‚or-style-traversal‘, wo der Nutzer sich für genau eine Navigationsrichtung entscheidet. Beim ‚Spatial Hypertext‘ spricht Rosenberg von ‚and-interactions‘, wo der Nutzer simultan mehrere Dokumentenknoten und Verzweigungen betrachten kann.

In der Regel lassen sich paarweise Ähnlichkeiten vieler Textdokumente nicht ohne Informationsverlust auf die wenigen Dimensionen einer grafischen Darstellung abbilden. Dimensionsreduzierende Verfahren (z.B. Mehrdimensionales Skalieren, MDS), die zur Abbildung auf wenige räumliche/grafische Darstellungsdimensionen eingesetzt werden, verfolgen unterschiedlichen Optimierungsziele, z.B. absolut Distanz- oder globale Topologie-Erhaltung, oder Performanz der Berechnung. Die Abbildung von inhaltlicher Ähnlichkeit auf geometrische Nähe von Dokumenten-Symbolen kann auf mannigfaltige Weise in Darstellungen umgesetzt werden (z.B. Scatterplots, Dokumentenlandkarten). Neben den geometrischen Dimensionen (Abstände in 2D oder 3D) kommen dafür zahlreiche grafische Dimensionen in Frage, z.B. Form, Farbe, Größe der Symbole oder Einfärbung des Hintergrunds – etwa durch Berg&Tal- oder Dichte-Metaphern. Die Interpretation insbesondere der grafischen Ähnlichkeits-Information durch den Benutzer ist im Allgemeinen jedoch subjektiv.

In SWAP-Systemen repräsentiert die *Nachbarschaftsfunktion* μ (vgl. Def. 5.1) die vom Benutzer interpretierte Ähnlichkeit jedes Textdokumentes zu allen anderen Texten. Sie ist abhängig von Darstellung und Benutzer. Im Folgenden wird zu Erklärungszwecken vereinfacht angenommen, dass sich die Nachbarschaften, also die wahrgenommenen paarweisen Ähnlichkeiten, ohne Informationsverlust in zwei räumlichen Dimensionen darstellen lassen.

Canter et al. studieren Navigationsmuster im Hypertext (Canter et al. 1985, siehe [St95], S.41). Sie unterscheiden vier Grundmuster: Path, ring, loop und spike. Eibl et al. [EM02] untersuchen *assoziative* Navigations-Strategien von Benutzern mit dem Ziel, Dokumentenkarten nutzenorientiert zu vergleichen. Dazu schlagen sie ein einfaches und idealisiertes Modell vor: Sie unterscheiden eine *Global Browsing Strategy* von einer *Local Browsing Strategy*: Bei der *Global Browsing Strategy* fixiert der Anwender ein Dokument und besucht in aufsteigender Reihenfolge des Abstandes alle dazu benachbarten Dokumente. Bei der *Local Browsing Strategy* wird von einem fixierten Dokument zum nächst-benachbarten navigiert, von diesem aus wieder zum nächsten (noch nicht besuchten) Nachbarn, usw. In diesem vereinfachten Modell bleiben einige bedeutsame Aspekte unberücksichtigt:

Im Zusammenhang mit Ähnlichkeitsdarstellungen zur Dokumentenanalyse wird, wie häufig wird in der Literatur, lediglich auf solche Schlüsse eingegangen, die ein Anwender ziehen kann, wenn er für ihn *interessante* Dokumente identifiziert. Zusätzlich bieten Ähnlichkeitsdarstellungen jedoch die wertvolle Möglichkeit, ausgehend von Einzeldokumenten rasch Gruppen von nicht näher zu betrachtenden Dokumenten auszumachen.

Es werden lediglich Navigations-Strategien zur Informations-Suche adressiert. Zudem sind Ähnlichkeitsdarstellungen gerade auch für solche Fragestellungen hilfreich, bei denen es weniger um einzelne Dokumente, sondern mehr um einen thematischen Überblick über eine Dokumentensammlung geht (vgl. Becks et al. [BSM02]).

Die betrachteten Strategien basieren auf der Grundannahme, dass ein Benutzer sich ausgehend von einigen fokussierten Texten in deren Nachbarschaft nach ähnlichen Texten umsieht. Bei der Entscheidung, welche Dokumente als nächstes zu betrachten sind, spielt jedoch auch der durch die gesamte Dokumentensammlung implizierte Themenraum eine Rolle, so dass ein Sprung in einen nicht unmittelbar benachbarten Bereich der Ähnlichkeitsdarstellung durchaus plausibel sein kann. Plausible Beispiele sind Sprünge zu komplementären Themen bzw. Sprünge zu benachbarten dichten oder besonders großen Ballungen von Dokumenten vor der Inspektion einzelner Texte auf dem Weg dorthin.

- Die Beziehungen zwischen Strategien (z.B. Übergänge, Ergänzung) in eine andere werden nicht diskutiert.
- Die *Local User Strategy* lässt sich als Folge von Schritten einer *Global User Strategy* auffassen und verliert somit die Berechtigung als eigenständiger Strategie-Typ auf der selben Ebene wie die *Global User Strategy*

Im Folgenden wird die assoziative Navigation in Ähnlichkeitsdarstellungen im Licht obiger Überlegungen diskutiert. Ziele sind die Klassifikation von Benutzerstrategien, aber auch die Handlungsanweisung zur Navigation oder der Entwurf neuer anwenderorientierter Interaktionsmechanismen.

Um neben Strategien der assoziativen Suche nach Dokumenten auch übersichtsorientierte Fragestellungen zu adressieren, werden im Folgenden drei Instrumentalziele der Textanalyse mit Ähnlichkeitsdarstellungen unterschieden (vgl. auch Becks, Seeling [BS01]):

- **Assoziative Suche:** Ausgehend von einer Menge von Texten weitere Texte mit ähnlichem Inhalt identifizieren
- **Assoziativer Überblick:** Themen, Themenzusammenhänge (z.B. Verfeinerung), sowie Themengröße (z.B. Schwerpunkte) erkennen
- **Assoziative Beschreibung:** Eine Menge von Texten im Themenraum der Textsammlung verorten (Kontext; u.a.: thematische Homogenität, Gruppen, Ausreißer)

Die Ziele werden hier bewusst polarisiert dargestellt. In einem typischen Arbeitsablauf sind Mischformen die Regel, d.h. die Übergänge zwischen diesen Instrumentalzielen und damit verbundenen Benutzerstrategien sind fließend. So kann auf eine Phase des *Assoziativen Überblicks* die Fokussierung eines Themas durch *Assoziative Suche* folgen. Die dann fokussierten Dokumente können durch *Assoziative Beschreibung* inhaltlich charakterisiert werden, usw.

Kernaktivitäten bei der Umsetzung der Strategien sind (vgl. Becks, Seeling [BS01]):

- physisch aktiv: Auswahl einer Menge von Texten; Öffnen von Texten
- phys. passiv; kognitiv: Fokus der Betrachtung verschieben oder ausweiten/ verkleinern

(1) Assoziative Suche

Geht es darum, zu einem interessanten Text weitere interessante Texte zu finden, so wird der Benutzer die wahrgenommenen Textähnlichkeiten nutzen, um benachbarte Texte in einer sinnvollen Reihenfolge zu untersuchen. Dabei erscheint es rational, zunächst die ähnlichsten Texte zu betrachten, die in der Darstellung potentiell am stärksten zum Ausgangspunkt benachbart sind. Somit wird er seinen Fokus sukzessive erweitern (**Wellenprinzip, konvexer Fokus**). Daher sind die Beziehungen zwischen dem betrachteten Text und allen anderen Texten von besonderem Interesse. Diese Strategie deckt sich mit der *Global User Strategy* von Eibl et al. [EM02]. Orthogonal zum Wellenprinzip funktioniert das **Ausschlussprinzip (konkaver Fokus)**. Hier wird von einem Text ausgegangen, der für die Fragestellung *nicht* interessant ist. Die dargestellten Nachbarschaften zwischen diesem und anderen Texten kann nun genutzt werden, um eine Reihe ähnlicher Dokumente ebenfalls „auf einmal“ von der Betrachtung auszuschließen. Der zu betrachtende Informationsraum wird dadurch um einen konvexen Ausschnitt verkleinert. Hearst unterstreicht den Wert des Ausschlussprinzips: „An underappreciated aspect of clusters is their utility for eliminating groups of documents from consideration. This result is supported by participant comments found in several studies...“ [He06]. Bei seiner Diskussion der Vorzüge grafischer Darstellungen zur Exploration semantischer Textähnlichkeiten stellt Doyle bereits 1961 fest: „What we are after at this stage of the retrieval process is speedy and confident rejection of irrelevant documents.“ [Do61].

Wellen- und Ausschlussprinzip sind komplementäre Navigations-Strategien zur systematischen Fokussierung von Texten mittels räumlicher Ähnlichkeitsdarstellungen. Bei beiden Prinzipien wird, von einer Menge interessanter Texte ausgehend, die Nachbarschaft dieser Texte zu anderen Texten genutzt, um deren Wichtigkeit grob abzuschätzen.

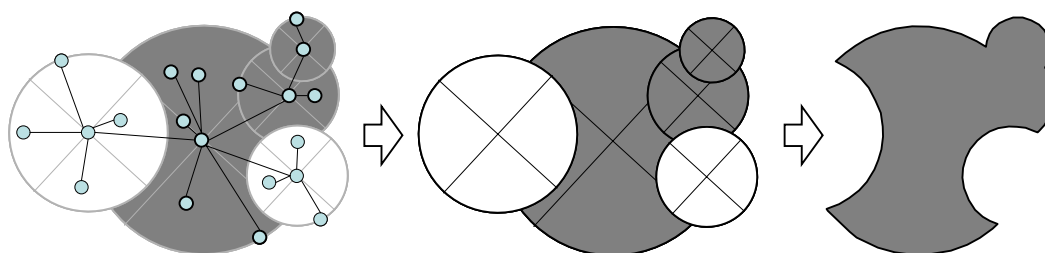


Abbildung 5.4: Welle und Ausschluss schematisch – grundlegende komplementäre Strategien zur assoziativen Fokussierung in Ähnlichkeitssichten

Es entsteht ein Navigationsmuster wie in Abbildung 5.4 schematisch dargestellt: Dunkle Kreise symbolisieren fokussierte Bereiche der Ähnlichkeitsdarstellung, während helle Kreise Bereiche kennzeichnen, die nach Prüfung des Dokumentes im Zentrum für unwichtig gehalten und daher von der weiteren Betrachtung ausgeschlossen werden. Die Radien hängen davon ab, wie deutlich sich das zentrale Dokument als interessant/uninteressant einstufen lässt und wie groß das Vertrauen in die Angemessenheit der Ähnlichkeitsdarstellung bzgl. der Fragestellung ist. Im rechten Teil der Abbildung ist nur der (positive) *Interessensfokus* dargestellt. Der linke Teil der Abbildung zeigt alle wichtigen Nachbarschaften als Verbindungslinien, und alle Texte, deren Betrachtung in Frage kommt, als Punkte. Die *Betrachtungsreihenfolge* wird bewusst nicht modelliert, da neben den Werten der Nachbarschaftsfunktion zahlreiche andere Einflussfaktoren in Frage kommen (Aufbau des Themenraums, verfügbare Zeit, Dichte und Größe von Ballungen, etc.).

(2) Assoziativer Überblick

Das Entdecken wesentlicher Konzepte in Dokumentensammlungen wird von Marchionini trefflich unterschieden von der Suche nach Dokumenten: „*Much of the search time in learning tasks is devoted to examining and comparing results and reformulating queries to discover boundaries of meaning for key concepts*“ [Ma06]. Er betont den Aspekt des Abgrenzens gegen andere Konzepte. Die Identifikation der von einer Dokumentensammlung abgedeckten Themen kann verschiedene Ziele haben. Zum einen kann es darum gehen, den Themenraum als Ganzes kennen zu lernen. Zum anderen kann die Orientierung in der Themenlandschaft beim groben Routing einer Textsuche hilfreich sein: „*Such design [of association maps] would have as its goal good over-all routing of searchers to arterioles [of the system]*“ (vgl. Doyle [Do61]).

Bei der Suche nach Themen werden den Benutzer die Lageverhältnisse aller Dokumente (Topologie des Dokumentenraums) stärker interessieren als einzelne Textähnlichkeiten. Themen werden repräsentiert durch Texte, die bezüglich der Nachbarschaften Einheiten bilden und sich vom Rest der Textsammlung gut abgrenzen lassen. Ähnlichkeitsdarstellungen von Dokumenten verzichten auf die explizite Zuordnung von Dokumenten zu Themen und überlassen dem Experten die Gruppierung.

Um Themen zu identifizieren, abstrahiert der Benutzer von Einzeldokumenten und untersucht die Nachbarschaftsverhältnisse der Dokumente. Wahrgenommene Gruppen können Stern- oder Kettenartige Ähnlichkeits-Strukturen, oder Mischformen aus beiden, besitzen (vgl. auch Eibl et al. [EM02]). Bei Sternartigen Strukturen existieren ein oder mehrere zentrale Texte, deren enge Nachbarschaft zu anderen Texten eine Gruppe konstituiert. Bei Kettenartigen Strukturen ist ein Text zu einem anderen benachbart, dieser wieder zu einem weiteren, usw.

Bei der Identifizierung von Themen in Ähnlichkeitsdarstellung geht der Anwender tolerant vor: Je größer (Themengröße) und kohärenter (Themendichte) ein Thema ist, desto eher wird das „Ausreißen“ einzelner (Brücken-)Dokumente in andere Themen akzeptiert, manifestiert sich das Thema doch bereits an der starken Nachbarschaft zahlreicher Dokumente. Besonders interessant für den Anwender sind im Allgemeinen sowohl Ballungen mit besonders starken (absoluten) paarweisen Ähnlichkeiten, als auch große Themen (viele Dokumente). Somit sind Themendichte und Themengröße unabhängige Faktoren der Themenqualität.

Neben den Themen selbst sind die *Themenzusammenhänge* von Interesse. Analog zu den Texten lässt die Nachbarschaft von Themen auf inhaltliche (semantische) Querbezüge schließen. Brücken- oder Übergangsdokumente helfen bei der Interpretation solcher Beziehungen. Auch die Verfeinerung von Themen in Unterthemen ist ein wichtiger zu beobachtender Themenzusammenhang.

(3) Assoziative Beschreibung

Geht es darum, eine Menge von Texten inhaltlich einzuordnen, steht die Suche nach gemeinsamen Merkmalen im Vordergrund: Existieren neben der vorangehenden expliziten Auswahl oder Spezifikation (z.B. Stichwortsuche oder Metadaten-Filter), welche zu der Menge fokussierter Texte geführt hat, weitere Gemeinsamkeiten der Texte? Welche Themen werden abgedeckt und wie stark?

Anders als beim *Assoziativen Überblick* wird also nicht von Nachbarschaften ausgegangen, um zusammenhängende Textgruppen zu identifizieren, sondern die Verteilung einer Menge von Texten im Themenraum wird adressiert. Zunächst lassen sich die Themen identifizieren, zu denen fokussierte Texte gehören. Zur Abwägung der Abdeckung eines Themas durch die Texte kann visuell innerhalb einer Themengruppe der Quotient aus fokussierten zu nicht fokussierten Texten gebildet werden. Auch die nicht fokussierten Texte in der Nachbarschaft helfen bei der thematischen Einordnung und Abgrenzung, da sie den dargestellten Themenraum mit aufspannen. Alle paarweisen Ähnlichkeiten der fokussierten Texte untereinander und der fokussierten Texte zu anderen Texten spielen also potentiell bei der Charakterisierung der fokussierten Texte eine Rolle.

Faktensicht

Zur Navigation in relationalen Datentabellen stehen meist interaktive Möglichkeiten zum Filtern, Sortieren und Verdichten/Gruppieren von Datensätzen (Zeilen in der tabellarischen Darstellung) zur Verfügung. Entsprechende Operationen der Relationenalgebra werden durch RDBMS⁶⁵, beispielsweise durch SQL, realisiert.

Geht es dem Anwender um das Entdecken von Mustern in den Daten (z.B. Trends, Größenordnungen, Verteilungen, Gemeinsamkeiten), oder um die Untersuchung von Zusammenhängen der Attribute, so wird er die Datensätze zunächst filtern und dann bzgl. der Werte in einer oder mehreren ausgewählten Attributen (Spalten) vergleichen. Dazu kann er mittels Sortierung bezüglich der Attribute Reihenfolgen vorgeben und die Daten visuell prüfen. Beim Prüfen hilft das Verdichten von Datensätzen. Dies kann unterstützt werden durch kompakte („dichte“) grafische Darstellung einzelner Werte und/oder durch Weglassen von Werten (z.B. Verschmelzen bei gleichen Werten). Solche Verdichtungsmöglichkeiten helfen dabei, einen Überblick auch über große Mengen von relationalen Daten zu gewinnen. Grafische Systeme, die clevere Algorithmen zum übersichtlichen Erkunden von Fakten in Tabellen anwenden, sind FOCUS (später InfoZoom[®] [SBB96] und MatrixBrowser [BK02]).

⁶⁵ Relationale Datenbank Management-Systeme

Kategoriensicht

Anders als bei Ähnlichkeits- oder Faktensicht ist die Navigation in der Kategoriensicht eng verbunden (zweckgebunden) mit den klassifizierten Texten. Aufgrund des Schema-Charakters der Kategoriensicht ist es unwahrscheinlich, dass ein Benutzer mit der Kategoriensicht arbeitet, um die Kategorien selbst bzw. ihre Zusammenhänge besser zu verstehen: Zum liegt der Zweck von Kategorien typischer Weise in der Beschreibung von Texten; zum anderen soll ein SWAP-Kategorienschema dem Benutzer vertraut bzw. speziell für ihn oder seine Aufgaben modelliert sein, dürfte also per Definitionem kaum Entdeckungscharakter für den Benutzer haben.

Die Navigation in Hierarchien besteht aus Auswahl (*select*), Verfeinerung (*expand*) und Vergrößerung (*collapse*) von Kategorien. Die hierarchiebasierte Navigation in klassifizierten Textsammlungen lässt sich nach folgenden Kriterien unterscheiden:

- **Trennung von Facetten:** Repräsentiert jede Teil-Hierarchie einen eigenen Gesichtspunkt, die durch die Sub-Kategorien beschrieben wird⁶⁶, oder werden spezielle Begriffe explizit als Kategorien repräsentiert?
- **Kardinalität der Zuordnung:** Wird ein Text durch genau eine Kategorie beschrieben oder kann er durch mehrere Kategorien beschrieben werden (Multi-Klassifikation)?
- **Inklusion der Zuordnung:** Impliziert die Beschreibung eines Textes durch eine Kategorie auch die Zuordnung aller Sub-Kategorien zu dem Text?
- **Inklusion der Auswahl:** Impliziert die Auswahl einer Kategorie die Auswahl aller Sub-Kategorien?
- **Fokus der Auswahl:** Lässt sich bei der Navigation genau eine Kategorie auswählen oder mehrere simultan (Multi-Fokus)⁶⁷?

Häufig werden Facettierte Kategorien-Hierarchien eingesetzt, die für Interaktive Boolesche Anfragen prädestiniert sind. Ziel ist, die Menge damit klassifizierter Texten zu identifizieren. Dies wird erreicht, indem der Benutzer durch Auswahl der Kategorien einen Interessensfokus spezifiziert. In Hearst [He06] repräsentiert die Auswahl mehrerer Kategorien innerhalb einer Hierarchie eine disjunktive Anfrage. Durch Verknüpfung der Hierarchien können daraus konjunktive Anfragen kombiniert werden.

Eine weitere interessante Strategie der Exploration in Kategorienhierarchien ist kognitiv statt physisch-aktiv: Der Benutzer sieht sich die Klassifikations-Struktur zu Texten an, mit dem Ziel, interessante Kategorien zu identifizieren oder die Texte besser zu verstehen. Diese Strategie ähnelt der Strategie Assoziativen Beschreibung in Ähnlichkeitssichten (Kap.5.6), wo es um die Verortung von Texten im Themenraum geht. Eine besondere Rolle spielt dann die übersichtliche, auf den Beschreibungsfokus adaptierte Darstellung der Kategorienhierarchie. In der Literatur werden zur übersichtlichen Darstellung von relevanten Ausschnitten eines Graphen zahlreiche Verfahren vorgeschlagen, die meist dem Prinzip „Focus & Context“ folgen. Treemaps [JS91] zeigen die Größenverhältnisse von Kategorien übersichtlich in rechteckigen Bereichen an – navigiert werden kann mittels Zoom in eine Kategorie. Hyperbolische Browser⁶⁸ [LRP95], inzwischen u.a. aus dem Mind-Mapping bekannt, nutzen den verfügbaren Raum aus, indem ausgehend von der fokussierten Kategorie in der Mitte in alle Richtungen verzweigt wird. Degree-of-Interest Trees (DOI [CN02]) variieren Position und Größe der Darstellung einer Kategorie gemäß einem Maß für deren Bedeutung im aktuellen Analysekontext.

⁶⁶ Bei facettierten Hier. schafft erst die Kombination der Ausprägungen der Facetten einen speziellen Begriff

⁶⁷ Ergänzungsfrage: Sind mehrere Auswahlen im selben Zweig zugelassen?

⁶⁸ Lamping ließ sich inspirieren von der hyperbolischen Geometrie und den Werken M.C. Eschers

5.7 Diskussion

5.7.1 Charakterisierung des Multiple View-Ansatzes

In Kap.1.2 wurde das MultiView-Paradigma vorgestellt. In Kap.3.3 und 3.4 wurden ausgewählte MultiView-Systeme und -Frameworks betrachtet. Der folgende Abschnitt dient der Klassifizierung des MultiView-Konzepts dieser Arbeit.

Die Klasse der SWAP-Systeme zur interaktiven Textanalyse verwendet multiple koordinierte Sichten, die verschiedene Arten von Information übersichtlich darstellen. Die Sichten sind interaktiv sternförmig verknüpft, d.h. es gibt eine zentrale Sicht (Ähnlichkeitssicht von Dokumenten), mit der die anderen Sichten (für strukturierte Daten) interaktiv gekoppelt sind. SWAP-Systeme verwenden dazu eine spezialisierte *Brush&Link*-Metapher: Gemäß dem Prinzip des „*navigational slaving*“ [BWK00] bewirken interaktive Auswahlen in einer Sicht unmittelbar Markierungen in damit verknüpften Sichten, d.h. die Auswahlen werden simultan in andere Sichten propagiert.

Gemäß der Klassifikation der MultiView-Koordination von North & Shneiderman [NS97] lässt sich das Konzept folgendermaßen beschreiben: SWAP-Systeme kombinieren die Analyse *verschiedener* Mengen von Informationsobjekten. Eine Auswahl von Dokumenten, Fakten oder Kategorien führt zu Markierungen in den anderen Sichten: „*selecting items to items*“. Außerdem führt eine Auswahl von Dokumenten zu einer komprimierten (gefilterten) Darstellung der Kategorien und Fakten. Daher handelt es sich bei der Koordination um uni-direktionales „*selecting items to navigating views*“.

Gemäß der Definition von Baldonado et al. [BWK00] sind Textdokumente die „*conceptual entities*“ des MultiView-Systems. Die Sichten stellen verschiedene Aspekte der Dokumente dar und lassen sich als Metadaten der Dokumente auffassen. Insofern wird der Benutzer in die Lage versetzt, interaktive Anfragen über Metadaten durch Auswahl der Informationsobjekte in den Sichten zu formulieren.

5.7.2 Semantik von Präsentation und Interaktion

Die Informationsobjekte aller betrachteten Informationsarten (Textdokumente, Datentupel, Kategorien) werden auf die Objektidentität reduziert. Auch die Beziehungen zwischen Informationsobjekten werden als Relationen repräsentiert, also als Verknüpfungen betrachtet, die ungerichtet, untypisiert und ungewichtet sind. Die Interaktion besteht in der Auswahl von Informationsobjekten durch den Benutzer, die in Markierungen von Informationsobjekten resultieren.

Die *formale Semantik* von SWAP-Systemen wurde in diesem Hauptkapitel beschrieben. Sie lässt sich beschreiben durch

- Mengen von Informationsobjekten,
- Zusammenhängen zwischen den Mengen (s. Koordinationsgleichungen, Def.5.8),
- Teilmengen von Informationsobjekten (Auswahlen u. Markierungen)

Die Reduktion auf rein strukturelle Eigenschaften von Information und Interaktion ist notwendig, weil die Informationsobjekte ohnehin nicht formal beschrieben sind. Die vereinfachte Repräsentation bietet zudem die Vorteile der Unabhängigkeit des Systemkonzepts von Anwendungsdomänen und der Einfachheit und Einheitlichkeit der Navigation. Allerdings hat der Systemdesigner die schwierige Aufgabe, Sichten und Verknüpfungen problemspezifisch so zu konfigurieren, dass Systemdesigner und Systemanwender das *gleiche Verständnis* von Sichten und Sichtenkoordination haben. Daher kommt dem Anpassungsprozess der Referenztechnologie (Kap.2.4 und 6.8) eine besonders wichtige Rolle zu, denn falsch interpretierte Informationen

und Verknüpfungen sind nicht nur nutzlos, sondern können sogar zu falschen Entscheidungsgrundlagen führen. Zur korrekten Interpretation der Systemkonfiguration trägt bei:

- Die (aus der Sicht des Anwenders) *intuitive Modellierung* der Sichten und Verknüpfungen
- Die *explizite Beschreibung* (z.B. natürlichsprachlich) der Semantik der Sichten und der zur Koordination verwendeten Verknüpfungen (z.B.: „Personen lesen Dokumente“)
- Die geeignete Schulung des Anwenders
- Werkzeuge zur Inhaltserschließung (z.B. Statistiken für Terme oder relationale Daten), die Informationsobjekte besser zu verstehen und zu interpretieren.

5.7.3 Abgrenzung

Bei Diskussionen und Präsentationen im Zusammenhang mit der Arbeit stellte sich oft die Frage, wie das Systemkonzept sich zu den Themen OLAP und Datenintegration abgrenzen lässt. Diese Frage wird im Folgenden kurz adressiert.

OLAP

Bei Analysen von Geschäftsprozessen werden Kennzahlen unter verschiedenen Gesichtspunkten (Dimensionen) wie Produkt, Zeit, Ort, Kunde betrachtet. OLAP (Online Analytical Processing) ist ein in den 90er Jahren entwickeltes und längst etabliertes Zugriffs-Paradigma mit dem Ziel, komplexe ad-hoc Anfragen über konsolidierten Geschäftsdaten interaktiv beantworten zu können. Cody et al. [CK+02] integrieren Kategorien zur Textbeschreibung in Multidimensionale Datenmodelle zur Kennzahlenanalyse. Anwender können OLAP-artige Anfragen im integrierten Datenmodell stellen. Im Folgenden werden Gemeinsamkeiten zwischen interaktiven SWAP-Anfragen und OLAP-Anfragen diskutiert.

Tabelle 5-3: Analogien von OLAP-Systemen zur Datenanalyse und SWAP-Systemen zur Textanalyse

| | OLAP | SWAP |
|--------------------|--|--|
| Ziele | „Information auf Knopfdruck“ (online; ad-hoc, explorativ) dispositiv (statt operativ) | |
| Daten | Daten aus DWH laden (speziell für Analyse angelegt, ETL) | Daten aus operativen Systemen laden (Sichten-Konfiguration; ggf. Materialisierung logischer Verknüpfungen) |
| Modell | Dimensionen | Sichten |
| | Kennzahlen, Werte, Einträge | Texte |
| | Kategorien | Kategorien, Datentupel |
| | Dimensions-Hierarchien | Sichten-interne Beziehungen zw. Informationsobjekten |
| | Stern-Schema (Faktentabelle vs. Dimensionstabellen) | Stern-Schema (Textsicht vs. andere Sichten) |
| Operationen | Report-Spezifikation (Multidimensionaler Spezifikationsvektor) | SWAP-Systemzustand (Auswahlen von Informationsobjekten in Sichten) |
| | Auswahl mit Multiplem Fokus | |
| | Auswahl in jeder Dimension | Auswahl in beliebig vielen Sichten |
| | Aggregation von Werten | Beschreibung von Textgruppen (Phrasen, Begriffe, verknüpfte IO anderer Sichten) |
| | Dimensions-interne Navigation (Roll-Up, Drill-Down) | Sichten-interne Navigation |

Wie beim OLAP-Paradigma zur Datenanalyse werden bei SWAP-Systemen zur Textanalyse interaktive ad-hoc Anfragen in mehreren Dimensionen (Sichten) unterstützt. Die in den Sichten dargestellten Zusammenhänge zwischen Informationsobjekten sind, analog zu Drill-Down- bzw. Roll-Up-Operationen innerhalb von OLAP-Dimensionen, zur Sichten-internen Navigation nutzbar (vgl. Kap.5.6.2). Ergebnis der Spezifikationsaktivitäten mit SWAP-Systemen sind keine Kennzahlenberichte, sondern Textauswahlen. In beiden Paradigmen geht es um die Datenverdichtung: Bei SWAP-Systemen werden keine Gruppen von Kennzahlen gemäß Verrechnungsfunktionen aggregiert, sondern Gruppen von Texten zusammengefasst. Tabelle 5-3 fasst konzeptionelle Analogien und Unterschiede zwischen OLAP- und SWAP-Systemen zusammen.

Datenintegration

Das SWAP-Konzept und das Konzept der Datenintegration sind in vielerlei Hinsicht komplementär, wie Tabelle 5-4 zeigt. Während das SWAP-Konzept integrierte Informationen für die interaktive Analyse voraussetzt, ist die Modellierung von Querbezügen zwischen heterogenen Daten gerade Gegenstand der Datenintegration. Die Verknüpfungen werden dabei im Gegensatz zu SWAP-Systemen auf Schema-Ebene beschrieben und werden meist in einer Logik formalisiert, während sich das SWAP-Systemkonzept auf einfache Relationen zwischen Instanzen (Informationsobjekten) konzentrieren. Der Anwender (Suchende) spielt bei der Datenintegration keine unmittelbare Rolle. Ziel ist die Unterstützung eines Anfragesystems und Verbergen der Heterogenität der zugrunde liegenden verteilt abgespeicherten Daten (Anfragetransformation). Bei SWAP-Systemen ist es umgekehrt: Der Anwender soll gerade in die Lage versetzt werden, Querbezüge zwischen zu erkennen und zur Exploration der Daten zu nutzen.

Tabelle 5-4: Die Bedeutung der klassischen Datenintegration für das SWAP-Konzept

| | Datenintegration | SWAP |
|----------------------|---|--|
| Ziele | Integrierter Zugriff auf heterogene Datenquellen | |
| | Automatisierte Anfrageausführung durch System | Interaktive Exploration durch Experten |
| Daten | Quellen bleiben unverändert und unabhängig bestehen | |
| | Schwerpunkt strukturierte (modellerte) Daten | Schwerpunkt Texte und deren Metadaten |
| Verknüpfungen | zwischen Datenmodellen (Schema-Ebene) | zw. Informationsobjekten (Instanz-Ebene) |
| | Berechnung ist Gegenstand | Berechnung ist Voraussetzung |
| | logisch | strukturell (Relationen) |
| | Formale Semantik durch Maschine interpret. | Verknüpfungen durch Experten interpretiert |
| | vor dem Anwender verbergen | dem Anwender aufzeigen |

Mit einem SWAP-System werden Querbezüge zwischen Texten und andersartigen Informationen untersucht. Sind diese Querbezüge nicht physikalisch vorhanden, so ist ein Datenintegrationsschritt zur Vorverarbeitung notwendig. Einige Methoden und Werkzeuge werden bei der Systembeschreibung diskutiert (vgl. Kap.6.7).

6 Referenztechnologie SWAPit

6.1 Entstehung

Das Werkzeug SWAPit baut auf den Prototyp DocMINER [Be01] zur Erforschung der Arbeit mit Dokumentenlandkarten⁶⁹ auf. Chronologie und Gegenstand der Entwicklung von SWAPit wurde bestimmt durch die zahlreichen Fallstudien, die nicht selten neue Anforderungen mit sich brachten (Kap.1.5 u. 7.2). Die Einführung eines MultiView-Konzeptes war motiviert die Beobachtung, dass der Zugriff auf Texte stets in einem Prozess- und Informationskontext geschieht, der in die Arbeit einbezogen werden muss. Damit SWAPit den Charakter eines „Werkzeugkastens“ für verschiedene Probleme der Textanalyse erhielt, mussten Sichten flexibel anpassbar sein. Matchingfunktionen zur Relevanzbewertung von Markierungen von Informationsobjekten (Kap. 5.4) wurden dem Auswahlkonzept hinzugefügt. Zur komfortablen Arbeit wurde eine Web-basierte Client-Server-Lösung benötigt, die sich leicht an die vorhandene Unternehmens-IT ankoppeln lässt.

6.2 Architektur

SWAPit/DocMINER ist als Client-Server Architektur realisiert. Der DocMINER-Webservice stellt Funktionalitäten Verwaltung, Text Mining und Retrieval zur Verfügung, die teilweise generisch und zum Teil SWAPit-spezifisch sind. Der Webservice ist in C++ entwickelt und wird als CGI-Skript deployed. SWAPitUp, das JSP⁷⁰-basierte Administrationswerkzeug, nutzt den DocMINER-Webservice zur Projekterstellung und -verwaltung. Insbesondere realisiert SWAPitUp ein kooperatives Autorisierungskonzept, das auf Gruppen, Rollen und Rechten basiert. SWAPit, die Client-Anwendung, ist als Java Applet oder Java-Programm nutzbar. SWAPit stellt die mit SWAPitUp oder einem anderen Administrationswerkzeug angelegten Projekte dar und bietet reichhaltige Interaktionsmöglichkeiten zur Exploration und Suche in Texten und in damit verknüpften strukturierten Daten.

Bei der gesamten Entwicklung wurde großer Wert auf die flexible technische Anpassbarkeit der Projekterstellung und -bereitstellung gelegt. So lassen sich Texte und Daten aus nahezu beliebigen Dokumenten- und Datenquellen verarbeiten, wobei Spezialsoftware entworfen werden kann, welche die Webservice-API nutzt. Anschließend kann SWAPit als „Viewer“ in Unternehmensportale oder Webseiten eingebunden werden. Für die Verteilung der Server und Datenbanken gibt es zahlreiche Möglichkeiten.

SWAPit ist *intern* organisiert in jeweils ein Package für jede Sichtenart, ein Package für die Kommunikation mit dem DocMINER Webservice, sowie Packages für Ereignisbehandlung und Sichtenkoordination. Hervorzuheben ist das erweiterbare Koordinationskonzept (Abbildung 6.2). Jede Sicht ist modular realisiert: Die Logik, welche Art von Informationsobjekten auf welche Weise angezeigt und interaktiv zugreifbar gemacht wird, ist darin verkapselt. Ebenso obliegt die Abbildung der Informationsobjekte einer Sicht auf URLs jeder Sicht.

⁶⁹ Dokumentenlandkarten sind Clusterdarstellungen von Textdokumenten, welche die paarweisen inhaltlichen Ähnlichkeiten der Dokument auf räumliche Nähe von Dokumentenpunkten abbilden.

⁷⁰ Java Server Pages™

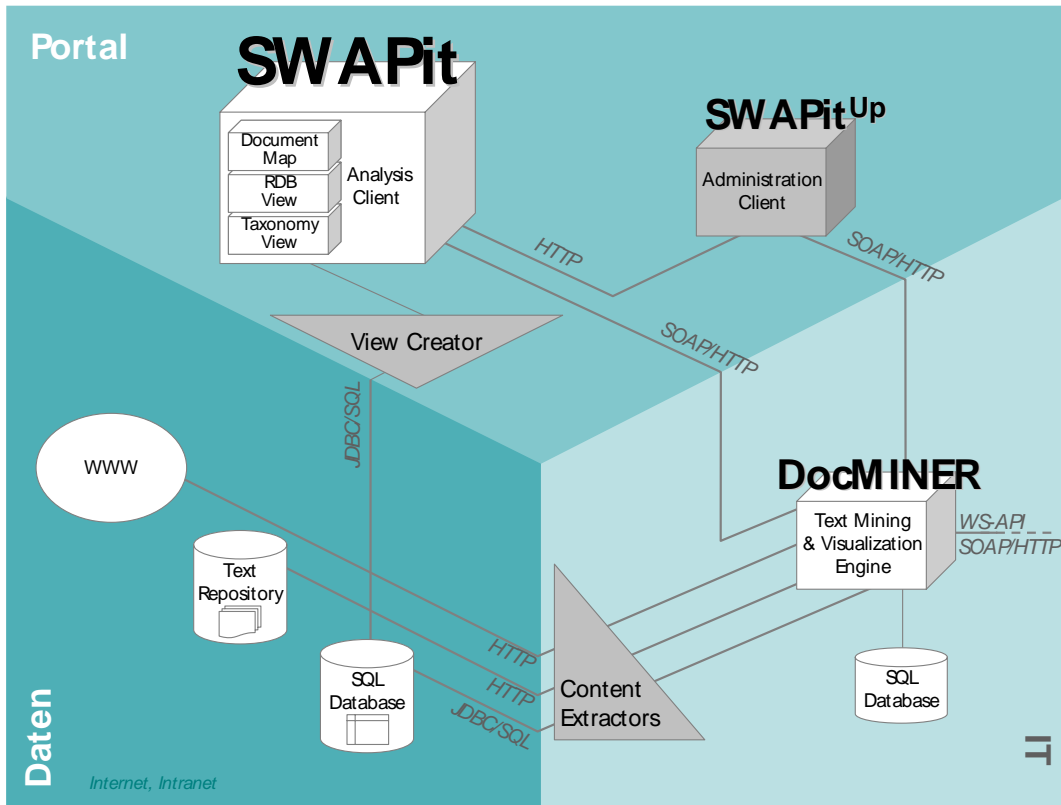


Abbildung 6.1: Systemarchitektur von DocMINER (Server) und SWAPit (Client)

Die Schnittstelle zum Gesamtsystem besteht im Austausch von URLs: Jede Sicht implementiert das Interface *MarkingComponent*. Beim Systemstart registriert sich jede Sicht beim statischen *ComponentManager*. Dabei wird jeder Sicht eine Farbe zugeordnet. Nimmt der Benutzer in einer Sicht eine interaktive Auswahl von Informationsobjekten vor, so wird dem *ComponentManager* Mitteilung über die betroffenen URLs gemacht. Dieser gibt die Information an alle als *MarkingComponent* registrierten Sichten weiter, die darauf entsprechend reagieren können. Die Klasse *SelectionManager* kann von jeder Sicht genutzt werden, um Informationsobjekte, ihre Verknüpfungen mit Dokumenten, sowie interaktive Auswahlzustände zu speichern und Aktivierungen von Dokumenten oder Informationsobjekten zu berechnen.

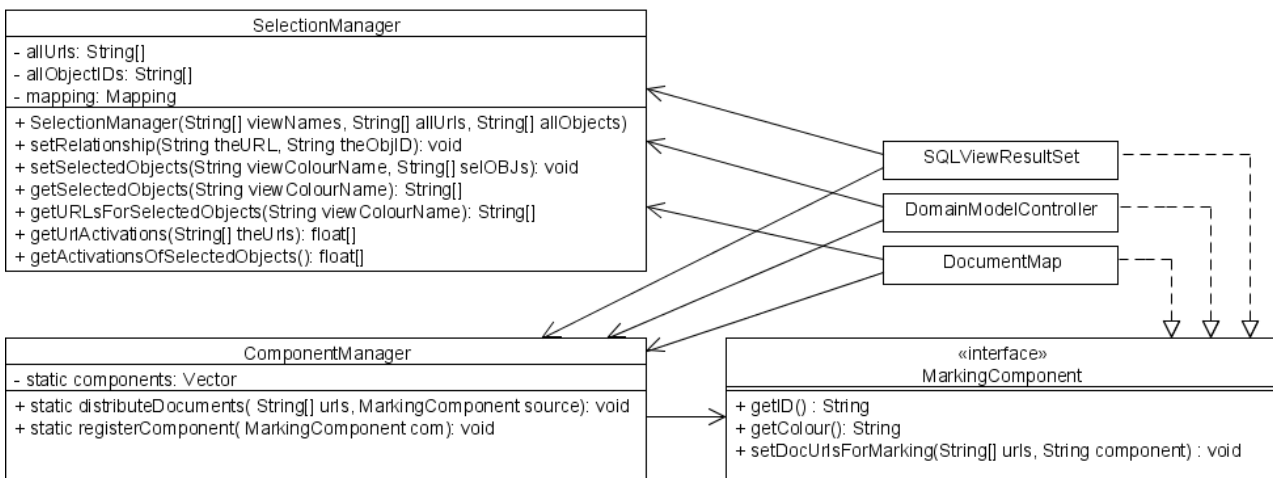


Abbildung 6.2: Sichtenkoordination in SWAPit (UML-Klassendiagramm)

Auch die Werkzeugleiste zur Suche in Texten ist wie eine Sicht realisiert. Die vorhandenen Sichten sind für verschiedenste Szenarien. Neue Sichten, beispielsweise die Darstellung von Themen durch Topic Maps, oder eine Darstellung von Personen, die kooperativ an Dokumenten arbeiten, lassen sich mit diesem Konzept ohne großen Programmieraufwand einfügen.

6.3 Administration mit DocMINER und SWAPitUp

SWAPitUp ist ein webfähiges kooperatives System zur Erstellung und Verwaltung von SWAPit-Projekten. Ein autorisierter Benutzer kann Mitglied in verschiedenen Arbeitsgruppen mit verschiedenen Rollen sein, andere Benutzer zu seinen Projekten einladen, sowie Projekte anlegen, bearbeiten und abrufen.

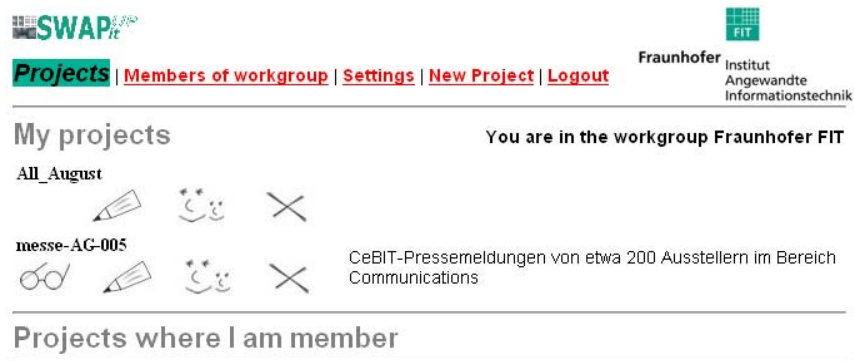


Abbildung 6.3 Administration von Projekten und Benutzern in SWAPitUp

Der Spezifikationsprozess für ein neues Projekt wird in SWAPitUp durch einen Fortschritts- und Navigationspfeil symbolisiert, dem der Benutzer folgen kann. Die Schritte sind: Projektname und –Beschreibung definieren, Textdokumente spezifizieren, Dokumentenanalyse und -Vergleich festlegen (u.a. Indexierungsparameter sowie interaktive Optimierung von Positiv- und Negativlisten), Sichten spezifizieren, Berechnungsparameter der Landkartenberechnung festlegen, Berechnung starten.

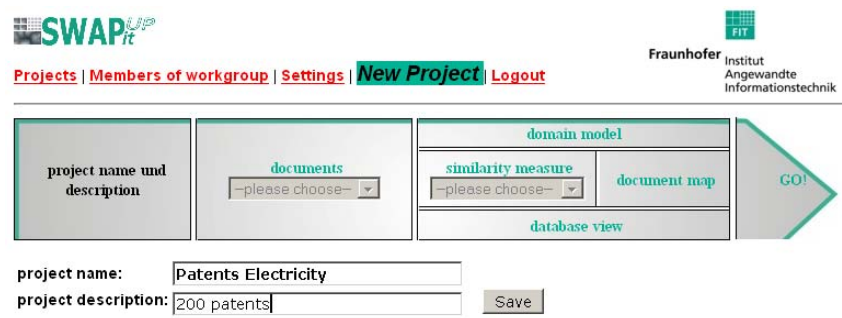


Abbildung 6.4: Erstellen neuer Projekte in SWAPitUp

Alternativ zu SWAPitUp können eigene Programme entworfen werden, die plattformunabhängig auf alle Funktionalitäten des DocMINER-Webservice zugreifen können. Diese umfassen die Funktionsbereiche Verwaltung (Projekte, Metadaten, Sichtenspezifikation), Text Mining (Indexierung, Textanalyse, Clustering) und Retrieval.

6.4 Koordination der Sichten

SWAPit ist ein SWAP-System gemäß den Spezifikationen aus Kap.5. Jede Sicht ist mit der Farbe assoziiert, mit der sie gekennzeichnet ist (Umrandung). Wählt der Benutzer Informationsobjekte der Sicht explizit aus, so wird die Auswahl in der Farbe der Sicht hervorgehoben. Gleichzeitig werden alle Informationsobjekte anderer Sichten, die mit ausgewählten Informationsobjekten (direkt oder transitiv über eine weitere Sicht) verknüpft sind, ebenfalls in der Farbe markiert. Bei Auswahlen in allen Sichten entsteht ein Bild wie in Abbildung 6.5, wo alle Informationsobjekte, die mit der Auswahl einer Sicht zu tun haben, an der gleichen Farbe erkennbar sind. Auch die Suchergebnisse haben eine eigene Farbe. In jeder Sicht wird jede Farbe separat dargestellt, so dass Überlappungen von Markierungen und Auswahlen möglich und erkennbar sind.

Die Ähnlichkeitssicht nimmt für die Sichtenkoordination eine zentrale Rolle ein, da Informationsobjekte aller anderen Sichten mit den Informationsobjekten dieser Sicht, den Texten, verknüpft sein können, aber nicht untereinander. Damit wie bei gerankten Suchergebnislisten eine schnelle Fokussierung auf die potenziell wichtigsten Informationsobjekte möglich ist, kann für jede Markierung in jeder Farbe eine Relevanzabstufung (Matching-Werte) angezeigt werden. Auf welche Weise die einzelnen Sichten diese Zusatzinformation visualisieren, wird in den jeweiligen Unterkapiteln zu den Sichtenarten erläutert.

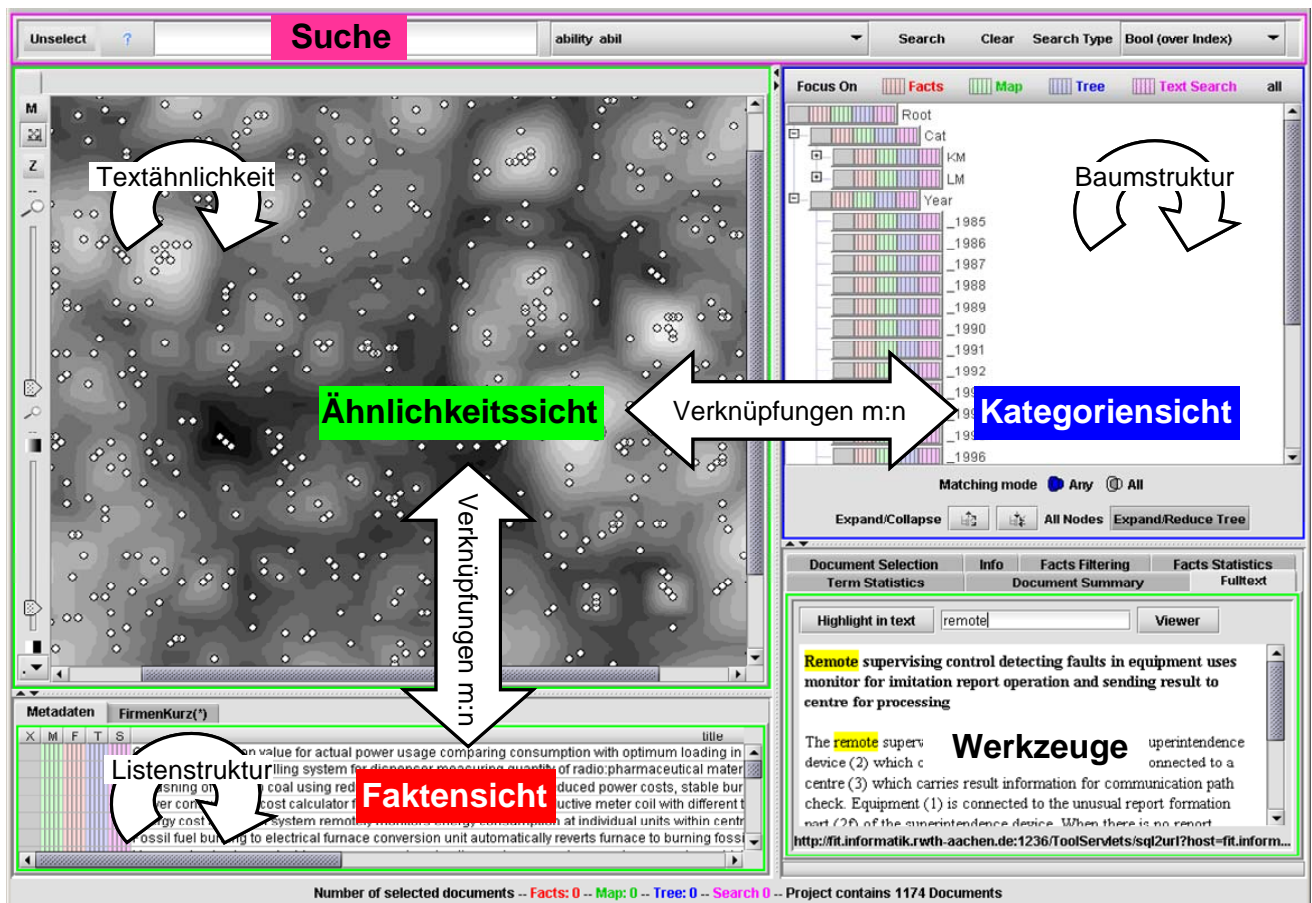


Abbildung 6.5: Sichtenkoordination in SWAPit: Verknüpfungen beziehen sich sternartig auf die zentrale Textsicht.

6.5 Repräsentation der Informationsarten in Sichten

6.5.1 Textähnlichkeiten

Darstellung

Für die Darstellung der paarweisen inhaltlichen Ähnlichkeit von Textdokumenten verwendet SWAPit Dokumentenlandkarten [Be01]. Jeder Punkt auf der Karte entspricht einem Textdokument. Je näher zwei Punkte räumlich benachbart sind, desto stärker ähneln sind die Textinhalte. Die Graustufeninformation gibt Aufschluss darüber, wie spezialisiert eine Region der Karte auf die darauf abgebildeten Dokumente ist und lässt sich folgendermaßen interpretieren: Helle Bereiche kennzeichnen Themen, also Gruppen von stark ähnlichen Dokumenten, während dunkle Bereiche als thematische Grenzen betrachtet werden können. Eine solche Darstellung (zwei räumlichen Dimensionen und eine grafische Dimension) wird als 2½D-Visualisierung bezeichnet. Anders als bei gängigen Clusterverfahren wurde bewusst auf die Partitionierung der Dokumentenmenge verzichtet zugunsten einer „weichen“ Darstellung, die in Abstufungen erkennen lässt, *wie homogen* Bereiche sind, und *wie sehr* sich Texte ähneln. Eine Verringerung der Anzahl von Abstufungen kann in SWAPit durch eine Art Kontrastregler erreicht werden: Abbildung 6.6 zeigt dieselbe Karte in drei Variationen der Darstellung.

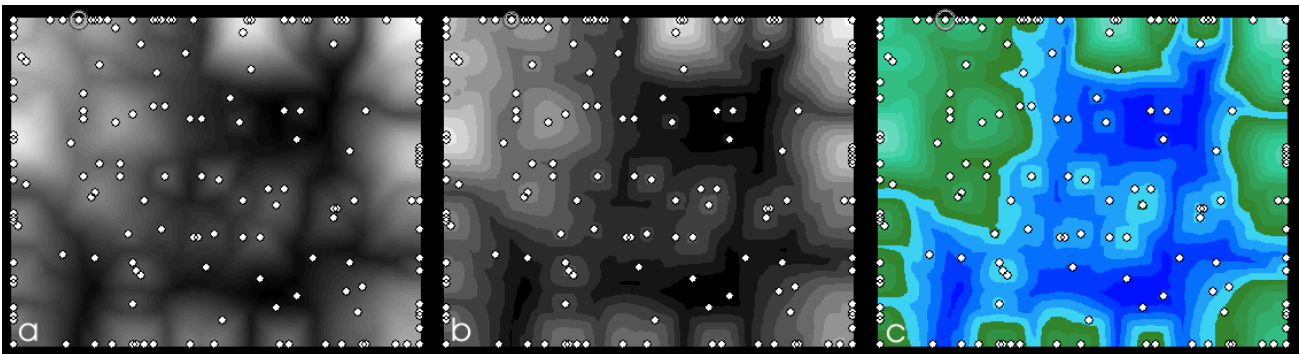


Abbildung 6.6: Darstellung von Dokumentenlandkarten: (a) 256 Graustufen, (b) 12 Graustufen, (c) 12 Farben

Die Erfahrung zeigt, dass bei Überblick-orientierten Aussagen wie der Identifikation von Kernthemen die diskretisierte Darstellung (b) gegenüber (a) helfen kann, den Vergleich der Helligkeit verschiedener Kartenbereiche zu erleichtern. Eingefärbte Karten (c), die von vielen ähnlichen Ansätzen⁷¹ verwendet werden, mögen zwar optisch ansprechend und intuitiv wirken, bergen aber die Gefahr der Fehlinterpretation, da beim Übergang einer Grundfarbe oft eine viel stärkere Trennung empfunden wird, als in den Daten tatsächlich vorhanden. Zudem wird es erheblich schwieriger und anstrengender, Farben von Auswahl- und Markierungssymbolen wahrzunehmen – ein wichtiger Aspekt in Anbetracht der Tatsache, dass in SWAPit jede Sicht durch mit einer Farbe assoziiert ist. Technisch wird die Clusterdarstellung mit Hilfe eines Neuronalen Kohonen-Netzes⁷² erreicht. Das Maß zur Bestimmung der Ähnlichkeiten ist austauschbar, so dass sowohl Dokumente sowohl Metadatenbasiert als auch Inhaltsbasiert miteinander verglichen werden können.

⁷¹ *Themescape* der Fa. Aureka (<http://www.micropatent.com/static/aureka.htm>), *VxInsight* der Fa. VisWave (http://www.metricsgroup.com/svc_3dplm.cfm)

⁷² Self Organising Feature Map (SOFM)

Auswahl und Markierung von Dokumenten

Ausgewählte Dokumentenpunkte werden grün ausgefüllt. Markierungen von Dokumenten werden durch farbige Symbole repräsentiert. Das ursprüngliche System war nicht in der Lage, Relevanzwerte für Markierungen von Informationsobjekten anzuzeigen, sondern stellte sich überlappende Markierungen von Texten auf der Karte als ineinander verschachtelte konzentrische Formen dar. Bei der Erweiterung auf die zusätzliche Darstellung von Matchingwerten wurde sehr sorgfältig vorgegangen, um den kognitiven Overhead trotz der Darstellung der zusätzlichen Information gering zu halten. So wurden drei prototypische Darstellungskandidaten entworfen (Abbildung 6.7).

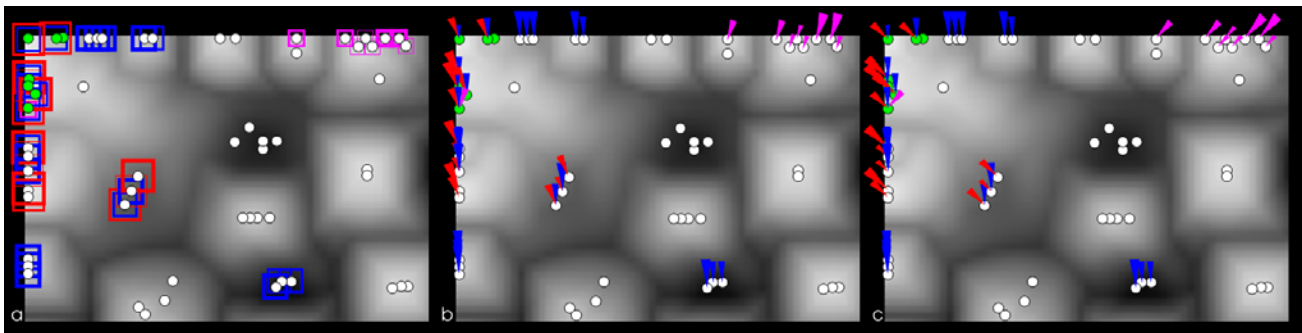


Abbildung 6.7: Drei Varianten der Darstellung von Matchingwerten: (a) konzentrische verschachtelte Rechtecke mit verschiedenen Linienstärken (b) Keile verschiedener Länge (kleine Winkelabstände) (c) Keile verschiedener Länge (große Winkelabstände)

Diese Varianten der Darstellung wurden im Rahmen einer Befragung von 10 Mitarbeitern des Lehrstuhls Informatik 5 (Informationssysteme) der RWTH Aachen unter verschiedenen Gesichtspunkten bewertet. Der Fragenkatalog (Anhang 9.5) adressierte folgende Fragen:

- Kann die zweiwertige Information (ausgewählt bzw. nicht ausgewählt) immer noch ebenso gut wahrgenommen werden wie zuvor?
- Übersteigt der wahrgenommene Nutzen der Matching-Werte die zusätzliche kognitive Belastung?
- Bei welcher Darstellung ist die Ausgewogenheit zwischen Übersichtlichkeit, Detailtreue (Zuordnung u. Abstufung) und kognitive Belastung am größten?
- Welche Darstellung ist optisch am ansprechendsten?

Zunächst musste jede Variante einzeln mit der ursprünglichen zweiwertigen Darstellung verglichen werden (Teil 1). Dann mussten die Varianten direkt verglichen werden (Teil 2). Zur Bewertung im ersten Teil standen den Testpersonen 3er-Skalen zur Verfügung. Im zweiten Teil sollten sie die drei Systeme bzgl. jeder Frage in eine Reihenfolge bringen.

Für Teil 1 ist die Anzahl der Antworten auf der 3er-Skala (schlechter / genauso gut / besser als die zweiwertige Darstellung) in Abbildung 6.8 dargestellt. Der Überblick über Auswahlen in verschiedenen Farben wird für Darstellung A (vgl. Abbildung 6.7) meist mindestens genauso gut eingeschätzt wie bei der Originaldarstellung, während bei den Keildarstellungen (B,C) die Meinungen polarisiert sind. Anders bei der Detaildarstellung, also der Frage, wie gut sich für einzelne Dokumente Farbzusordnungen erkennen lassen, wo die Keildarstellungen insgesamt positiv gewertet werden, insbesondere Variante C. Die kognitive Belastung ist erwartungsgemäß bei allen Darstellungen höher als bei der zweiwertigen Originaldarstellung. Am geringsten wird die zusätzliche Belastung bei Variante C empfunden.

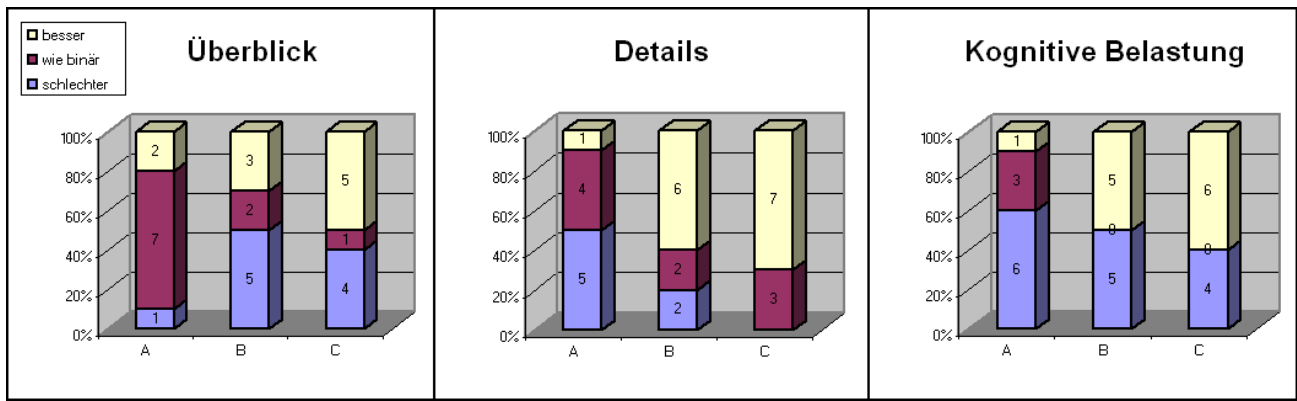


Abbildung 6.8: Befragung zur Darstellung von Matching-Werten, Teil 1 (separate Einschätzung jeder Darstellung)

Bei Teil 2 wurden die von den Teilnehmern für jedes Kriterium angegebenen Reihenfolgen der Varianten mit Punkten bewertet (-1, 0, +1) und in Abbildung 6.9 dargestellt. Bei fast allen Kriterien wird Variante C am besten eingestuft. Allerdings wird C als etwas unruhiger empfunden als A u. B.

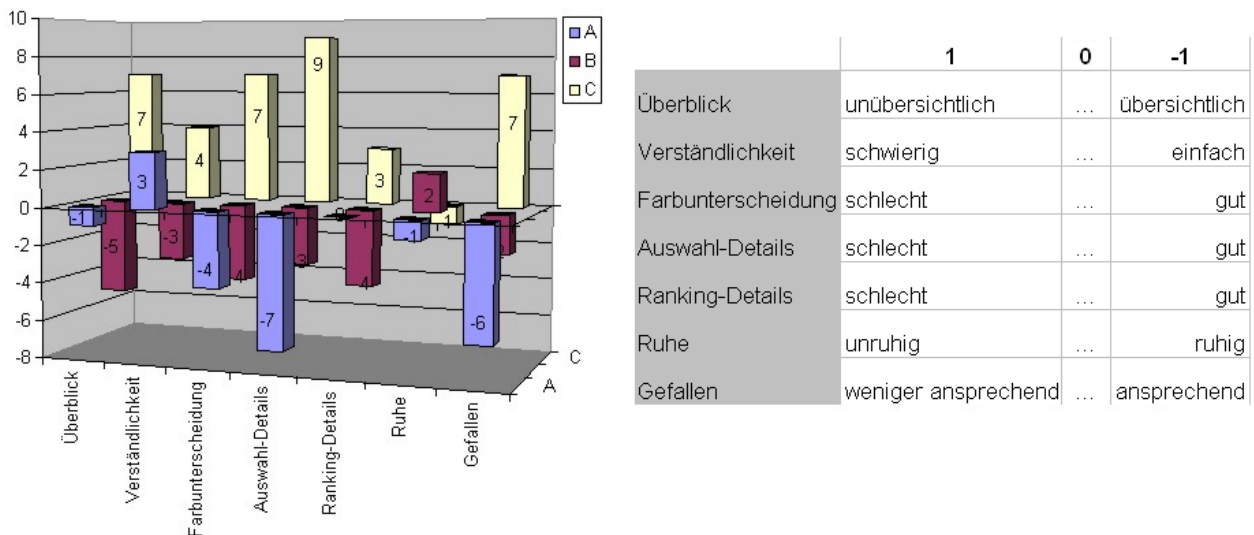


Abbildung 6.9: Befragung zur Darstellung von Matching-Werten, Teil 2 (vergleichende Einschätzung jeder Darstellung)

Die Ergebnisse der Umfrage bieten eine differenzierte Betrachtung und bestätigen auch, dass große Unterschiede in der Wirkung grafischer Informationsdarstellungen wahrgenommen werden, die solche Erhebungen notwendig machen. Insgesamt legt das Umfrageergebnis die Wahl von Variante C (Keile mit großen Winkelabständen) zur Darstellung nahe zugunsten der besonders wichtigen Kriterien Detailwahrnehmung und geringe kognitive Belastung, unter der Inkaufnahme von Einschränkungen in den Kriterien Übersichtlichkeit und empfundene Ruhe der Darstellung.

- Im Freitext-Teil der Umfrage wurden einige hilfreiche Kommentare erhoben, die zum Teil Zwecks Optimierung von Variante C umgesetzt wurden:
- Vergleich und Unterscheidung der Länge von Keilen ist für das Auge schwierig (Jeder Keil wurde daraufhin in 5 Fragmente unterteilt)
- Überlappende Keile sind schwer zu unterscheiden (Daraufhin wurde eine helle Umrandung jedes Keils eingeführt)

Die an den Kartenrändern vermehrt auftretenden Überlappungen erschweren die Unterscheidung insbesondere vertikaler bzw. horizontaler Keile (Daraufhin wurden die Winkel so gewählt, dass die Winkel zur hor./vert. Achse maximiert wurden)

Die optimierte Darstellung verwendet Keile (Abbildung 6.10), die in 90°-Abständen zueinander und maximalem Achsabstand um den zu kennzeichnenden Dokumentenpunkt angeordnet sind. Die Matchingwerte jeder Farbe werden durch die Länge der Keile dargestellt, wie im rechten Teil der Abbildung beispielhaft gezeigt wird.

Übersicht

Auf einer Dokumentenlandkarte sind alle Dokumente zugleich sichtbar. Zur Fokussierung eines bestimmten Kartenbereichs stehen verschiedene Zoom-Möglichkeiten zur Verfügung.

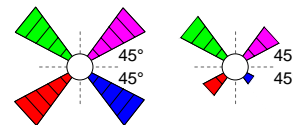


Abbildung 6.10: Darstellung von Markierungen mit Matching-Werten

6.5.2 Beschreibungskategorien

Darstellung

Für die Darstellung von hierarchisch gegliederten Beschreibungskategorien zu Texten verwendet SWAPit eine ‚klassische‘ lineare Baumdarstellung, in der sich mehrere Knoten gleichzeitig auswählen lassen. Die Beschreibungskategorien stellen eine Multi-Klassifikation der Dokumente dar: Jeder Kategorie können mehrere Dokumente zugeordnet sein und jedem Dokument mehrere Kategorien. Bei der Verknüpfung jedes Knotens mit Textdokumenten gilt keine Inklusion, d.h. Dokumente, die einer Kategorie zugeordnet sind, sind nicht automatisch den Oberkategorien zugeordnet. Grund für die Realisierung dieses Systemverhaltens ist der Wunsch unterscheiden zu können, ob eine Dokument einer Kategorie *explizit* zugeordnet ist oder einer der Unterkategorien (z.B. Zuordnung von Gesetzestexten zu Bund oder Ländern). Allerdings gibt es viele Fälle, in denen eine Kategorie genau die Menge aller Unterkategorien repräsentiert (z.B. Gruppe von Personen). Letztlich hängt die Frage, ob Inklusion gelten soll oder nicht, von der Anwendung und der Modellierung der Kategorienhierarchie ab. Um beide Möglichkeiten zuzulassen, werden bei der Auswahl einer Kategorie alle Unterkategorien standardmäßig mit ausgewählt, können aber deaktiviert werden.

Auswahl und Markierung von Kategorien

Die Beschriftung ausgewählter Kategorien wird blau hervorgehoben (Abbildung 6.11). Jedem Knoten ist eine Farbleiste zur Darstellung von Markierungen zugeordnet, in der für jede Sichtenfarbe ein Platzhalter vorgesehen ist. Matchingwerte zu Markierungen werden durch 5-stufige Amplituden angezeigt. Mehrere untereinander angeordnete Farbleisten lassen sich aufgrund der festen Struktur der Farbleiste gut miteinander vergleichen.

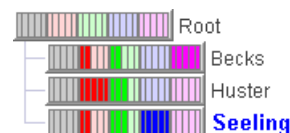


Abbildung 6.11: Auswahl und Markierung von Beschreibungskategorien

Übersicht

Je größer die Hierarchien von Beschreibungskategorien sind, desto unübersichtlich wird die Darstellung. Ein interaktiv aktivierbarer Algorithmus zur Baumkompression [Br04] wirkt als Filter für Baumknoten und entfernt unbenötigte Knoten: Im Baum verbleiben nur die Knoten, die für eine Menge vorgegebener Farben mindestens eine Farbmarkierung aufweisen, sowie alle Knoten auf dem Pfad von der Wurzel dorthin. So lassen sich beispielsweise leicht alle Baumkategorien anzeigen, die ein Suchergebnis beschreiben. Während die klassische Baumdarstellung der Vorteil der guten Vertrautheit bei den meisten Nutzern hat, stehen zahlreiche grafische Alternativen zur Verfügung, die alle Kategorien auf einen Blick darstellen können, z.B. TreeMaps [Sn92] oder Hypergraphen. DOI⁷³-Bäume führen situationsbezogen eine Bewertung der Relevanz von Kategorien durch und stellen wichtige Kategorien entsprechend größer dar. Eine lohnende Untersuchung wäre die Beantwortung der Frage, inwiefern solche Darstellungen die vorhandene Darstellung ergänzen oder ersetzen können.

⁷³ Degree-of-Interest, vgl. Card et al. [CN02]

Motiviert durch die dritte Fallstudie dieser Arbeit wird als Erweiterung der Ähnlichkeitssicht ein Prototyp für eine grafische Sicht entwickelt, die jede Tabellenspalte als eigene Facette repräsentiert (FacetTable, Kap.7.5.5).

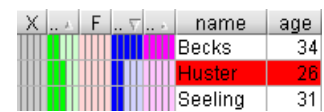
6.5.3 Relationale Daten

Darstellung

Relationale Daten werden als Tabelle dargestellt: Spalten repräsentieren Datenfelder und Zeilen repräsentieren Datensätze. Jeder Zeile können mehrere Dokumente zugeordnet sein und jedem Dokument mehrere Tabellenzeilen. Spalten lassen sich aus-/einblenden und nach Werten sortieren.

Auswahl und Markierung von Datentupeln

Ausgewählte Tabellenzeilen werden rot hervorgehoben. Ähnlich der Baumdarstellung wird jeder Tabellenzeile eine Farbleiste zugeordnet, um Markierungen und deren Aktivierungen anzuzeigen (Abbildung 6.12). Jede Farbamplitude zeigt den Matchingwert als Pegel an. Die Besonderheit der Tabelle gegenüber dem Baum ist, dass die Zeilenreihenfolge verändert werden kann. Diese Möglichkeit macht sich die Multi-Farbsortierung zunutze: Jede Farbe oder Farbkombination kann als Sortierkriterium verwendet werden. Dadurch entstehen flexible Rankings. So lässt sich beispielsweise leicht herausfinden, welche Firmen die meisten der thematisch ausgewählten Patente hervorgebracht haben.



| X | F | name | age |
|---|---|---------|-----|
| | | Becks | 34 |
| | | Huster | 28 |
| | | Seeling | 31 |

Abbildung 6.12: Auswahl und Markierung von Relationalen Datentupeln

Übersicht

Je mehr Datensätze vorhanden sind, desto unübersichtlicher wird die Tabelle. Hilfestellung bietet einem Tabellen-Filter, mit dem sich Wertebereiche von Feldern einschränken und somit uninteressante Datensätze ausschließen lassen. Die Farbsortierung hilft, fokussierte Datensätze in den sichtbaren Betrachtungsbereich zu befördern. Graphische Metaphern zur übersichtlichen Darstellung großer Tabellen könnten helfen, die Exploration von Relationalen Daten weiter zu verbessern. In Frage kommen dabei alle Arten von Grafiken, die im Data Mining eingesetzt werden oder Systeme, die für die interaktive Erkundung von Datentabellen entworfen wurden (Kap.4.4.4).

6.6 Interaktive Werkzeuge

Ergänzend zur Auswahl von Texten oder strukturierten Daten lassen sich mit SWAPit Texte und Daten spezifizieren (suchen/filtern) und beschreiben. Kompakte grafische Darstellungen oder Zusammenfassungen von relationalen Daten sind allgemein bekannt. Analog dazu lassen sich in SWAPit auch Texte beschreiben, beispielsweise durch Termstatistiken oder Zusammenfassungen. Abbildung 6.13 zeigt die Dialoge von SWAPit: Grün umrandet sind Werkzeuge zur Arbeit mit Texten (a-d). Die Werkzeuge für Relationale Daten (e-f) kennzeichnet ein grüner Rand und der Suchdialog ist Magenta eingefasst (g). Die Dialoge (a, f-g) dienen der Spezifikation von Texten bzw. Daten durch den Benutzer und zielen auf eine Verkleinerung der zu betrachtenden Informationsmenge ab. Die Dialoge (b-e) helfen dabei, Text und Daten zu charakterisieren und zielen darauf ab, die Notwendigkeit der detaillierten Untersuchung einzelner Texte oder Datensätze durch Hilfsmittel zum „Querlesen“ aufzuheben.

Dokumentenlegende (a): Gruppen von Dokumenten können interaktiv beschriftet und bei Bedarf wieder angezeigt oder abgerufen werden. Beim Anlegen von Legenden-Einträgen kann die fokussierte Farbmarkierung kann in einer Auswahl-Werkzeuggeste bestimmt werden. Jedes Dokument kann mit mehreren Farben und dazugehörigen Beschriftungen gekennzeichnet werden. Dokumentengruppen lassen sich anzeigen (Na-

delsymbol auf der Landkarte) und in eine (grüne) Dokumentenauswahl auf der Karte transformieren, wodurch die Dokumente zur Interaktion mit anderen Werkzeugen reaktiviert werden. Legendensymbole werden persistent im DocMINER-Webservice gespeichert und stehen damit auch als Annotation für die kooperative Textanalyse zur Verfügung. Die Legende hilft beispielsweise dabei, Suchergebnisse zu speichern, besonders interessante Dokumente zu merken, oder Dokumente mit bekannten Merkmalen in der Ähnlichkeitsdarstellung hervor zu heben.

Volltext (b): Ein Html-Browser stellt Textinhalte zum Lesen bereit. Begriffe oder Begriffsfolgen lassen sich suchen und farbig hervorheben. Dadurch lassen sich besonders interessante Passagen in Texten rasch finden und lesen.

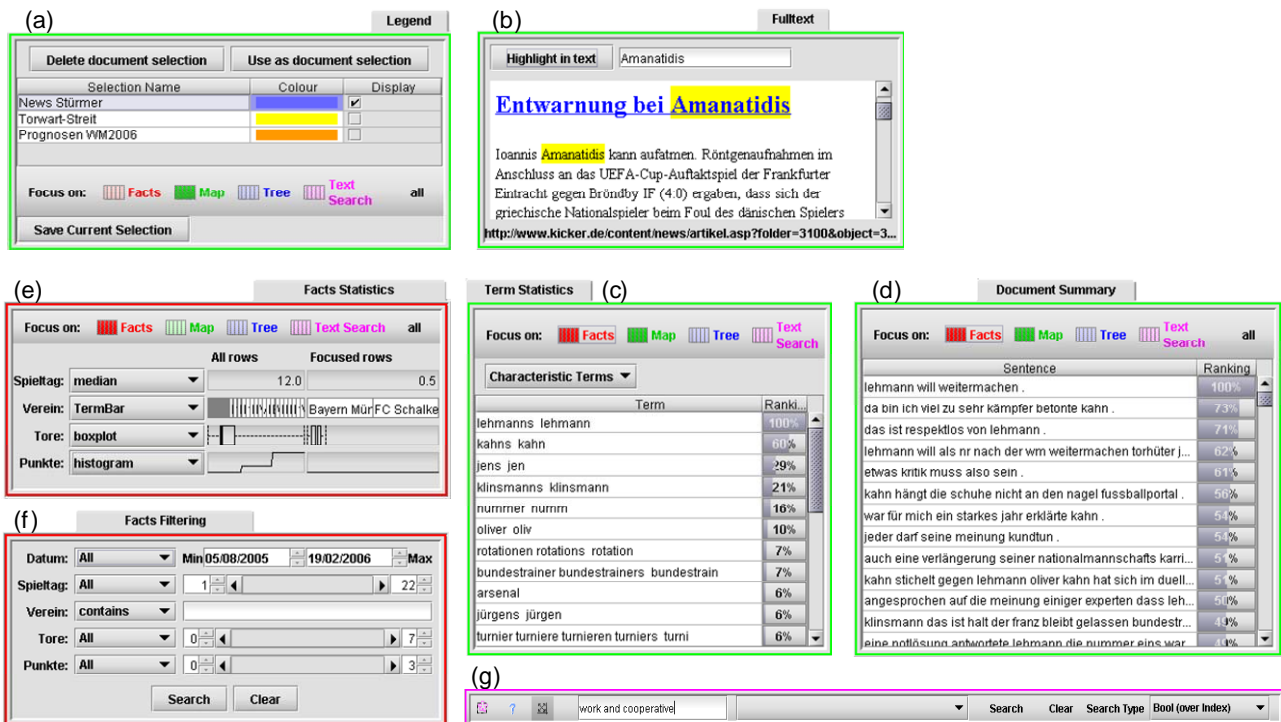


Abbildung 6.13: Interaktive Werkzeuge zur Analyse von Texten und Relationalen Daten (Fakten)

Termstatistik (c), Zusammenfassung (d): Anders als die Volltext-Darstellung ermöglichen (c) und (d) nicht das Lesen, sondern das Quer-Lesen von Texten. Die Termstatistik zeigt wichtigste Terme einer Gruppe von Dokumenten. Die Gruppe kann dabei in einer Farbauswahl-Werkzeugleiste bestimmt werden. Der Benutzer kann zwischen zwei Relevanzmaßen für Terme wählen: *Charakteristische Terme* sind für eine Gruppe wichtig und gleichzeitig für den Rest der Texte nicht so wichtig [Be01]. Gemeinsame Terme kommen in jedem Text der Gruppe vor. Die identifizierten Terme können alphabetisch oder nach Relevanz-Wert sortiert werden. Ein Doppelklick sucht einen Term in allen Dokumenten. Die Zusammenfassung bewertet Sätze bzgl. des Vorkommens wichtiger Terme. Besonders wichtige Sätze werden aufgelistet. Ein Doppelklick öffnet das betreffende Dokument und hebt den Satz hervor. Termstatistik und Zusammenfassung helfen beispielsweise dabei, die inhaltlichen Gemeinsamkeiten von ausgewählten Texten zu erkennen.

Faktenstatistik (e): Für jede Datenspalte der angezeigten Datentabelle ist eine Zeile in der Faktenstatistik vorgesehen. Für jede Spalte können die Werte aller Datensätze den Werten der Datensätze mit einer bestimmten Farbmarkierung statistisch gegenübergestellt werden. Die fokussierte Farbmarkierung kann in einer Auswahl-Werkzeugleiste bestimmt werden. Die statistischen Maße und Darstellungen lassen sich aus einem Auswahlfeld wählen. Zur Beschreibung von Zeichenketten können TermBars eingesetzt werden, die

Worthäufigkeiten durch Balken symbolisieren. Zur Beschreibung von numerischen Werten stehen Kennzahlen (Mittelwert, Median, Varianz, Standardabweichung, Min., Max., Summe, Signifikanztest) und Diagramme (Boxplot, Histogramm, Verteilung) zur Verfügung. Mit der Faktenstatistik lässt sich zum Beispiel die Altersverteilung von Kunden untersuchen, deren Beschwerden thematisch ausgewählt wurden.

Faktenfilter (f): Für jede Datenspalte der angezeigten Datentabelle ist eine Zeile im Faktenfilter vorgesehen. Die in der Faktensicht angezeigten Datensätze können durch Einschränkung der Werte jeder Spalte ausgewählt werden. Analog zur Faktenstatistik werden Datentypen unterschieden. Zeichenketten lassen sich auswählen nach dem Vorkommen von Zeichenketten (enthält, enthält nicht, ist gleich), während numerische oder Zeit-Daten nach Wertebereichen gefiltert werden können.

Inhaltssuche (g): Texte können nach Zeichenketten durchsucht werden. Das Suchergebnis wird als Markierung von Treffer-Dokumenten auf der Karte dargestellt. Dabei wird unterschieden nach Verfahren, die auf dem Dokumentenindex arbeiten und solchen, die im Volltext suchen. Bei der Suche im Index wird die linguistische Vorverarbeitung berücksichtigt (z.B. Wortstammreduktion und Stoppworteliminierung). Die *Boolesche Suche* erlaubt Boolesche Kombination von Suchbegriffen. Die *Vektorsuche* behandelt eine Anfrage wie ein Dokument und sucht mit Vektor-Vergleichsmaßen nach den ähnlichsten Dokumenten. Die *Kartensuche* fasst eine Anfrage ebenfalls als Dokumenteninhalt auf und bildet dieses virtuelle Dokument auf die Clusterkarte ab. Ergebnis der Kartensuche ist also nicht eine Markierung von Dokumenten auf der Karte, sondern ein Startpunkt für die weitere Exploration. Die Suche im Volltext ist mittels regulären Ausdrücken möglich. Die Suche wird oft im Wechsel mit der Textbeschreibung verwendet: Gefundene Dokumente werden analysiert, daraus ergeben sich neue Suchbegriffe, usw. Die Suche hilft beispielsweise, Texte mit bekannten Inhalten wieder zu finden oder interessante Themen auf der Landkarte einzuordnen.

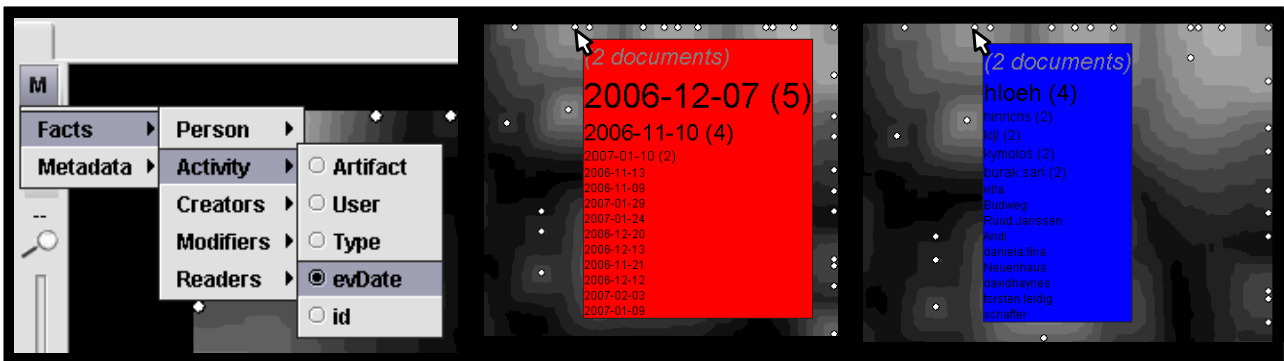


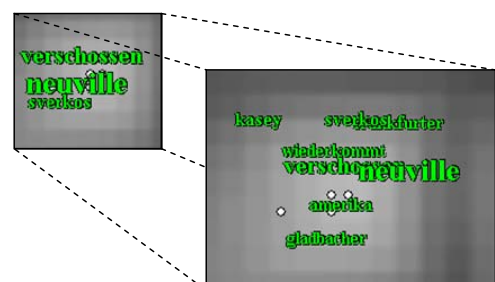
Abbildung 6.14: MultiView-Tooltips zeigen für einzelne Dokumente damit verknüpfte strukturierte Daten an. Im Beispiel (Mitte) werden Tage an ein Dokument annotiert, an denen Benutzeraktivitäten stattgefunden haben. Die Daten stammen aus einer Faktentabelle und werden daher rot hinterlegt. Alternativ (rechts, blau) lassen sich aus dem Kategorienbaum die Namen von Benutzern extrahieren, die das Dokument modifiziert haben.

Neben den beschriebenen interaktiven Basiswerkzeugen zur Suche und Beschreibung von Texten und Daten lädt Das MultiView-Konzept dazu ein, die mit Texten zusammenhängenden

strukturierten Daten zur Beschreibung von Texten und Textgruppen zu nutzen. Dazu bietet SWAPit zwei Interaktionskonzepte an: Tooltips für einzelne Dokumente und Labels für Dokumentengruppen. Dabei handelt es sich um alternative Realisierungen des Konzepts der *MultiView-Annotation* aus Kap.5.5.

MultiView-Tooltips ermöglichen die Annotation jedes einzelnen Dokuments mit Beschreibungskategorien oder Werten aus Da-

Abbildung 6.15: MultiView Labels zeigen für Dokumentengruppen Daten an, die mit den Dokumenten verknüpft sind.



tentabellen (Abbildung 6.14). Dabei lässt sich interaktiv bestimmen, aus welcher Tabelle und welcher Spalte die Daten verwendet werden sollen. So lassen sich beispielsweise zu kooperativ bearbeiteten Dokumenten nicht nur der Titel, sondern auch daran arbeitende Personen, Bearbeitungszeiträume oder Inhaltskategorien durch einfache Mausbewegung auf der Dokumentenlandkarte einblenden.

MultiView-Labels ermöglichen die Annotation von Dokumentengruppen auf der Landkarte mit wichtigen Termen, Beschreibungskategorien oder Werten aus Datentabellen (Abbildung 6.15). Wird ein Kartenausschnitt vergrößert dargestellt, so werden auch mehr Begriffe eingeblendet. Die Kartenbeschriftung hilft beispielsweise, Dokumentengruppen rasch quer zu lesen.

6.7 Sichtenspezifikation, Inhaltsextraktion und Informationsverknüpfung

Das vorgestellte MultiView-Konzept adressiert nicht die Herkunft der dargestellten Daten und Beziehungen. Für das Werkzeug und dessen Anwendung ist die Frage wichtig, wie sich Daten und Beziehungen konzeptionell sowie technisch beschaffen, möglicherweise extrahieren und geeignet transformieren lassen. Im Folgenden werden verschiedene Tools vorgestellt, die nicht in SWAPit integriert sind, sondern von SWAPit verwendet werden können.

Textinhalte liegen in der Praxis in zahlreichen Dokumentenformaten vor, darunter die Formate der Microsoft Office-Produktfamilie, Acrobat PDF und HTML. Neben der Extraktion der Textinhalte muss die Sprache identifiziert und das Dokument in Sinnabschnitte zerlegt werden. Für die beiden erstgenannten Formate existieren einige OpenSource-Projekte zur Inhaltsextraktion⁷⁴. Einige davon wurden in Java Servlets verkapselt. Diese Servlets dienen als Textfilter und können, je nach MIME-Typ eines Dokuments, der zu verarbeitenden URL als Präfix voran gestellt werden (Proxy-Konzept). Für Textformate wie HTML oder XML wurde ein Servlet *RegExtract* entwickelt, das mittels zu definierender regulärer Ausdrücke relevante Textbausteine erkennt und extrahiert. Stehen die Texte in relationalen Datenbanken, so hilft das Servlet *SQL2URL* bei deren Bereitstellung. Es erhält Zugangsdaten und Identifikationsschlüssel für einen Datensatz und liefert den Textinhalt in Http-verfügbare Form. Die Verwendung solcher Servlet-Filter zur Verkapselung von Dokumenten-Extraktion und -Transformation lässt Veränderungen und Erweiterungen flexibel zu.

Relationale Daten müssen in einer oder mehreren relationalen Datenbanken vorliegen. Daraus werden per SQL-Statements Tabellen definiert, die von SWAPit geladen werden. Auch *Beschreibungskategorien* können in einer Datenbank abgelegt sein. Die Angabe eines geeigneten SQL-Statements definiert dann eine standardisierte Tabelle, welche die gewünschten hierarchischen Strukturen repräsentiert. Alternativ kann für Beschreibungskategorien eine XML-Datei verwendet werden.

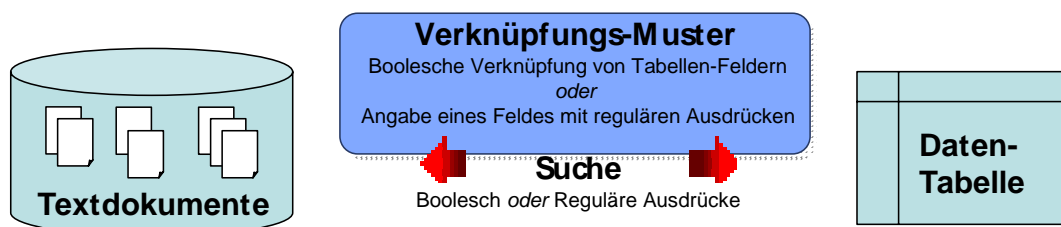


Abbildung 6.16: Werkzeuge für die automatische Verknüpfung von Texten und Strukturierten Daten

⁷⁴ Swish-e zur Indexierung u. Suche zahlreicher Dokumenten-Formate (<http://swish-e.org/>), Jakarta-POI zur Arbeit mit Office-Dokumenten: (<http://jakarta.apache.org/poi/>).

Liegen die *Verknüpfungen* zwischen den strukturierten Daten und den Texten explizit vor, können sie mittels SQL-Statements oder XML-Spezifikation unmittelbar für die Navigation nutzbar gemacht werden. In vielen Fällen müssen diese Verknüpfungen allerdings zunächst berechnet werden. Dabei helfen zwei Werkzeuge zur Wissensgenerierung auf Basis der Inhaltssuche in den Texten (Abbildung 6.16).

Enthalten die zu verknüpfenden Datensätze Felder, deren Werte direkt als Schlüsselbegriffe bei der Inhaltsuche verwendet werden können (z.B. Eigennamen von Personen oder Firmen), so lässt sich für die Verknüpfung ein Boolescher Ausdruck über den Tabellenfeldern verwenden. Das Verknüpfungswerkzeug setzt für jeden Datensatz die Feldwerte in diesen Ausdruck ein und sucht im Textbestand nach dem so instanziierten Booleschen Ausdruck. Die Ergebnisdokumente werden mit dem Datensatz verknüpft.

Sind die Tabellenfelder keine Schlüsselattribute, so kann für jeden Datensatz in der Datenbank ein regulärer Ausdruck hinterlegt werden, mittels welchem Texte gesucht werden. Bei dieser Variante hängt die Qualität der Verknüpfungen von der Güte der Regulären Ausdrücke, nicht von den Daten selbst ab. Andererseits muss für jeden Datensatz erst ein Ausdruck definiert werden. Dieser nachteilige Aspekt ist kaum von Bedeutung, wenn wenige Datensätze mit sehr vielen Textdokumenten verknüpft werden sollen, was in vielen Praxisbeispielen der Fall ist (z.B. Konkurrenzfirmen zu Patenten oder Produkte zu Artikeln der Fachpresse).

Soll die Informationsbeschaffung automatisiert werden, so lassen sich Monitoring-Werkzeuge einsetzen. Im SEWASIE-Projekt⁷⁵ wurde das Werkzeug MonA (Monitoring Agent) entwickelt, das in regelmäßigen Zeitabständen im Web nach Dokumenten sucht, die zum Interessensprofil eines Benutzers passen [KSB06]. Gefundene Dokumente werden automatisch klassifiziert und mit SWAPit visualisiert.

6.8 Stellschrauben zur Systemanpassung

SWAPit versteht sich als Werkzeugkasten zur Textanalyse. Indexierung, Dokumentenvergleich, Sichten und Funktionalitäten können *ohne Programmieraufwand* für spezielle Anwendungen angepasst werden. Darüber hinaus erlaubt der modulare Aufbau von Software und Systemkonzept die Entwicklung und Verwendung eigener Module zur Ähnlichkeitsberechnung sowie eigener Sichten für unstrukturierte oder strukturierte Daten.

6.8.1 Informationszugriff

Der Informationszugriff wird bestimmt durch Information- und Interaktionsdesign. Während die Grundelemente (Sichtenarten, Sichtenkoordination, Funktionalitäten) durch SWAPit realisiert werden, kann und soll die Ausgestaltung (Sichtdefinition) von der Anwendung abhängen. Faktensichten und Kategoriensicht lassen sich aus vorhandenen Daten definieren (Kap.6.7). Beliebige Datentabellen können in SWAPit eingebendet und, Verknüpfungen zu den Dokumenten vorausgesetzt, zur Navigation genutzt werden. Dazu bedarf es eines Spezifikationsdatensatzes für jede Sicht, der folgende Elemente umfasst: Zugang zur Datenquelle⁷⁶, Aufbau der Datenquelle, Darzustellende Daten, Verknüpfungen zu den Dokumenten.

Wird eine alternative Darstellungsmetapher (z.B. Treemap oder Hyperbolic Tree zur kompakten Darstellung von Beschreibungskategorien) oder eine zusätzliche Informationsart (z.B. Geo-Koordinaten zu Veranstaltungsinformationen) benötigt, so kann eine neue Sicht entwickelt und in SWAPit verwendet. Die Übersetzung zwischen den Informationsobjekten der Sichten und den Textdokumenten übernimmt die Sicht selbst. Erfüllt eine Sicht diese Voraussetzung, kann sie leicht an SWAPit's Koordinationskonzept angekoppelt werden, indem sie sich bei SWAPit registriert (Kap.6.2). Dadurch wird der Sicht eine bestimmte Farbe zugeord-

⁷⁵ SEMantic Webs and AgentS in Integrated Economies, 5. Rahmenprogramm, IST-2001-34825.

⁷⁶ Relationaler Datenbank-Server oder XML-Datei

net. Sie kann die Auswahl von Informationsobjekten dem System mitteilen und wird über Auswahlen in den anderen Sichten informiert.

Die interaktiven Werkzeuge zur Suche und Text- bzw. Datenanalyse lassen sich über eine XML-Datei anpassen. So können beispielsweise dem Experten auch komplexe Hilfsmittel an die Hand gegeben werden, die dem gelegentlichen Nutzer nicht zugemutet werden können.

6.8.2 Dokumentenvergleich

Das Ähnlichkeitsmaß für die Berechnung von Dokumentenlandkarten ist austauschbar, da die Berechnung auf Ähnlichkeitswerten aufbaut und unabhängig von der vorgelagerten Berechnung dieser Werte ist [Be01]. SWAPit's Prinzip eines Werkzeugkastens, der für konkrete Anwendungen spezialisiert werden kann, profitiert von der Möglichkeit, ein Aufgabenangemessenes Ähnlichkeitsmaß definieren zu können (Kap.4.4.2), beispielsweise den Vergleich von Literatur bzgl. gemeinsamer Referenzen, den Vergleich von Patenten über gemeinsame Beschreibungskategorien aus standardisierten Schemata wie dem IPC⁷⁷, den Vergleich von XML-Seiten über gemeinsame Konzepte einer Ontologie, oder den Vergleich von Zeitungsartikeln über statistische Inhaltsanalyse. Wird SWAPit in bestehende Retrieval- oder Text Mining-Umgebungen integriert, so lassen sich die bewährten Verfahren zur Inhaltsanalyse und Suche nutzen.

Die statistische Inhaltsanalyse wird, unabhängig vom verwendeten Ähnlichkeitsmaß, immer durchgeführt und liefert eine Standard-Ähnlichkeitsberechnung und die Möglichkeit der Suche im Terminde. Dokumente werden im Vektorraum-Modell [SM87] repräsentiert und verglichen. Das realisierte Ähnlichkeitsmaß bietet zahlreiche Möglichkeiten der Konfiguration für spezielle Anwendungen. So lassen sich Stoppworte angeben, die von der Inhaltsanalyse ausgeschlossen werden. Dies sind meistens besonders häufige Begriffe, welche wenig zur Differenzierung der Dokumente beitragen. Aber auch selten vorkommende Begriffe können durchaus für die mit dem System untersuchte Fragestellung bedeutungslos sein. Alternativ zur Stoppwortliste lässt sich eine Positivliste von Termen definieren, die den Index vorgibt. Dokumente werden dann ausschließlich bezüglich dieser Begriffe repräsentiert und verglichen. Diese Vorgehensweise bietet sich an bei präzise beschreibbaren Fragestellungen oder schwer zu erschließenden Textsammlungen wie Korrespondenz oder Prosa.

Zur linguistischen Verfeinerung der statistischen Analyse dient ein konfigurierbarer Thesaurus. Dort lassen sich Regeln zur Begriffs-Dekomposition angeben und Phrasen beliebiger Länge auf Basiskonzepte reduzieren. Dadurch können höherwertige Sprachkonzepte als freie lexikalische Morpheme erkannt werden. Beispielsweise lassen sich die Konzepte „power network“ und „electricity network“ adäquat unterscheiden. Auch Synonyme lassen sich als atomare Phrasen definieren, wodurch „Geld“, „Mäuse“, „Knete“ und „Kohle“ auf dasselbe Konzept abgebildet werden können. Die Zusammenfassung von verschiedenen Wortformen wie Flektionen von Verben erledigt eine sprachspezifische Komponente zur Wortstammreduktion (Stemmer).

Allerdings stellen die genannten Verfahren lediglich eine einfache, wenn auch im Allgemeinen mächtige und praxistaugliche Erweiterung der statistischen Inhaltsanalyse dar. Freilich lassen sich durch sprachspezifische Verfahren wie Grammatikanalyse (z.B. Part-of-Speech Tagging), lexikalische Datenbanken wie WordNet⁷⁸ oder Wörterbücher noch bessere Ergebnisse der Inhaltsanalyse von natürlichsprachlichen Texten erreichen. Wie gesagt, lassen sich beliebige Methoden zur Inhaltsanalyse anwenden, weshalb das Augenmerk bei dieser Arbeit und bei der Entwicklung von SWAPit nicht primär auf der Optimierung der linguistischen Dokumentenanalyse lag.

⁷⁷ *International Patent Classification*, ein international gültiges Klassifikationssystem für Patente.

⁷⁸ <http://wordnet.princeton.edu>, ein Online-Referenzsystem für lexikalische Konzepte

7 Evaluation in Fallstudien

Die drei Fallstudien sind nach dem Verfahren zur Systemanpassung und –bewertung aus Kap.2.4 aufgebaut. Ist-, Potenzial- und Aufgabenanalyse wurden bereits in der Problemanalyse der Arbeit beschrieben (Kap.3). Die Evaluation in diesem Kapitel knüpft daran an: Für jede Fallstudie werden Systemdesign, konzeptionelle und technische Integration, Evaluationsdesign, Durchführung der Studie und Ergebnisse dargestellt. Evaluieren werden Systemkonzept (Nutzen, Anpassbarkeit/ Erweiterbarkeit, Integrierbarkeit) und Bewertungsframework.

7.1 Allgemeine Bewertungsziele

Die drei Anwendungsbeispiele aus der Praxis, die schon zur Problemanalyse dienten (Kap.3), werden hier zur Bewertung in verschiedenen industrienahen Umgebungen wieder aufgegriffen (Kap.1.5). Die Referenztechnologie SWAPit zielt darauf ab, *unterschiedliche* Szenarien der Textanalyse zu bedienen. Für die Untersuchung der Eignung der Referenztechnologie wird daher zugunsten der Erprobung in mehreren verschiedenen Anwendungskontexten auf die detaillierte Bewertung von Teilaspekten eines Spezialsystems in nur einem Fall verzichtet. Empirische Evidenz für die Nützlichkeit der Referenztechnologie entsteht durch die wiederholte Verwendung in unterschiedlichen Anwendungskontexten.

Die durchgeführten Fallstudien sind qualitative Untersuchungen mit dem Ziel die Eignung und Anpassbarkeit von Konzept und Referenztechnologie für verschiedene Anwendungsfelder zu bewerten (Abbildung 7.1). Die Fallstudien untersuchen weder den Vergleich solcher MV-Systeme zu alternativen Informationssystemen, noch verfolgen sie das primäre Ziel, einzelne Aspekte zu isolieren (z.B. MultiView-Konzept, Navigation, Grafische Darstellung), zumal bei explorativen Suchsystemen ohnehin gerade das Zusammenspiel der einzelnen Funktionalitäten von entscheidender Bedeutung ist [WMM06]⁷⁹. Vielmehr wird der Nutzen des für spezielle Anwendungen optimierten Systems in realen Arbeitsumgebungen bewertet. Der Mehrwert der Systemunterstützung wird an der Verbesserung gegenüber der Ausgangssituation gemessen, also im Vergleich zum Nichtvorhandensein der Systemunterstützung.

Die empirische Bewertung von interaktiven Informationssystemen erfordert immer konkrete Aufgaben ([St95], S.77). Die Ergebnisse von Beobachtungen, Befragungen und Messungen bei der Bearbeitung dieser Aufgaben werden i.d.R. auf eine *Klasse von Aufgaben* übertragen, also verallgemeinert. In der vorliegenden Arbeit ist neben den Aufgaben eine weitere Verallgemeinerungsdimension zu berücksichtigen, die Systemdimension: Eigentlicher Erkenntnisgegenstand ist nicht das konkrete System, sondern die *Klasse von Systemen*, die sich aus der Referenztechnologie herleiten lässt. Die Verallgemeinerung bzgl. System und Aufgaben wird gestützt durch eine Charakterisierung von Systemen und Aufgaben (Kap.1.4), durch Plausibilitätsbetrachtungen, und durch empirische Evidenz, also ähnliche Erkenntnisse in mehreren Beispielanwendungen. Abbildung 7.1 stellt die Unterscheidung dar: In den einzelnen Fallstudien (Kap.7.3-7.5) geht es um die Eignung eines konkreten Systems für ein Fallbeispiel. Abschließend (Kap.8.2) wird diskutiert, welche Erkenntnisse sich aus der Gesamtheit der Fallstudien für die Eignung der Systemklasse in einer Anwendungsdomäne und für die Textanalyse insgesamt ableiten lassen. Zusätzlich kann aus der exemplarischen Systemanpassung im Rahmen eines Fallbeispiels für die Anpassbarkeit der Referenztechnologie gelernt werden.

⁷⁹ “...it may often be the case that the sum of the features in an ESS may lead to a different experience than the individual features in isolation (i.e., the interactions between features may be just as important as the features themselves)” [WMM06]

Benötigt wird ein „roter Faden“ für die Bewertung, der sicherstellt, dass die Fragen zur Systemklasse (Referenztechnologie) bei der Bewertung in sehr unterschiedlichen Fallstudien eine zentrale Rolle spielen.

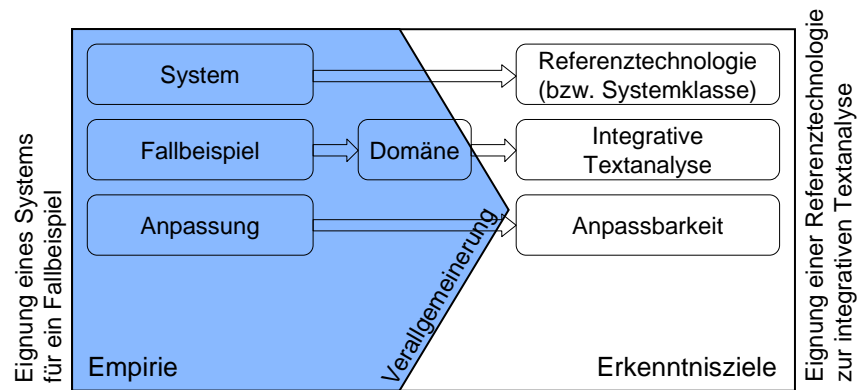


Abbildung 7.1: Unterscheidung Empirischer Beobachtungen und verallgemeinerten Aussagen.

Das folgende Bewertungsschema versteht sich als „roter Faden“ bei der Evaluation:

Anwendung

- Ergebnisqualität, Nutzen
- Prozesseffizienz, Benutzbarkeit
- Benutzerakzeptanz

Referenztechnologie

- Anpassbarkeit/Anwendbarkeit (Funktionsumfang vollst. u. präzise)
- Integrierbarkeit, Wartbarkeit (APIs, Datenformate, Update-Prozess)
- Erweiterbarkeit (Neue Sichten bzw. Funktionalitäten)

Das Schema unterscheidet die Kriterien der konkreten Fallstudie (Anwendung) von den Kriterien der Referenztechnologie. Die *Kriterien der konkreten Anwendung* sind Ergebnisqualität, Prozesseffizienz, Benutzerakzeptanz und Usability. Sie dienen nicht nur unmittelbar der Nutzen-Bewertung des Systems, sondern auch mittelbar der Bewertung der Klasse von Systemen, die sich aus der Referenztechnologie ableiten lassen. Diese Kriterien werden in der einheitlich verwendeten Scorecard zur Nutzenbewertung berücksichtigt (Kap.2.1). *Parameter der Referenztechnologie* adressieren dagegen Vielseitigkeit und Verwendbarkeit (Anpassbarkeit, Integrierbarkeit, Wartbarkeit, Erweiterbarkeit) und werden jeweils diskutiert. Die meisten spezialisierten Fragestellungen der Fallstudien leiten sich aus den allgemeinen Bewertungskriterien ab und werden am Anfang jedes Fallstudienabschnittes konkretisiert.

7.2 Systematik der Fallstudien

Die durchgeführten Fallstudien, in Tabelle 7-1 übersichtlich charakterisiert, bauen aufeinander auf (Kap.1.5): Jede Weiterentwicklung von Konzept und Referenztechnologie wurde für die nächste Fallstudie wieder nutzbar gemacht und ging in die nächste Bewertung ein. Befasst sich die erste Fallstudie noch mit einer Basis-Technologie, so werden die folgenden Fallstudien mit einem immer vielseitiger werdenden System durchgeführt. Daher ist die Chronologie der Fallstudien bedeutsam.

Zunächst wurde in einem EU-Projekt untersucht, wie sich hierarchische Modelle aus den verschiedenen Aufgabenkontexten Business Intelligence und Textzugriff integriert darstellen und zur Navigation verwenden lassen [Br04]. Die Anforderung, komplementär zur Textsicht eine Kategoriensicht einzuführen, war die „Geburtsstunde“ des Sichtenkonzepts von SWAPit. Unternehmensinterne und unternehmensexterne Daten wurden in einem zur Patentierung angemeldeten Verfahren verknüpft [Fr05] und integriert verfügbar gemacht. Eine Neuerung aus technischer Sicht war der Webservice-basierte Datenaustausch zwischen dem

Reporting-Werkzeug und dem Textanalyse-Werkzeug. Diese Fallstudie hatte eher entwicklungshistorische Bedeutung denn Aussagekraft über die Eignung des späteren Systemkonzepts. Sie wird daher lediglich in Anhang 9.3 kurz dargestellt.

In einer Diplomarbeit [Eg05] wurde gemeinsam mit einer großen deutschen Versicherung untersucht, wie sich Kundenbeschwerden mit Kunden- und Vertragsdaten integriert darstellen und zur Navigation verwenden lassen. Dazu wurden Daten aus den separaten Bereichen Stornoprohylaxe (erkennen und vermeiden von Kündigungen mit Data Mining Verfahren) und Beschwerdemanagement (bearbeiten und analysieren von Kundenbeschwerden) gebündelt und eine tabellarische Darstellung von relationalen Daten entwickelt. Die Fallstudie motivierte die Einführung eines Ranking-Konzeptes für interaktive Auswahlen (Kap. 5.4), wodurch sich - ähnlich wie bei der Ranking-Darstellung von Suchergebnissen - besonders wichtige Texte rasch erkennen lassen.

Im Gegensatz zu den beiden ersten Fallstudien, aus denen jeweils Konzept und Realisierung einer neuen Datensicht hervorging, war das Ziel der letzten beiden Fallstudien, Gesamtkonzept und Werkzeug im operativen Betrieb verschiedener Branchen zu erproben.

In einer Diplomarbeit [Hi06] wurde gemeinsam mit einem großen deutschen Systemhaus untersucht, wie sich ein etabliertes Suchwerkzeug zur Meldungsrecherche durch ein exploratives Werkzeug zur Textrecherche gewinnbringend ergänzen lässt. Dazu wurden alle im Suchprozess relevanten Daten integriert dargestellt und navigierbar gemacht. Neben gewonnenen Erkenntnissen zu Vereinfachungen in der Benutzbarkeit und zu Optimierungsmöglichkeiten der Suchfunktionalität wurde ein Ähnlichkeitsmaß auf Basis eines vorgegebenen Stichwortindex entwickelt, das mit den Top-Begriffen aus den Hinweistexten gespeist wurde.

Tabelle 7-1: Wichtigste Merkmale der Fallstudien im Überblick

| | Beschwerdemgmt. | Produktsupport | Koop. Arbeitsumgebung |
|---|--|--|---|
| Rechercheprozess | Kundenanalyse | Meldungsrecherche | Aktivitätsanalyse |
| Domäne | Versicherung | Systemhaus | Forschung |
| Prozess-Art | direkt & indirekt | direkt | direkt & indirekt |
| Prozess-Existenz | neuer Prozess | Prozessoptimierung | neuer Prozess |
| Rahmen & Jahr | Diplomarbeit (2005) | Diplomarbeit (2006) | EU-Projekt (2006/7) |
| Empirie-Modus | Expertengespräche | Testbetrieb (Feld) | Testbetrieb (Labor) |
| Daten | 80+ Beschwerden, Kunden, Verträge, Storno-Scores | <1000 Meldungen pro Systemmodul, Metadaten, Hinweisdokumente | 80 Word-Texte aus dem projekt-Workspace, Aktivitätsdaten |
| Testpersonen | Experten | 6 MA Meldungsbearbeitung (Kerngrp.: 3) | 5 MA im Projekt EcoSpace |
| Informations-Integration organisatorisch | Informationen über Abteilungsgrenze verknüpft | Inf. aus Suchunterstützung gebündelt für Meldungsrecherche; Hinweise u. Meldungen verknüpft! | Informationen für operative koop. Arbeit gebündelt für Aktivitätsanalyse |
| Integration technisch | Datenimport in SWAPit (Compile-Time) | Datenimport in SWAPit (Compile-Time) | Echte Integration BSCW-SWAPit (Run-Time mit Pre-processing) |
| Einfluss auf Konzept & Werkzeug | Tabell. Darst. relationaler Daten; entsprechende Interaktionsmögl., MV-Relevanzmaße; Positivlisten | Informiertes Clustering; Vereinfachung | Zeitbasierte Navigation; Handhabung großer Tabellen für Aktivitäten (Zeit x Artefakt) |
| Befragungsmethodik | Expertenworkshops | Strukt. Interviews, Expertenworkshops | Strukturierte Interviews, Expertenworkshops |
| Datenerhebung | Befragung | Befragung u. Messung | Befragung und Messung |

Schließlich wurde in einem EU-Projekt untersucht, wie die Analyse von Kooperationsaktivitäten in virtuellen Arbeitsumgebungen durch ein MultiView-System zur Textanalyse unterstützt werden kann [SPB07]. Außerdem sollten neue Zugriffspfade für die Navigation in kooperativen Dokumenten von Teams untersucht werden. Dazu wurden Daten über Akteure, Ablagestrukturen, Artefakte, Themen und Aktivitäten integriert dargestellt und navigierbar gemacht. Eine konzeptionelle Herausforderung war die adäquate Berücksichtigung der Zeit als wichtige Dimension bei der Aktivitätsanalyse. Die letzte Fallstudie motivierte die zukünftige Entwicklung von Konzepten und Werkzeugen zur besseren zeitbasierten Navigation und für die übersichtliche Gruppierung von Daten in umfangreichen Tabellen (Kap.8.3).

Obwohl ein einheitliches Bewertungsframework (Kap.2.1) verwendet wurde, lassen sich die Ergebnisse der 2. und 3. Fallstudie (Abbildung 7.9 u. Abbildung 7.19) nicht *quantitativ* miteinander vergleichen, weil verschiedene Maße verwendet werden mussten, die die individuellen Anforderungen jedes Anwendungsszenarien repräsentieren. Auch der *quantitative* Vergleich pro Quadrant der Scorecard ist problematisch, weil verschiedene Skalen verwendet wurden. Während die Skalen in der 2. Studie auf dem ordinalen Schulnotensystem basieren, wurden in der 3. Studie 5er-Skalen mit negativen und positiven Ausprägungen präferiert. Allerdings lassen sich bei übergreifender Betrachtung qualitative Vergleiche anstellen und Tendenzen erkennen.

7.3 Fallstudie 1: Beschwerdemanagement in Versicherungen

In Kap.4.1 wurden die Prozesse der Beschwerdebearbeitung/-berichterstattung und Stornovermeidung analysiert und ein zusätzlicher Prozess zur Nutzung von SWAPit vorgeschlagen mit dem Ziel, aus den Kundenbeschwerden weitere wertvolle Erkenntnisse zu extrahieren (Kündigungsgründe, Zweitbeschwerdegründe und Indikatoren für die Kündigungsgefährdung von Beschwerden) und dabei Abteilungsgrenzen zu überwinden.

Nach aufwändiger Digitalisierung und Verknüpfung aller Daten wurden mehrere Expertenworkshops mit Teilnehmern der Versicherung (Vertreter aus Vertrieb, IT, Beschwerdemanagement und Beschwerdebearbeitung) und von Fraunhofer-FIT durchgeführt. Dabei wurden mehrere Zyklen prototypischer Realisierung und Diskussion durchlaufen und ein Systemdesign entwickelt, bewertet und optimiert. Unter anderem war die konzeptionell-technische Frage zu beantworten, wie das bisherige System SWAPit um die Darstellung und Navigation strukturierter Information aus relationalen Datenbanken erweitert werden kann. Insofern hat diese Fallstudie, im Gegensatz zu den beiden anderen, nicht nur bewertenden sondern auch stark Systemgestaltenden Charakter.

In der Fallstudie ging es primär um erste Erfahrungen mit der Kombination operativer Informationen aus verschiedenen Organisationseinheiten sowie das Ausarbeiten von wichtigen Fragestellungen und Analysestrategien. Die Fallstudie versteht sich als erster Schritt in Richtung einer spezialisierten Systemunterstützung zur Beschwerdeanalyse und bezieht daher ein breitegefächerte Datenmenge ein: So betrifft das Datensample (die betrachteten Beschwerden) zahlreiche Versicherungssparten und bezieht sowohl gekündigte als auch nicht gekündigte Verträge ein, um Hinweise auf Unterschiede und Merkmale in den Daten aufzudecken. 84 Beschwerdetexte wurden für die Fallstudie anonymisiert und digitalisiert. Aufgrund des geringen verfügbaren Datenvolumens waren keine repräsentativen Aussagen zu Gründen oder Merkmalen zu Beschwerden oder Kündigungen zu erwarten. Der tatsächliche Nutzen der Systemunterstützung kann empirisch erst bewertet werden, wenn größere Mengen von Beschwerdevorgängen zur Verfügung stehen.

7.3.1 Bewertungsziele

Neben den allgemeinen Bewertungszielen (Kap.7.1) standen die folgenden Fragestellungen bei dieser Fallstudie im Vordergrund:

- Welches Wissen steckt in den Texten und wie lässt es sich extrahieren / nutzbar machen?
- Welches Potential hat die Nutzung von Kundenbeschwerden für die Stornoprognose?
- Welches Potential hat die Nutzung von Kundenbeschwerden für den Beschwerdebereich (z.B. Zweitbeschwerdegründe)?

Die folgenden Fragen betreffen Konzept und Technologie:

- Wie lassen sich Relationale Kunden- und Vertragsdaten zu Beschwerdetexten darstellen und interaktiv zur Navigation in der Textsammlung nutzbar machen?
- Welches Ähnlichkeitsmaß ist zu verwenden, um die Kundenkorrespondenz angemessen zu verarbeiten (große Sprachvielfalt; auch Umgangssprache)?
- Wie lassen sich bei vielen fokussierten Informationen schnell die wichtigsten erkennen (Große Auswahlmengen; Benutzerführung)?

7.3.2 Systemdesign

Aus der Problemanalyse in Kap.4.1 ging hervor, dass für die Analyse von Kundenbeschwerden neben den Anschreiben auch Kunden- und Vertragsdaten, Kundenklassen und Stornoscores, sowie die Metadaten der Beschwerden kombiniert zugreifbar gemacht werden sollten. Besondere Anforderungen für den Systemprototyp waren:

Informatorisch:

- Informationen aus verschiedenen Funktionsbereichen (Beschwerdebearbeitung, Beschwerdeberichterstattung/-management, Stornoprognose) und ihre Verknüpfungen sollen kombiniert zugreifbar sein

Funktional:

- Die inhaltlich und sprachlich sehr unterschiedlichen Beschwerden sollen inhaltlich gruppiert werden
- Beschwerdeschreiben sollen über Eigenschaften von Kunde und Vertrag selektiert werden können.
- Beschwerdeschreiben sollen über Eigenschaften von Kunden und Verträgen zu einer Menge von Beschwerdeschreiben sollen statistisch analysiert werden können.

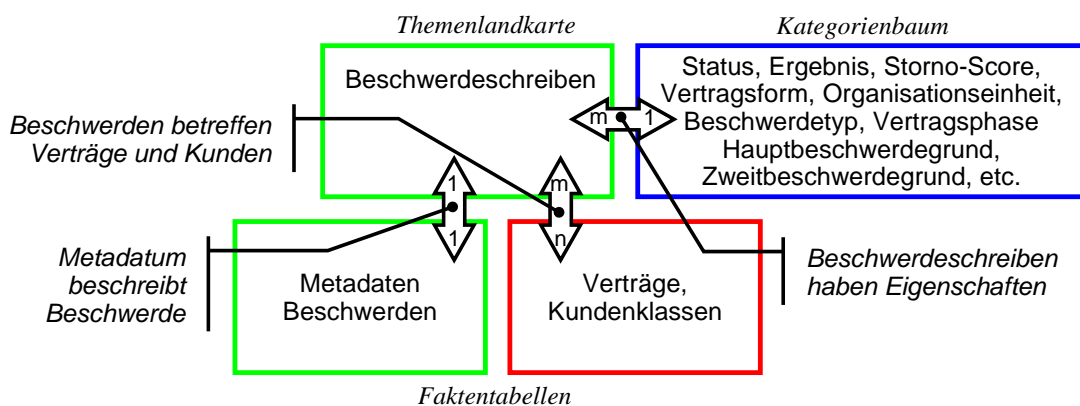


Abbildung 7.2: Systemskizze von SWAPit für die Beschwerdeanalyse.

Im mehreren Iterationen wurde eine SWAPit-Systemskizze für dieses Szenario entwickelt (Abbildung 7.2). Alle Kundenbeschwerden werden auf einer Themenkarte dargestellt. Wichtige Metadaten (z.B. Status, Ergebnis, Beschwerdegrund) werden als Kategorien zum schnellen Zugriff bereitgestellt. Verträge und Kundenklassen werden tabellarisch dargestellt und sind mit den entsprechenden Beschwerdeschreiben verknüpft.

Das *Ähnlichkeitsmaß* für die Themenkarte war eine besondere Herausforderung, da die Kundenbeschwerden ein großes sprachliches und inhaltliches Spektrum aufweisen. Eine einfache statistische Indexierung, wie sie für das erste Systemdesign verwendet wurde, repräsentiert Themen und Kommunikationsartefakte (z.B. Anrede) gleichermaßen. Da eine Vermischung wenig sinnvoll scheint, wurde eine Indexierung und Gruppierung nach einem vorgegebenen Vokabular benötigt. So wurde die Vorgabe eines Indexierungsvokabulars (Positivliste; Feature-Index) als zusätzliche Anforderung identifiziert und in SWAPit konzeptionell und technisch realisiert. Dazu wurden die wichtigsten 96 Terme (Wortstamm-Konzepte) aus insgesamt 1.895 Termen der Textsammlung ausgewählt (Anhang 9.6). Die Spezifikation einer Positivliste hängt vom Analyseziel ab. Soll beispielsweise gerade die Anrede untersucht werden, so ist ein entsprechendes Vokabular vorzugeben. Positivlisten sind eine Alternative zu den gebräuchlichen Negativlisten (Stoppwortlisten), die ebenfalls das Indexierungsvokabular beschränken. Man unterscheidet zwischen Extraktionsindexierung und Zuteilungsindexierung ([Fu99], S.107–133). In jedem Fall fordert Bertram ([Be05], S.73) für Indexierungsterme die folgenden Eigenschaften: Wiedergabetreue, Eindeutigkeit, Prägnanz, Relevanz, Gebräuchlichkeit und Verlässlichkeit. Für die Suche in den Kundenbeschwerden wurde in jedem Fall der extrahierte Term-Index verwendet, um dem Benutzer auch die Suche nach Begriffen zu ermöglichen, die nicht durch die Positivliste vorgegeben sind. Abbildung 7.3 zeigt das für die Beschwerdeanalyse konfigurierte SWAPit-System.

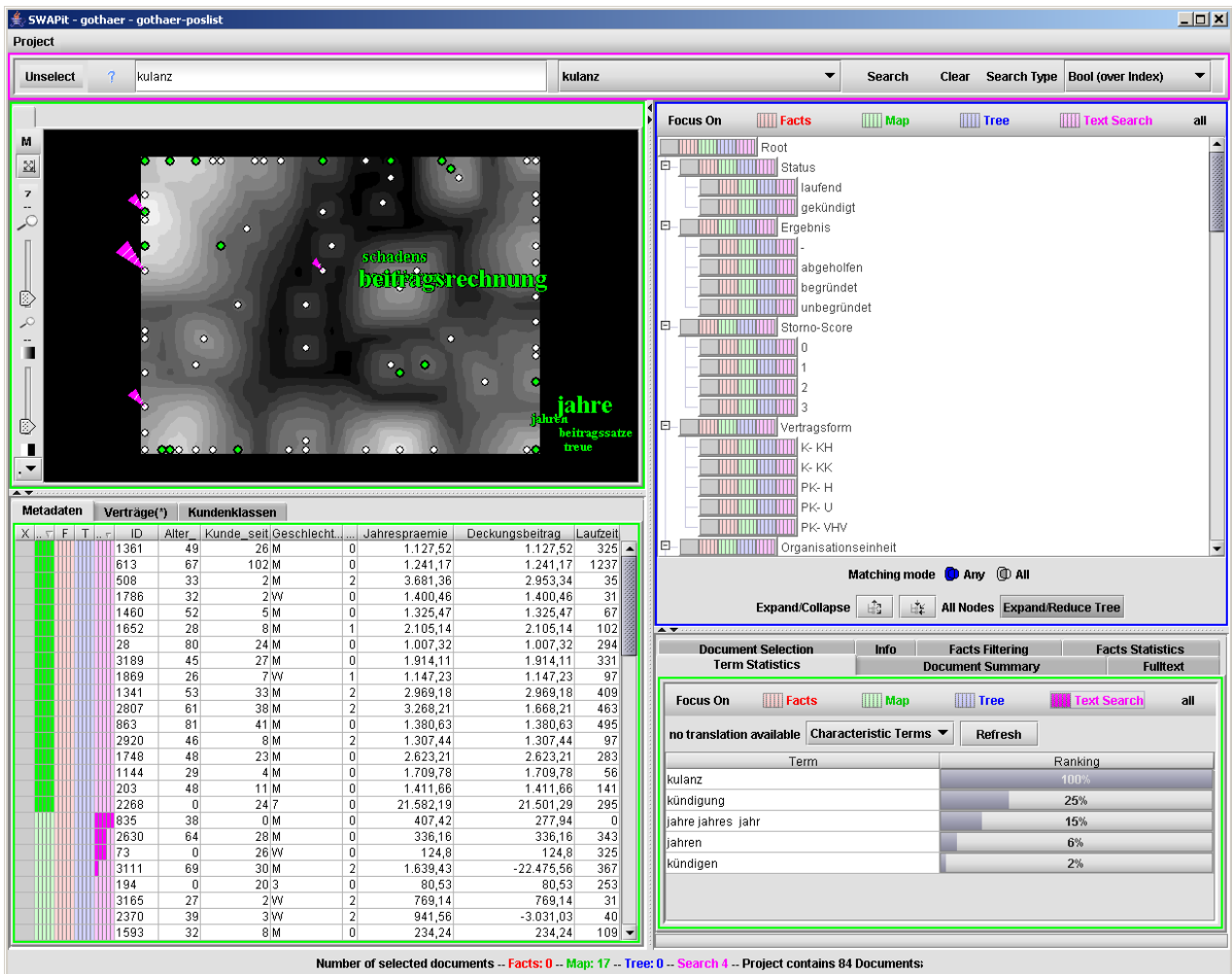


Abbildung 7.3: Screenshot des SWAPit-Systems für die Beschwerdeanalyse mit 84 Kundenbeschwerden und 3 Sichten (Ähnlichkeitssicht links oben, Faktensicht links unten, Kategoriensicht rechts oben).

7.3.3 Integration

Die Texte und die strukturierten Daten wurden aus verschiedenen Funktionsbereichen, in denen sie operational erstellt und gepflegt werden, zusammen getragen und in Datenbank bzw. Dateisystem abgelegt. Dann wurden alle Informationen von SWAPit geladen und verarbeitet.

7.3.4 Versuchsdesign und Durchführung

Zum Zeitpunkt der Fallstudie wurden die Beschwerden noch in Papierform bearbeitet, mussten also für die maschinelle Verarbeitung zunächst digitalisiert werden. Eine Umstellung auf papierlose Datenhaltung war jedoch bereits geplant. Die aufwändigen Vorgänge der Anonymisierung (Schwärzung) und anschließenden Digitalisierung (OCR) und Aufbereitung limitierten die Anzahl der untersuchten Beschwerdetexte.

Wie bereits erwähnt, wurden Daten aus verschiedenen Abteilungen exemplarisch zusammengeführt. Als Stichprobe wurden zur Hälfte Beschwerden von Kunden ausgewählt, die in einem gewissen Zeitraum nach der Beschwerde den betreffenden Vertrag aufkündigten und zur anderen Hälfte Beschwerden zu Verträgen, die noch bestanden. Ausgewählt wurden Beschwerden aus verschiedenen Versicherungssparten, bei denen die Kundenbindung eher niedrig und das Kündigungsrisiko eher hoch ist: Privat-Haftpflicht, Privat-Unfall, Hausrat, Kfz-Haftpflicht, Kfz-Kasko.

Nach der Schwärzung aller kundenbezogenen Informationen in Beschwerden durch die Versicherung und anschließender Digitalisierung der Texte wurden auch die strukturierten Daten zu den betreffenden Kunden zusammengestellt. Im September 2004 fand eine Spiegelung der ermittelten Systemanforderungen in Form eines Expertenworkshops statt. Im November 2004 wurde die prototypische Systemunterstützung in einem weiteren Workshop gemeinsam bewertet.

In [Eg05] wurde die Benutzbarkeit des Systemdesigns exemplarisch bezüglich des Arbeitsablauf der Bestimmung der Kündigungsrelevanz einer Beschwerde systematisch überprüft durch einen *Cognitive Walkthrough* [DF+04]. Dabei werden die einzelnen Aktionen, die zum Erfüllen einer Aufgabe notwendig sind, anhand ihrer Wirkung auf den Benutzer analysiert. Der Durchlauf attestierte dem System eine grundsätzlich gute Benutzbarkeit. Allerdings gab es Hinweise, dass die vielfältigen Aktionsmöglichkeiten die Auswahl der am besten geeigneten Aktion nicht einfach machen – ein potenzielles Problem von vieler MultiView-Systeme [BWK00].

7.3.5 Ergebnisse

Die Ziele der Fallstudie betrafen zum einen die Untersuchung des Potenzials der Verknüpfung von Kundenbeschwerden mit Kunden- und Vertragsdaten (z.B. Storno-Scores). Andererseits sollten Konzept und Referenztechnologie erweitert werden, u.a. um die Navigation in Relationalen Daten.

Zur Darstellung der relationalen Kunden- und Vertragsdaten wurde eine tabellarische Darstellung gewählt. Die Spezifikation und die Verknüpfung mit den Dokumenten basiert auf SQL (6.4). Bei der Navigation hilft die (multiple) Sortierung nach Werten und nach Auswahlfarben (Kap.6.5.3). Die dargestellten Datensätze jeder Faktentabelle können über einen interaktiven Wertefilter als Teilmenge aller Datensätze der Tabelle spezifiziert werden (Abbildung 7.4). Dadurch lassen sich beispielsweise Kundengruppen schnell nach diversen Merkmalen auswählen und die damit verknüpften Beschwerden inhaltlich analysieren. Umgekehrt ermöglicht eine Faktenstatistik (Abbildung 7.5) die statistische Beschreibung von Kunden- oder Vertragsmerkmalen zu ausgewählten Kundenbeschwerden.

Eine weitere Frage war, wie die inhaltlich und sprachlich heterogenen Beschwerden der Kunden mit automatischen Verfahren indiziert und thematisch gruppiert werden können. Eine Positivliste aus ca. 100 ausgewählten Termkonzepten der Textsammlung lieferte gute Ergebnisse.

Eine weitere technische Herausforderung war die Benutzerführung bei großen Mengen hervorgehobener Informationsobjekte. Dazu wurde ein interaktives Ranking-Konzept entwickelt, das konfigurierbare Berechnung von Relevanzwerten (Kap.5.4) und die grafische Darstellung in den Sichten (Kap.6.5) umfasst. So sieht der Benutzer beispielsweise auf einen Blick, welchen Kunden er sich zu einer Menge von Beschwerden als erstes ansehen sollte, weil dieser viele der Beschwerden verfasst hat.

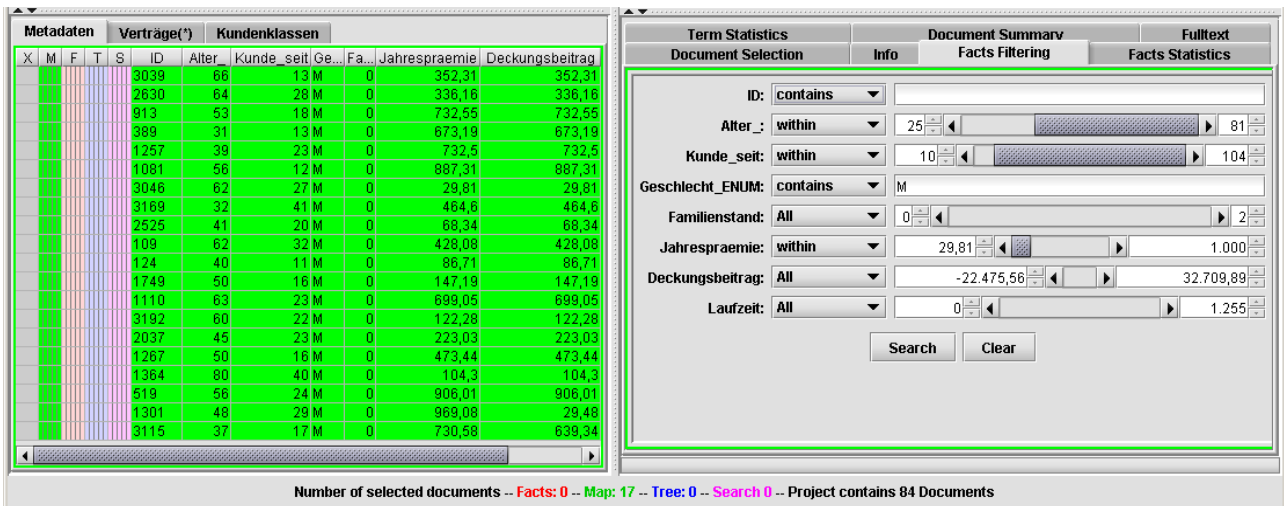


Abbildung 7.4: Werkzeuge zur Analyse Relationaler Daten: Faktenfilter. Im Beispiel werden Beschwerden von über 24-jährigen Männern spezifiziert, die seit mindestens 10 Monaten versichert sind. Die Texte zu den so spezifizierten Beschwerden lassen sich nun im System inhaltlich analysieren.

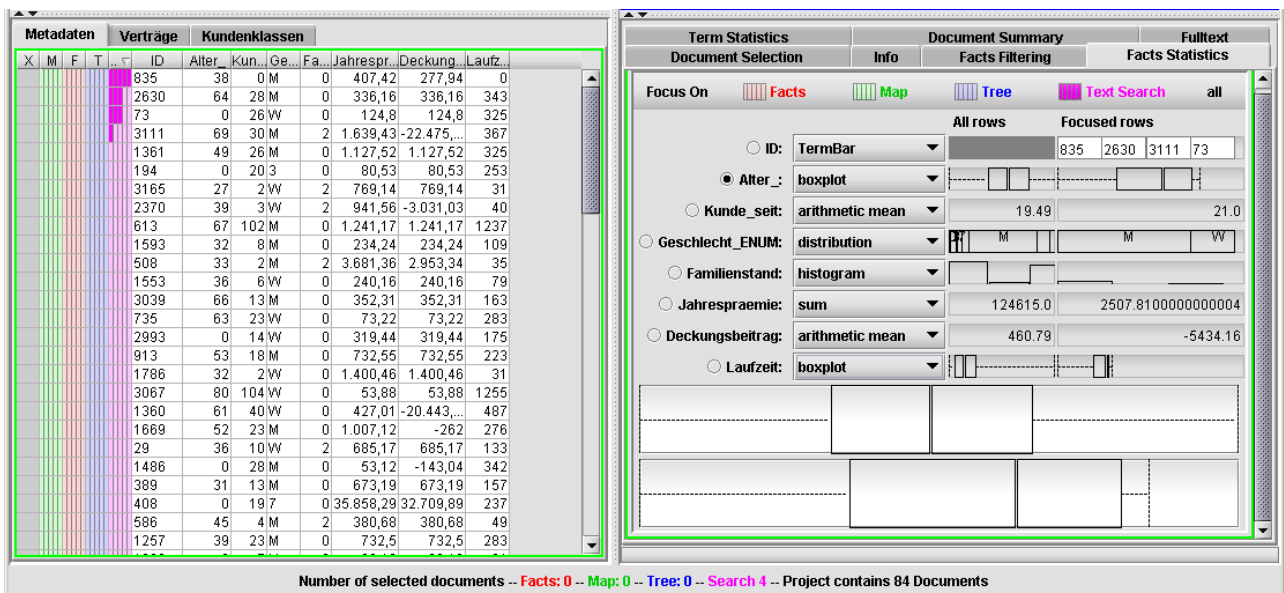


Abbildung 7.5: Werkzeuge zur Analyse Relationaler Daten: Faktenstatistik. Im Beispiel werden die Eigenschaften von 4 Beschwerden, die den Begriff "Kulanz" enthalten (Spalte „Focused rows“), den Eigenschaften aller 84 Beschwerden (Spalte „All rows“) gegenüber gestellt.

Erste Erfahrungen mit dem System zeigen, wie vorgegangen werden kann, um Zweitbeschwerdegründe oder Kündigungsmerkmale in den Kundenbeschwerden zu identifizieren. Ein exemplarischer Arbeitsablauf soll dies verdeutlichen. Im Folgenden sollen Hinweise auf Beschwerdemerkmale gefunden werden, die zur Bereitschaft zur Kündigung beitragen.

- Mit einem Faktenfilter wird die Faktensicht auf gekündigte Verträge eingeschränkt. Die Beschwerden zu gekündigten Verträgen werden in der Themenkarte hervorgehoben
- Auf der Themenkarte werden Gruppen von Beschwerden identifiziert, in denen auffällig viele Beschwerden hervorgehoben sind.
- Eine solche Gruppe von Beschwerden wird auf der Karte ausgewählt und auf charakteristische Merkmale hin untersucht: Die Termstatistik zeigt charakteristische Begriffe der Gruppe. Die Fakten- und Kategoriensicht heben wichtige Merkmale von Kunden und Verträgen hervor (z.B. Vertragsdauer).
- Wird ein auffälliges Merkmal identifiziert, lässt es sich in Bezug auf die gesamte Datenmenge überprüfen (z.B. Suche nach einem auffälligen Term oder Auswahl der Vertragsdauer)

Mit der beschriebenen Strategie lassen sich, ein entsprechendes Datenvolumen vorausgesetzt, Indikatoren identifizieren, die eine Kündigung begünstigen. Damit kann gezielt versucht werden gegenzusteuern, bspw. durch Rundschreiben, Rabattaktionen oder Werbung. Außerdem kann, analog zu anderen Score-Systemen der Versicherungen, ein zusammengesetzter Score dafür entwickelt werden, wie kündigungskritisch eine Beschwerde ist. So ließe sich für einzelne Beschwerden bereits bei der Beschwerdebearbeitung ein Warnsystem für den Sachbearbeiter einrichten. Die Wissensbasis wächst mit jeder eingehenden Beschwerde und lässt auf eine damit verbundene Steigerung der Ergebnispräzision hoffen.

Einige Tendenzen konnten bereits an der kleinen und heterogenen betrachteten Datenmenge beobachtet werden und waren von großem Interesse für die Spezialisten der Versicherung:

- Wird in der Beschwerde mit Kündigung gedroht, erfolgt häufig tatsächlich die Kündigung
- Kunden, die in der Beschwerde von „Entschädigung“ sprechen, sind relativ alt (im Vergleich zu allen Kunden)
- Kunden, in deren Beschwerde das Wort „freundlich“ vorkommt, haben einen höheren Deckungsbeitrag
- Wird eine Beschwerde von der Versicherung intern als unbegründet eingestuft, so ist eine Kündigung deutlich wahrscheinlicher
- Die Storno-Prognose funktioniert für einige Organisationseinheiten besser als für andere (Einfluss der Organisationseinheit auf Qualität der Storno-Scores)
- Internetbeschwerden ziehen deutlich seltener Kündigungen nach sich als Beschwerden, die mit der Post verschickt werden
- Auf eine langen Bearbeitungsdauer der Beschwerde folgt häufig eine Kündigung (unabhängig von der Einstufung in begründet/unbegründet!)

In einer Speziallösung für Versicherungsgesellschaften wäre es zusätzlich wertvoll, die Beschwerden oder sogar Verträge und Kunden nicht nur nach Termhäufigkeiten, sondern auch nach weiteren ausgewählten Merkmalen gruppieren zu können, wozu eine Erweiterung des Systems nötig wäre ([Eg05], S.60).

7.4 Fallstudie 2: Produktsupport in Systemhäusern

In Kap.4.2 wurde ausführlich auf die Problemstellung der Meldungsrecherche im Produktsupport von Systemhäusern eingegangen und auf Basis von Daten- und Systemanalysen ein verbesserter Prozess der Meldungsrecherche erarbeitet. Nach Spezialisierung von SWAPit auf die Anforderungen des Supports wurde das System drei Wochen lang im Support-Center in Wien ergänzend zum vorhandenen Suchsystem als exploratives Recherchewerkzeug eingesetzt und in Workshops und Befragungen von den Mitarbeitern bewertet [Hi06].

7.4.1 Bewertungsziele

Neben den allgemeinen Bewertungszielen (Kap.7.1) standen die folgenden Fragestellungen bei dieser Laborstudie im Vordergrund:

- Bieten die thematische Gruppierung (explorative Ähnlichkeitssuche) und die integrierte Navigierbarkeit von Suchkriterien *Vorteile* für die operative Meldungsrecherche?
- Welches *Ähnlichkeitsmaß* ist für die inhaltlich heterogenen und komplexen Meldungen aufgabengemessen?
- Wie aufwändig sind für die Meldungsbearbeiter *Einarbeitung und Handhabung* der für sie neuen Systemunterstützung?
- Inwiefern wird das neue Werkzeug durch die Meldungsbearbeiter als Ergänzung des jahrelang etablierten und jährlich geschulten Suchsystems *akzeptiert*?

7.4.2 Systemdesign

Aus den ausführlichen Analysen in Kap.4.2 ging hervor, welche Informationen für die Meldungsrecherche verfügbar und wichtig sind und daher geeignet zugreifbar gemacht werden sollten. Im ersten Iterationsschritt wurden an einer SWAPit-Systemskizze verschiedene Szenarien der Meldungsrecherche ausprobiert. Daraus ergaben sich Anforderungen für den zweiten Iterationsschritt, in dem auch ein funktionierender Systemprototyp basierend auf SWAPit entwickelt wurde.

Die *informatrischen Anforderungen* orientieren sich an den in der Suchmaschine bereitgestellten Informationen, da in diesem Fallbeispiel ein bestehender Suchprozess optimiert und ergänzt werden soll. Hinzu kommen die folgenden informatrischen Anforderungen:

- Bearbeiter sind oft spezialisiert auf bestimmte Systemkomponenten. Da der Support kleinen Gruppen arbeitet, ist die *Information des Bearbeiters besonders wichtig* und sollte flexibel zugreifbar sein
- Für eine Strukturierung der Suchergebnisse müssen Gruppierungen dargestellt werden, die die *Meldungstexte nach inhaltlicher Ähnlichkeit bezüglich Kernthemen miteinander in Beziehung setzen* (Themenübersicht)
- Um den Einfluss der Qualität einer Suchanfrage und guter Kenntnisse der Meldungssammlung auf den Erfolg des Rechercheprozesses zu relativieren, sollten inhaltliche *Querbezüge zwischen den Meldungen im Suchergebnis und anderen Meldungen* zur Systemkomponente dargestellt werden

Die funktionalen Anforderungen...

- *Suchparameter* sollen *interaktiv flexibel und effizient variiert* werden können. Umgekehrt sollten die *Ergebnisdokumente bezüglich der Suchparameter charakterisiert* werden können (z.B.: Welcher Kollege hat die meisten der gefundenen Meldungen bearbeitet?)
- Möglichkeiten zur Kombination von *zielgerichteter Suche und Explorationsstrategien*

Die sonstigen Anforderung:

- Eine geschickte Vorberechnung ist sinnvoll, da Meldungsrecherchen schnell starten müssen und sich die Teilmenge der Meldungen eines Systemmoduls nicht allzu schnell ändert
- Um *Einprägungseffekte bzgl. der Themenkarten* nutzbar zu machen, sollten Themenkarten fortgeschrieben werden, damit sich Lageverhältnisse nicht drastisch ändern

Alle Meldungen einer Systemkomponente werden auf einer Themenkarte dargestellt. Die strukturierten Daten werden teilweise tabellarisch, dann hierarchisch, manchmal auch redundant abgebildet. In zwei Iterationsschritten⁸⁰ wurde ein Sichtendesign entwickelt und optimiert (Abbildung 7.6).

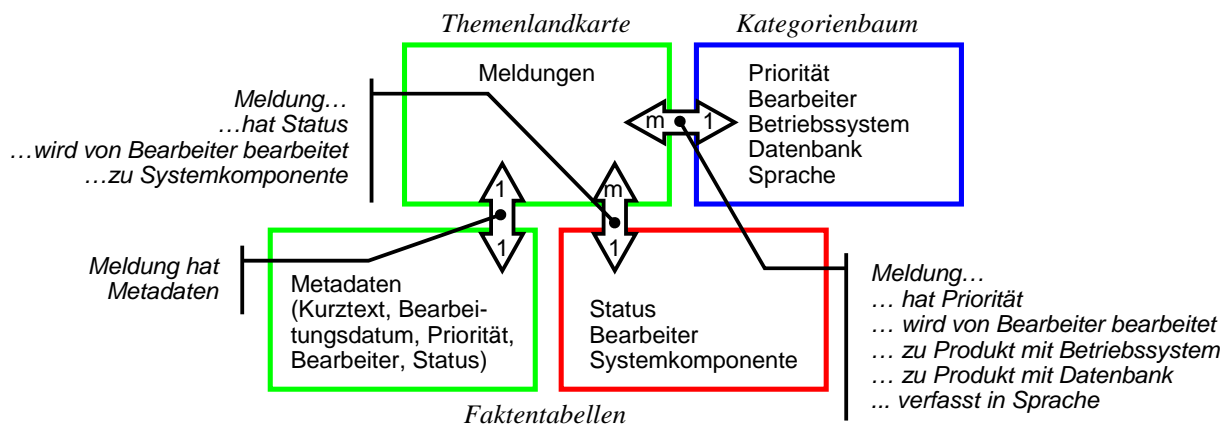


Abbildung 7.6: Systemskizze von SWAPit für den Produktsupport.

Die redundante Darstellung eröffnet verschiedene Möglichkeiten, auf die Information zuzugreifen. So taucht der Meldungsbearbeiter als eigenständige Datentabelle, als Attribut in den Metadaten der Meldungen, sowie als Zweig im Baum auf. Im Baum lassen sich alle Bearbeiter leicht ein- und ausblenden und übersichtlich darstellen, während die Bearbeiter-Tabelle das Filtern, z.B. anhand des Namens, erlaubt. Außerdem kann zu einer Auswahl von Meldungen eine Statistik über die damit befassten Meldungsbearbeiter dargestellt werden.

Die Wahl eines adäquaten *Ähnlichkeitsmaßes* stellte eine besondere Herausforderung dar. Ähnlich wie die Kundenbeschwerden in der ersten Fallstudie (Kap.7.3) umfassen die Meldungstexte die Korrespondenz zwischen Meldungsbearbeiter und Kunden und dokumentieren mögliche Lösungswege. Eine einfache statistische Indexierung, wie sie für das erste Systemdesign verwendet wurde, repräsentiert Themen und Kommunikationsartefakte gleichermaßen. Gewünscht war eine Gruppierung von Meldungen nach den thematischen Schwerpunkten einer Systemkomponente. Daher wurde ein Indexierungsvokabular vorgegeben (Positivliste; Feature-Index). Um der relativ kurzen Gültigkeitsdauer von Meldungen aufgrund Produktveränderungen Rechnung zu tragen, wurde von einer statischen Themenvorgabe abgesehen. Stattdessen wurden die qualitativ hochwertigeren Hinweisdokumente für die betrachtete Systemkomponente für eine automatische Extraktion der wichtigsten Themen verwendet: Aus den Hinweisen wurden die jeweils 300 häufigsten Terme extrahiert und gesichtet. Diese Terme dienten als Indexierungsvokabular für die Meldungsdokumente. Die Strukturen der entstandenen Dokumentenlandkarten wurden von Mitarbeitern des Systemhauses sehr gut bewertet und folglich für die Studie verwendet. Für die Suche in den Meldungen wurde jedoch nach wie vor der unbeeinflusste Term-Index verwendet. Abbildung 7.7 zeigt das für die Meldungsbearbeitung konfigurierte SWAPit-System.

⁸⁰ Für eine ergebnisorientierte Darstellung werden beide Zyklen des Systemdesigns gemeinsam dargestellt.

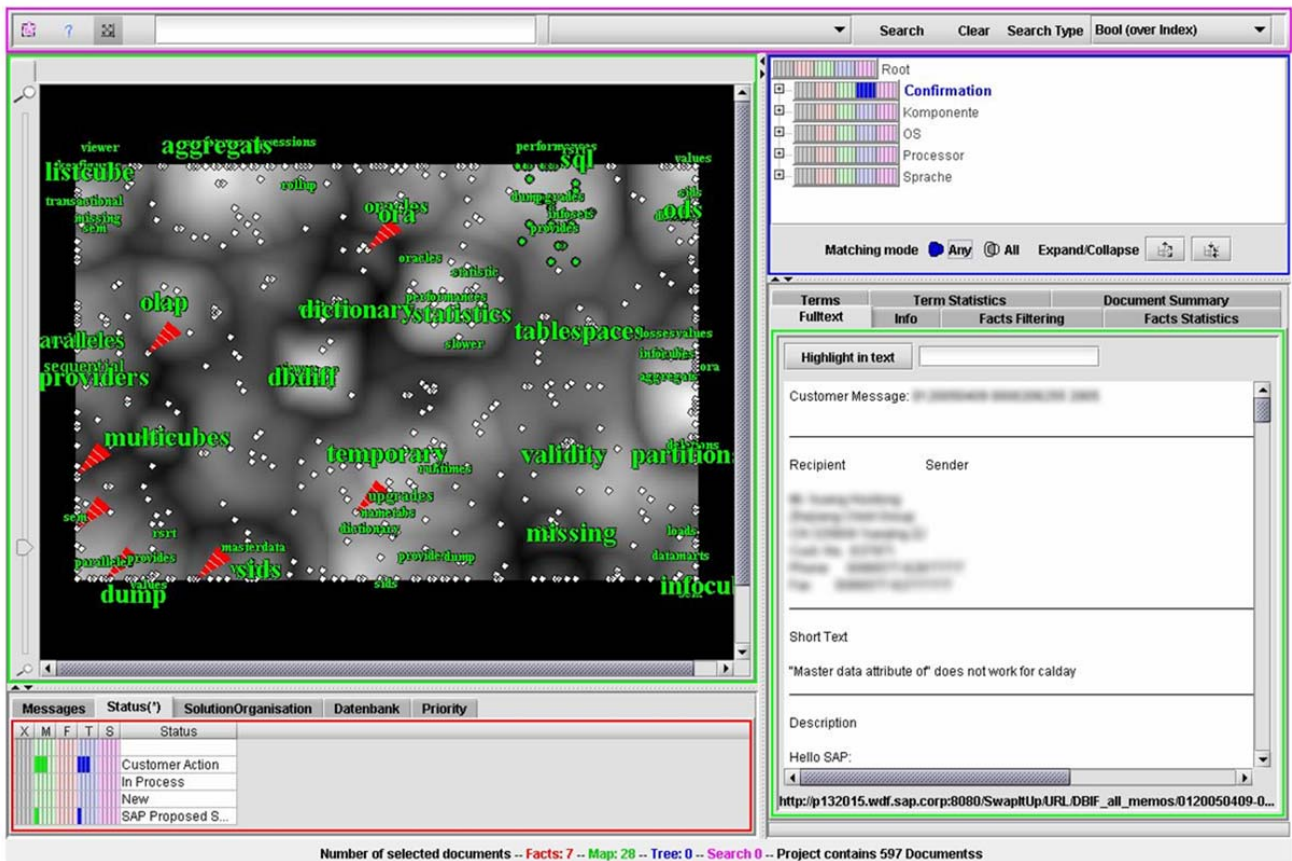


Abbildung 7.7: Screenshot des SWAPit-Systems für die Meldungsrecherche mit ca. 600 Meldungsdocumenten und 3 Sichten (Ähnlichkeitssicht links oben, Faktensicht links unten, Kategoriensicht rechts oben).

7.4.3 Integration

Die SWAPit-Lösung für den Kundensupport war als Ergänzung der bewährten Schlagwortsuche konzipiert. Für eine technische Integration beider Systeme ergeben sich einige Anforderungen:

- Dokumentenvergleich und Suche in SWAPit verwenden spezialisierte Suchmaschine
- Suchergebnisse werden nach vorberechneten Karten segmentiert. Diese Karten stellen den Einstiegspunkt für die explorative Recherche dar.
- In den Themenkarte werden die Suchergebnisse hervorgehoben und dienen als Einstiegspunkte
- Die zu bearbeitende Kundenmeldung wird auf die Themenkarte abgebildet und dient als Einstiegspunkt
- Für jede Systemkomponente der dritten oder vierten Stufe in der Komponentenhierarchie wird ein eigenes SWAPit-Projekt aufgesetzt (jeweils rund 400 Meldungen)
- Das Update der SWAPit-Projekte (Preprocessing) erfolgt alle 14 Tage oder bei mehr als zehn Prozent neu hinzugekommener Meldungen

7.4.4 Versuchsdesign und Durchführung

Im Rahmen der durchgeführten Studie konnten die Systeme technisch nicht so vollständig integriert werden, dass ein durchgängiger Kontrollfluss möglich gewesen wäre. Die *Prototypische Systemintegration* bestand in einem Import der Meldungen aus dem CSN-System (Customer Support Network). Kurz vor Beginn der Testphase wurden zu fünf ausgewählten Systemkomponenten der 3. bzw. 4. Stufe der Komponentenhierar-

chie jeweils zwischen 100 und 1.000 Meldungen mit einem Alter von max. 2 Jahren herunter geladen. Zu diesen Komponenten wurde je eine Dokumentenlandkarte berechnet und zusammen mit den Metadaten der Meldungen in SWAPit dargestellt.

Rahmenbedingungen und Versuchsdesign

Nach einer ausführlichen Anforderungsanalyse (informelle, dann strukturierte Interviews) sollte eine Werkzeuganpassung und anschließend die prototypische Integration in einem Primary Support Center des Systemhauses erfolgen, wo SWAPit im Mai 2006 drei Wochen von den Meldungsbearbeitern genutzt und bewertet werden würde. Sechs Mitarbeiter des Business Warehouse Teams erklärten sich dazu bereit. Drei von ihnen wurden als Fokusgruppe ausgewählt, da sie SWAPit für mehr als zehn Meldungsrecherchen verwendeten und somit über die meiste Erfahrung mit dem System verfügten. Von den anderen drei Mitarbeitern wurden lediglich Kommentare, Kritik oder Anregungen berücksichtigt.

SWAPit sollte als Ergänzung für Recherchen mit vagem Informationsbedürfnis, nicht als Alternative zum Suchsystem eingesetzt werden, weil beide Werkzeuge komplementäre Stärken besitzen. Dementsprechend wurde den in beiden Systemen geschulten Meldungsbearbeitern die Entscheidung überlassen, mit welchem Werkzeug sie eine Recherche durchführen wollen. Zur Nutzenmessung beider Werkzeuge wurden die gleichen Maße verwendet, um eine gute Vergleichbarkeit der Ergebnisse zu ermöglichen.

Eine Feldstudie wurde durchgeführt, um die prinzipielle Eignung von Konzept und Werkzeug für die Meldungsbearbeitung zu untersuchen. Im Gegensatz zu Laborstudien, die auf statistische Verallgemeinerbarkeit durch Kontrolle aller unabhängigen Einflussfaktoren abzielen, steht die kontrollierte Beobachtung realer Problem- und Lösungssituationen im Vordergrund. Testnutzer und Arbeitsaufgaben sind durch die reale Arbeitsumgebung festgelegt und nicht zwangsläufig homogen, bestimmen jedoch repräsentativ die faktischen Anforderungen eines Unternehmens an eine Technologie. Ergebnis sind konkrete Erkenntnisse über Potentiale und Defizite einer Technologie im Anwendungskontext.

Die *Schulung* der Mitarbeiter erfolgte, einem festen Ablauf folgend, individuell und problemorientiert. Beide Werkzeuge wurden nacheinander für die Recherche in Bezug auf eine aktuelle Kundenmeldung angewendet, das neue System erklärt und erprobt (u.a. Interpretation und Interaktion mit der Dokumentenlandkarte, Sichten, Auswählen, Schlagwortsuche), und das korrekte Ausfüllen der Bearbeitungsbögen geübt. Die ersten Fälle der selbständigen Werkzeugnutzung wurden separat betrachtet und als Migrationsphase gewertet. Vor der Migrationsphase stand die Erwartungsphase, die mit der Schulung begann. Auf die Migrationsphase folgte die Erfahrungsphase – der eigentliche Betrachtungsfokus des Experiments. Jeweils nach der Erwartungs- und der Erfahrungsphase wurden die Bewertungsbögen ausgefüllt und es fand eine Feedbackrunde statt.

Bewertungsschema, Maße und Datenerhebung

Für die Bewertung beider Systeme sollte das gleiche Maßsystem verwendet werden, um vergleichbare Ergebnisse zu erhalten. Die Leistungsfähigkeit klassischer Retrieval-Systeme, wie der Meldungssuchmaschine, lassen sich weitgehend an der Ergebnisqualität messen. Die etablierten Retrievalmaße, darunter Precision und Recall, helfen bei der adäquaten Bewertung und ermöglichen den Vergleich zwischen Retrievalsystemen⁸¹. Bei interaktiven Suchsystemen wie SWAPit ist die Bewertung komplizierter. Schließlich besteht gerade eine potentielle Stärke solcher Systeme darin, menschliche Erfahrung, Intuition und kognitive Fähigkeiten für den explorativen Suchprozess nutzbar zu machen. Bei Einbeziehung des Benutzers als zentrale Komponente des zu bewertenden Gesamtsystems besteht die Herausforderung darin, Rückschlüsse auf

⁸¹ Die Text REtrieval Conference (<http://trec.nist.gov/>) verfolgt seit vielen Jahren das Ziel, Suchsysteme adäquat zu vergleichen.

das Werkzeug zu ziehen. Dabei spielen die „Kontext“-Dimensionen wie Funktionalität und Benutzbarkeit naturgemäß eine besondere Rolle.

Für eine differenzierte Nutzenbetrachtung wurden, in Anlehnung an die Idee der Balanced Scorecard zur vielschichtigen Bewertung der Performanz von Unternehmen, vier Bewertungsdimensionen unterschieden [Hi06]: Wahrgenommene Funktionalität, wahrgenommene Bedienbarkeit, Prozesseffizienz und Ergebnisqualität (Abbildung 7.9). Diese Dimensionen verstehen sich als komplementäre Betrachtungsperspektiven und wurden jeweils in Maße herunter gebrochen. Die Diversifizierung der Dimensionen Funktionalität, Bedienbarkeit und Prozesseffizienz ist zum Teil angelehnt an die Norm ISO/IEC 9126 zur Bewertung der Softwarequalität. Bietet ein Werkzeug die richtigen Funktionalitäten und hilft, gute Ergebnisse zu produzieren, so wird es als nützlich eingestuft. Wenn gleichzeitig der Aufwand durch Bedienung des Werkzeugs gering ist und dessen Eingliederung in den Arbeitsprozess gelingt, so ist dies ebenfalls dem Nutzen zuträglich. Das skizzierte Bewertungsschema basiert auf dem DART⁸²-Ansatz für die qualitative indikatorbasierte Nutzenmessung von Informationssystemen unter besonderer Berücksichtigung der Benutzerakzeptanz [AHS03].

Alle Maße wurden in Form einer Einschätzung durch die Meldungsbearbeiter zweimal, zum Anfang und zum Ende der Studie, durch Befragung erhoben. Dabei wurden, analog zum Bewertungssystem der Schulnoten, ordinale 6er-Skalen verwendet. Die Ergebnisqualität und die Prozesseffizienz wurden zusätzlich durch Messung bestimmt. Dazu wurde während jeder Recherche von den Meldungsbearbeitern ein Protokollierungsbogen ausgefüllt, u.a. zur Dokumentation von verwendetem Werkzeug, Meldungsnummer, Suchzeiten, gefundenen Dokumenten und deren Relevanz, sowie Erfolg der Recherche. Alle wichtigen Interaktionen mit SWAPit wurden in einem System-Log aufgezeichnet mit dem Ziel, Unstimmigkeiten auf den Erhebungsbögen zu klären und Interpretationen zu stützen.

7.4.5 Ergebnisse

Die Ergebnisse der beiden Mitarbeiterbefragungen und der Messungen sind in Abbildung 7.9 übersichtlich aufgeführt und werden weiter unten ausführlich diskutiert. Sie beziehen sich auf die Fokusgruppe der drei Nutzer mit besonders intensiver Erfahrung bzgl. Beider Systeme. Insgesamt wurden 60 Meldungen bearbeitet. Bei der Hälfte der Recherchen entschieden die Mitarbeiter sich für SWAPit, weil sie die Fragestellung für recht offen einschätzten und die Formulierung einer geeigneten Suchanfrage in der Suchmaschine nicht offensichtlich war. Die Mitarbeiter zeigten großes Interesse und waren für die Studie motiviert. Die prototypische Integration von SWAPit in die operative Umgebung der Meldungsbearbeitung verlief problemlos, wengleich Datenexport und Transformation mit großem Aufwand verbunden waren. Die Anpassung der Sichten und des Ähnlichkeitsmaßes an die Prozesse und Daten der Meldungsrecherche waren mit geringem Konfigurationsaufwand durchführbar.

SWAPit (Abbildung 7.9 a, b, c, e)

In der Erwartungsphase (a) lassen sich große Unterschiede in der Einschätzung des neuen Systems erkennen. Ausgenommen ist die durchweg hohe Erwartung an die Gruppierungsfunktion. Bezüglich Bedienbarkeit, Funktionalität und Prozesseffizienz gibt es überaus optimistische, aber auch eher pessimistische Erwartungen. Bei der Ergebnisqualität sind alle Teilnehmer eher zurückhaltend bei der Bewertung. Die Erfahrungsphase zeigt insgesamt eine deutliche Festigung der Ergebnisse: Die Standardabweichung nimmt, gemittelt über alle Maße, vom 1,13 auf 0,92 ab. Betrachten wir die Mittelwerte (c), so sind positive Korrekturen bei den Maßen *Einarbeitungsaufwand* und *Interaktionsintensität* (nur wenige Interaktionsschritte nötig) zu erkennen, wohingegen bei den Maßen *Suchfunktion* und *Recall/Precision* (Einschätzung der Vollständigkeit u.

⁸² Dynamic Acceptance Modell for Re-Evaluating Technologies

Genauigkeit der Rechercheergebnisse) die Erwartungen enttäuscht wurden. Diese beiden Negativ-Korrekturen hängen zusammen, wie auch die anderen Ergebnisse bestätigen: Die Schlagwortsuche von SWAPit entsprach, wie sich erst während der Studie heraus stellte, allen Anforderungen der Mitarbeiter. So werden häufig bei der Beschreibung technischer Terme Sonderzeichen oder Zahlen bei der Suche benötigt, die in der Werkzeugrealisierung von SWAPit nicht zur Verfügung standen. Der negative Effekt auf die Rechercheergebnisse wird noch verstärkt durch die Tatsache, dass jede Recherche mit einer Schlagwortsuche beginnt.

Betrachten wir die absolute Bewertung von SWAPit in der Erfahrungsphase (b), so wird eine sehr gute bis gute Bedienbarkeit bestätigt. Allerdings besteht bei der *Intuitivität* noch Optimierungsspielraum (Mittelwert: 3). Die Prozesseffizienz wird im Durchschnitt als gut angesehen. Alle Ergebnis-bezogenen Maße werden allerdings nur ungefähr mit 4 bewertet, was zum Teil auf die fehlenden Suchfunktionalitäten zurück zu führen ist. Möglicherweise mitverantwortlich für die schlechten absoluten *Precision*-Werte ist die Tatsache, dass in zahlreichen Fällen gar keine relevanten Meldungen in der Datenbank vorhanden waren. Die SWAPit-spezifischen Funktionalitäten wurden angenommen und zwischen 2 und 3 bewertet.

SWAPit vs. herkömmliche Suchmaschine (Abbildung 7.9 b, d, f)

Vergleicht man die Streuung der Einschätzungen der Systeme SWAPit (b) und der Suchmaschine (d), so fällt die größere Uneinigkeit bei der Suchmaschine auf: Die mittlere Standardabweichung über alle Maße beträgt 1,33 – für SWAPit lediglich 0,92. Ganz besonders im Bereich Bedienbarkeit, bei der *Suchdauer* und den Ergebnismaßen *Precision/Recall* scheiden sich die Geister bei der Suchmaschine. Die Prozesseffizienz wird insgesamt zwischen 2 und 3 bewertet.

Während an SWAPit (gegenüber der Suchmaschine) die Visualisierung, die Navigation und die Suchdauer geschätzt werden (f), wird bei der Suchmaschine (gegenüber SWAPit) die Suchfunktion mit ca. 2 weit besser eingestuft. Die Ergebnismaße werden für die Suchmaschine ungefähr eine Note besser bewertet.

Erwartungen aus der Problemanalyse (Kap.4.2)

Die Systemkomponentenweise Darstellung von Meldungen (Kontextkarte) wurde als thematische *Orientierungshilfe* bei der Recherche sehr geschätzt, was sich in der guten Bewertung der Visualisierung zeigt und auch bei den Workshops geäußert wurde. Die gute Akzeptanz von *Themen- bzw. Suchkriterien-orientiertem Zugriff* auf die Meldungen wird belegt durch die geäußerte Zufriedenheit mit der Navigation. Wie erwartet, wurde auch die *Recherchedauer* als gut eingestuft. Die *Ergebnisqualität* allerdings ist nicht zufrieden stellend. Allerdings lässt sich dies auf die nicht unspezialisierte Schlagwortsuche in SWAPit zurückführen – ein offensichtliches Optimierungspotential.

Nutzung der Interaktionsmöglichkeiten

Fast ein Drittel der aufgezeichneten Benutzeraktivitäten fand in der Themenkarte statt (Abbildung 7.8). Offensichtlich haben die Benutzer diese zur Exploration genutzt. Ein weiteres knappes Drittel bilden die durchgeführten Suchanfragen. Die Vertrautheit der Arbeit mit dem anfrageorientierten Suchsystem mag zu dieser hohen Zahl beigetragen haben. Termstatistik und Kategoriensicht wurden dagegen kaum genutzt.

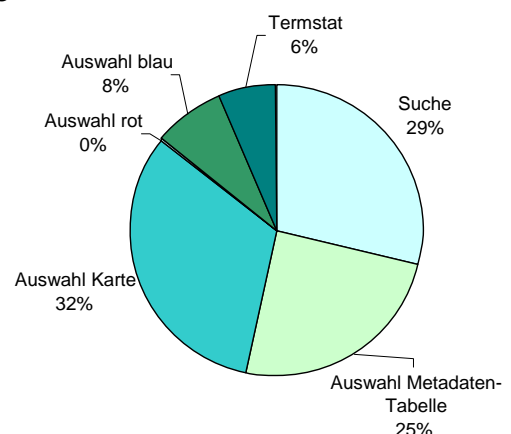


Abbildung 7.8: Verwendung des SWAPit-Systems

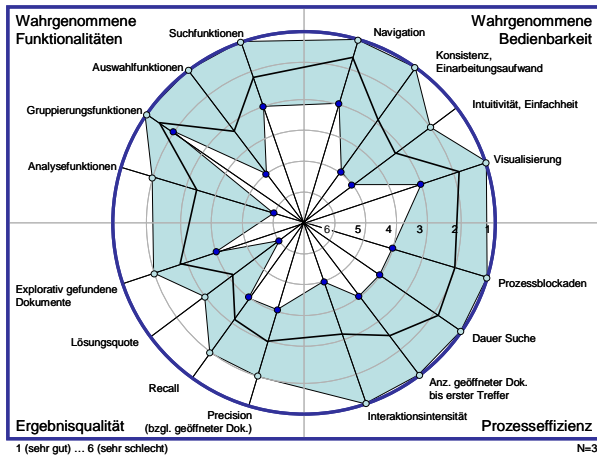
Fazit

Die sorgfältige Differenzierung von Funktionalität, Bedienbarkeit, Prozessen und Rechercheergebnis ermöglicht einen dezidierten Erkenntnisgewinn. So lässt sich festhalten, dass die exzellente Akzeptanz und Erlernbarkeit die Erwartungen übertrafen, zumal das Expertenwerkzeug SWAPit

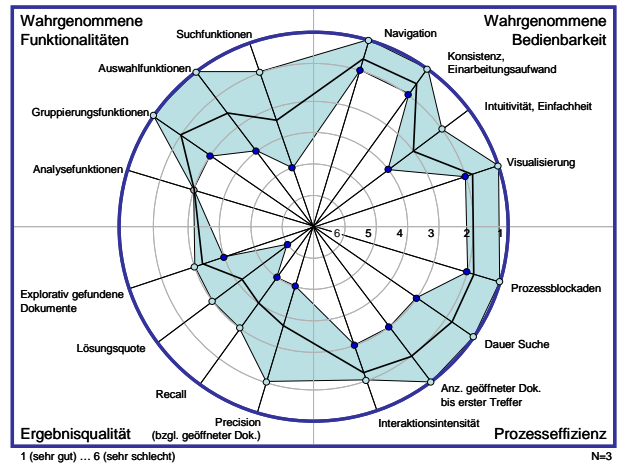
ohne Einschränkung der Vielfalt an Funktionalitäten und ohne Anpassung der Software eingesetzt wurde. Wie bereits diskutiert, spielen in einem stark interaktiven System wie SWAPit, wo der Mensch eine Kernkomponente darstellt, die Aspekte Benutzbarkeit, Prozess- und Funktionsangemessenheit eine dominierende Rolle. Dass diese Aspekte schon beim prototypischen Einsatz gute Bewertungen erhielten, unterstreicht sowohl die Flexibilität von Konzept und Werkzeug als auch die generelle Eignung für den Einsatz bei der Meldungsrecherche.

Zugleich muss festgestellt werden, dass die Systemunterstützung ihr Gesamtpotential aufgrund der bereits genannten technischen Hindernisse bei der Schlagwortsuche nicht entfalten konnte – somit ist die erreichbare Ergebnisqualität eine offene Frage. Als Perspektive zur Optimierung sind zwei Wege denkbar: Eine Möglichkeit ist, die traditionelle Recherchestrategie, welche immer von einer Suchergebnisliste ausgeht, besser durch SWAPit zu unterstützen – dazu muss SWAPit nicht nur prototypisch, sondern real mit der spezialisierten Suchmaschine integriert werden, was durch SWAPit's modulares und erweiterbares Systemkonzept ermöglicht wird. Alternativ, oder aber zusätzlich, könnte eine möglicherweise adäquatere Recherchestrategie damit beginnen, zur bearbeiteten Meldung automatisch, d.h. ohne Angabe von Schlagworten durch den Bearbeiter, einen geeigneten Startpunkt für die explorative Recherche in der Themenkarte zu bestimmen. Erste Experimente in diese Richtung waren bereits viel versprechend. Durch eine „echte“ Integration mit einem automatischen Update-Prozess für die Themenkarten und die Metadaten, würde außerdem die Aktualität der Meldungen erhöht, ein kritischer Faktor in Anbetracht der kurzen Gültigkeitszeiten von Kundenmeldungen. Durch Einbeziehung der Referenzen auf Hinweise oder der Meldungsbearbeiter in die Ähnlichkeitsbestimmung von Meldungen könnte die Qualität der thematischen Karten weiter optimiert werden. Während die Meldungsbearbeitung in Wien als spezialisiert und ausführlich gilt, wird in Niedriglohnländern oft eine andere Strategie gefahren – Meldungen werden schneller bearbeitet und der Recherche in bereits gelösten Fällen kommt somit ein noch höherer Stellenwert zu. Eine Erprobung des Werkzeugs in einem solchen Supportcenter wäre eine weitere interessante Studie. Von der differenzierten Analyse überzeugt, hat das Systemhaus entschieden, im Rahmen eines Praktikums eine Wirtschaftlichkeitsanalyse für den Einsatz von SWAPit im Support durchzuführen, die als Entscheidungsgrundlage für das Management über ein kommerzielles Folgeprojekt dienen wird.

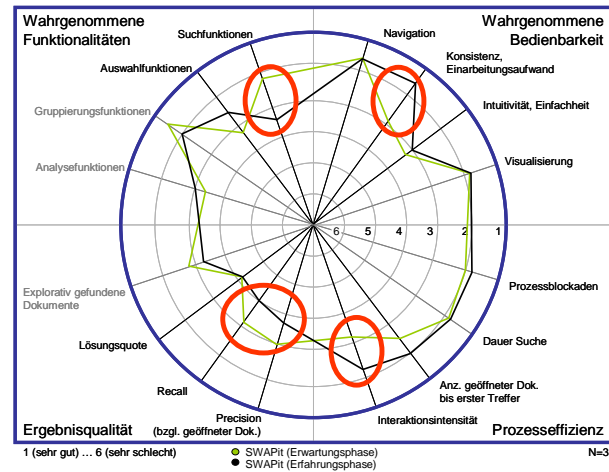
(a) SWAPit Erwartungsphase



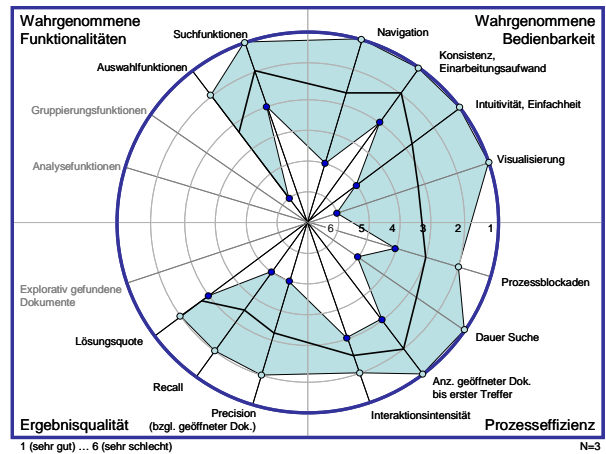
(b) SWAPit Erfahrungsphase



(c) SWAPit (Erwartungs- vs. Erfahrungsphase, Mittelw.)



(d) Suchmaschine (Erfahrungsphase)



(e) Messergebnisse

| Prozessblockadezeiten | N | Summe | Mittelwert |
|------------------------|----|-------------------|-----------------|
| SWAPit | 30 | 3 | 0,1 |
| TREX | 21 | 1 | 0,05 |
| Suchzeit | N | Min/Max | Mittelwert |
| SWAPit | 30 | 0,5 / 33,3 (Min.) | 7,44 (Min.) |
| TREX | 21 | 1 / 24 (Min.) | 5,56 (Min.) |
| Geöff. Dok. | N | Min/Max | Mittelwert |
| SWAPit | 11 | 1 / 5 (Dok.) | 3,18 (Dok.) |
| TREX | 10 | 1 / 4 (Dok.) | 2,1 (Dok.) |
| Interaktionsintensität | N | Min/Max | Mittelwert |
| SWAPit | 30 | 0 / 15 (Schritte) | 4,37 (Schritte) |
| TREX | 21 | 1 / 8 (Schritte) | 2,81 |
| Precision (Off. Dok.) | N | Min/Max | Mittelwert |
| SWAPit | 30 | 0 / 1 | 0,12 |
| TREX | 21 | 0 / 1 | 0,32 |
| Schritte pro Minute | N | | Mittelwert |
| SWAPit | 30 | | 0,59 |
| TREX | 21 | | 0,51 |
| Lösungsquote | N | Summe | Mittelwert |
| SWAPit | 30 | 8 | 0,27 |
| TREX | 21 | 7 | 0,33 |

(f) SWAPit vs. Suchmasch. (Erfahrungsph., Mittelwerte)

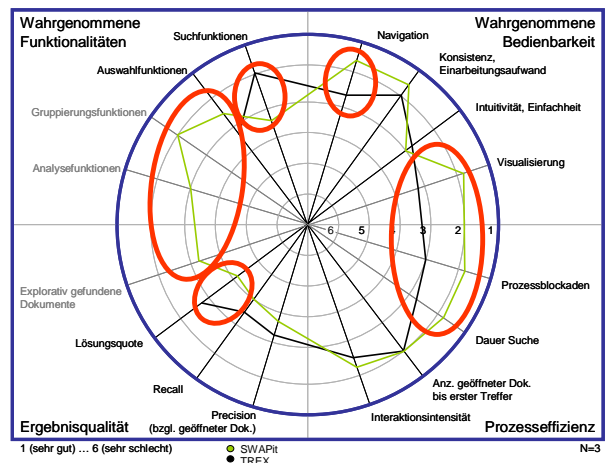


Abbildung 7.9: Ergebnisse der Feldstudie im Produktsupport des Systemhauses (Mittelwerte, Min. und Maxima)

7.5 Fallstudie 3: Aktivitätsanalyse in Kooperationsumgebungen

In Kap.0 wurde ausführlich auf die Problemstellung der Aktivitätsanalyse in Virtuellen Kooperationsumgebungen eingegangen und das Potential einer Systemunterstützung durch SWAPit abgeleitet. Nach Anpassung des Systems auf den Informations- und Funktionalitätsbedarf der Aktivitäts- und Inhaltsanalyse in kooperativen Arbeitsumgebungen wurde das System in einem Workshop von 5 Forschern des Fraunhofer Instituts FIT (Forschungsbereich Kooperationssysteme) in Kombination mit dem BSCW-System eingesetzt und bewertet [SP+07]. In der Laborstudie wurden drei Klassen von Aufgaben gestellt. Als Ergebnisse wurden Fragebögen, Aufzeichnungen der Interaktionsschritte, und Diskussionsbeiträge ausgewertet. Später hatten andere interessierte Konsortialpartner des EcoSpace-Projektes ebenfalls Gelegenheit, das System für ihre Arbeitsaufgaben auszuprobieren.

7.5.1 Bewertungsziele

Neben den allgemeinen Bewertungszielen (Kap.7.1) standen die folgenden Fragestellungen bei dieser Benutzerstudie im Vordergrund:

- Hilft SWAPit durch umfassende Zugriffsmöglichkeiten, kompakte Übersichtsdarstellungen, inhaltsbasierten Zugriff und Bereitstellung der gesamten Aktivitätshistorie, *Informationszugriff und Awareness zu optimieren?*
- Wie kann SWAPit in eine Kooperationsumgebung *integriert* werden?
- (Wie) werden die *neuen Möglichkeiten* zur Personen-, Themen- und Aktivitätsorientierten Recherche und *akzeptiert* und *genutzt?*
- Wie schwierig finden die Forscher *Einarbeitung und Handhabung* der für sie neuen Systemunterstützung?

7.5.2 Systemdesign

Die Szenarioanalyse (Kap.4) hat gezeigt, welche Informationen in kooperativen Arbeitsumgebungen verfügbar und relevant für Informationszugriff und Aktivitätsanalyse sind. Dabei sollen *alle* in BSCW verfügbaren Informationen in Sichten angezeigt werden. In Abbildung 7.10 werden die Informationsarten durch Knoten und die Zugriffspfade als Kanten eines Graphen dargestellt. Jeder Pfeil symbolisiert eine mögliche interaktive Anfrage. Ein Pfeil von Akteuren zu Themen bedeutet beispielsweise, dass sich für eine Menge von Personen herausfinden lässt, mit welchen Themen sich diese Gruppe beschäftigt hat (genauer: Themen, die in Artefakten eine Rolle spielen, welche von den Personen bearbeitet oder gelesen werden). Die aufgeführten Informationsarten erinnern an die Dimensionen von [PPS04] zur Unterscheidung von Awareness-Werkzeugen (Kap.4.3.2), ergänzt um die Informationsarten „Wissensstrukturen“ (z.B. Ordnerstruktur des BSCW-Systems) und „Themen“ für den inhaltsbasierten Zugriff. Das intendierte System zeichnet sich gegenüber existierenden Spezialwerkzeugen (Kap.4.3.2) durch seine Vielseitigkeit aus, da *alle* Informationen *parallel* zur Navigation genutzt werden können. Insbesondere die *inhaltliche Analyse* von Artefakten stellt eine wichtige Neuerung dar.

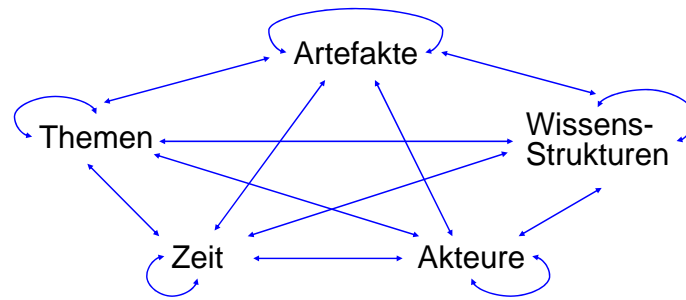


Abbildung 7.10: Informationen und Zugriffspfade für Interaktive Anfragen bei der Aktivitätsanalyse

Anforderungen an das Systemdesign:

- Die interaktiven Anfragen aus Abbildung 7.10 sollen mit Hilfe des Systems gestellt werden können. Insbesondere soll die gesamte Aktivitätshistorie dargestellt und zur Exploration nutzbar gemacht werden.
- SWAPit soll in den Arbeitsablauf mit dem BSCW-Werkzeug integriert werden. Insbesondere soll BSCW-Benutzern nicht zugemutet werden, auf längere Berechnungen warten zu müssen.
- Die Zugriffsrechte der kooperativen Arbeitsumgebung definieren für jeden Benutzer eine individuelle Perspektive auf den Bestand aller Informationen. Diese individuelle Perspektive soll auch in SWAPit berücksichtigt werden.

Nach einem initialen Sichtendesign für den Workshop am Fraunhofer-Institut FIT (Kap.4.3.3) wurde ein verbessertes Sichtendesign entwickelt⁸³ (Abbildung 7.11). Jede Informationsart wird in einer geeigneten Sicht dargestellt. Sichten sind verknüpft durch Aktivitäten. Die Artefakte werden auf der Themenkarte dargestellt und sind mit tabellarischen Metadaten (z.B. Name und Beschreibung) verknüpft. Die Aktivitäten, die sich in drei Arten unterscheiden lassen (Erstellen, Modifizieren und Lesen von Artefakten) werden explizit in drei Tabellen dargestellt und sind mit den entsprechenden Artefakten verknüpft. Die Projektteilnehmer (Akteure) werden als Kategorien dargestellt und (redundant) in einer Faktensicht tabellarisch aufgelistet, damit sich Fragen nach Kooperation bzw. gemeinsamen Interessen adressieren lassen. Die Ordnerstrukturen des BSCW, in welchen die Artefakte abgelegt sind, werden hierarchisch in einer Kategoriensicht dargestellt.

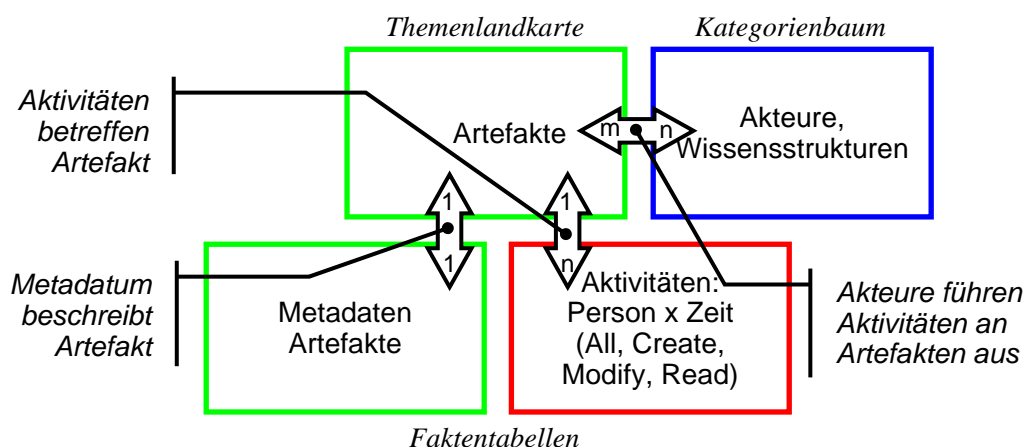


Abbildung 7.11: SWAPit-Systemskizze für die Aktivitätsanalyse in kooperativen Arbeitsumgebungen.

⁸³ Für eine ergebnisorientierte Darstellung werden beide Zyklen des Systemdesigns gemeinsam dargestellt.

Abbildung 7.12 zeigt das SWAPit-System, konfiguriert für Aktivitäts- und Inhaltsanalysen in kooperativen Arbeitsumgebungen. In diesem Beispiel werden Informationen aus dem EU-Forschungsprojekt EcoSpace darstellt und flexibel explorierbar gemacht.

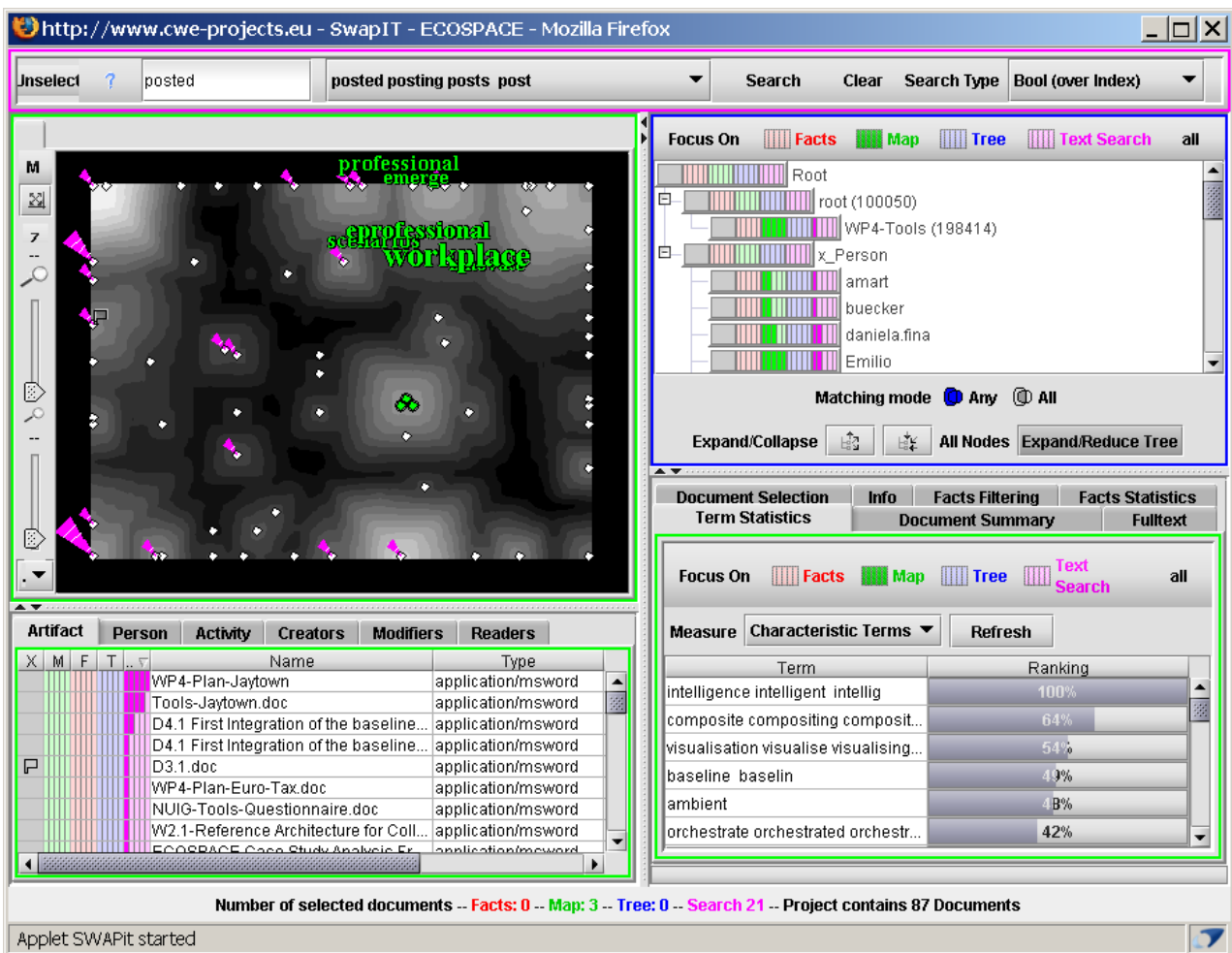


Abbildung 7.12: Screenshot des SWAPit-Systems für die Aktivitätsanalyse mit über 87 Word-Dokumenten und 3 Sichten (Ähnlichkeitssicht links oben, Faktensicht links unten, Kategoriensicht rechts oben). Im Beispiel kann der Benutzer nicht nur auf einen Blick erkennen, in welchen Dokumenten der Suchbegriff vorkommt, sondern auch, zu welchen Workpackages die Suchtreffer gehören, welche Personen damit zu tun hatten, usw.

Zur Berechnung der Themenkarte wurde SWAPit's Standardmodul zur statistischen Ähnlichkeitsanalyse von Dokumenten verwendet. Dokumente, für die Aktivitäten stattgefunden haben, die für den angemeldeten Benutzer neu sind, werden durch Fähnchen in Themenkarte und Metadattentabelle gekennzeichnet.

7.5.3 Integration

Die Berücksichtigung der benutzerspezifischen Zugriffsrechte sowie die Notwendigkeit einer kurzen Startzeit von SWAPit legen ein Preprocessing-Konzept nahe, das folgendermaßen ausgestaltet wurde: Ein BSCW-Administrator legt durch ein neues BSCW-Attribut fest, welcher BSCW-Ordner als SWAPit-Projekte bereit gestellt werden sollen. Diese SWAPit-Projekte werden in regelmäßigen Abständen (z.B. jede Nacht) automatisch erstellt bzw. aktualisiert. Dabei werden alle verfügbaren Dokumente des betreffenden Ordners und aller Unterordner und alle verfügbaren Informationen (Dokumente, Personen, Aktivitäten, Ordnerstrukturen) verarbeitet. Wird ein vorberechnetes SWAPit-Projekt von einem Benutzer aufgerufen, so wird es beim Start personalisiert:

- Nur die für den Benutzer zugelassenen Informationen werden angezeigt und von den interaktiven Werkzeugen (z.B. Suche, Statistiken) berücksichtigt
- Artefakte, an denen für den Benutzer neue Aktivitäten ausgeführt wurden, werden in allen Sichten hervorgehoben (durch Fahnen-Symbole)

SWAPit sollte technisch mit dem BSCW-System gekoppelt werden. Ein optionales Symbol in der Orientierungsleiste zeigt an, dass für einen BSCW-Ordner ein vorbereitetes SWAPit-Projekt existiert (Abbildung 7.13). Ein Klick auf dieses Symbol startet das Projekt in einem neuen Browser-Fenster.

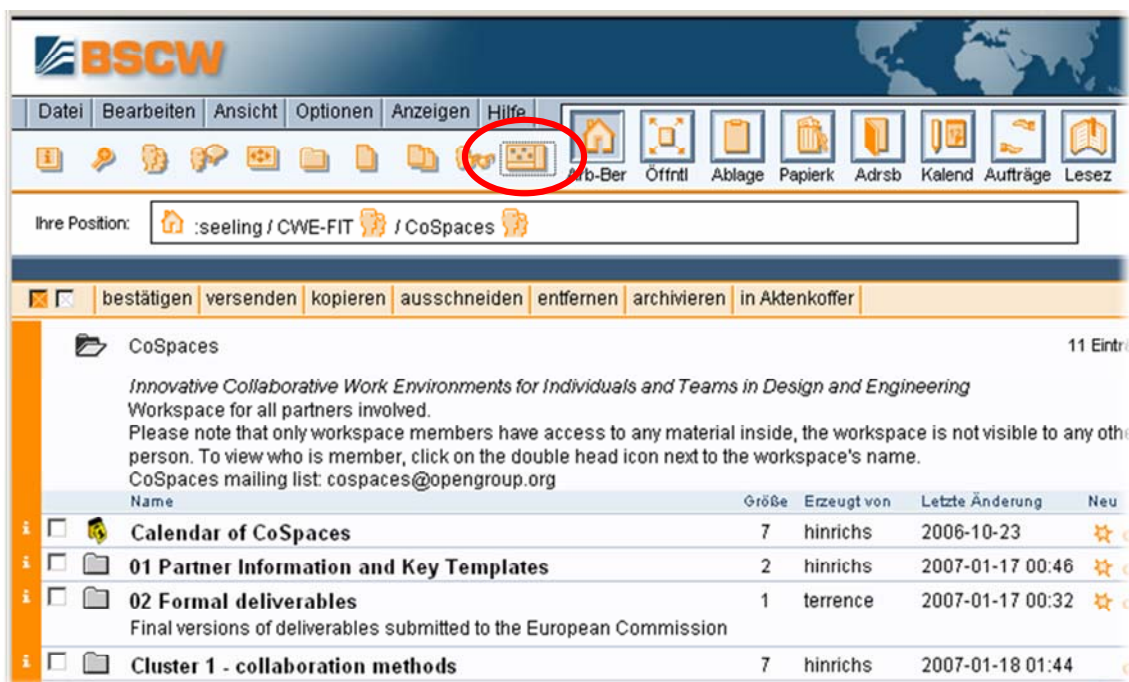


Abbildung 7.13: Einstiegspunkte für SWAPit in der Benutzerschnittstelle des BSCW-Systems

Die Vorberechnung kann durch folgende Maßnahmen noch performanter und Ressourcenschonender gestaltet werden (allerdings sind Realisierungsaufwand und Lastenverteilung der Berechnungen beider Systeme dabei kritische Faktoren):

- Vermeidung von Kommunikationsoverhead durch Installation des DocMINER-Webservice und von BSCW auf demselben Server. Informationen könnten dann von beiden Anwendungen durch gemeinsamen Zugriff auf Dateisystem und Datenbank genutzt werden.
- Indexierung und Themenkarten könnten inkrementell an Veränderungen der Informationen in der kooperativen Arbeitsumgebung angepasst werden.

In der aktuellen Realisierung, in der das BSCW-System und DocMINER auf separaten Servern ausgeführt werden, umfasst das Preprocessing von SWAPit-Projekten die folgenden Schritte:

1. Abrufen und Speichern von *Artefakten* (Dokumenten) u. Metadaten beim BSCW-System (BSCW-Suche wird zum Filtern der Informationen genutzt)
2. Abrufen und Speichern von *Aktivitäten und Ablagestrukturen* zu den Artefakten beim BSCW-System
3. Anlegen und Berechnen eines SWAPit-Projekts mit dem DocMINER-WebService
 - Puffern und Indexieren der Artefakte
 - Berechnen der paarweisen Ähnlichkeiten der Artefakte sowie der Themenkarte

7.5.4 Versuchsdesign und Durchführung

Rahmenbedingungen und Versuchsdesign

Die Benutzerstudie fand am 09.02.2007 von 14 bis 17 Uhr in St. Augustin statt. An der Studie nahmen 5 BSCW-Experten des Fraunhofer-Instituts FIT (Forschungsbereich Koordinationssysteme) teil. Die Studie war wie folgt aufgebaut:

Zunächst wurden die Teilnehmer geschult (ca. 1,5 Std.). Ziel der Schulung war, die Versuchspersonen gut auf die Bearbeitung der Aufgaben und die Benutzung des Werkzeugs vorzubereiten. Die Schulung umfasste einen Folienvortrag, eine Demonstration des Werkzeugs, und eine praktische Übung. Zur Erfolgskontrolle des theoretischen Teils der Schulung wurden den Teilnehmern nach der Demonstration verschiedene Systemzustände präsentiert, die sie interpretieren sollten.

Danach wurden von jedem Teilnehmer 3 Aufgabenblöcke (je ca. 25 Min.) bearbeitet. Für jede Aufgabe gab es ein Aufgabenblatt mit Aufgabenstellung, möglicher Lösungsstrategie mit dem SWAPit-System und Platz für Notizen (Anhang 9.8). Nach Bearbeitung der Aufgabe wurde ein Fragebogen ausgeteilt, in dem die Teilnehmer die Qualität der Aufgabenbearbeitung mit SWAPit (als Add-On zum BSCW-System) im Vergleich zur Qualität der Bearbeitung mit BSCW (ohne SWAPit) beurteilen sollten. Für die Fragebögen wurden 5er-Skalen verwendet (schlecht/niedrig = -2, -1, 0, +1, +2 = gut/hoch).

In einem abschließenden Fragebogen wurden einige Fragen aus dem Aufgabenteil wiederholt und einige aufgabenunabhängige Fragen (z.B. zu den interaktiven Relevanzmaßen) gestellt. Außerdem wurde ein Profil-Fragebogen zur Expertise der Teilnehmer ausgefüllt.



Abbildung 7.14: Die Teilnehmer der Benutzerstudie bei der Arbeit.

Die Studienteilnehmer verfügen alle über mehrere Jahre Berufserfahrung am Institut. In der Selbsteinschätzung (Fragebogen „User Profile“, Anhang 9.8, Abbildung 7.15) schätzten die Teilnehmer ihre Fähigkeiten in den Bereichen Kooperatives Arbeiten, Handhabung von BSCW und Bedienung von SWAPit besonders stark ein. Auch die eigene Innovationsfreude wurde hoch eingestuft. Die Kenntnis der Daten und die Fähigkeiten der Informationsrecherche wurden durchschnittlich etwas schwächer bewertet.

Bewertungsschema, Maße und Datenerhebung

Auch bei dieser Studie wird die allgemeine Bewertungs-Scorecard aus Kap. 2.1 als grundlegendes und umfassendes Bewertungsschema verwendet. Neben der IT-Perspektive spielt die Unternehmens-Perspektive eine wichtige Rolle. Außerdem kommt bei der Bewertung explorativer Recherche-Systeme dem Benutzer als „intelligentem Treiber“ der Recherche eine wichtige Bedeutung zu. Die Dimensionen der Scorecard verstehen sich als komplementäre Aspekte des Gesamtnutzens und wurden jeweils in Maße herunter gebrochen. Die Benutzereinschätzung bzgl. dieser Maße wurde mittels Fragebögen für jede Aufgabe erhoben (Anhang 9.8), teilweise abschließend noch einmal. Alle wichtigen Benutzer-Interaktionen (Auswahlen und Suche) wurden mit Zeitstempel in einem System-Log aufgezeichnet, um die tatsächliche Verwendung des SWAPit-Systems zu untersuchen.

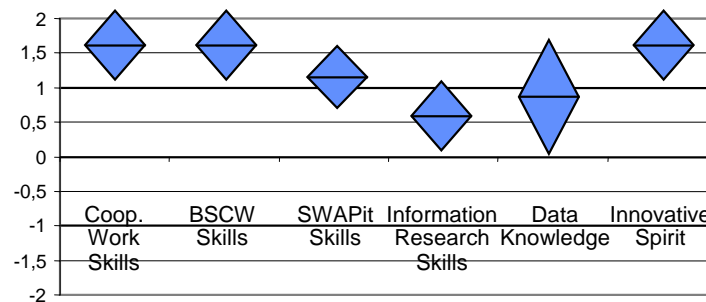


Abbildung 7.15: Auswertung des Fragebogens "Benutzerprofil", der am Ende der Studie ausgefüllt wurde (Mittelwerte und Standardabweichungen, N=5).

Während die Teilnehmer als BSCW-Experten die Qualität der Systemunterstützung durch BSCW einschätzen sollten, ohne das System tatsächlich zu verwenden, wurde SWAPit aktiv zur Aufgabebearbeitung eingesetzt. Beide Systeme wurden bezüglich der Maße vergleichend bewertet.

7.5.5 Ergebnisse

Die dritte Fallstudie motivierte die Entwicklung einer *Übersichtsdarstellung* von Tabellen für den Zugriff auf umfangreiche Datensätze zu Benutzeraktivitäten in kooperativen Arbeitsumgebungen. Die prototypische Realisierung, dargestellt in Abbildung 7.16, basiert auf Facetten (Kap.4.4.3) und zeigt für jedes Attribut der Datentabelle die vorkommenden Werte als Aufzählung an. Durch Auswahl kann der Benutzer interaktive Filter auf den Daten spezifizieren. Die Auswahl in einer Facette bewirkt die Beschränkung der anderen Facetten auf damit verknüpft vorkommende Werte. Die Anzeige von Werthäufigkeiten durch Balken informiert den Benutzer über Muster in den Daten und liefert Hinweise für interessante interaktive Anfragen. Unten werden die gefilterten Datensätze als Tabelle angezeigt.

In Abbildung 7.19 (a) sind die Mittelwerte der Antworten aller Teilnehmer nach den Aufgaben unterschieden. Da die Kurven beinahe gleich verlaufen wird, auch auf Grund der geringen Größe der Stichprobe, auf eine Differenzierung der Aufgabenarten verzichtet, denn die Antworten sind in allen Maßen unabhängig vom Aufgabentyp. Abbildung 7.19 (b) stellt Mittelwerte und Standardabweichungen zu SWAPit und BSCW *aufgabenübergreifend* dar. Zunächst einige Beobachtungen:

- Die Mittelwerte aller Maße zu SWAPit liegen im positiven Bereich (>0). Die meisten Mittelwerte liegen bei ungefähr 1. Ausnahmen sind: „Klarheit des Vorgehens“, „Empfundene Anstrengung“, „Erreichte Vollständigkeit“. Hier liegen die Werte bei ungefähr 0,5.
- Die Mittelwerte der meisten Maße zu SWAPit (7 von 11) liegen sogar deutlich im positiven Bereich (d.h. $m-\sigma > 0$)

- Besonders einig bzgl. SWAPit waren sich die Teilnehmer bei den Maßen „Nützlichkeit Sichtenkonzept“, „Erreichte Übersicht“ und „Funktionsumfang“, besonders uneinig bei „Präferenz“, „Empfundene Anstrengung“, „Angemessenheit Informationszugriff“.
- Der Abstand zwischen der Bewertung des SWAPit-Systems und des BSCW-Systems ist insgesamt groß: Die meisten Mittelwerte der BSCW-Bewertung liegen im negativen Bereich (<0).

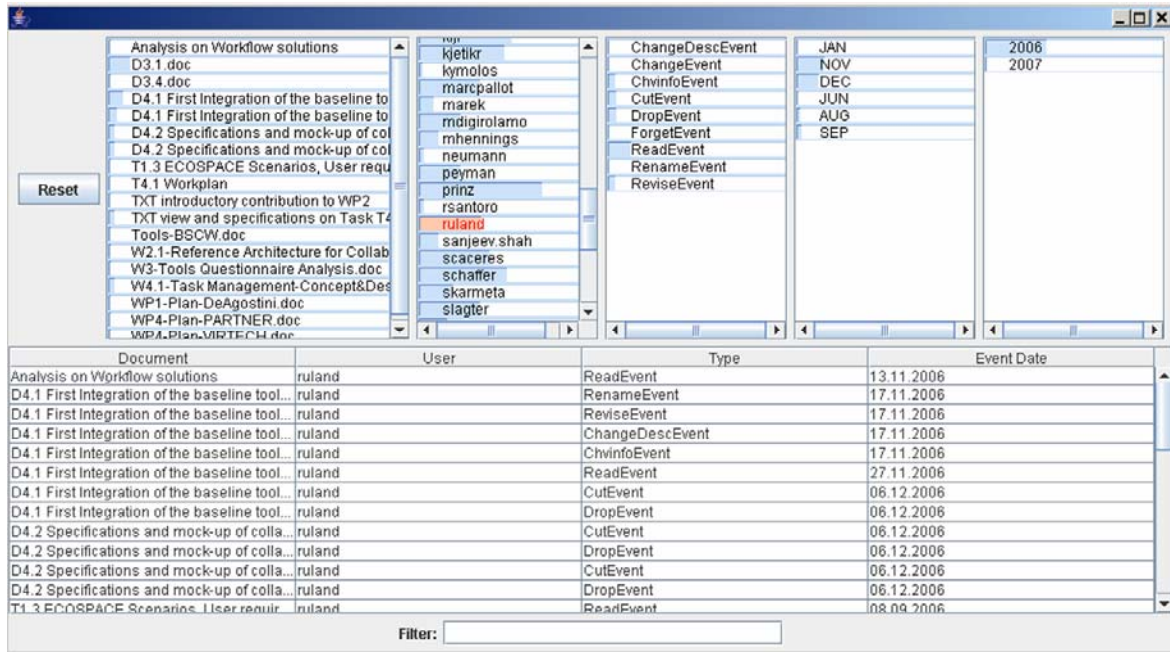


Abbildung 7.16: Eine neue Übersichtsdarstellung für Tabellen („FacetTable“) basiert auf Facetten und zeigt Werthäufigkeiten durch Balken an. Im Beispiel sind einige tausend Benutzeraktivitäten in einer kooperativen Arbeitsumgebung dargestellt.

Die Uneinigkeit und etwas schlechtere Beurteilung in den Maßen zur Benutzbarkeit von SWAPit („Klarheit des Vorgehens“, „Empfundene Anstrengung“) wird unterstrichen durch den Vergleich mit dem BSCW-System: Die beiden genannten Maße (und „Erreichte Übersicht“) sind die einzigen, für die das BSCW-System im Mittel eine Bewertung >0 erhält. Anzunehmende Gründe für die leichten Defizite von SWAPit in den Maßen zur Benutzbarkeit sind die inhärent hohe Systemkomplexität (durch die parallele Bereitstellung aller entscheidungsrelevanten Informationen) und die Vermutung, dass die Lernkurve der Teilnehmer bei einmaliger Benutzung des Systems noch wenig ausgeprägt ist (insbes. gegenüber der besonders guten Expertise mit dem BSCW-System, s. Abbildung 7.15). Die Ergebnisse in den Maßen zur Benutzbarkeit sind zwar durchaus noch akzeptabel, jedoch auffällig und liefern daher wertvolle Hinweis für die zukünftige Entwicklung von Konzept und Referenztechnologie (ggf. Systemkomplexität verringern und Usability weiter optimieren). In Bezug auf den Bewertungsprozess bestätigt die Differenziertheit der Antworten, wie bei der zweiten Fallstudie, die Nutzenbewertung nach komplementären Perspektiven und Maßen!

Die Aufzeichnung wichtiger Benutzerinteraktionen erlaubt einen Einblick in die tatsächliche Verwendung der Interaktionsmöglichkeiten von SWAPit (Abbildung 7.17). Fast die Hälfte der Benutzeraktivitäten waren Auswahlen in der Themenkarte (45%). Kategorien (Personen, WP-Ordner, 28%) wurden etwas häufiger zur Navigation genutzt als Tabellen (22%; Metadatentabelle: Dokumententitel; Faktentabellen: Personen und Aktivitäten). Besonders beim Vergleich explorativer (95%) und anfrageorientierter Aktivitäten (5%, zwischen 8 und 27 Suchanfragen pro Benutzer) wird deutlich, dass

- Suche und Exploration komplementär verwendet wurden
- Benutzer einen klaren Schwerpunkt auf explorative Recherchestrategien gelegt haben

Die Aufschlüsselung nach einzelnen Benutzern zeigt, dass die Aktivitätsarten von allen Benutzern ungefähr gleich häufig eingesetzt wurden.

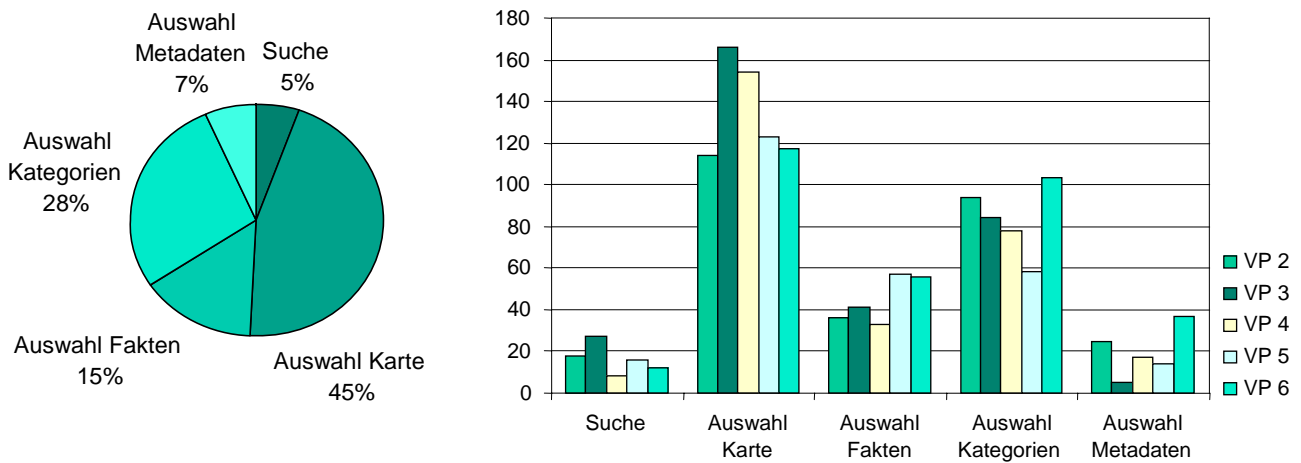


Abbildung 7.17: Verwendung des SWAPit-Systems – Aktivitäten der Benutzer aufgeschlüsselt (Anzahl der Aktivitäten pro Kategorie und Nutzer; N=5).

Abschließend wurden den Teilnehmern einige aufgabenunabhängige Fragen gestellt. Sowohl die Übersicht über vorhandene Information und als auch die Fokussierung interessanter Teilmengen von Information wurde im Mittel mit 1 bewertet (Abbildung 7.18, rechts). Bei der Fokussierung waren sich alle Teilnehmer einig – eine Bestätigung der Eignung Farbkonzepts, das gerade der Fokussierung von Information durch verschiedene interaktive Auswahl-Filter unterstützt. Die Frage nach den Relevanzmaßen für Markierungen (Interaktive Rankings, Abbildung 7.18, links) war zweigeteilt: Die Nützlichkeit wurde sehr hoch eingestuft, während beim Verständnis der Ranking-Information eine geringere Einstufung und eine geringere Einigkeit vorliegen: Benutzer finden es sehr hilfreich, durch die Ranking-Information geleitet zu werden, die Transparenz der zugrunde liegenden Maße sollte jedoch verbessert werden.

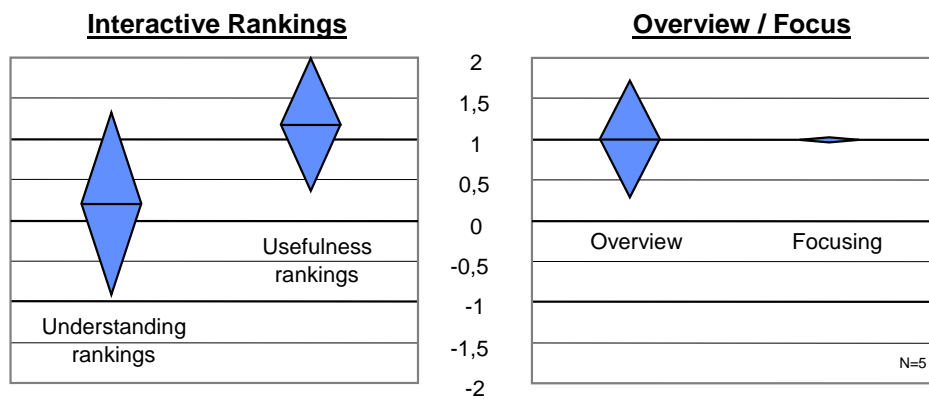
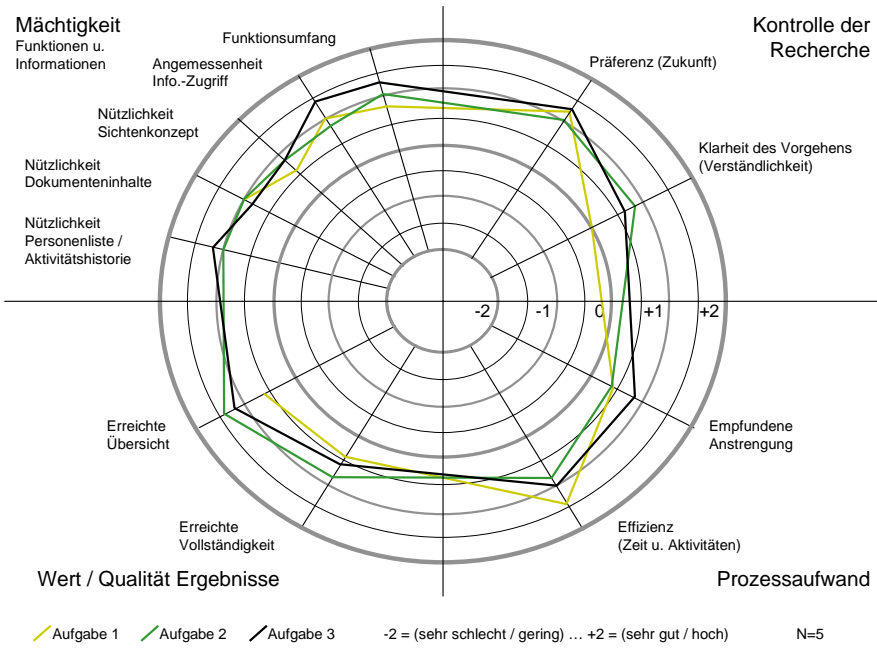
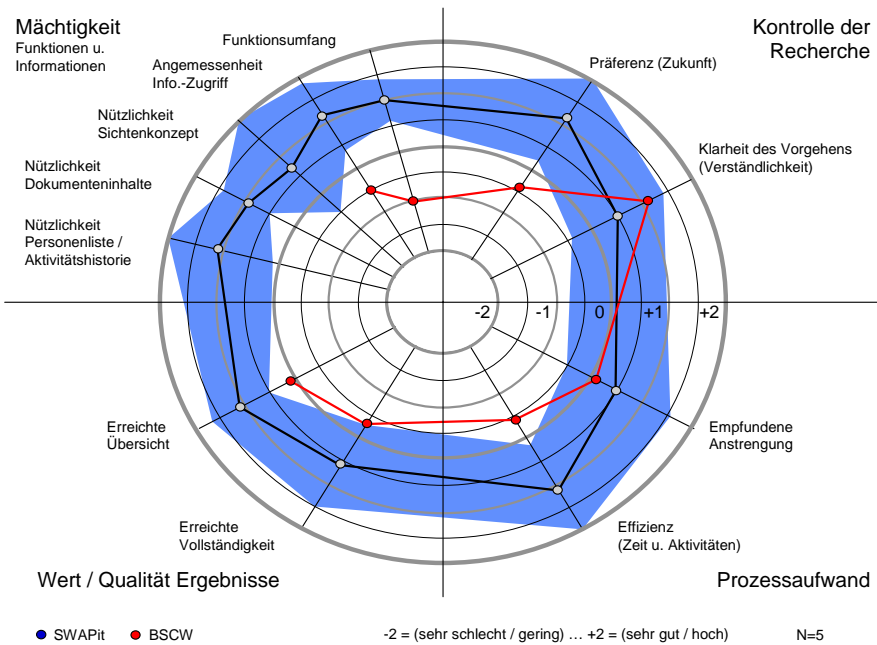


Abbildung 7.18: Weitere Maße der abschließenden Einschätzung (Mittelwerte und Standardabweichungen, N=5).

Beobachtungen bei der Fallstudie motivieren einige Weiterentwicklungen von SWAPit: Zeitbezogene Daten sollten intuitiver navigiert und dargestellt werden können, wozu in Zukunft interaktive Zeitstrahle („Brushing Histograms“, [LN03]) eingesetzt werden könnten. Große Tabellen, typisch bei der Darstellung kooperativer Aktivitäten, sollten interaktiv nach allen Feldern gruppiert werden können. Dazu wird nun ein Konzept entwickelt, das Gruppierungsoperationen auf Tabellenspalten erlaubt, sogenannte „FacetTables“. Das SWAPit-Derivat soll im Forschungsprojekt zu einem Produkt-unabhängigen Betrachtungswerkzeug für kooperative Arbeitsumgebungen weiter entwickelt werden.



(a) SWAPit (Aufgaben einzeln; Mittelwerte)



(b) SWAPit vs. BSCW (aufgabenübergreifend; Mittelwerte u. Standardabweichungen)

(c) Ergebnisse detailliert (aufgabenbegleitende Fragebögen)

| | Aufgabe 1: | | | | | | | | | | | Aufgabe 2: | | | | | | | | | | | Aufgabe 3: | | | | | | | | | | | | | | | | | | | |
|-------------------------------|------------|----|----|----|----|-----|------|----|----|----|----|------------|------|-----|---|----|----|------|---|-----|-----|----|------------|----|----|---|------|-----|------|---|----|---|---|-----|-----|----|---|----|----|----|------|-----|
| | swapit | | | | | | bscw | | | | | swapit | | | | | | bscw | | | | | swapit | | | | | | bscw | | | | | | | | | | | | | |
| | 2 | 3 | 4 | 5 | 6 | m | σ | 2 | 3 | 4 | 5 | 6 | m | σ | 2 | 3 | 4 | 5 | 6 | m | σ | 2 | 3 | 4 | 5 | 6 | m | σ | 2 | 3 | 4 | 5 | 6 | m | σ | | | | | | | |
| Erreichte Übersicht | 1 | 0 | 0 | 1 | 1 | 0,6 | 0,5 | 0 | 0 | 1 | 0 | 1 | 0,4 | 0,5 | 2 | 1 | 1 | 1 | 2 | 1,4 | 0,5 | -1 | 0 | -1 | -1 | 0 | -0,6 | 0,5 | 2 | 1 | 1 | 1 | 1 | 1,2 | 0,4 | 0 | 0 | 2 | 0 | 0 | 0,4 | 0,9 |
| Erreichte Vollst. | 1 | 1 | 0 | 0 | 0 | 0,4 | 0,5 | -1 | 2 | -1 | -1 | 1 | 0,0 | 1,4 | 2 | 1 | 1 | 0 | 0 | 0,8 | 0,8 | -2 | 0 | 1 | -1 | 0 | -0,4 | 1,1 | 2 | 0 | -1 | 0 | 2 | 0,6 | 1,3 | -1 | 0 | 0 | -1 | 0 | -0,4 | 0,5 |
| Effizienz | 2 | 1 | 1 | 1 | 2 | 1,4 | 0,5 | -1 | -1 | -1 | -1 | 0 | -0,4 | 0,9 | 2 | 0 | 1 | 0 | 1 | 0,8 | 0,8 | -1 | 0 | 2 | -1 | 0 | 0,0 | 1,2 | 2 | 1 | -1 | 1 | 2 | 1,0 | 1,2 | -2 | 0 | -1 | -1 | 0 | -0,8 | 0,8 |
| Anstrengung | 1 | 1 | 1 | -1 | 0 | 0,4 | 0,9 | 0 | 1 | -1 | 1 | 1 | 0,4 | 0,9 | 2 | -1 | 1 | -1 | 1 | 0,4 | 1,3 | -1 | 0 | 2 | 1 | 0 | 0,4 | 1,1 | 2 | 1 | -1 | 1 | 1 | 0,8 | 1,1 | -2 | 1 | -1 | 0 | 0 | -0,4 | 1,1 |
| Klarheit Vorgehen | 1 | 0 | 0 | 0 | -1 | 0,0 | 0,7 | 1 | 2 | 1 | 1 | 1 | 1,2 | 0,4 | 2 | 1 | 1 | 0 | 0 | 0,8 | 0,8 | 2 | 2 | 2 | 1 | 0 | 1,4 | 0,9 | 2 | 0 | -1 | 1 | 1 | 0,6 | 1,1 | 2 | 2 | 1 | -1 | 0 | 0,8 | 1,3 |
| Präferenz | 2 | 1 | 1 | 1 | 1 | 1,2 | 0,4 | 1 | 0 | -2 | 0 | 0 | -0,2 | 1,1 | 2 | -1 | 2 | 0 | 2 | 1,0 | 1,4 | -1 | 0 | 0 | 0 | 0 | -0,2 | 0,4 | 2 | 1 | 0 | 2 | 1 | 1,2 | 0,8 | -2 | 0 | -1 | -1 | 0 | -0,8 | 0,8 |
| Informationszugriff | 1 | 0 | 1 | 1 | 2 | 1,0 | 0,7 | 0 | 1 | -1 | -1 | 0 | -0,2 | 0,8 | 2 | 0 | 1 | 0 | 1 | 0,8 | 0,8 | -2 | 0 | -1 | 0 | 0 | -0,6 | 0,9 | 2 | 1 | 0 | 2 | 2 | 1,4 | 0,9 | -2 | 0 | -1 | 0 | 0 | -0,6 | 0,9 |
| Funktionsumfang | 1 | 0 | 1 | 1 | 1 | 0,8 | 0,4 | -1 | 0 | -1 | -1 | 0 | -0,6 | 0,5 | 1 | 1 | 1 | 1 | 1 | 1,0 | 0,0 | -2 | -1 | -1 | -1 | 0 | -1,0 | 0,7 | 2 | 1 | 1 | 1 | 1 | 1,2 | 0,4 | -2 | 0 | -1 | -2 | -2 | -1,4 | 0,9 |
| Nützl. Personen / Aktivitäten | | | | | | | | | | | | | | | 2 | 1 | 0 | 1 | 1 | 1,0 | 0,7 | | | | | | | | 2 | 1 | -1 | 2 | 2 | 1,2 | 1,3 | | | | | | | |
| Nützl. Inhalte | 1 | 1 | 1 | 1 | 1 | 1,0 | 0,0 | | | | | | | | 1 | 0 | 1 | 1 | 2 | 1,0 | 0,7 | | | | | | | | 1 | 1 | 0 | 1 | 1 | 0,8 | 0,4 | | | | | | | |
| Nützl. Sichtenkonzept | 2 | -1 | -1 | 2 | 1 | 0,6 | 1,5 | | | | | | | | 2 | 0 | -1 | 2 | 1 | 0,8 | 1,3 | | | | | | | | 2 | 1 | -1 | 1 | 1 | 0,8 | 1,1 | | | | | | | |

Abbildung 7.19: Ergebnisse der Laborstudie zu Kooperationsumgebungen (Mittelwerte u. Standardabweichungen).

7.6 Systemverwendung

Für die 2. und 3. Fallstudie wurden die Benutzeraktivitäten untersucht (Abbildung 7.8 u. Abbildung 7.17). Bei der Gegenüberstellung fällt auf, dass beim Einsatz für den Produktsupport (Studie 2) sehr häufig Suchanfragen gestellt wurden. Auch die Äußerungen der Anwender machten deutlich, dass ein anfrageorientiertes Vorgehen für sie eine hohe Priorität hat. Mutmaßliche Ursache ist die eingeschränkte Fähigkeit, Suchaufgaben explorativ zu lösen, selbst wenn ein solches Vorgehen erfolgsversprechender ist. Ein weiteres Indiz, dass die Benutzer bei der 2. Studie ihre gewohnten Arbeitsprozesse beibehalten haben, ist die Tatsache dass viel mit den Metadaten gearbeitet wurde, jedoch wenig mit der Kategoriensicht, die bestimmte Metadaten adäquater bereit stellt. Bei der 3. Fallstudie (Kooperatives Arbeiten) wurde daher von der Schulung bis zur Aufgabenstellung noch stärker darauf geachtet, explorative Recherchestrategien für verschiedene Fragestellungen aufzuzeigen und insbesondere die geeignete Nutzung der Sichten zu erläutern. So machten Suchanfragen bei der dritten Fallstudie nur unter 5% der Aktivitäten aus. Die zentrale Ähnlichkeitssicht mit Themenkarte wurde in beiden Studien intensiv genutzt. Allerdings verwendeten die Anwender die zur Beschreibung von Dokumentenmengen wichtige Termstatistik-Funktion nur bei der letzten Fallstudie.

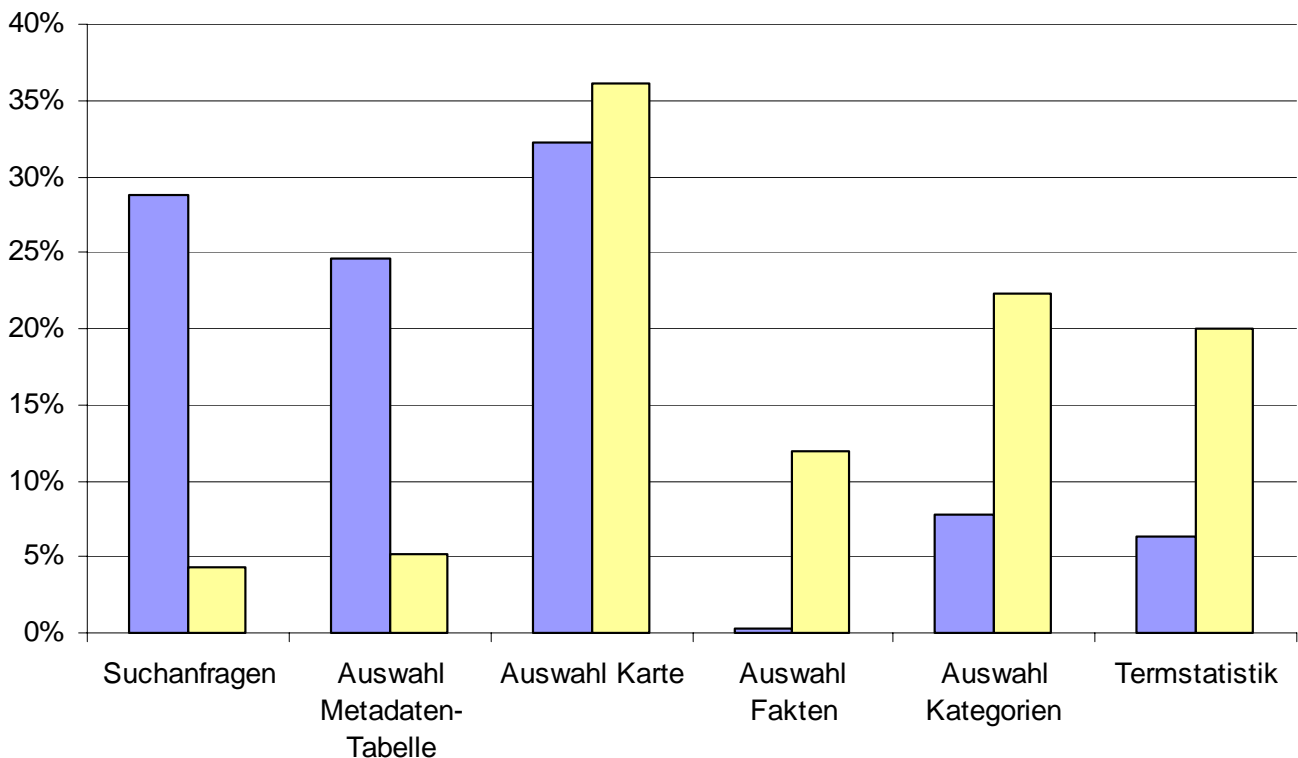


Abbildung 7.20: Vergleich der Verwendung des SWAPit-Systems in der 2. bzw. 3. Fallstudie

8 Fazit

8.1 Zusammenfassung und Beiträge

In der Arbeit wurde untersucht, wie gut ein auf die Textanalyse spezialisiertes aber von Anwendungsdomänen unabhängiges MultiView-Konzept für Aufgaben der explorativen Text- und Datenanalyse geeignet ist. Dazu wurde der methodisch-technische Rahmen für die empirische Erforschung des Nutzens in unterschiedlichen wissensintensiven Anwendungsfeldern geschaffen. Für Anforderungsanalyse, Systemanpassung und Nutzenbewertung wird ein Vorgehensmodell vorgeschlagen, das mit Modellen für Information, Aufgaben und Prozesse hinterlegt ist (Kap.2).

Das in der Arbeit entwickelte *Systemkonzept* zielt darauf ab, die entscheidungsrelevanten Informationen in *einer* übersichtlichen Gesamtdarstellung zugreifbar zu machen, die sich problemspezifisch konfigurieren lässt: Strukturierte Daten (genauer: Metadaten, relationale Daten und Beschreibungskategorien) lassen sich in verknüpften Sichten anzeigen und zur Navigation in der Textsammlung nutzen (Kap.5). Dabei liegt der Schwerpunkt auf Darstellungs- und Interaktionskonzepten, während Algorithmen und Werkzeuge zur Analyse und Verknüpfung von Information Randthemen sind. Das SWAP-Systemkonzept beschreibt Interaktionswerkzeuge, die sich für MultiView-Systeme zur explorativen Datenanalyse im Allgemeinen einsetzen lassen (Kap.5.4-5.5). Das *SWAPit-System* (Kap.6) ist eine flexible Referenztechnologie, auf Basis derer sich durch Konfiguration (und ggf. Erweiterung) spezialisierte Systeme für unterschiedliche Rechercheaufgaben realisieren lassen. Dazu verfügt SWAPit über eine offene Architektur, in der sich neue Sichten zur Darstellung weiterer Arten von Informationsobjekten leicht einbinden lassen, welche auch weitere Darstellungsmetaphern verwenden können.

Für die betrachteten Anwendungsfelder wurde untersucht, welche Informationsarten und Verknüpfungen zwischen den Informationsarten für die explorative Recherche in Textsammlungen eine Rolle spielen und durch welche Eigenschaften die Informationsarten und Verknüpfungen gekennzeichnet sind (Kap.4). Zudem wurde gezeigt, wie SWAPit in bestehende Geschäftsprozesse und Arbeitsumgebungen integriert werden kann. Dazu wurden Werkzeuge zur Inhaltsextraktion, Informationsverknüpfung und Administration (u.a. Sichtenspezifikation), entwickelt (Kap.6.3 und 6.7).

Schließlich wurde in drei empirischen Studien bewertet, welcher Nutzen unter technischen sowie betriebswirtschaftlichen Gesichtspunkten durch die neuartige Systemunterstützung erzielt werden kann (Kap.7). Das Ziel der Nutzenbewertung in realen Arbeitsprozessen unter differenzierten Gesichtspunkten impliziert die empirische Bewertung in Feldstudien, wo die zahlreichen Einflussfaktoren weder vollständig kontrolliert noch isoliert, wohl aber beobachtet, systematisch erfasst und diskutiert werden. Ergebnis der Studien ist daher nicht eine statistisch gesicherte Überprüfung der Nützlichkeit des Systemkonzepts. Vielmehr ermöglichen die in Kap.7 und Kap.8.2 verdichteten Beobachtungen ein differenziertes und zugleich einheitliches Bild der Potentiale und kritischen Aspekte des Einsatzes von MultiView-Konzepten für die Textanalyse. Insgesamt belegen die Fallstudien die Vielseitigkeit und Nützlichkeit von MultiView-Konzept und Werkzeug.

8.2 Ergebnisse und Schlussfolgerungen

8.2.1 SWAP-Systemkonzept und SWAPit-Referenztechnologie

Die empirischen Studien und damit verbundene Modifikationen an Konzept und Referenztechnologie bauen aufeinander auf (Kap.1.5). Während die ersten Studien mit SWAPit-Derivaten (Kap.7.3 u. Anhang 9.3) noch

wichtige Entwicklungen auslösten, erforderten die zweite und dritte Fallstudie der Arbeit keine Modifikationen an der Referenztechnologie, sondern lediglich die obligatorische Konfiguration für die konkreten Anwendungen. In diesem Sinne ist der Entwicklungsprozess gegen den aktuellen Zustand konvergiert und die letzten Fallstudien dienten im Wesentlichen der Bewertung von Konzept und System. Betrachtet man die Evolution von Konzept und System genauer, so lassen sich vier maßgebliche Meilensteine (Bewertungs-/Entwicklungszyklen gemäß Kap.1.5) ausmachen, die in Kap.7.2 ausführlicher beschrieben sind.

Die Ergebnisse der Fallstudien belegen insgesamt die gute Eignung des Systems für Rechercheaufgaben. Der erfolgreiche Einsatz für so unterschiedliche Aufgaben wie Meldungsrecherche im Produktsupport und Aktivitätsanalyse in kooperativen Arbeitsumgebungen belegt die Vielseitigkeit des Konzepts und die gute Anpassbarkeit der Referenztechnologie SWAPit (s. Anhang 9.3 für zahlreiche weitere Anwendungsbeispiele). In der letzten Fallstudie wurde SWAPit sogar technisch vollständig in die bestehende Systemlandschaft integriert (Informations- und Kontrollflüsse). Ein Vergleich der Aufgabenbeschreibungen im Aufgabenmodell (Kap.3) zeigt, dass die auf den ersten Blick sehr unterschiedlichen Rechercheaufgaben vieles gemeinsam haben – teilweise bedingt dadurch, dass das Aufgabenmodell für die explorative Textrecherche im allgemeinen konzipiert wurde, für den potentiell gewinnbringenden Einsatz des SWAPit-Systems aber Szenarien mit bestimmten Voraussetzungen ausgewählt wurden. Zu diesen Auswahlkriterien gehören beispielsweise die Existenz mehrerer unterschiedlicher Informationsarten und explorativ ausgerichtete Recherchestrategien. Die größten Unterscheidungen weisen die Aufgaben bzgl. der Kriterien auf, die den betriebswirtschaftlichen Arbeitskontext beschreiben (unterer Teil der Nutzen-Scorecard, Kap.2.1).

Die Methode zur differenzierten Evaluation des Systemnutzens hat sich bewährt – insbesondere bei der zweiten Studie, wo die anfrageorientierte Suchfunktion beim praktischen Einsatz nicht alle Anforderungen erfüllte und daher einige Kriterien aus den Bereichen Funktionalität und Ergebnisqualität negativ bewertet wurden. Dennoch war das Nutzenpotential des Systems deutlich erkennbar (Abbildung 7.9). Die Abstraktion von anwendungsspezifischen Maßen durch die Unterscheidung von Perspektiven der Scorecard war förderlich für die Diskussion und den Vergleich der Studienergebnisse.

Bei der Frage, ob sie die Technologie auch in Zukunft gerne einsetzen würden, äußerten die Benutzer in allen Studien eine sehr starke Präferenz für SWAPit. Unterstrichen wird diese Einschätzung durch die noch andauernde Fortsetzung der Projektaktivitäten in den Bereichen kooperatives Arbeiten und Produktsupport.

Die Systembenutzung wurde als schwierig eingestuft⁸⁴. Dafür kommen zahlreiche Gründe in Betracht (Kap.7.5.5), beispielsweise die inhärent hohe Systemkomplexität durch die parallele Bereitstellung aller entscheidungsrelevanten Informationen „auf einen Blick“ und die freie Interaktion, die als potenzielles Risiko von MV-Systemen bereits in Kap.1.2 diskutiert wurde. Allerdings wurde die Navigation im System in beiden Studien ausgesprochen gut bewertet⁸⁵, woraus sich möglicherweise ableiten lässt, dass die Schwierigkeit darin besteht, aus den bestehenden Navigationsmöglichkeiten komplexe Arbeitsabläufe mit konkreter Zielerreichung zu kombinieren. Vereinfachungen am System und Optimierung der Benutzerschulung können das Phänomen kompensieren helfen. Die Analyse der Verwendung von SWAPit (Kap.7.6) legt die Vermutung nahe, dass die Ausnutzung der *explorativen* Recherchemöglichkeiten des Systems, insbesondere des Sichtenkonzepts, eine schwierige Aufgabe darstellt. Anwender tendieren offenbar dazu, die gewohnten Strategien der anfrageorientierten Informationsrecherche auch in Rechercheumgebungen anzuwenden, die neuartige explorative Werkzeuge für den Informationszugriff anbieten. Bei verbesserter Anleitung sind die Anwender jedoch zu einer stärkeren Nutzung der interaktiven Sichten und Werkzeuge bereit (s. dritte Fallstudie).

⁸⁴ Bei der 2. Fallstudie wurde SWAPit im Kriterium „Intuitivität/Einfachheit“ etwas schlechter eingestuft als die Suchmaschine. Bei der 3. Studie wurde SWAPit im Kriterium „Klarheit des Vorgehens (Verständlichkeit)“ gegenüber BSCW etwas schlechter bewertet.

⁸⁵ Kriterium „Navigation“ bei Studie 2; Kriterium „Informationszugriff“ bei Studie 3.

Die Relevanzmaße für Markierungen von Informationsobjekten (Kap.5.4), die durch Hervorhebung besonders relevanter Information helfen, den Benutzer bei der Recherche zu leiten, wurden als sehr nützlich eingestuft (Abbildung 7.18). Allerdings besteht bei der Nachvollziehbarkeit der Maße Verbesserungspotential.

Der einfache und nachvollziehbare Wechsel zwischen Übersicht und Detail ist ein grundsätzlich wichtiges Qualitätskriterium explorativer Recherchesysteme [CC05]. Die Tatsache, dass sowohl Übersicht als auch Fokussierung von Informationsobjekten gut bewertet werden (Abbildung 7.18), belegt die Eignung der Farbcodierung von Auswahlen und Markierungen in SWAPit (Kap.6.4).

Insgesamt belegen die zum SWAP-Konzept durchgeführten Studien die Eignung des Ansatzes, bei der Textrecherche die Texte und alle relevanten strukturierten Daten übersichtlich und verknüpft darzustellen und zur komplementären Navigation nutzbar zu machen.

8.2.2 Ergebnisse für die Explorative Textrecherche

Den Kern der Arbeit bilden das neuartige Systemkonzept zur Textrecherche und seine Bewertung für den Einsatz in der Praxis. Allerdings wird in Kap.2.4.1 gefordert, dass anpassbare technische Lösungen (Referenztechnologien) begleitet werden müssen von Vorgehensmodellen und Qualitätskriterien zur problemspezifischen Herleitung *aufgabenangemessener* Spezialsysteme – ein Aspekt, der in den meisten Arbeiten zu Referenztechnologien vernachlässigt wird (Kap.3.4). Stellen sich doch Fragen wie: Welche Informationen werden benötigt, wie hängen sie zusammen und wie können sie geeignet navigierbar und durchsuchbar gemacht werden? Die Herleitung *aufgabenangemessener* Recherchesysteme muss die in Kap.2.4.2 beschriebenen Phasen Prozess-/Aufgabenanalyse, Systemdesign und Bewertung durchlaufen. Für die technische Umsetzung kann die Verwendung einer bewährten Referenztechnologie dann natürlich einen enormen Geschwindigkeits- und Qualitätsvorteil bedeuten.

Die methodischen Beiträge zur Herleitung und Bewertung aufgabenangemessener Recherchelösungen sind unabhängig vom Systemkonzept der Arbeit und lassen sich daher für explorative Rechercheaufgaben in Dokumentensammlungen allgemein verwenden:

- **Das Modell zur Charakterisierung von Aufgaben der explorativen Textrecherche** (Kap.2.2 und Anhang 9.2) repräsentiert als kontinuierliche Erweiterung des früheren Basismodells [BSM02] durch zahlreiche Fallstudien (Kap.3 und Anhang 9.3) ein breites Erfahrungswissen.
- **Die Methoden zur Bewertung des Nutzens explorativer Recherchesysteme und zur Anpassung einer Referenztechnologie an konkrete Szenarien der Textrecherche** (Kap.2.1 und Kap.2.4) sind ein wichtiges Hilfsmittel für die Einführung innovativer Werkzeuge zur Unterstützung bestehender Geschäftsprozesse. Neben den IT-orientierten Kriterien können auch betriebswirtschaftliche Kriterien bei Aufgabenbeschreibung und Nutzenbewertung abgebildet werden.

Die Beobachtungen in den Fallstudien deuten darauf hin, dass Benutzer die ihnen vertrauten Strategien der anfrageorientierten Suche beibehalten, auch wenn sie Werkzeuge zur explorativen Recherche zur Verfügung gestellt bekommen, die eine aufgabenadäquatere Unterstützung der Aufgabe versprechen (Kap.7.6). Erst die *problemorientierte* Schulung der Anwendung und Kombination der einzelnen Interaktionsmöglichkeiten führt zur tatsächlichen Nutzung der explorativen Recherchewerkzeuge.

8.2.3 Interaktive Werkzeuge für MultiView-Systeme

Einige der funktionalen Erweiterungen, die im SWAPit-System realisiert, erprobt und optimiert wurden, sind generisch einsetzbar für MultiView-Systeme zur explorativen Informationsrecherche:

- **Die Unterscheidung von Auswahl und Markierung bei der Sichtenkoordination** (überlappendes Brush&Link, Kap.5.3) erlaubt zum einen die Darstellung einer Interaktionshistorie, da eine neue Auswahl in unterschiedlichen Sichten sich nicht gegenseitig „überschreiben“. Andererseits werden dadurch kombinierte Anfragen möglich, die zusammen mit den Relevanzmaßen „weiche“ interaktive konjunktive und disjunktive Anfragen zulassen (Kap.5.6)
- **Relevanzmaße zur Gewichtung interaktiver Auswahlen in MultiView-Systemen** (Berechnung Kap.5.4; Darstellung Kap.6.5). Die bekannten Systeme, die mit multiplen grafischen Sichten und Brush&Link-Metaphern arbeiten, verwenden ungewichtete (flache) Auswahlen, die dem Benutzer keine Hinweise darüber liefern, welche weiteren Informationen sich aufgrund seiner letzten Auswahlen anzusehen lohnen.
- **Die Annotation ausgewählter Informationsobjekte mit verknüpften Informationsobjekten** aus anderen Sichten dienen der Beschreibung von Informationsobjekten durch ihren Informationskontext (Kap.5.5, realisiert als MultiView-Tooltips bzw. –Labels, Kap.6.6). Durch die räumliche Nähe lassen sich Zusammenhänge besser erkennen und machen zusätzliche Interaktionsschritte überflüssig.

8.3 *Ausblick*

Die Forschung an MultiView-Systemen hat erst vor wenigen Jahren begonnen. Interaktive Übersichtsdarstellungen von Informationsräumen etablieren sich allmählich auch für die Suche im WWW. Beide Bereiche bieten noch viel Raum für Forschung und Entwicklung. Mit zunehmender Konsolidierung und Standardisierung explorativer Recherchesysteme für Texte und Daten kann gehofft werden, dass solche Systeme eine ähnliche Erfolgsgeschichte wie OLAP in den 90er Jahren schreiben können.

In der Arbeit wurde betont, dass das *aufgabenangemessene* Design von explorativen Recherchesystemen schwierig ist und daher die Wiederverwendbarkeit von qualitätsgeprüften Konzepten, Prozessen und Werkzeugen eine wichtige Rolle für die Effizienz bei der Entwicklung spielt und die Qualität der resultierenden Systeme gewährleisten hilft. Die Wiederverwendung ist limitiert durch die Schwierigkeit, Vorgehensweisen der explorativen (also insbesondere datengetriebenen) Recherche a priori feingranular als Prozess oder Workflow zu modellieren, weil ja gerade sehr *flexible* Zugriffspfade ermöglicht werden sollen. Aufgabenmodelle wie das der Arbeit bieten eine schwache Strukturierung der Informationszugriffsprozesse, ohne allerdings Abläufe vorzugeben. Solche Informationszugriffsprozesse müssen weiter erforscht werden mit dem Ziel, Teilaufgaben, Phasen oder Schritte zu identifizieren und mit Systemmodulen zu hinterlegen, die sich rekombinieren lassen („information research as a service“). Lassen sich Aufgaben und Strategien besser semantisch beschreiben, so kann die Lücke zwischen Vorgehensweise/Informationsbedarf einerseits und Systemfunktionalitäten/Werkzeugen andererseits weiter geschlossen werden. Das Werkzeug IN-SPIRE (Kap.3.3.3) beispielsweise geht einen Schritt in diese Richtung, indem es den vordefinierten Prozess des Überprüfens von Hypothesen an einer Textsammlung sowie interaktive Makros unterstützt.

Für die kontinuierliche Verbesserung und Erweiterung des SWAPit-Systems lassen sich nach dieser Arbeit die nächsten Entwicklungsschritte skizzieren. So hat die letzte Fallstudie eine spezialisierte Sicht für die zeitbasierte Navigation motiviert; Brushing Histograms [LN03] erscheinen dafür interessant. Außerdem wird eine kompakte und interaktive Übersichtsdarstellung für große Tabellen benötigt. Eine Eigenentwicklung, die auf dynamischer Gruppierung durch interaktive Filter basiert, wird derzeit an Beispieldaten evaluiert.

Das in der Arbeit vorgestellte Ranking-Konzept für interaktive und Markierungen sollte für MultiView-Systeme verallgemeinert werden. Die Nützlichkeit und Nachvollziehbarkeit verschiedener Maße ist in vergleichenden empirischen Laborstudien zu bewerten. Gleiches gilt für die in Kap.5.5 eingeführten MultiView-Werkzeuge zur Beschriftung von Informationsobjekten.

In den Fallstudien hat sich eine hohe Integrationstiefe in bestehende IT-Umgebungen als wichtig erwiesen. So hätte sich für die Arbeit im Produktsupport beim Systemhaus angeboten, die dort etablierte Suchmaschine in SWAPit einzubinden. Auch für den Beitrag zum Informatikjahr 2007 (Informatiklandkarte, Anhang 9.3) musste SWAPit interaktiv mit anderen Werkzeugen der Informationsrecherche gekoppelt werden. Konzept und modulare Systemarchitektur von SWAPit lassen die Integrierbarkeit anderer Technologien zu - allerdings verbunden mit einem größeren Entwicklungsaufwand. Daher sollten weitere externe Schnittstellen, z.B. für Suchmaschinen oder Zeit- bzw. Geo-basierte Navigation entwickelt werden.

Eine wichtige Erkenntnis aus den Fallstudien war die schwierige Benutzbarkeit von MultiView-Systemen für die Textrecherche. Daher stellt sich die Frage, wie die Systemkomplexität reduziert bzw. bezüglich der Benutzerexpertise skalierbar gestaltet werden kann. So könnte die Anzahl der Auswahlfarben verringert werden – allerdings unter Inkaufnahme eingeschränkter Anfragemöglichkeiten. Roberts et al. [RW06] schlagen ein Farbkonzept vor, das der Benutzer selbst spezifiziert. Allerdings bleibt herauszufinden, ob eine gleich bleibende systemdefinierte Farbsemantik oder eine dynamisch durch den Benutzer festgelegte Farbsemantik für Überblick und Navigation vorteilhaft ist.

Die existierenden MultiView-Frameworks sollten die Faktoren Benutzbarkeit und Systemqualität ebenso stark adressieren wie die bereits gute Systematisierung von Verknüpfungs- und Darstellungsarten. Adaptive Darstellungen und Benutzerführung (Wizards) sind hier mögliche Optionen. Die Transparenz der Verknüpfungen zwischen Sichten ist ein wichtiger Faktor der Benutzbarkeit von MultiView-Systemen: Benutzer müssen erkennen und nachvollziehen können, warum und inwiefern Informationen zusammenhängen (Kap.5.7.2).

9 Anhang

9.1 Notationen

9.1.1 GQM: Goal, Question, Metric

Das GQM-Modell (Goal, Question, Metric) [BW84] ist ein Vorgehensmodell um Ziele zu operationalisieren und wird häufig für Softwaredesign bzw. -evaluation eingesetzt. Wenn beispielsweise das Ziel ist, ein Unternehmen besser kennen zu lernen, so könnte gefragt werden nach Partnern, Zuverlässigkeit der Firma und öffentlicher Meinung. Fragen können wiederum Unterfragen haben. Zu den Fragen lassen sich Maße definieren. Die Methode impliziert eine Top-Down-Dekomposition zur Verfeinerung und Operationalisierung des Ziels. Für eine Evaluation bedeutet GQM folgendes:

- Welches Ziel soll durch die Messung erreicht werden? (Goal)
- Was soll gemessen werden bzw. welche Fragen soll die Messung beantworten? (Question)
- Welche Metrik(en) sind in der Lage, die notwendigen Eigenschaften zu beschreiben? (Metric)

Das folgende Vorgehen bietet sich für die Bewertung mit GQM an:

1. Definiere die Auswertungsziele für alle entwicklungspezifischen Qualitätsmerkmale
2. Leite alle Fragestellungen ab, die zu einer Quantifizierung dieser Auswertungsziele beitragen können
3. Leite alle Maße ab, die Informationen zur Beantwortung der Fragestellungen beitragen können
4. Entwerfe einen Mechanismus zur Meßwerterfassung der Maße.

Verwandte Vorgehensmodelle zur Messung der Systemqualität sind das FCM-Modell (Factor Criteria Metric) und das QIP (Quality Improvement Paradigm).

9.1.2 K3: Koordination, Kooperation, Kommunikation

K3 ist eine Notation für Zustände und Aktivitäten, die als Erweiterung von UML am Institut für Arbeitswissenschaft (iaw) der RWTH Aachen entwickelt wurde ([LBS01], S.1-13). Ziel war die partizipative Erhebung und Darstellung kooperativer, schwach strukturierter Arbeitsprozesse sowie die Beschreibung von Soll-Prozessen. K3 ist einfach genug, um als Kommunikationsbasis für Kunden und Systementwickler zu dienen und fokussiert die Modellierung von Arbeitsprozessen. Neben den Modellierungselementen für Arbeitsschritte und Übergängen dazwischen umfasst K3 auch Elemente zur Beschreibung benötigter Information und eingesetzter Systeme.

9.1.3 EPK: Ereignisgesteuerte Prozesskette

EPK ist eine Flowchart-Notation für Geschäftsprozesse, die von Scheer 1992 im Rahmen des ARIS-Konzepts, einer Modellarchitektur für integrierte Informationssysteme, entworfen wurde. EPK stellen Arbeitsprozesse in einer semiformalen Modellierungssprache grafisch dar. Dadurch sollen betriebliche Vorgänge systematisiert und parallelisiert werden. Objekte werden in gerichteten Graphen mit gerichteten Verknüpfungslinien verbunden. In einer solchen Verknüpfungskette wechseln die Objekte sich in ihrer Bedeutung zwischen Ereignis und Funktion ab. Jede Funktion kann zusätzlich mit einem Informationsobjekt verbunden werden, aus dem Informationen geladen oder in das Informationen gespeichert werden [Te01].

9.1.4 SWAPit-Systemskizze

SWAPit-Systemskizzen sind eine proprietäre Notation dieser Arbeit. Sie beschreiben, welche Sichten dargestellt werden, welche Informationsarten die Sichten repräsentieren und wie die Sichten zusammen hängen (Kap.9.1). Abbildung 9.1 zeigt schematisch, wie die Skizzen aufgebaut sind. Jedes Rechteck repräsentiert eine Sicht im System. Die Art der dargestellten Information und die Art der Darstellung werden ober- oder unterhalb des Rechtecks notiert. Die dargestellten Informationsmengen werden im Rechteck notiert. Manche Sichten können mehrere Informationsmengen darstellen (z.B. als Zweige einer Baumdarstellung oder als gestapelte Tabellen). Die Farben der Rechtecke zeigen an, welche Sichten sich eine Auswahlfarbe „teilen“: Auswahlen in zwei Sichten mit gleicher Farbe überschreiben ihre Auswahlen gegenseitig. Die Verknüpfungen von Daten in verschiedenen Sichten werden durch Pfeile repräsentiert, in denen die Kardinalität der Verknüpfung notiert wird. Im Beispiel (Abbildung 9.1) liest sich ein Pfeil folgendermaßen: „Die Auswahl eines Objekts der Informationsmenge 1 in der Themenlandkarte kann zu einer Hervorhebung von mehreren Objekten der Informationsmenge 5 in der Faktentabelle führen“. Ein letztes wichtiges Element sind die Beschreibungen der Pfeile durch Randnotizen. Schließlich können zwei Informationsweisen auf viele Weisen verknüpft sein (z.B. Patente und Firmen durch „ist-Urheber-von“ oder „hat-Interesse-an“, Dokumente und Personen durch „hat-abonniert“ oder „hat-gelesen“, usw.).

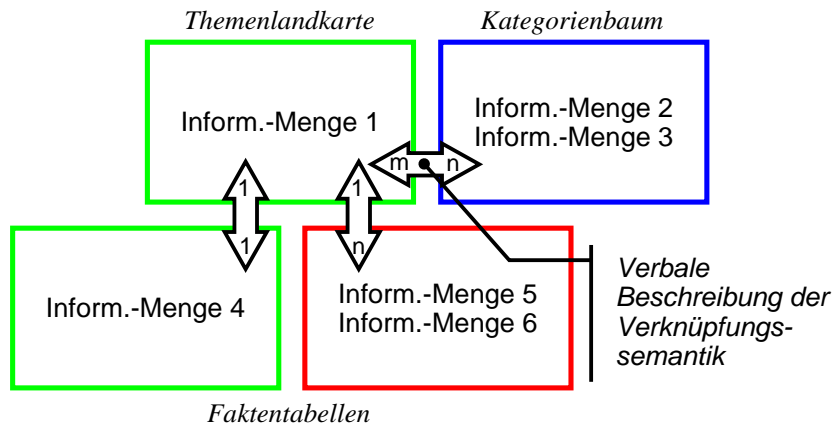


Abbildung 9.1: SWAPit-Systemskizzen - eine Notation für das Sichtendesign.

Zahlreiche Beispiele für die Verwendung der beschriebenen Systemskizzen finden sich in Kap.4 sowie Anhang 9.3. Die Notation könnte auch für MultiView-Systeme im Allgemeinen nützlich sein, für deren Beschreibung kein Standard-Formalismus verfügbar ist.

9.2 Detaillierte Beschreibung des Aufgabenmodells

In Kap.2.2 wurde die Entstehung des Aufgabenmodells zur Beschreibung von Rechercheaufgaben in Unternehmen beschrieben und das Aufgabenmodell vorgestellt. Hier wird jedes Aufgabenmerkmal mit seinen Ausprägungen erläutert und durch Beispiele illustriert. Dazu wird das Aufgabenmodell zunächst noch einmal übersichtlich dargestellt:

| Funktionale und Informatorische Mächtigkeit | | | | | |
|---|---------------------|------------------|-------------------------|-------------------------|------------------|
| Ziel der Info.-Recherche | lernen | Info. verdichten | Info. auswählen | Qualität prüfen | Qualität sichern |
| Betrachtete Info.-Arten | Text-Dok. | Dok.-Attribute | Beschreibungskategorien | Relationale Information | |
| Info.-Quellen | intern | | | extern | |
| Relat. Größe d. Info.-Teilmenge | gesamtes Repository | | große Teilmenge | kleine Teilmenge | |
| Dynamik der Info.-Teilmenge | konstant | | unterschiedlich | ad-hoc | |
| Betrachtete Verknüpfungen | Dok.-Spezifikation | Dok.-Dok. | Dok.-Thema | Dok.-Kategorie | Dok.-Tupel |
| Existenz der Verknüpfungen | logisch | | organisatorisch | physisch | |
| Verwendung der strukt. Info. | irrelevant | prüfen | verwenden | Dok. zuweisen | |
| Vokabular | allgemein | | Fachsprache | Korrespondenz | |
| Spektrum der Inhalte | allgemein | | | domänenspezifisch | |

| Kontrolle der Informationsrecherche | | | |
|-------------------------------------|---------------------------|------------------------------|-----------------------------|
| Interaktionsmethode | zielgerichtete Suche | explorative Recherche | Stöbern u. Entdecken |
| Interaktionsstrategie | filtern | | ausschließen |
| Granularität | Übersicht | | Details |
| Dynamik des Interesses | statisch | | wechselnd |
| Flexibilität Interaktion | frei | angeleitet | Workflow |
| Kommunikationsmodus | wahrnehmen (phys. passiv) | | spezifizieren (phys. aktiv) |
| Erwartetes Antwortverhalten | ohne Verzögerung | kurze Verzögerung akzeptiert | lange Berechnung akzeptiert |

| | | | | |
|-----------------------------------|---------------------------------|---------|-----------------------|--|
| Relevanz im Unternehmen | Kernprozess | | Unterstützungsprozess | |
| Absolute Bedeutung | „nice to have“ | wichtig | kritisch | |
| Intendierte Wertsteigerung | Steigerung der Effizienz | | neue Aufgabe | |
| Primäres Qualitätskrit. | Genauigkeit und Vollständigkeit | | grober Überblick | |
| Qualität und Wert des Service | | | | |

| | | | |
|-------------------------------|-----------------------|-----------------------|---------------------|
| Vorhersagbarkeit | gelegentlich | wiederholt | regelmäßig |
| Durchlaufzeit | Zeit ist irrelevant | Zeit ist wichtig | Zeit ist kritisch |
| Expertise in Domäne | Laie | | Domänen-Experte |
| Expertise in Korpus | Information unbekannt | | Information bekannt |
| Expertise in Recherche | gelegentl. Nutzer | Experte der Recherche | System-Experte |
| Aufwand der Arbeitsprozesse | | | |

9.2.1 Funktionale und Informatorische Mächtigkeit

Dieser Aspekt des Aufgabenmodells beschreibt wichtige Informationen und Systemfunktionalitäten, die bei der interaktiven Recherche eine Rolle spielen.

ZIEL DER INFORMATIONS-RECHERCHE:

Das Ziel leitet sich aus der Motivation für die Bearbeitung der Aufgabe ab. Will ein Analyst etwas von oder über die Informationen LERNEN, so beabsichtigt er die Internalisierung (Informationsaufnahme) von explizitem dokumentiertem Wissen. Will der Analyst INFORMATION VERDICHTEN, so betreibt er die Recherche um neue Information zu generieren (z.B. Ergänzung, Annotation, Zusammenfassung), also eine Teilmenge der Informationsobjekte zu synthetisieren. Will der Analyst INFORMATION AUSWÄHLEN, so möchte er Information nach bestimmten Merkmalen identifizieren und aus der Sammlung extrahieren. Die beiden folgenden Ziele betreffen die Informationsqualität, die Informationsobjekte selbst sind also Gegenstand des Interesses. Bei QUALITÄT PRÜFEN soll die Gültigkeit von Qualitätskriterien validiert werden. Bei QUALITÄT SICHERN werden Daten ggf. aktiv verändert, damit sie den Qualitätskriterien genügen.

BEISPIELE:

LERNEN: Analyse einer Sammlung von Patenten, um sich ein Bild von der Konkurrenz zu machen.

INFORMATION VERDICHTEN: Für eine Präsentation werden Projektdokumente zu einem Bericht verdichtet.

INFORMATION AUSWÄHLEN: Suche nach Texten, die Zusammenfassungen sind.

QUALITÄT PRÜFEN: Prüfen, ob technische Dokumentation in sich abgeschlossen ist.

QUALITÄT SICHERN: Dokumentation von Software ergänzen um die Verlinkung verwandter Abschnitte.

BETRACHTETE INFORMATIONSARTEN:

Die betrachteten Informationsarten unterscheiden verschiedene Repräsentationsformen für Information (siehe auch Kap.5.1). TEXTDOKUMENTE enthalten natürliche Sprache und sind nicht oder semi-strukturiert (gliedert). DOKUMENTENATTRIBUTE sind strukturierte Metadaten zu Dokumenten. BESCHREIBUNGSKATEGORIEN repräsentieren Konzepte. Sie können miteinander in Beziehung stehen und Dokumenten zugewiesen werden (Klassifikation). RELATIONALE INFORMATION besteht aus Werte-Tupeln mit festgelegter Struktur.

BEISPIELE:

TEXTDOKUMENT: Patente, E-Mail, Berichte, Protokolle, Blogs, Firmenprofile, Produktbeschreibungen

DOKUMENTENATTRIBUTE: Erstellungsdatum, Länge, Sprache, Quelle, Bewertung, Schlagworte (Tags)

BESCHREIBUNGSKATEGORIEN: Patentklassen, Unternehmensbereiche (Organigramm), Länder

RELATIONALE INFORMATION: Institutionen, Veranstaltungen, Transaktionen, Kundenstammdaten

INFORMATIONSQUELLEN:

Information lässt sich nach ihrer Herkunft unterscheiden. Damit ist nicht der Ort der Speicherung und Verwaltung gemeint, sondern der Urheber der Information. INTERNE Informationen werden vom Unternehmen erstellt, EXTERNE Informationen nicht.

BEISPIELE:

INTERN: Bericht, Produktkatalog, E-Mail, Memo, Dokumentation, Produktions- und Absatzzahlen, Lieferanten

EXTERN: Fachpresse, Statistiken der GfK⁸⁶ oder des Stat. Bundesamts, Whitepaper von Konkurrenzprodukten

RELATIVE GRÖÖE DER INFORMATIONEN-TEILMENGE:

Die Größe der betrachteten Informationsteilmenge in Relation zur gesamten verfügbaren Menge von Information ist ein Ergebnis der Selektivität. Die Selektivität beeinflusst u.a. Informationsdarstellung und Vorgehensstrategie. Unterschieden wird, ob das GESAMTE REPOSITORY betrachtet wird, eine GROÖE TEILMENGE oder eine KLEINE TEILMENGE. Die Größe der Teilmenge ist relativ zur Gesamtmenge zu beurteilen. Wird beispielsweise das Internet als Informationsgesamtheit betrachtet, so ist der fokussierte Ausschnitt in der Regel klein.

BEISPIELE:

GESAMTES REPOSITORY: Uninformierte Patentrecherche; Übersicht über Projektdokumentation

GROÖE TEILMENGE: Informierte Patentrecherche zu einer ausgewählten Patentklasse

KLEINE TEILMENGE: Ergebnisse einer Internetsuche auswerten

DYNAMIK DER INFORMATIONEN-TEILMENGE:

Die Dynamik der Informations-Teilmenge hängt zusammen mit dem vorangehenden Merkmal zusammen. Der zu betrachtende Ausschnitt ist KONSTANT, wenn immer genau die gleichen Informationen betrachtet werden. Wird immer die gleiche Partitionierung durchgeführt, beispielsweise durch Klassifikation nach festen Kriterien oder Annotation, ist der Ausschnitt UNTERSCHIEDLICH. Wechseln die Kriterien zur Auswahl der Teilmenge, so ist sie AD-HOC.

BEISPIELE:

KONSTANT: Zugriff auf die Dokumentation eines abgeschlossenen Kundenprojektes

UNTERSCHIEDLICH: Wiederholte Recherche im Internet nach Informationen zu einem Thema

AD-HOC: Suchergebnisse im Intranet auswerten

⁸⁶ Gesellschaft für Konsumforschung

BETRACHTETE VERKNÜPFUNGEN:

Dieses Merkmal unterscheidet die für eine Aufgabe relevanten Arten von Verknüpfungen von Informationen. Bei DOKUMENT-SPEZIFIKATION spielt eine Rolle, inwiefern Dokumente zu einer expliziten Spezifikation wie beispielsweise Filter, Anfrage, Klassenbeschreibung passen. Bei DOKUMENT-DOKUMENT ist die (Un-)Ähnlichkeit von Dokumenten bezüglich eines oder mehrerer Merkmale bedeutsam. Bei DOKUMENT-THEMA ist die Abstraktion von Dokumenten zu Themen wichtig. Dabei können Themen intensional (durch Eigenschaften) oder extensional (durch eine Menge von Dokumenten) gegeben sein. Die Ausprägungen DOKUMENT-KATEGORIE und DOKUMENT-TUPEL bezeichnen die Zugehörigkeit von Strukturierten Daten zu Dokumenten.

BEISPIELE:

DOKUMENT-SPEZIFIKATION: Suchergebnisse auswerten: Inwiefern passen die Ergebnisse zur Anfrage?

DOKUMENT-DOKUMENT: In einer Fall-Datenbank zu einer relevanten Problembeschreibung weitere finden

DOKUMENT-THEMA: Prüfen des Single-Source-Kriterium für technische Dokumentation

DOKUMENT-KATEGORIE: Prüfen der Konsistenz einer Klassifikation.

DOKUMENT-TUPEL: Vertragsmerkmale zu auffälligen Kundenbeschwerden identifizieren

EXISTENZ DER VERKNÜPFUNGEN:

Die Verknüpfungen von Informationen können in verschiedenen Formen vorliegen. Sind die Verknüpfungen in digital gespeichert, so existieren sie PHYSISCH. In manchen Fällen sind die Informationen nicht explizit verknüpft abgelegt, aber es gibt einen ORGANISATORISCHEN Zusammenhang aufgrund von Arbeitsflüssen. Als LOGISCH verknüpft werden Informationen bezeichnet, wenn sie bislang weder organisatorisch noch informationstechnisch verknüpft sind, aber in Bezug auf die Aufgabe aufgrund ihrer Bedeutung zusammen gehören.

BEISPIELE:

LOGISCH: Kundenbeschwerden und Kündigungswahrscheinlichkeiten zu Verträgen

ORGANISATORISCH: Kennzahlen zur Unternehmensperformanz aus Sicht von Marketing und Controlling

PHYSISCH: Blog-Eintrag und Schlagwort-Beschreibung (Tags)

VERWENDUNG DER STRUKTURIERTEN DATEN:

Das Merkmal beschreibt, welche Rolle Kategorien und Relationale Tupel bei der Aufgabe spielen. Sind die strukturierten Daten IRRELEVANT, so sollen nur die reinen Textdokumente analysiert werden. Oft sind die strukt. Informationen aber wichtig zur Beschreibung der Texte und sollen zur Recherche VERWENDET werden, z.B. zur Filterung, Suche oder Navigation. Die strukt. Informationen können auch selbst Gegenstand des Interesses sein. Zum einen kann es darum gehen, diese nach Qualitätskriterien wie Konsistenz oder Vollständigkeit zu PRÜFEN. Zum anderen kann das Ziel der Aufgabe sein, gerade die Verknüpfung mit den Texten herzustellen: Man möchte die strukt. Informationen den DOKUMENTEN ZUWEISEN.

BEISPIELE:

IRRELEVANT: Untersuchen der Wortverwendung in einer Menge kulturwissenschaftlicher Publikationen

VERWENDEN: Analyse von Beschwerdeschreiben zu Kunden mit bestimmten demografischen Merkmalen

PRÜFEN: Stimmt die vergebene Wertung mit den Aussagen in Kundenrezensionen zu Büchern überein? Repräsentiert ein Klassifikationsschema die klassifizierten Textinhalte gut (z.B. Diversifikation, Symmetrie)?

DOKUMENTEN ZUWEISEN: Klassifikation von Zeitungsartikeln im Themenkatalog des Verlags

VOKABULAR:

Die Wortverwendung spielt eine große Rolle bei der automatischen Inhaltsanalyse, die durch ein präzise Begrifflichkeit und ein abgeschlossenes Vokabular erleichtert wird. Bei einem ALLGEMEINEN Vokabular ist keine solche Spezifizierung der Begrifflichkeiten gegeben. FACHSPRACHE wird von Spezialisten einer Branche, eines Technologiebereichs, einer Unternehmung oder ein Produkt (etc.) gebraucht und verstanden. KORRESPONDENZ zeichnet sich durch eine Vermischung von Kommunikationsvokabular (z.B. Begrüßung, Drohung) mit den eigentlichen Inhalten der Mitteilungen aus.

BEISPIELE:

ALLGEMEIN: Archiv einer Presseagentur, Gedichte aus dem Projekt Gutenberg

FACHSPRACHE: Beiträge im Unternehmensblog zu einem Produkt, techn. Dokumentation, Börsennachrichten

KORRESPONDENZ: Kundenbeschwerden, Kundenmeldungen

SPEKTRUM DER INHALTE:

Obgleich das Inhaltsspektrum von Texten häufig mit dem Merkmal „Vokabular“ zusammenhängt, ist eine Unterscheidung berechtigt. Sind die Inhalte ALLGEMEIN, so lässt sich der Themenraum nicht klar begrenzen, wie dies bei DOMÄNENSPEZIFISCHEN Inhalten möglich ist.

BEISPIELE:

ALLGEMEIN: Dokumente im Intranet

DOMÄNENSPEZIFISCH: Kundenbeschwerden, Kundenmeldungen

9.2.2 Kontrolle der Informationsrecherche

Dieser Aspekt des Aufgabenmodells beschreibt interaktive Strategien und Vorgehensweisen, die beim Recherche-Prozess eine Rolle spielen.

INTERAKTIONSMETHODE:

Die Interaktionsmethode unterscheidet nach dem Vorgehen bei der Recherche. Bei der ZIELGERICHTETEN SUCHE können die gesuchten Informationen explizit beschrieben werden, da sie selbst oder ihre Eigenschaften bereits bekannt sind. Bei der EXPLORATIVEN RECHERCHE ist die Fragestellung vager und es wird assoziativ vorgegangen (Kap.5.1): Interessante Verknüpfungen zwischen Dokumenten oder zwischen Dokumenten und strukturierten Daten können zur Orientierung bei der Navigation genutzt werden. Wenn das Ziel unspezifiziert ist und die Aufgabe starken Entdeckungscharakter hat, also schlicht dem Kennenlernen der Information aus nicht näher festgelegten Gründen dient, wird nicht so systematisch vorgegangen. In solchen Fällen geht es um STÖBERN UND ENTDECKEN.

BEISPIELE:

ZIELGERICHTETE SUCHE: Stichwortsuche im DMS mit dem Ziel einen Brief wieder zu finden

EXPLORATIVE RECHERCHE: Suche nach Erfolgsfaktoren in der Dokumentation eines erfolgreichen Projekts; Ein Chemiekonzern sucht nach Patent-Nischen um lohnende Experimentierfelder zu entdecken.

STÖBERN UND ENTDECKEN: Tägliches Querlesen der Finanznachrichten

INTERAKTIONSSTRATEGIE:

Dieses Merkmal beschreibt komplementär zum Merkmal „Interaktionsmethode“ das Vorgehen. Beim FILTERN werden die interessanten Informationen direkt ausgewählt bzw. spezifiziert. Beim AUSSCHLIEBEN wird umgekehrt versucht, die uninteressanten Informationen auszuwählen oder zu spezifizieren. Welches Vorgehen geeigneter ist, hängt neben der der Interaktionsmethode auch von der Größe der Informationsmenge (Kap.9.2.1) und dem primären Qualitätskriterium (Kap.9.2.3) ab.

BEISPIELE:

FILTERN: Suche nach Dokumenten zu einem bestimmten Thema im Intranet

AUSSCHLIEBEN: Patente einer Klasse als unkritisch für eine technische Erfindung bestätigen

GRANULARITÄT:

Dieses Merkmal hängt meist mit dem primären Qualitätskriterium (Kap.9.2.3) zusammen, beschreibt jedoch im Gegensatz dazu den Detailgrad des Vorgehens, nicht des Ziels. Bei vielen Aufgaben bedarf es der groben ÜBERSICHT über Information, um interessante Information zu finden, deren DETAILS genauer zu betrachten sind, wie B. Shneiderman in seinem Mantra zur Informationssuche feststellt [CC05].

BEISPIELE:

ÜBERSICHT: Kernthemen im Kunden-Blog eines Unternehmens u. ihre Entwicklung über die Zeit analysieren

DETAILS: Wie wurde ein erneut in ähnlicher Weise aufgetauchtes Kundenproblem die letzten Male gelöst?

DYNAMIK DES INTERESSES:

Ist der Interessensfokus der Recherche unabhängig von den Informationen und der Arbeit damit, also vorgegeben und STATISCH, so ist ein besonders systematisches Vorgehen möglich. Oftmals ist diese Ausprägung mit einer guten Kenntnis der Sammlung verbunden und beschreibt Routineaufgaben. Es ist möglich, Informationswunsch und Dokumentensammlung weitgehend isoliert zu betrachten. Ist der Interessensfokus eher vage und grob, so kann es während der Recherche verfeinert, präzisiert und angepasst werden, ist also WECHSELND. Dokumentensammlung und Informationswunsch sind eng miteinander verquickt. Während der Analyse gewonnene Erkenntnisse können dazu führen, dass sich der Informationswunsch verschiebt und dem neuen Erkenntnisstand anpasst. Der Analyst möchte ein Informationsbedürfnis durch die Analyse konkretisieren und erfüllen.

BEISPIELE:

STATISCH: Suche in einer Kooperationsumgebung nach Personen, die Experten für ein Thema sind

WECHSELND: Suche in der Fachpresse nach Erklärungen für eine Entwicklung an der Börse

FLEXIBILITÄT DER INTERAKTION:

Das Merkmal beschreibt, wie intensiv ein Analyst durch eine Systemunterstützung betreut und geführt werden soll. Soll sich der Analyst FREI zwischen den interaktiven Möglichkeiten der Recherche entscheiden können, wünscht er sich eine dynamische ANLEITUNG, (z.B. Empfehlungen, Hinweise), oder folgt der Rechercheprozess sogar einem klaren Muster, das sich als WORKFLOW beschreiben und unterstützen lässt?

BEISPIELE:

FREI: Literaturrecherche zum State-of-the-Art im Rahmen einer Forschungsarbeit

ANGELEITET: Informationssuche mit einigen Feedback-Zyklen durch den Benutzer

WORKFLOW: Standardisierte Erstellung von Quartalsberichten („briefing books“) für das Management.

KOMMUNIKATIONSMODUS:

Der Kommunikationsmodus bezeichnet die dominierende Richtung der Kommunikation zwischen Benutzer und System. Ist die Kommunikation geprägt von kognitiven Reaktionen oder Assoziationen, ausgelöst durch die Präsentation von Information, so geht es um das WAHRNEHMEN zur Ausbildung eines mentalen Modells. Steht die aktive Kontrolle und Auswahl von Information und Präsentation im Vordergrund, so geht es um das SPEZIFIZIEREN.

BEISPIELE:

WAHRNEHMEN (PHYS. PASSIV): Einprägen der inhaltlichen Struktur einer Dokumentensammlung

SPEZIFIZIEREN (PHYS. AKTIV): Entwurf eines Kategorisierungs-Schemas für ein Zeitungsarchiv

ERWARTETES ANTWORTVERHALTEN:

Das Merkmal erfasst die maximale Zeit, die Nutzer bereit sind bei der Recherche auf eine Reaktion des unterstützenden Systems zu warten. Besonders bei zeitkritischen Prozessen wird oft eine Antwort OHNE VERZÖGERUNG erwartet. In anderen Fällen wird eine KURZE VERZÖGERUNG hingenommen, oder Zeit spielt keine große Rolle und sogar LANGE BERECHNUNGEN werden vom Benutzer akzeptiert.

BEISPIELE:

OHNE VERZÖGERUNG: Ein neues Recherchesystem wird gerade eingesetzt, um die Suche zu beschleunigen
KURZE VERZÖGERUNG AKZEPTIERT: E-Mail-Klassifikation nach benutzerdefinierten Kriterien
LANGE BERECHNUNG AKZEPTIERT: Berechnung eines übersichtlichen Scatterplots zur Patentanalyse

9.2.3 Qualität und Wert des Service

Dieser Aspekt des Aufgabenmodells beschreibt die Auswirkung und Verwendung der Ergebnisse im Unternehmen.

RELEVANZ IM UNTERNEHMEN:

Die Relevanz einer Rechercheaufgabe ist für das Unternehmen besonders hoch, wenn es sich dabei um eine Kernkompetenz des Unternehmens handelt, die in einem KERNPROZESS realisiert wird. Andererseits kann eine Recherche als UNTERSTÜTZUNGSPROZESS den Kernprozessen des Unternehmens zuarbeiten. Die Unterscheidung, die das Merkmal liefert, legt nicht den absoluten Wert des Prozesses fest: Ein Unterstützungsprozess eines besonders werthaltigen Kernprozesses kann dem Unternehmen zeitweise wichtiger sein als andere Kernprozesse.

BEISPIELE:

KERNPROZESS: Patentrecherche wird selbst als Dienstleistung angeboten
UNTERSTÜTZUNGSPROZESS: Lösungsdatenbank wird durchsucht um eine Kundenanfrage zu beantworten

ABSOLUTE BEDEUTUNG:

Dieses Merkmal ergänzt das Merkmal „Relevanz im Unternehmen“ um die absolute Bewertung der Aufgabe im Unternehmen. Ist eine Aufgabe NICE-TO-HAVE, so kann das Unternehmen notfalls darauf verzichten, ohne dass die Prozesse in Qualität und Effizienz wesentlich beeinträchtigt werden. Bei WICHTIGEN Aufgaben kann das Unternehmen ohne die Aufgabe alle Geschäftstätigkeiten fortsetzen, wird aber wesentlich beeinträchtigt. Bei KRITISCHEN Aufgaben ist das Unternehmen auf die Rechercheaufgabe angewiesen.

BEISPIELE:

NICE-TO-HAVE: Überblick über die Publikationen einer Wissenschafts-Community hilft bei der Vernetzung
WICHTIG: Gründe von Kundenbeschwerden untersuchen, um Produkte und Services zu optimieren
KRITISCH: Recherche eines Sportjournalisten für einen Abschluss-Bericht zur Weltmeisterschaft

| INTENDIERTE WERTSTEIGERUNG: |
|---|
| Das Merkmal unterscheidet die Art der Wertsteigerung. Wird eine STEIGERUNG DER EFFIZIENZ angestrebt, so soll ein Rechercheprozess in Qualität oder Effizienz verbessert werden. Andererseits kann die Recherche eine NEUE AUFGABE sein. |
| BEISPIELE: |
| STEIGERUNG D. EFFIZIENZ: Suchmaschine wird um Cluster-Darstellung zur Ergebnis-Exploration ergänzt NEUE AUFGABE: Beschwerdetexte werden jetzt auf Kundenstimmung und Nachdrucksmittel hin untersucht |

| PRIMÄRES QUALITÄTSKRITERIUM: |
|--|
| Das Merkmal beschreibt den Fokus der Analyse und hängt oft mit dem Merkmal „Granularität“ (Kap.9.2.2) zusammen. Geht es um GENAUIGKEIT UND VOLLSTÄNDIGKEIT, so spielt potentiell jede Information eine Rolle und muss ggf. genau untersucht werden. Geht es um den GROBEN ÜBERBLICK, so liegt der Fokus auf besonders wichtigen bzw. repräsentativen Informationen und aggregierter Information. |
| BEISPIELE: |
| GENAUIGKEIT UND VOLLSTÄNDIGKEIT: Patente identifizieren, die für Erfindung kritisch sind GROBER ÜBERBLICK: Themenentwicklung einer Fachzeitschrift untersuchen |

9.2.4 Aufwand der Arbeitsprozesse

Dieser Aspekt des Aufgabenmodells beschreibt die Rahmenbedingungen und Nutzer des Rechercheprozesses im Unternehmen.

| VORHERSAGBARKEIT: |
|--|
| Die Vorhersagbarkeit beschreibt, wie genau sich eine Aufgabe planen lässt. Eine GELEGENTLICH auftretende Aufgabe wird ausgelöst durch ein Ereignis, welches die Recherche notwendig macht. Bei einer WIEDERHOLT auftretenden Aufgabe steht lange im Voraus fest, dass die immer wieder auftretende Recherche erneut erforderlich wird. Für eine REGELMÄßIG auftretende Aufgabe lassen sich genaue Termine angeben. |
| BEISPIELE: |
| GELEGENTLICH: Eine Absatzkennzahl weicht von der Zielvorgabe ab und die Ursachenforschung erfordert eine umfangreiche Recherche WIEDERHOLT: Patentanalyse vor Einführung einer neuen Produktionsanlage REGELMÄßIG: Recherche im Bestand der Kundenbeschwerden für den Beschwerdebericht des Quartals. |

DURCHLAUFZEIT:

Die maximale Zeit, die eine Recherche in Anspruch nehmen darf, spielt eine wichtige Rolle für die Gestaltung der Systemunterstützung. Zeitunkritische Aufgaben sind meist besonders werthaltig, weshalb es viel stärker auf das Ergebnis als auf den Rechercheprozess selbst ankommt und der Faktor Zeit nahezu IRRELEVANT ist. Ist Zeit WICHTIG, so soll die Recherche zeiteffizient durchführbar sein. Ist Zeit KRITISCH, so ist die Recherche unbedingt zeitlich zu optimieren. Oft limitieren sogar maximale Durchlaufzeiten die Dauer der Recherche.

BEISPIELE:

ZEIT IST IRRELEVANT: Im Presseclipping wird das Presse-Echo auf eine Werbekampagne ausgewertet

ZEIT IST WICHTIG: Ein Sportjournalist bereitet den täglichen Sportkommentar vor

ZEIT IST KRITISCH: Die Recherche in der Lösungsdatenbank im Call-Center

EXPERTISE IN DOMÄNE:

Das Merkmal unterscheidet die fachliche Vertrautheit des Analysten mit dem Fachgebiet, dem die betrachteten Informationen zuzurechnen sind. Ein LAIE kennt sich in der Domäne kaum aus, während ein EXPERTE über gutes Fachwissen verfügt.

BEISPIELE:

LAIE: Person recherchiert in einer medizinischen Datenbank nach Krankheitsbildern zu ihren Symptomen

EXPERTE: Wissenschaftler recherchiert Publikationen in seinem Spezialgebiet

EXPERTISE IN KORPUS:

Das Merkmal unterscheidet die Vertrautheit des Analysten mit der Informationssammlung. Sind die Informationsobjekte UNBEKANNT oder BEKANNT? Sind sie bekannt, so hat der Analyst die meisten Informationsobjekte zu einem früheren Zeitpunkt bereits einmal betrachtet. Das Merkmal hat wichtigen Einfluss auf die Systemunterstützung. So ist eines der größten Probleme mit anfrageorientierten Retrievalsystemen die Unkenntnis der Sammlung und die daraus resultierende potentielle Unfähigkeit, Schlagworte zu finden, die das gesuchte Thema optimal von anderen Themen der Sammlung abgrenzen ([St95], S.39).

BEISPIELE:

INFORMATION UNBEKANNT: Recherche in den Kundenrezensionen eines Online-Buchhandels

INFORMATION BEKANNT: Patentrecherche in einer vertrauten Patentklasse

EXPERTISE IN RECHERCHE:

Das Merkmal unterscheidet die Vertrautheit des Analysten mit Verfahren und Systemen der Informationsrecherche. Die Recherchekompetenz eines GELEGENTLICHEN NUTZERS ist gering. Ein EXPERTE DER RECHERCHE verfügt über eine hohe Recherchekompetenz. Ein SYSTEM-EXPERTE ist vertraut mit dem System, das eingesetzt und bewertet werden soll.

BEISPIELE:

GELEGENTLICHER NUTZER: Mitarbeiter sucht sporadisch nach Information im Intranet

EXPERTE DER RECHERCHE: Presseabteilung bereitet wöchentlich Clippings vor

SYSTEM-EXPERTE: siehe Ausprägung „Experte der Recherche“

9.3 Weitere Anwendungsbeispiele

Analyse wissenschaftlicher Publikationen der Informatik

Seit mehreren Jahren wird SWAPit auf den Webseiten des Lehrstuhls Informationssysteme (Informatik 5, Prof. Jarke) als Alternative zur anfragebasierten Suche eingesetzt. Mit SWAPit lassen sich die wissenschaftlichen Publikationen des Lehrstuhls (englische Veröffentlichungen seit 1998) umfassend analysieren. Das SWAPit-Projekt wird monatlich aktualisiert, um die neuen Abstracts einzubeziehen. Die strukturierten Daten sind stets aktuell, weil SWAPit direkt auf die Publikationsdatenbank des Lehrstuhls zugreift⁸⁷.

Publications

Some papers are available for download from one of the following locations:

- [Annual report 2006](#)
- Papers related to [ConceptBase](#)
- [CREWS](#) Report Series
- [NATURE](#) Report Series
- [Technical Reports](#) (Aachener Informatik-Berichte)

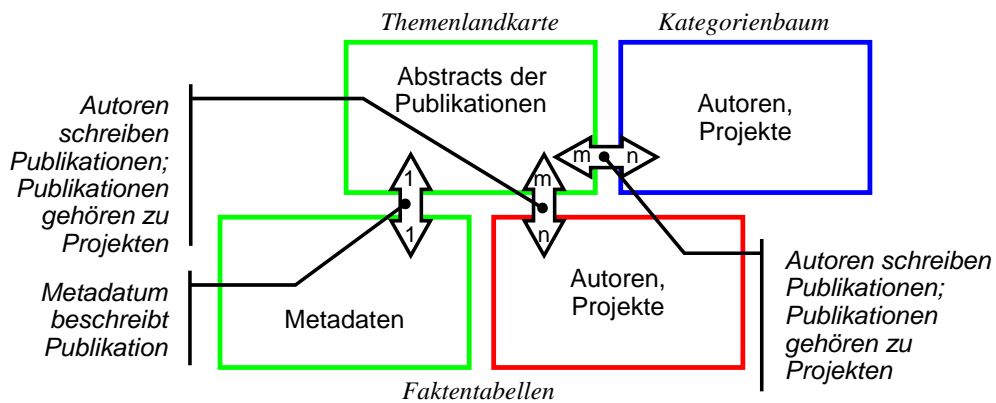
[Search our Publications Database](#)

[Browse our Publication Database with SWAPit \(English abstracts since 1998\)](#) [[Help](#)]

Books and conference proceedings (sorted by year of publication)

- [2007](#)
- [2006](#)
- [2005](#)
- [2004](#)
- [2003](#)

⁸⁷ <http://www-i5.informatik.rwth-aachen.de/lehrstuhl/publications/swapit.html>



negotiation negotiation negotiations negotiators negoti Search Clear Search Type Bool (over Index)

Focus on: Facts Map Tree Text Search all

- Root
- Author
- Project
 - CREWS
 - enegotiation
 - GCO
 - MEMO
 - Negotiation Group
 - SEWASIE

Matching mode Any All

Info Facts Filtering Facts Statistics
Term Statistics Document Summary Fulltext Legend

Highlight in text negotiation

LAP and Semantic Web: A Language Action Perspective on Electronic Contracts

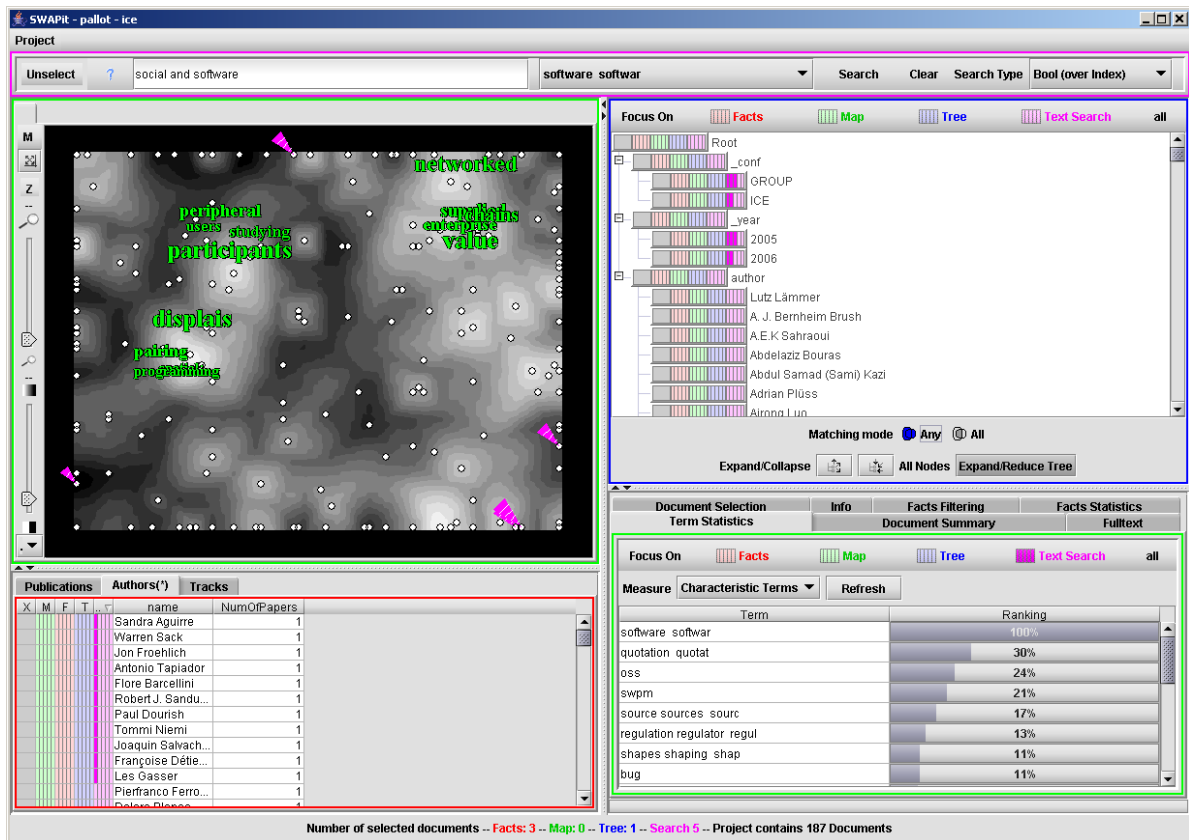
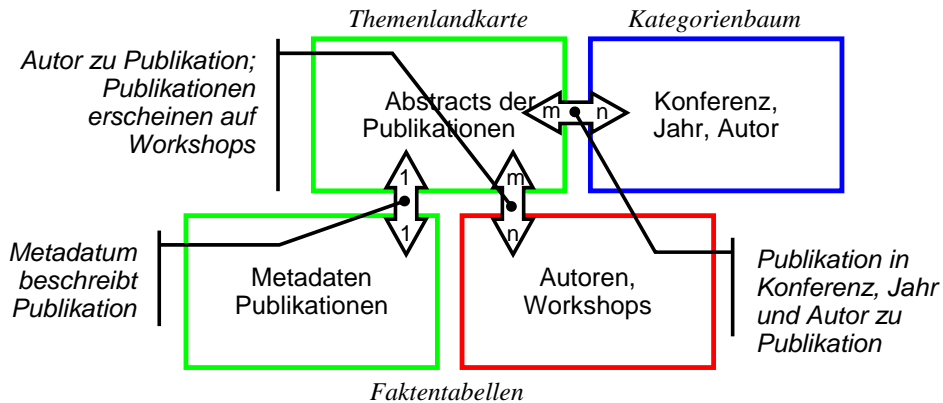
An electronic business transaction consists of three phases, namely search, **negotiation**, and fulfilment. The **negotiation** phase is the most communication-intensive of the three. In order to enable electronic **negotiations**, the complex communicative exchanges need to be supported by means of information technology. The Language-Action Perspective can provide a suitable theoretical and conceptual basis. In addition to the communicative exchanges, document management also plays a vital role for **e-negotiations**. Semantic web ideas can be most useful for this part of a **negotiation**. We will present an integrated approach implemented in the **negotiation** support system **Nepocist** that combines LAP and <http://fit.informatik.rwth-aachen.de:1236/ToolServlets/sql?url?host=fit.inf...>

| Publications | Authors(*) | Projects | |
|--------------|------------|-------------------|---|
| X M F T | name | fullname | homepage |
| | Schoop | Mareike Schoop | http://www-i5.informatik.rwth-aachen.de/le |
| | Quix | Christoph Quix | http://www-i5.informatik.rwth-aachen.de/le |
| | Jarke | Matthias Jarke | http://www-i5.informatik.rwth-aachen.de/le |
| | Becks | Andreas Becks | http://www-i5.informatik.rwth-aachen.de/le |
| | Haumer | Peter Haumer | http://www-i5.informatik.rwth-aachen.de/le |
| | Huster | Jessica Huster | http://www-i5.informatik.rwth-aachen.de/le |
| | Jertila | Aida Jertila | http://www-i5.informatik.rwth-aachen.de/le |
| | Jeusfeld | Manfred Jeusfeld | http://www-i5.informatik.rwth-aachen.de/m |
| | Kensche | David Kensche | http://www-i5.informatik.rwth-aachen.de/le |
| | Köller | Jörg Köller | http://www-i5.informatik.rwth-aachen.de/le |
| | List | Thomas List | http://www-i5.informatik.rwth-aachen.de/le |
| | Nissen | Hans Nissen | http://www-i5.informatik.rwth-aachen.de/le |
| | Pohl | Klaus Pohl | http://www-i5.informatik.rwth-aachen.de/pc |
| | Seeling | Christian Seeling | http://www-i5.informatik.rwth-aachen.de/le |

Number of selected documents -- Facts: 0 -- Map: 0 -- Tree: 0 -- Search 13 -- Project contains 134 Documents

People-Concept Mapping

In einem Forschungsprojekt wurde untersucht, in welchem Zusammenhang Personen und Themen stehen („people-concept mapping“). Dazu wurden die Tagungsbände dreier wichtiger Tagungen aus dem Bereich Kooperativen Arbeitens (ICE 2005, ICE 2006, GROUP 2005) mit SWAPit untersucht. Einige Autoren publizierten bei mehr als einer Tagung. Fragen waren u.a., welche Themen die Wissenschaftler besetzen, welche Experten es für bestimmte Themen gibt, und ob Publikationen, die inhaltlich verwandt sind, in den gleichen Workshops publiziert werden [SP+07].



Geo- u. Themenbasierter Zugriff auf Veranstaltungen im Informatikjahr

Im Rahmen des deutschen Informatikjahrs 2006 fand in Berlin am 22.09.2006 der „Tag der Innovationen“ statt. Bei dieser Veranstaltung wurde mit Unterstützung von Microsoft Deutschland die Website „Informatiklandkarte“ vorgestellt, welche die deutschen Forschungsstandorte der Informatik und die Veranstaltungen des Informatikjahrs zugreifbar machte⁸⁸. Zur Visualisierung wurde ein Geo-Informationssystem mit SWAPit zur themenbasierten Navigation gekoppelt.

- Home
- News & Blogs
- Angebote des ZIS Aachen
- Termine
- Downloads
- Informatiklandkarte**
- Veranstaltungskalender
- Universitäten
- Fachhochschulen
- Forschungsinstitute
- Route planen
- Wir über uns
- Kontakt
- Newsletter

Veranstaltungskalender

Veranstaltungen anzeigen

Zeitraum: 19 September 2006 bis 24 September 2006 Anzeigen

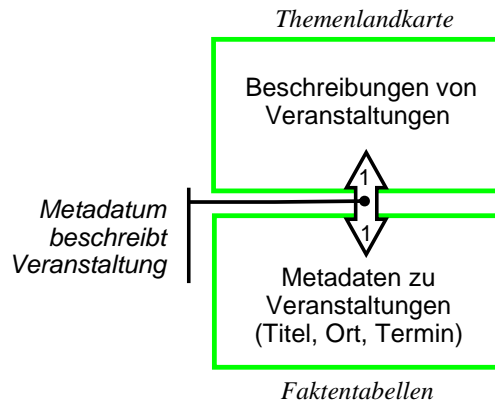
Ort: Umkreis: 10 km Anzeigen

Veranstaltungen von Dienstag, 19. September 2006 bis Sonntag, 24. September 2006

| Nr | Veranstaltung |
|----|--|
| 1 | Tag der Innovationen 19.9.2006, Berlin |
| 2 | bwcon_Seminar_Web 2.0- Was steckt im Web der nächsten Generation? 20.9.2006, Stuttgart |
| 3 | Informationstour "barrierefrei kommunizieren" 21.9.2006, Magdeburg |
| 4 | Technikführung durch das rock'n'popmuseum 21.9.2006, Gronau |
| 5 | Schüler-Kryptotag 22.9.2006, Paderborn |
| 6 | Informationstour "barrierefrei kommunizieren" 22.9.2006, Fulda |
| 7 | START-Messe: Messe für Existenzgründung, Franchising und junge Unternehmen 22.9.2006, Essen |
| 8 | Hightech zum Anfassen 23.9.2006, Karlsruhe |
| 9 | Hightech zum Anfassen ? Der BGV einmal anders! 23.9.2006, Karlsruhe |

In Themenkarte anzeigen

⁸⁸ <http://zfs.informatik.rwth-aachen.de/InformatikLandkarte/Default.aspx>



Metadaten

| X | M | F | T | S | Titel | Ort | Termin / |
|---|---|---|---|---|---|-------------|------------|
| | | | | | Jahrhundert des Computers | Groß Glogow | 18.09.2006 |
| | | | | | Tag der Innovationen | Berlin | 19.09.2006 |
| | | | | | bwcon: Seminar Web 2.0- Was steckt im Web d. | Stuttgart | 20.09.2006 |
| | | | | | Technikführung durch das rock'n'popmuseum | Gronau | 21.09.2006 |
| | | | | | Schüler-Kryptotag | Paderborn | 22.09.2006 |
| | | | | | START-Messe: Messe für Existenzgründung, Fr... | Essen | 22.09.2006 |
| | | | | | Hightech zum Anfassen | Karlsruhe | 23.09.2006 |
| | | | | | Hightech zum Anfassen □ Der BGV einmal and... | Karlsruhe | 23.09.2006 |
| | | | | | Bankprodukte erreichen ihre Kunden | Paderborn | 23.09.2006 |
| | | | | | Kongress der Informatik für Schülerinnen und S... | Berlin | 25.09.2006 |
| | | | | | Abschlussveranstaltung IT-Gründungsnetzwerk | Dortmund | 28.09.2006 |

Anzahl ausgewählter Dokumente -- Fakten: 0 -- Karte: 6 -- Baum: 0 -- Suche: 0 -- Projekt enthält 131 Dokumente

Blog-Analyse

Soziale Software, wie sie das Web 2.0 propagiert, erweckt zunehmend Aufmerksamkeit in Unternehmen. Der Blog für Mitarbeiter oder Kunden wird als ein leichtgewichtiges Medium für Wissensmanagement geschätzt. Aber Nutzen für das Unternehmen, der über den operativen Wissensaustausch hinausgeht, entsteht erst durch Analyse der Inhalte eines Blogs. Sogenannte Tag-Clouds visualisieren Themenschwerpunkte grob. SWAPit ist als Werkzeug zur Analyse von Blog-Beiträgen vielversprechend: Themen und Treiber lassen sich ebenso detailliert identifizieren und analysieren wie Netze von Mitarbeitern oder häufige Kundenprobleme. Außerdem lassen sich die inhaltlich meist homogenen Beiträge mit TextMining-Methoden besonders gut verarbeiten. In weiteren Arbeiten soll die Eignung von SWAPit für Analysen im Web 2.0-Umfeld untersucht werden. Die Idee wird in [SP+07] dargestellt.

Home

ECOSPACE CWE-News Blog

This Blog reports on interesting news and facts about Collaborative Working Environments.

Actions
[Add entry](#)
[Show all articles](#)

Tagcloud
[ajax](#) [authentication](#)
[blog](#) [bscw](#) [calendar](#)
[cfp](#) [cms](#) [collaboration](#)
[community](#)
[conferenc](#)
[cwe](#) [ecospace](#) [ecscw](#)
[html](#) [ipr](#) [jax-rpc](#) [linux](#)
[netweaver](#) [newsletter](#)
[office](#) [oin](#) [openid](#) [patent](#)
[portal](#) [proceedings](#) [rss](#)
[sid](#) [soa](#) [soap](#) [standard](#)
[usermanagement](#) [vre](#)
[web](#) [web2.0](#) [wiki](#)
[workshop](#) [ws](#) [wsdl](#) [xml](#)
[zoho](#)

User
[amart](#) [Budweg](#) [kjetikr](#)
[marcpallo](#)
[prinz](#) [tkallai](#)
[vonrueden](#)

Dates
[2006-08](#) [2006-09](#)
[2006-10](#)
[2006-11](#)

Freitag, 6. April 2007

Oracle Joins Linux Patent Commons

Steven J. Vaughan-Nichols, Linux-Watch
 Oracle and the Open Invention Network (OIN) announced that Oracle will become an OIN licensee. By doing this, Oracle opens the doors to making some of its paten ...more

18:31 by marcpallot 20 readers  0 comments
 Tags: [oin](#) [patent](#) [linux](#)

Freitag, 6. April 2007

Intellectual Property Rights and Standard Setting

Andrew Updegrove, Consortium Standards Bulletin
 Intellectual property rights (IPR), and particularly patent claims, provide special challenges to standards developers. Following adoption of standard, ...more

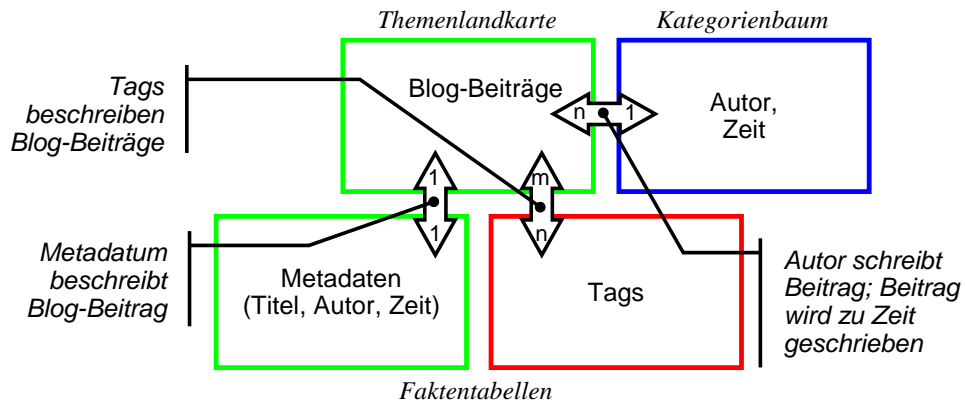
18:17 by marcpallot 20 readers  0 comments
 Tags: [ipr](#) [standard](#)

Freitag, 6. April 2007

Best Practices for Service Interface Design in SOA

Mikhail Genkin, IBM developerWorks
 This article Part 1: Exploring the Development, Interfaces, and Operation Semantics of Services is the first in a series that focuses on best practices for service ...more

18:12 by marcpallot 26 readers  0 comments
 Tags: [soa](#) [xml](#) [sid](#) [soap](#) [jax-rpc](#) [wsdl](#)



The screenshot shows the SWAPit interface with the following components:

- Project:** Unselect, conf, conference conferencing confer, Search, Clear, Search Type, Bool (over Index).
- Map:** A topographic map with red and green markers.
- Category Tree:** A hierarchical tree structure showing nodes like 'author', 'prinz', 'vonrueden', 'time', '2006', '2006-10', '2006-11', '2007', and '2007-01'.
- Metadata Table:** A table with columns X, M, F, T, S, name, and occurrences.

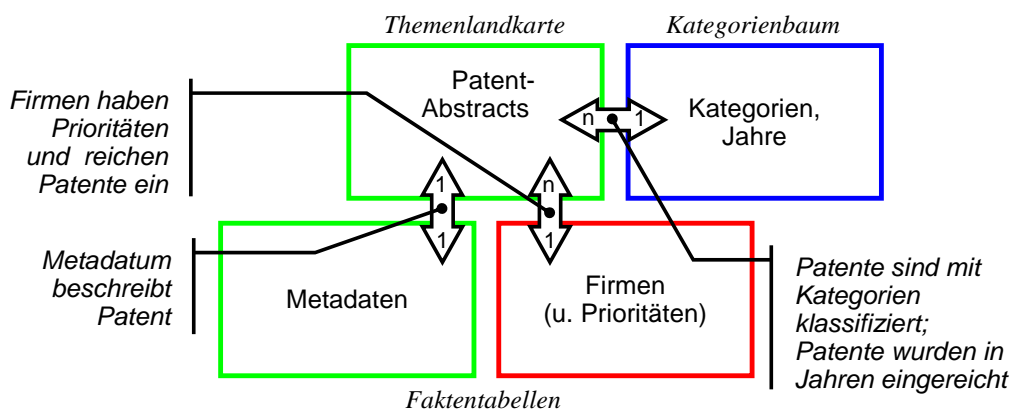
| X | M | F | T | S | name | occurrences |
|---|---|---|---|---|----------------|-------------|
| | | | | | conference | 6 |
| | | | | | cfp | 5 |
| | | | | | web2.0 | 4 |
| | | | | | blog | 3 |
| | | | | | bscw | 2 |
| | | | | | cwe | 2 |
| | | | | | ecospace | 2 |
| | | | | | wiki | 2 |
| | | | | | authentication | 1 |
| | | | | | calendar | 1 |
| | | | | | community | 1 |
| | | | | | ecscw | 1 |
| | | | | | newsletter | 1 |
| | | | | | office | 1 |
| | | | | | openid | 1 |
- Term Statistics Table:** A table showing characteristic terms and their rankings.

| Term | Ranking |
|----------------------|---------|
| suite suited suit | 100% |
| zoho | 68% |
| intel | 62% |
| office offic | 55% |
| wikis wiki | 51% |
| comes coming com | 16% |
| german | 16% |
| informatics informat | 16% |
| symposium | 16% |
| knowledge knowledg | 15% |
| cowmaths cowmath | 10% |

Number of selected documents -- Facts: 4 -- Map: 10 -- Tree: 0 -- Search 0 -- Project contains 41 Documents

Patentanalyse

Die Patentanalyse mit SWAPit zielt darauf ab, vor der Produkteinführung eines Patents im Bereich Energie die Patentsituation zu prüfen, also potentiell kritische Patente zu identifizieren. Prioritätsklassen zur Einstufung von Konkurrenz- bzw. Partnerfirmen waren gegeben. Die Kombination aus Themenkarte und damit verknüpfter Prioritätsliste und Patentklassenkatalog half interaktiv dabei, aus den Ergebnissen einer breiten Suchanfrage (ca. 2000 Patente) schnell große Mengen irrelevanter offensichtlich Patente auszuschließen.



Das Screenshot zeigt die SWAPit-Software-Oberfläche. Die Hauptansicht ist eine Themenlandkarte (Map) mit einer grauen Hintergrundkarte und weißen Punkten. Rechts daneben ist ein Kategorienbaum (Tree) mit einer hierarchischen Struktur von Root über Cat, kM, LM bis hin zu den Jahren von 1985 bis 1996. Unten links ist eine Tabelle mit Metadaten und Firmennamen zu sehen. Unten rechts ist ein Textfenster mit einer Suchanfrage 'remotel' und dem entsprechenden Patenttext.

Die Suchanfrage lautet: **remotel**.
 Der Text des Patents lautet: **Remote** supervising control detecting faults in equipment uses monitor for imitation report operation and sending result to centre for processing.

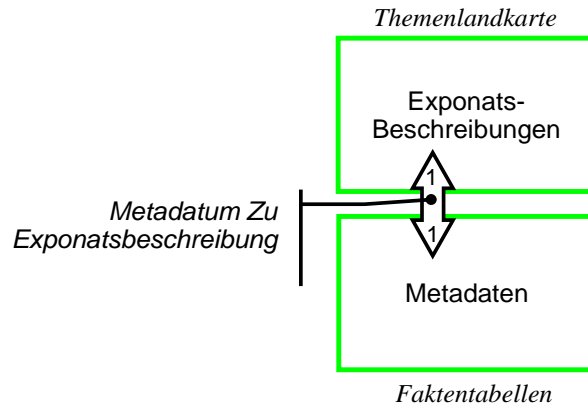
Das Patent beschreibt ein **remote** supervisory control device, das aus einer Superintendenz-Einrichtung (2) besteht, die eine Berichtsoperation ausführt. Es ist mit einem Zentrum (3) verbunden, das die Ergebnisinformationen für die Kommunikationspfadprüfung überträgt. Die Ausrüstung (1) ist mit dem ungewöhnlichen Berichtsformatteil (7) der Superintendenz-Einrichtung verbunden. Wenn es keine Berichts...

Die URL des Patents ist: <http://fit.informatik.rwth-aachen.de:1236/ToolServlets/sql?url?host=fit.inform...>

Die Statusleiste am unteren Rand zeigt: **Number of selected documents -- Facts: 0 -- Map: 0 -- Tree: 0 -- Search: 0 -- Project contains 1174 Documents**

CeBIT 2006: Messepräsentation

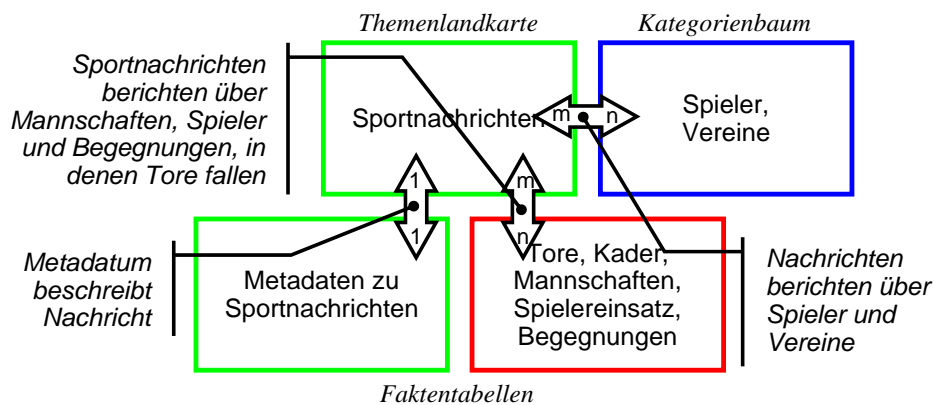
Zur Bewertung von Technologietrends wurden 2006 die Themenschwerpunkte der größten Computermesse CeBIT anhand der Exponatsbeschreibungen untersucht. Ziele waren die Messe-Berichterstattung, sowie die Optimierung der zukünftigen Präsentation der CeBIT-Themen.



| X | M | F | T | titel | url |
|---|---|---|---|---|--|
| | | | | Hören statt lesen - linguatex präsentiert das Vorleseprogramm Voice | http://ft.informatik.rwth-aachen.de:1236/ToolServlet |
| | | | | Das Handy als Nebenstelle im IP-Firmennetzwerk | http://ft.informatik.rwth-aachen.de:1236/ToolServlet |
| | | | | Crealog erstmals mit Interactive Video Response auf der CeBIT 2006 | http://ft.informatik.rwth-aachen.de:1236/ToolServlet |
| | | | | Mensch-Maschine Kommunikation bei dms | http://ft.informatik.rwth-aachen.de:1236/ToolServlet |
| | | | | Massive Markimpulse durch VoIP | http://ft.informatik.rwth-aachen.de:1236/ToolServlet |
| | | | | Digitale Arbeitswelten im modernen Dialogmanagement | http://ft.informatik.rwth-aachen.de:1236/ToolServlet |
| | | | | IP-Telefon für extreme Einsätze - innovaphone und Funke und Huster | http://ft.informatik.rwth-aachen.de:1236/ToolServlet |
| | | | | Das erste Internet-Telefon mit Zugang zu E-Mails und Nachrichten | http://ft.informatik.rwth-aachen.de:1236/ToolServlet |
| | | | | BlueFire 800 - VoIP-ISDN Gateway mit acht ISDN-Schnittstellen | http://ft.informatik.rwth-aachen.de:1236/ToolServlet |
| | | | | AVM FRITZBox: Fon WLAN 7170 | http://ft.informatik.rwth-aachen.de:1236/ToolServlet |
| | | | | AVM VDSL-Endgeräte | http://ft.informatik.rwth-aachen.de:1236/ToolServlet |
| | | | | AVM VoIP-Gateway 5188 - VoIP-Gateway 5144 | http://ft.informatik.rwth-aachen.de:1236/ToolServlet |
| | | | | AVM zeigt Technik-Highlights bei VoIP und Breitband | http://ft.informatik.rwth-aachen.de:1236/ToolServlet |
| | | | | Gewinner Deutscher Innovationspreis 2006 - der Initiative Mittelstand | http://ft.informatik.rwth-aachen.de:1236/ToolServlet |
| | | | | IT-Schulung Mandelstam und Prosser auf dem CeBIT 2006 | http://ft.informatik.rwth-aachen.de:1236/ToolServlet |

CeBIT 2006: Fußballnachrichten

Zur Präsentation von SWAPit auf der CeBIT 2006 wurden im Jahr der Fußballweltmeisterschaft in Deutschland ca. 500 Sportnachrichten sowie Daten zu Mannschaften, Spielern und Begegnungen mit den Methoden aus Kap.6.7 weitgehend automatisch verknüpft. Fragestellungen betrafen die Berichterstattung zu ausgewählten Kaderspielern und Begegnungen, die Kernthemen der Sportnachrichten, sowie die Recherche nach Hintergrundinformationen zu Begegnungen.



Das Screenshot zeigt die SWAPit-Software zur Darstellung von Fußballnachrichten und Faktentabellen. Die Software ist in mehrere Bereiche unterteilt:

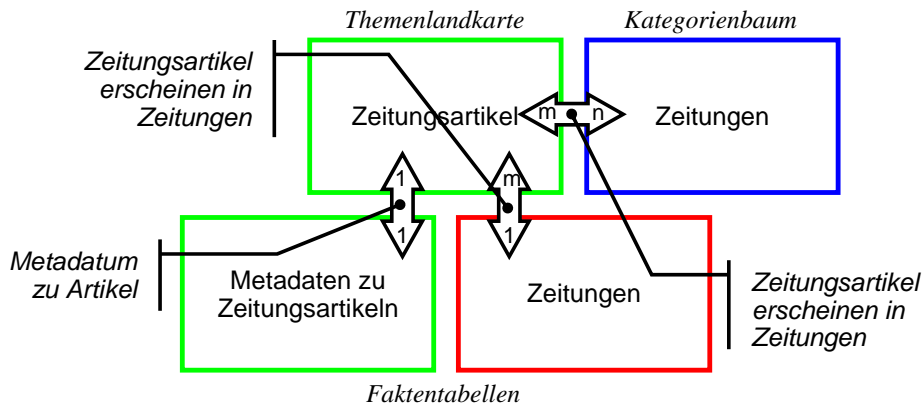
- Projekt:** "kahn's kahn".
- Themenlandkarte:** Ein 2D-Plot, der die Themen der Nachrichten darstellt.
- Kategorienbaum:** Ein Baumdiagramm, das die Kategorien "Spieler" und "Vereine" zeigt. Die Kategorien "Spieler" sind weiter unterteilt in "Abwehr", "Mittelfeld" und "Tor". Die Kategorie "Vereine" listet verschiedene Fußballvereine auf, wie "1. FC Kaiserslautern", "1. FC Köln", "Arminia Bielefeld", "Bayer Leverkusen", "Bayern München", "Borussia Mönchengladbach", "FC Schalke 04", "Hannover 96" und "VfB Stuttgart".
- Metadaten News:** Eine Tabelle, die die Metadaten der Nachrichten darstellt. Die Spalten sind "Datum", "Spieltag", "Verein", "Tore" und "Punkte".
- Metadaten Spielberichte:** Eine Tabelle, die die Metadaten der Spielberichte darstellt. Die Spalten sind "Kader", "Spielereinsatz" und "Mannschaften".
- Begegnungen:** Eine Tabelle, die die Begegnungen darstellt. Die Spalten sind "Datum", "Spieltag", "Verein", "Tore" und "Punkte".
- Term Statistics:** Eine Tabelle, die die Term Statistics darstellt. Die Spalten sind "Datum", "Spieltag", "Verein", "Tore" und "Punkte".
- Document Selection:** Eine Tabelle, die die Document Selection darstellt. Die Spalten sind "Datum", "Spieltag", "Verein", "Tore" und "Punkte".
- Info:** Eine Tabelle, die die Info darstellt. Die Spalten sind "Datum", "Spieltag", "Verein", "Tore" und "Punkte".
- Facts Filtering:** Eine Tabelle, die die Facts Filtering darstellt. Die Spalten sind "Datum", "Spieltag", "Verein", "Tore" und "Punkte".
- Facts Statistics:** Eine Tabelle, die die Facts Statistics darstellt. Die Spalten sind "Datum", "Spieltag", "Verein", "Tore" und "Punkte".

Die Tabelle "Metadaten News" zeigt folgende Daten:

| X | M | F | T | Datum | Spieltag | Verein | Tore | Punkte |
|---|---|---|---|------------|----------|--------------------------|------|--------|
| | | | | 07.08.2005 | 1 | Bayer Leverkusen | 4 | 3 |
| | | | | 05.08.2005 | 1 | Bayern München | 3 | 3 |
| | | | | 07.08.2005 | 1 | FC Schalke 04 | 2 | 3 |
| | | | | 13.08.2005 | 2 | Bayer Leverkusen | 2 | 0 |
| | | | | 13.08.2005 | 2 | Bayern München | 5 | 3 |
| | | | | 13.08.2005 | 2 | Borussia Dortmund | 1 | 0 |
| | | | | 13.08.2005 | 2 | Borussia Mönchengladbach | 1 | 1 |
| | | | | 13.08.2005 | 2 | FC Schalke 04 | 2 | 3 |
| | | | | 13.08.2005 | 2 | Hannover 96 | 1 | 1 |
| | | | | 14.08.2005 | 2 | VfB Stuttgart | 2 | 0 |
| | | | | 27.08.2005 | 3 | Bayern München | 3 | 3 |
| | | | | 10.09.2005 | 4 | Bayern München | 2 | 3 |
| | | | | 10.09.2005 | 4 | Werder Bremen | 5 | 3 |
| | | | | 20.09.2005 | 6 | Bayern München | 1 | 3 |
| | | | | 24.09.2005 | 7 | Bayern München | 0 | 0 |
| | | | | 01.10.2005 | 8 | Bayern München | 2 | 3 |
| | | | | 02.10.2005 | 8 | Borussia Dortmund | 0 | 1 |
| | | | | 02.10.2005 | 8 | VfB Stuttgart | 0 | 1 |
| | | | | 01.10.2005 | 8 | VfL Wolfsburg | 0 | 0 |
| | | | | 15.10.2005 | 9 | Borussia Dortmund | 3 | 1 |
| | | | | 29.10.2005 | 11 | Bayern München | 2 | 3 |
| | | | | 06.11.2005 | 12 | VfB Stuttgart | 1 | 3 |
| | | | | 19.11.2005 | 13 | Arminia Bielefeld | 1 | 0 |

Presseclipping

Bei vielen großen Unternehmen werden Zeitungsartikel zum Unternehmen zentral gesammelt, ausgewertet, aufbereitet und als Pressebericht verteilt. Für die Fraunhofer Gesellschaft wurden solche Presseclippings mit SWAPit analysiert und zugreifbar gemacht.



The screenshot shows the SWAPit - roetzer - PresseClipping software interface. The interface is divided into several sections:

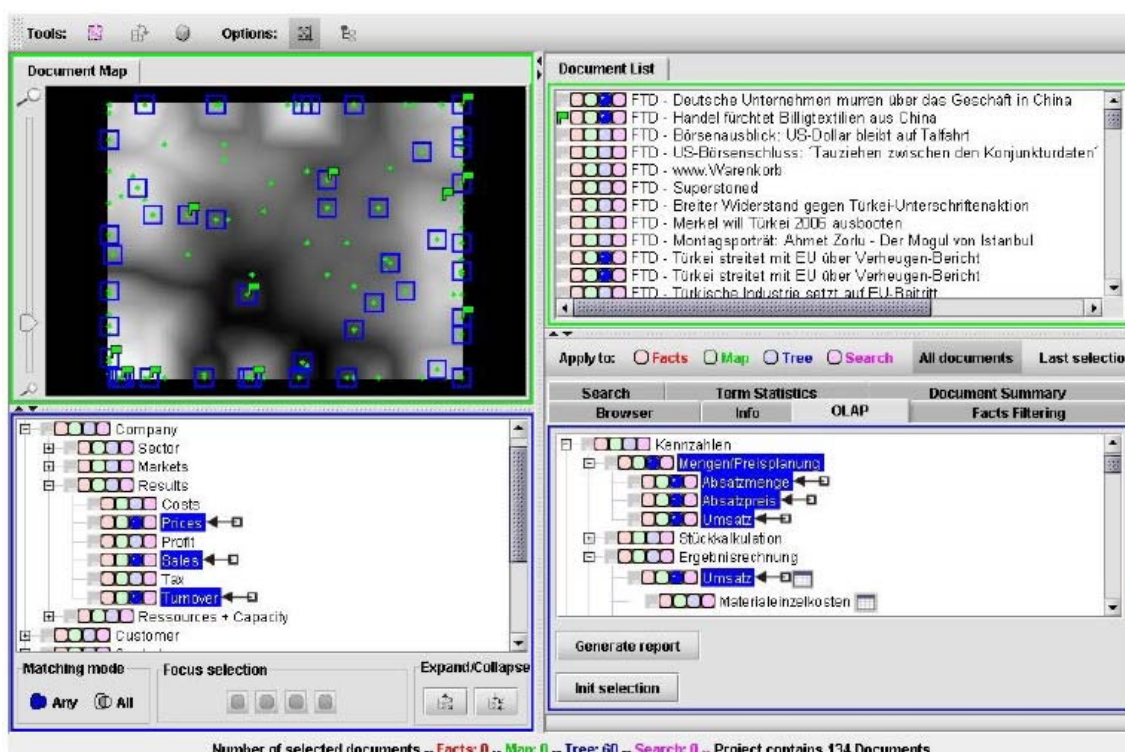
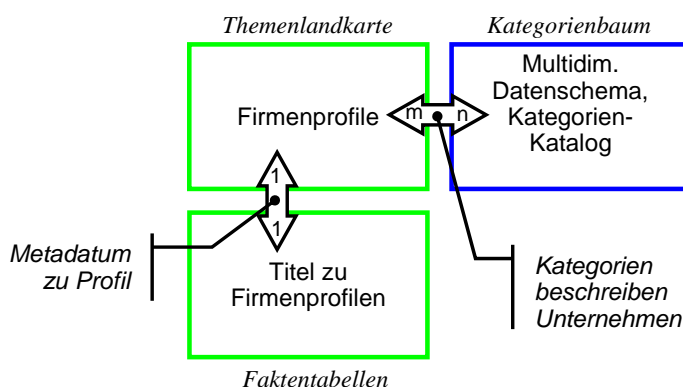
- Project:** Shows the current project name and search options.
- Focus On:** A panel for navigating through the data, showing a tree view of 'Zeitung' (Newspaper) categories.
- Metadaten Artikel:** A table listing newspaper articles with columns for 'Titel', 'Datum', and 'Zeitung'.
- Document Selection:** A panel for selecting documents to view.
- Viewer:** A window displaying the content of a selected document, showing a snippet of text about 'Fraunhofer-Gesellschaft'.

| Metadaten Artikel | Zeitungen(*) | Titel | Datum | Zeitung |
|-------------------|--------------|---|------------|------------------------|
| | | MP3-Erlöse als Stiftungskapital | 2006-05-15 | wiwo.de |
| | | Gesellschaft will Stiftung gründen | 2006-05-15 | Handelsblatt.com |
| | | Fraunhofer-Gesellschaft will Stiftung gründen | 2006-05-16 | Tagesspiegel online |
| | | Allein 2005 100 Millionen Euro eingenommen | 2006-05-16 | FZDFheute.de |
| | | Fraunhofer-Gesellschaft gründet mit MP3-Lizenz-Erlösen eine Stit... | 2006-05-16 | Münchner Merkur |
| | | Fraunhofer-Gesellschaft will mit MP3-Lizenz-Erlösen Stiftung gründ... | 2006-05-15 | dpa |
| | | Fraunhofer-Gesellschaft will mit MP3-Lizenz-Erlösen Stiftung gründ... | 2006-05-15 | dpa |
| | | Brennstoffzellen: Neue Materialien für hohe Temperaturen | 2006-05-23 | gfp |
| | | Eis auf Lupinen-Basis | 2006-05-15 | ntr online |
| | | Eis ohne Milchzucker | 2006-05-16 | Bietheimer Zeitu... |
| | | Statt Milchzucker | 2006-05-16 | Hamburger Abend... |
| | | Fraunhofer-Institut entwickelt Hochtemperatur-Brennstoffzellen | 2006-05-16 | IWR |
| | | Neues Stack-Design aus Keramikplatten sorgt für langes Leben | 2006-05-16 | ZDNetNews |
| | | Lizenzen für MP3 brachten 100 Mio. | 2006-05-16 | westfälische Rund... |
| | | Eis essen trotz Laktose-Allergie | 2006-05-16 | Pforzheimer Zeitung |
| | | Milch-Eis trotz Laktose-Unverträglichkeit | 2006-05-16 | Westdeutsche Zeit... |
| | | Speiseeis auf rein pflanzlicher Basis | 2006-05-16 | Nordsee-Zeitung |
| | | Keramik bringt Brennstoffzelle voran | 2006-05-27 | Leipziger Volkszeit... |
| | | Statt Milchzucker Speiseeis aus Lupinensamen | 2006-05-16 | Hamburger Abend... |
| | | - | 2006-05-28 | Euro am Sonntag |
| | | 100 Millionen aus MP3-Erlösen | 2006-05-16 | Weser-Kurier |
| | | Stiftung dank MP3-Erlösen | 2006-05-16 | Münchner Merkur |
| | | Millionen aus MP3-Erlösen | 2006-05-16 | Delmenhorster Kr... |
| | | MP3 | 2006-05-16 | Norddeutscher K... |

Controlling und Planung in der Textilindustrie (EU-Projekt SEWASIE)

Im Europäischen Forschungsprojekt SEWASIE⁸⁹ wurde untersucht, wie Multidimensionale Datenmodelle für Unternehmenskennzahlen und Kataloge zur Textklassifikation integriert dargestellt und exploriert werden können, um externe Hintergrundinformationen zu OLAP-Kennzahlenberichten abrufen und kontextbezogen zugreifen zu können. Die Kernherausforderung bestand in der Entwicklung eines Darstellungs- und Navigationskonzeptes, das den OLAP-Analysten in die Lage versetzt, die annotierten Textdokumente auf Basis der ihm vertrauten Kategorien zu explorieren. Ein europäisches Patent zum Verfahren der Annotation von OLAP-Berichten wurde für Europa erfolgreich angemeldet. Die folgenden Fragestellungen standen bei SEWASIE im Vordergrund:

- Wie kann die Exploration strukturierter Daten bei der interaktiven Textanalyse helfen?
- Welches Potential hat die Nutzung von Kundenbeschwerden für die Stornoprognose?
- Wie lassen sich Taxonomien von Beschreibungskategorien und hierarchische OLAP-Datenmodelle darstellen und interaktiv zur Navigation in der Textsammlung nutzbar machen?



⁸⁹ SEmantic Webs and AgentS in Integrated Economies, 5. Rahmenprogramm, IST-5-34825.

9.4 Veröffentlichen zur Arbeit

9.4.1 Wissenschaftliche Publikationen

| Autoren, Jahr | Titel, Tagung |
|--|--|
| Seeling, Hirschmeier, Becks 2007 | Acceptance and Usefulness of Explorative Search Systems in Business Contexts IEEE InfoVis, Sacramento, CA, USA (eingereicht) |
| Seeling, Prinz, Ruland, Becks 2007 | Kooperatives Arbeiten unter der Lupe 7. Fachübergreifende Konferenz Mensch & Computer (MC) |
| Seeling, Prinz, Becks 2007 | Providing Awareness in Complex Cooperation Networks IADIS Int. Conf. on Web Based Communities (WBC07) |
| Seeling, Prinz, Becks 2007 | Transparency via Activity Visualization in Professional Cooperation Environments 4. Konferenz Professionelles Wissensmanagement, Workshop on Collaborative Knowledge Management (CoKM) |
| Seeling, Becks 2005 | Semantische Interaktionspfade: Ein MultiView System für betriebswirtschaftliche Entscheidungsprobleme 35. Jahrestagung der Gesellschaft für Informatik (GI) |
| Becks, Huster, Jarke, Jertila, Kensche, Quix, Seeling 2005 | Value-Added Services enabling Semantic Web Technologies for SMEs International Semantic Web Conference (ISWC) |
| Seeling, Spaniol, Becks, Klamma 2004 | Discourse Visualization Strategies for a Comprehensive Medial Analysis of Cultural Science Communities; 4th International Conference on Knowledge Management (I-KNOW) |
| Becks, Seeling 2004 | SWAPit - A Multiple Views Paradigm for Exploring Associations of Texts and Structured Data; 9th Int. Conf. on Advanced Visual Interfaces (AVI) |
| Seeling, Becks 2004 | Analysing Associations of Textual and Relational Data with a Multiple Views System 2nd International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV) |
| Seeling, Becks 2003 | Exploiting Metadata for Ontology-Based Visual Exploration of Weakly Structured Text Documents; 7th Int. Conf. on Inf. Visualisation (IV) |
| Becks, Jarke, Seeling 2003 | Integration von Geschäftskennzahlen und Textdokumenten mit Semantic Web Technologien; Zeitschrift für Künstliche Intelligenz (KI) |
| Becks, Seeling, Minkenberg 2002 | Benefits of Document Maps for Text Access in Knowledge Management: A Comparative Study; ACM Symposium on Applied Computing (SAC) |
| Becks, Seeling 2001 | A Task-Model for Text Corpus Analysis in Knowledge Management 8th International Conference on User Modeling (UM), Workshop on User Modeling, Machine Learning and Information Retrieval |

9.4.2 Pressespiegel

- *Ein Röntgenauge für Textarchive*, Aachener Zeitung, 2002
- *Computing der Zukunft – Die neue Art der Informationssuche*. Chip, Mai 2006
- *Warum die Torhüter jeden Ball kassieren: Softwareprogramm kann eigenständig Zusammenhänge aus Vielzahl von Daten erkennen*. Rhein-Sieg-Anzeiger, 21.2.2006
- *Software erkennt die Käuflichkeit von Ballack & Co. – Text- und Data Mining für Fans*. 27.01.2006, silicon.de
- *Dokumente vielschichtig durchleuchten*. Pressemitteilung, 26. Januar 2006; erschienen u.a. in: idw – Informationsdienst Wissenschaft, Innovationsreport – Forum für Wissenschaft, Industrie und Wirtschaft⁹⁰
- *Visual Text Mining: non è una magia*. Forum P.A., l'altra p.a. – percorsi di innovazione nell'amministrazione che cambia (Magazin der italienischen öffentlichen Verwaltung), Januar 2006.

⁹⁰ www.innovations-report.de

9.5 Fragebogen zur Darstellung von Matching-Werten

Darstellung O zeigt für jeden Dokumentenpunkt und jede Farbe an, ob der Punkt in der Farbe ausgewählt ist (binär). Die Darstellungen A, B, C zeigen zusätzlich an, zu welchem Grad ein Dokumentenpunkt in einer Farbe ausgewählt ist (Abstufung).

I. Bitte vergleiche Darstellung O mit den anderen Darstellungen A, B, C

Betrachte das **Bild als Ganzes**. Ist der Überblick, welche Gruppen von Punkten mit einer Farbe eingefärbt sind, noch genauso deutlich wie in Darst. O?

| | schlechter als O | gleich gut wie O | besser als O |
|---|------------------|------------------|--------------|
| A | | | |
| B | | | |
| C | | | |

Nun schaue Dir bitte **einzelne Punkte** an. Kann die Information, ob ein Punkt in einer Farbe ausgewählt ist, noch mindestens genauso gut wahrgenommen werden wie in Darst. O?

| | schlechter als O | gleich gut wie O | besser als O |
|---|------------------|------------------|--------------|
| A | | | |
| B | | | |
| C | | | |

Empfindest Du es als **anstrengender**, eine der Darstellungen A, B, C zu betrachten als die binäre Darstellung O?

| | anstrengender O | genauso wie O | weniger anstrengend als O |
|---|-----------------|---------------|---------------------------|
| A | | | |
| B | | | |
| C | | | |

Jede Farbe stellt ein anderes Suchergebnis dar. Dabei zeigt die Abstufung einer Farbe die Relevanz eines Dokumentes für das Suchergebnis an. Wie hoch schätzt Du den **Nutzen einer solchen Abstufung** sich ein (wie in den Darst. A, B, C)?

| unwichtig | Egal | nützlich |
|-----------|------|----------|
| | | |

II. Bitte vergleiche im folgenden nur die Darstellungen A, B, C (direkter Vergleich). Als Antwort auf die Fragen bringe bitte die Darstellungen A, B, C in eine Reihenfolge und trage die Buchstaben in das jeweilige Kästchen ein.

Wo ist die **Übersicht** über Einfärbungen von Gruppen von Dokumentenpunkten am besten zu erkennen?

übersichtlich - ____ - unübersichtlich

Welche Darstellung ist **am einfachsten** zu verstehen?

einfach - ____ - schwierig

Wo lassen sich die **Farben** am besten **unterscheiden**?

Gut - ____ - schlecht

Wo kannst Du am besten im **Detail** erkennen, welcher Dokumentenpunkt mit einer Farbe eingefärbt ist?

gut - ____ - schlecht

Wo kannst Du am besten im Detail erkennen, welcher Dokumentenpunkt mit einer Farbe **wie stark** eingefärbt ist (Abstufungen)?

gut - ____ - schlecht

Welche Darstellung wirkt **am ruhigsten**?

ruhig - ____ - unruhig

Welche Darstellung **spricht Dich am stärksten an**?

ansprechend - ____ - weniger ansprechend

9.6 Positivlisten zur Fallstudie 1 (Beschwerdemgmt.)

| | | |
|------------------------|----------------------|---------------------|
| Ablehnungsbescheid | geduldig | Schaden |
| Anwalt | Gerichtsverfahren | Schäden |
| Anwälte | Geschäftsgebahren | Schadenersatz |
| aufgehoben | hochgestuft | Schadenersatzklage |
| aufheben | Hochstufung | Schadenersatzklagen |
| Aufhebung | Ignoranz | Schwierigkeit |
| Aufsichtsamt | Jahre | Schwierigkeiten |
| Aufsichtsbehörde | Jahren | storniere |
| Beitrag | juristisch | stornieren |
| Beiträge | juristische | treu |
| Beitragserhöhung | juristischen | treue |
| Beitragsrechnung | Klage | treuem |
| Beitragsrechnungen | Klagen | treuen |
| Beitragsrückerstattung | Korrespondenz | treuer |
| Beitragssatz | Kulanz | Unverschämtheit |
| Beitragssätze | Kulanzzahlung | unzureichend |
| Beschwerde | kundenfreundlich | verärgert |
| Beschwerden | Kundenfreundlichkeit | Verärgerung |
| beschwere | kündige | Verzögerung |
| beschweren | kündigen | Verzögerungen |
| beschwert | Kündigung | Verzögerungstaktik |
| beschwerte | Kündigungen | Widerruf |
| beschwerten | langjährig | widerrufe |
| betrogen | langjährige | widerrufen |
| Betrug | langjähriger | |
| betrügen | Mahnbescheid | |
| Bundesanstalt | Mahnung | |
| Bundesaufsichtsamt | Mahnungen | |
| erwarte | Manipulation | |
| erwarten | Manipulationen | |
| Erwartung | negativ | |
| fordere | negative | |
| fordern | negativer | |
| Forderung | negatives | |
| Forderungen | Niederlassung | |
| Geduld | Regulierung | |

9.7 Fragebögen zur Fallstudie 2 (Produktsupport)

Fragebogen zur Meldungsbearbeitung

Zeitaufwand: ca. 12 Minuten

Nach welchen Kategorien hältst Du es i.A. für sinnvoll, ein Meldungssuchergebnis weiter zu selektieren? Die Skala geht von 1 (sehr nützlich) bis 6 (unbrauchbar).

| | | | | | | | |
|--|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|-------------------------------------|
| Zeit (Jahr, Monat, Tag) | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |
| Release, Support Package | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |
| Verwandte Komponenten | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |
| Priorität | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |
| Lösungsinstanz (IMS, Primary Support, Bearbeiter) | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |
| Fehlerkategorie (Codeerror, Consulting, Performance) | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |
| Meldungsstatus (Autom. geschlossen oder quittiert) | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |

Kommentar: _____

Welche der folgenden Optionen nutzt Du bei der Notesearch?

Die Skala geht von 1 (sehr oft) bis 6 (gar nicht)

| | | | | | | | |
|--|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|-------------------------------------|
| Die Selektionsmöglichkeit „Application Area“ bei „Restricted by Attributes“? | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |
| Die Selektionsmöglichkeit “More terms” bei “Restricted by Attributes”? | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |
| Die Selektionsmöglichkeit “Priority” bei “Restricted by Attributes”? | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |
| Die Selektionsmöglichkeit “Category” bei “Restricted by Attributes”? | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |
| Das Kontextmenü “Find similar notes”? | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |

Kommentar: _____

Welche Spalten benötigst Du bei der Anzeige von Suchergebnissen?

Die Skala geht von 1 (benötige ich häufig) bis 6 (uninteressant).

| | | | | | | | |
|------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|-------------------------------------|
| Meldungsnummer | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |
| Short Text | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |
| Sprache | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |
| Jahr | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |
| Status | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |
| Letzte Änderung | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |
| Version | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |
| Application Area | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |
| Processor | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 5 | <input type="checkbox"/> 6 | <input type="checkbox"/> weiß nicht |

Kommentar: _____

Bei welchen Fehlerkategorien glaubst Du, dass eine Lösung eher in abgeschlossenen Meldungen als in Notes liegt?

| | eher Meldungen | | | | | eher Notes |
|--------------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Program Error | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Consulting | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Customizing | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| AdvanceDevelopment | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Special Development | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Documentation Error | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Translation Error | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Installation Information | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| FAQ | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Performance | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Release Planning Information | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Upgrade Information | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Workaround for Missing Functionality | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| External Error | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Help for External Errors | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Wenn Du eine abgeschlossene Meldung gefunden hast, die recht lang ist, wie findest Du schnell die Essenz des Textes heraus? (Mehrfachnennung möglich)

- Ich lese ausschließlich den Short Text.
- Ich lese vom LongText nur die erste und die letzte Seite.
- Ich suche nach „Vielen Dank“, „Thank you“, um an wichtige Textbausteine zu gelangen.
- Ich suche nach „helpful“ oder „hilfreich“.
- Ich lese insbesondere die internen Memos.
- Ich lese insbesondere die Replies.
- Ich überfliege die Meldung komplett.
- Eigene Strategie: _____

Welche Textanalysemöglichkeiten fändest Du bei der Lösungssuche interessant?

Die Skala geht von 1 (würde mir sehr gefallen) bis 6 (bräuchte ich gar nicht)

Autozusammenfassung einer einzelnen Meldung

1 2 3 4 5 6 weiß nicht

Anzeige der ähnlichsten Meldung(en), die schon in der Datenbank ist/sind.

1 2 3 4 5 6 weiß nicht

Anzeige, worin diese Ähnlichkeit besteht (z.B. in Form von ähnlichsten Worten oder Sätzen)

1 2 3 4 5 6 weiß nicht

Vorschläge für geeignete Suchworte

1 2 3 4 5 6 weiß nicht

Andere: _____

Zeiteinsparung

Ab _____ % durchschnittlicher Zeiteinsparung würde ich die Benutzung eines Text Mining Tools zur Meldungsbearbeitung schon als „sehr gut“ erachten.

Bis _____ % durchschnittlicher Zeiteinsparung würde ich die Benutzung eines solchen Tools noch als „befriedigend“ erachten.

Unter _____ % durchschnittlicher Zeiteinsparung würde ich die Benutzung als „ungenügend“ einstufen.

Dauer der Vergleichsanfrage

Bitte gib nun deine Erwartungshaltung bezüglich der Dauer an, die vergeht, um einen von Dir gewählten Textabschnitt mit anderen Meldungen aus der Datenbank zu vergleichen.

Unter _____ Sekunden würde ich den Vergleich schon als sehr gut einschätzen.

Bis _____ Sekunden wäre die Vergleichsanfrage grade noch „befriedigend“.

Ab _____ Sekunden würde ich die Anfragedauer mit „ungenügend“ bewerten und gar nicht erst benutzen wollen.

Offene Fragen zur Hinweissuche

Wie gehst Du üblicherweise bei der Hinweissuche vor?

In wie viel Prozent der Fälle führt eine Hinweissuche zur Lösung (geschätzt)? _____%

In wie viel Prozent führt schon das erste Suchergebnis nach Notes direkt zur Lösung? _____%

Wie lange brauchst Du für die Hinweissuche

Im besten Fall _____ Minuten

Durchschnittlich _____ Minuten

Im schlimmsten Fall _____ Minuten

Offene Fragen zur Meldungssuche

Wie gehst du üblicherweise bei der Meldungssuche vor?

Wenn eine Hinweissuche nicht zum Erfolg führt, in wie viel % der Fälle hältst Du dann eine Meldungssuche noch für sinnvoll und führst sie dann auch durch? _____

In wie viel % führt das erste Suchergebnis nach Meldungen direkt zur Lösung? _____

Wie lange brauchst Du für die Meldungssuche?

Im besten Fall _____ Minuten

Durchschnittlich _____ Minuten

Im schlimmsten Fall _____ Minuten

Allgemeine Kommentare, Ideen, Anregungen, etc.

Bewertungsbogen SwapIt

Datum: _____ Name: _____

Benutzte Komponenten: RT QDEF WB WEB DBIF

- Wie brauchbar fandest Du die Suchfunktionen? 1 2 3 4 5 6
(1 bedeutet „sehrbrauchbar“, 6 bedeutet „unbrauchbar“)
- Wie bewertest du die Textanalysefunktionen (Term Statistics, Document Summary)? 1 2 3 4 5 6
(1 bedeutet „sehr nützlich“, 6 bedeutet „unnützlich“)
- Wie sinnvoll waren ähnliche Meldungen gruppiert (Cluster bzgl. Themen und Sprache)? 1 2 3 4 5 6
(1 bedeutet „sinnvolle Gruppierung“, 6 bedeutet „keine sinnvolle Gruppierung“)
- Wie nützlich fandest du die Selektions- und Filterfunktionen /Einschränkung nach Attributen? 1 2 3 4 5 6
(1 bedeutet „sehr nützlich“, 6 bedeutet „unbrauchbar“)
- Wie bewertest du die Visualisierung der Oberfläche? 1 2 3 4 5 6
(1 bedeutet „sehr gut“, 6 bedeutet „sehr schlecht“)
- Wie gefiel Dir die Navigation? 1 2 3 4 5 6
(1 bedeutet „gefiel mir sehr gut“, 6 bedeutet „gefiel mir gar nicht“)
- Wie gut konntest Du dich beim Erneuten benutzen wieder zurechtfinden? 1 2 3 4 5 6
(1 bedeutet „sofort wieder zurecht gefunden“, 6 bedeutet „musste mich erst neu orientieren“)
- Ist das Tool intuitiv zu bedienen? 1 2 3 4 5 6
(1 bedeutet „sehr intuitiv“, 6 bedeutet „sehr hohen Einarbeitungsaufwand“)
- Hattest du den Eindruck, dass die von dir angeschauten Dokumente oft relevant waren? 1 2 3 4 5 6
(1 bedeutet „ich habe fast nur hochwertige Dokumente geöffnet“, 6 bedeutet „ich habe viel Mist anschauen müssen“)
- Wie viele der Dokumente, die dir wirklich weitergeholfen hätten, glaubst du, hast du gesehen? 1 2 3 4 5 6
(1 bedeutet „ich glaube alle“, 6 bedeutet „vermutlich nur einen kleinen Teil“)
- Wie schnell bist du mit dem Tool zu ersten Treffern gelangt? 1 2 3 4 5 6
(1 bedeutet „angenehm schnell“, 6 bedeutet „relevante Treffer fand ich erst sehr spät“)
- Wie oft hast du eine Lösung finden können? 1 2 3 4 5 6

(1 bedeutet „Suche sehr oft erfolgreich“, 6 bedeutet „leider nur selten eine Lösung gefunden“)

Hast du viele relevante zusätzliche Meldungen in der Karte finden können?

1 2 3 4 5 6

(1 bedeutet „Ja ich bin oft auf weiterführende Aspekte gestoßen“, 6 bedeutet „dadurch bin ich meistens nicht weiter gekommen“)

Wie warst du mit der Suchzeit im Allgemeinen zufrieden?

1 2 3 4 5 6

(1 bedeutet „war schnell genug“, 6 bedeutet „hat viel zu lange gedauert“)

Wie warst du mit der Bootzeit zufrieden, bis das Tool startbereit ist?

1 2 3 4 5 6

(1 bedeutet „fand ich völlig in ordnung“, 6 bedeutet „verhindert effizienten Arbeiten“)

Wie zuverlässig war das Tool?

1 2 3 4 5 6

(1 bedeutet „lief tadellos“, 6 bedeutet „hing oft sehr lange oder fiel komplett aus“)

Hast du viele Schritte in der Suche unternehmen müssen?

1 2 3 4 5 6

(1 bedeutet „Wenige Schritte reichten aus“, 6 bedeutet „ich musste sehr viel klicken“)

Kommentar:

Erhebungsboden Suchmaschine

Datum: _____ Name: _____

Benutzte Komponenten: RT QDEF WB WEB DBIF

- Wie brauchbar fandest Du die Suchfunktionen? 1 2 3 4 5 6
(1 bedeutet sehr gut, 6 bedeutet unbrauchbar)
- Wie geeignet fandest du die Selektionsfunktionen
/Möglichkeit zur Einschränkung nach Attributen? 1 2 3 4 5 6
(1 bedeutet sehr geeignet, 6 bedeutet unbrauchbar)
- Wie bewertest du die Gestaltung und Anordnung der
Oberfläche? 1 2 3 4 5 6
(1 bedeutet sehr geeignet, 6 bedeutet unbrauchbar)
- Wie gefällt Dir die Menüführung? 1 2 3 4 5 6
(1 bedeutet „war sehr einfach für mich“, 6 bedeutet „fiel mir sehr schwer“)
- Wie gut kannst Du dich beim Erneuten benutzen wieder
zurechtfinden? 1 2 3 4 5 6
(1 bedeutet „sofort wieder zurecht gefunden“, 6 bedeutet „musste mich erst neu orientieren“)
- Ist das Tool intuitiv zu bedienen? 1 2 3 4 5 6
(1 bedeutet „sehr intuitiv“, 6 bedeutet „sehr hohen Einarbeitungsaufwand“)

Kommentar:

Questionnaire „Task 1“

[SWAPit on top of BSCW] vs. [BSCW]

| | very bad/little | | | | very good/much |
|---|-----------------|----|---|----|----------------|
| I believe that the overview of the workpackages' topics I got using SWAPit (on top of BSCW) is ... | -2 | -1 | 0 | +1 | +2 |
| I believe that the overview of the workpackages' topics I could get using BSCW is ... | -2 | -1 | 0 | +1 | +2 |
| How confident are you that you found all important topics using SWAPit (on top of BSCW)? | -2 | -1 | 0 | +1 | +2 |
| How confident would you be using BSCW? | -2 | -1 | 0 | +1 | +2 |
| How efficiently (time/activities) could you solve this task using SWAPit (on top of BSCW)? | -2 | -1 | 0 | +1 | +2 |
| How efficiently would you solve it using BSCW? | -2 | -1 | 0 | +1 | +2 |
| How stressless did you find the work with SWAPit (on top of BSCW) for this task? | -2 | -1 | 0 | +1 | +2 |
| How stressless would you find the work using BSCW? | -2 | -1 | 0 | +1 | +2 |
| How well did you know how to use the BSCW+SWAPit for solving this task? | -2 | -1 | 0 | +1 | +2 |
| How well would you know how to use BSCW? | -2 | -1 | 0 | +1 | +2 |
| How much would you like to use SWAPit (on top of BSCW) in the future for tasks of this kind? | -2 | -1 | 0 | +1 | +2 |
| How much would you like to use BSCW in the future for tasks of this kind? | -2 | -1 | 0 | +1 | +2 |
| Did introduction, practice & task description prepare you well for using SWAPit (on top of BSCW) for this task? | -2 | -1 | 0 | +1 | +2 |
| How well are you prepared for estimating how to use BSCW for this task? | -2 | -1 | 0 | +1 | +2 |
| How appropriate was the access to information (overviews) of SWAPit (on top of BSCW) for this task? | -2 | -1 | 0 | +1 | +2 |
| How appropriate is the access to information of BSCW for this task? | -2 | -1 | 0 | +1 | +2 |
| Did SWAPit (on top of BSCW) offer all desired interaction facilities / functionalities for efficient solution of this task? | -2 | -1 | 0 | +1 | +2 |
| Does BSCW offer all desired interaction facilities / functionalities for efficient solution of this task? | -2 | -1 | 0 | +1 | +2 |

| | | | | | |
|--|----|----|---|----|----|
| How useful was SWAPit's analysis of document contents for this task? | -2 | -1 | 0 | +1 | +2 |
|--|----|----|---|----|----|

| | | | | | |
|---|----|----|---|----|----|
| How useful did you find SWAPit's concept of multiple coordinated views for this task? | -2 | -1 | 0 | +1 | +2 |
|---|----|----|---|----|----|

| |
|--|
| <p><u>Comments</u> (What was particularly useful? What did you miss? ...)</p> |
|--|

Questionnaire „Task 2“

[SWAPit on top of BSCW] vs. [BSCW]

| | | | | | |
|-----------------|--|--|--|--|----------------|
| very bad/little | | | | | very good/much |
|-----------------|--|--|--|--|----------------|

| | | | | | |
|--|----|----|---|----|----|
| I believe that the overview of relationships between persons and topics I got using SWAPit (on top of BSCW) is ... | -2 | -1 | 0 | +1 | +2 |
| I believe that the overview of relationships between persons and topics I could get using BSCW is ... | -2 | -1 | 0 | +1 | +2 |

| | | | | | |
|---|----|----|---|----|----|
| How confident are you that you found all important persons for a topic using BSCW+SWAPit? | -2 | -1 | 0 | +1 | +2 |
| How confident would you be using BSCW? | -2 | -1 | 0 | +1 | +2 |

| | | | | | |
|--|----|----|---|----|----|
| How efficiently (time/activities) could you solve this task using SWAPit (on top of BSCW)? | -2 | -1 | 0 | +1 | +2 |
| How efficiently would you solve it using BSCW? | -2 | -1 | 0 | +1 | +2 |

| | | | | | |
|--|----|----|---|----|----|
| How stressless did you find the work with SWAPit (on top of BSCW) for this task? | -2 | -1 | 0 | +1 | +2 |
| How stressless would you find the work using BSCW? | -2 | -1 | 0 | +1 | +2 |

| | | | | | |
|---|----|----|---|----|----|
| How well did you know how to use the BSCW+SWAPit for solving this task? | -2 | -1 | 0 | +1 | +2 |
| How well would you know how to use BSCW? | -2 | -1 | 0 | +1 | +2 |

| | | | | | |
|--|----|----|---|----|----|
| How much would you like to use SWAPit (on top of BSCW) in the future for tasks of this kind? | -2 | -1 | 0 | +1 | +2 |
| How much would you like to use BSCW in the future for tasks of this kind? | -2 | -1 | 0 | +1 | +2 |

| | | | | | |
|---|----|----|---|----|----|
| Did introduction, practice & task description prepare you well for using SWAPit (on top of BSCW) for this task? | -2 | -1 | 0 | +1 | +2 |
| How well are you prepared for estimating how to use BSCW for this task? | -2 | -1 | 0 | +1 | +2 |

| | | | | | |
|---|----|----|---|----|----|
| How appropriate was the access to information (overviews) of SWAPit (on top of BSCW) for this task? | -2 | -1 | 0 | +1 | +2 |
| How appropriate is the access to information of BSCW for this task? | -2 | -1 | 0 | +1 | +2 |

| | | | | | |
|---|----|----|---|----|----|
| Did SWAPit (on top of BSCW) offer all desired interaction facilities / functionalities for efficient solution of this task? | -2 | -1 | 0 | +1 | +2 |
| Does BSCW offer all desired interaction facilities / functionalities for efficient solution of this task? | -2 | -1 | 0 | +1 | +2 |

| | | | | | |
|---|----|----|---|----|----|
| How useful was SWAPit's list of persons (i.e. tree and/or table) for this task? | -2 | -1 | 0 | +1 | +2 |
|---|----|----|---|----|----|

| | | | | | |
|--|----|----|---|----|----|
| How useful was SWAPit's analysis of document contents for this task? | -2 | -1 | 0 | +1 | +2 |
|--|----|----|---|----|----|

| | | | | | |
|---|----|----|---|----|----|
| How useful did you find SWAPit's concept of multiple coordinated views for this task? | -2 | -1 | 0 | +1 | +2 |
|---|----|----|---|----|----|

| |
|--|
| Comments (What was particularly useful? What did you miss? ...) |
|--|

Task 3 „Search Results & Activities“

Please imagine that you are especially interested in the topic “CoCoS” (*composite collaboration service*). You aim to find documents and persons related to this topic.

Which documents and WPs are important for this topic?

One possible way of using SWAPit: Use the search function to find relevant documents. Important documents are highlighted in the metadata table. Important WPs are highlighted in the tree.

Do you find documents which have not been found by your search but which might be of interest for the topic, anyway? (Please only *identify* candidate documents, i.e. you don’t have to verify their actual relevance for the topic)

One possible way of using SWAPit: Look at the document map. Are there documents close to the search results which haven’t been found by your search?

Who has read many of the documents? Who has actively contributed to this topic? What is new (i.e. are there documents for this topic with new user activities)? When were there high activities regarding this topic?

One possible way of using SWAPit: For identifying readers / modifiers, use the respective tables and sort by the selection colour. For analysis of time & activity, use the activity table in the same way. You can also sort by time.

For inspecting activities, use the activity tables and sort by selection colour and/or time. For identifying persons with special interest, also look at the person table.

Please note what you can find out about the topic “CoCoS”:

| | |
|--------------------------------|--|
| relevant WPs | |
| relevant documents | |
| modifiers | |
| readers | |
| months with high user activity | |
| ... | |
| | |
| | |
| | |

Questionnaire „Task 3“

[SWAPit on top of BSCW] vs. [BSCW]

| | very bad/little | | | | very good/much |
|---|-----------------|----|---|----|----------------|
| I believe that the overview of relevant documents and activities for the topic I got using SWAPit (on top of BSCW) is ... | -2 | -1 | 0 | +1 | +2 |
| I believe that the overview of relevant documents and activities for the topic I could get using BSCW is ... | -2 | -1 | 0 | +1 | +2 |
| How confident are you that you found all important documents and activities using BSCW+SWAPit? | -2 | -1 | 0 | +1 | +2 |
| How confident would you be using BSCW? | -2 | -1 | 0 | +1 | +2 |
| How efficiently (time/activities) could you solve this task using SWAPit (on top of BSCW)? | -2 | -1 | 0 | +1 | +2 |
| How efficiently would you solve it using BSCW? | -2 | -1 | 0 | +1 | +2 |
| How stressless did you find the work with SWAPit (on top of BSCW) for this task? | -2 | -1 | 0 | +1 | +2 |
| How stressless would you find the work using BSCW? | -2 | -1 | 0 | +1 | +2 |
| How well did you know how to use the BSCW+SWAPit for solving this task? | -2 | -1 | 0 | +1 | +2 |
| How well would you know how to use BSCW? | -2 | -1 | 0 | +1 | +2 |
| How much would you like to use SWAPit (on top of BSCW) in the future for tasks of this kind? | -2 | -1 | 0 | +1 | +2 |
| How much would you like to use BSCW in the future for tasks of this kind? | -2 | -1 | 0 | +1 | +2 |
| Did introduction, practice & task description prepare you well for using SWAPit (on top of BSCW) for this task? | -2 | -1 | 0 | +1 | +2 |
| How well are you prepared for estimating how to use BSCW for this task? | -2 | -1 | 0 | +1 | +2 |
| How appropriate was the access to information (overviews) of SWAPit (on top of BSCW) for this task? | -2 | -1 | 0 | +1 | +2 |
| How appropriate is the access to information of BSCW for this task? | -2 | -1 | 0 | +1 | +2 |
| Did SWAPit (on top of BSCW) offer all desired interaction facilities / functionalities for efficient solution of this task? | -2 | -1 | 0 | +1 | +2 |
| Does BSCW offer all desired interaction facilities / functionalities for efficient solution of this task? | -2 | -1 | 0 | +1 | +2 |

| | | | | | |
|--|----|----|---|----|----|
| How useful was SWAPit's access to the complete history of user activities for this task? | -2 | -1 | 0 | +1 | +2 |
|--|----|----|---|----|----|

| | | | | | |
|--|----|----|---|----|----|
| How useful was SWAPit's analysis of document contents for this task? | -2 | -1 | 0 | +1 | +2 |
|--|----|----|---|----|----|

| | | | | | |
|---|----|----|---|----|----|
| How useful did you find SWAPit's concept of multiple coordinated views for this task? | -2 | -1 | 0 | +1 | +2 |
|---|----|----|---|----|----|

| |
|--|
| Comments (What was particularly useful? What did you miss? ...) |
| |

Questionnaire „User Profile“

| | | | | | |
|-----------------|--|--|--|--|----------------|
| very bad/little | | | | | very good/much |
|-----------------|--|--|--|--|----------------|

| | | | | | |
|--|----|----|---|----|----|
| How experienced are you using tools for coop. work? | -2 | -1 | 0 | +1 | +2 |
| How experienced are you using BSCW? | -2 | -1 | 0 | +1 | +2 |
| Did the introduction and practice enable you to understand and control SWAPit? | -2 | -1 | 0 | +1 | +2 |
| How experiences are you in performing information researches? | -2 | -1 | 0 | +1 | +2 |
| How familiar are you with the text collection and data treated in the user study? | -2 | -1 | 0 | +1 | +2 |
| How much do you like experimenting with innovative systems you did not use before? | -2 | -1 | 0 | +1 | +2 |

Questionnaire „General Statement“

| | very bad/little | | | | very good/much |
|---|-----------------|----|---|----|----------------|
| How easy was learning how to use SWAPit? | -2 | -1 | 0 | +1 | +2 |
| How much are you now capable of understanding, controlling and using SWAPit? | -2 | -1 | 0 | +1 | +2 |
| How much do you like using SWAPit? | -2 | -1 | 0 | +1 | +2 |
| How well is the usability (SWAPit does what you expect and helps you to do the work effectively & efficiently)? | -2 | -1 | 0 | +1 | +2 |
| How appropriate is the technical integration of SWAPit into BSCW | -2 | -1 | 0 | +1 | +2 |
| How well does a tool like SWAPit fit into your work processes? | -2 | -1 | 0 | +1 | +2 |
| How useful was the system paradigm (comprehensive information in multiple coordinated views)? | -2 | -1 | 0 | +1 | +2 |
| How useful was the analysis and overview of document contents? | -2 | -1 | 0 | +1 | +2 |
| How useful was the compact access to the whole history of user activities? | -2 | -1 | 0 | +1 | +2 |
| How useful was the compact access to persons? | -2 | -1 | 0 | +1 | +2 |
| How do you assess the SWAPit's maturity (stability, usability, interaction ...) | -2 | -1 | 0 | +1 | +2 |
| How good is the information overview one can achieve using SWAPit? | -2 | -1 | 0 | +1 | +2 |
| How well does SWAPit support you in focusing information (i.e. identifying interesting subsets)? | -2 | -1 | 0 | +1 | +2 |
| How easy to understand were the interactive rankings? | -2 | -1 | 0 | +1 | +2 |
| How useful were the interactive rankings? | -2 | -1 | 0 | +1 | +2 |

10 Literaturverzeichnis

- [AHS03] **Amberg, M., Hirschmeier, M., Schobert, D.**, *An Acceptance Model for the Analysis and Design of Innovative Technologies*, 7th Int. Conference on Synergetics, Cybernetics and Informatics (SCI), Orlando, Florida, USA, 2003.
- [AK+02] **Abramowicz, W., Kalczynski, P., Wecel, K., Hunt, J.**, *Filtering the Web to Feed Data Warehouses*, Springer, 2002.
- [Ap99] **Appelt, W.**, *WWW Based Collaboration with the BSCW System*, Proc. of SOFSEM'99 (Milovy, Czech Republic, Nov.26-Dec.4) Springer Lecture Notes in Computer Science 1725, Heidelberg, pp. 66-78, 1999.
- [As04] **Aschenbrenner, N.**, *Elektronische Schatzsuche*. In: Pictures of the Future - Die Zeitschrift für Forschung und Innovation, 2004.
- [Ba04] **Bange, C.**, *Business Intelligence aus Kennzahlen und Dokumenten*, Studien zur Wirtschaftsinformatik, Bd.13, Verlag Dr. Kovac, 2004.
- [Ba96] **Bauer, M.**, *Altbekanntes in neuer Verpackung?*, In: Business Computing o. Jg., 1996/4, S. 46, 1996.
- [BBN02] **Barr, P., Biddle, R., Noble, J.**, *A Taxonomy of User-Interface Metaphors*, Proceedings of SIGCHI-NZ Symposium On Computer-Human Interaction, 2002.
- [Be01] **Becks, A.**, *Visual knowledge management with adaptable document maps*, GMD research series, 15, ISBN 3-88457-398-5, 2001.
- [BR+03] **Boukhelifa, N., Roberts, J. C., Roberts, P. J., and Rodgers, P. J.**, *A Coordination Model for Exploratory Multi-View Visualization*, in: Proc. of the Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV'03), 2003.
- [BS01] **Becks, A., Seeling, C.**, *A Task-Model for Text Corpus Analysis in Knowledge Management*. In: Proceedings of 8th International Conference on User Modelling (UM'01), Workshop on User Modeling, Machine Learning and Information Retrieval, Sonthofen (Germany), July 2001.
- [BS06] **Batini, C., Scannapieco, M.**, *Data Quality*, Springer-Verlag Berlin Heidelberg, ISBN: 978-3-540-33172-8, 2006.
- [Be94] **Belkin, N.J., Stein, A., Thiel, U.**, *Cases, Scripts, and Information-Seeking Strategies: On the Design of Interactive Information Retrieval Systems*. Arbeitspapiere der GMD No. 875, Sankt Augustin, 1994.

-
- [BG05] **Burmester, L., Goeken, M.:** *Benutzerorientierter Entwurf von unternehmensweiten Data-Warehouse-Systemen.* Wirtschaftsinformatik, S.1421-1440, 2005.
- [BK02] **Botsch, V., Kunz, C.,** *Visualization and navigation of networked information spaces: the Matrix Browser,* In: Proceedings of the WWDU'02, Berchtesgarden, 2002.
- [Br04] **Brunner, A.,** *Ontologiebasierte Navigation in textuell annotierten Geschäftskennzahlen,* Diplomarbeit Informatik, RWTH Aachen, 2004.
- [BR02] **Brusilovsky, P., Rizzo, R.,** *Map-Based Horizontal Navigation in Educational Hypertext,* Journal of Digital Information (JoDI), Volume 3 Issue 1, 2002.
- [BR07] **Becks, A., Rose, T.,** *Portals for business intelligence,* In: A. Tatnall, Hrsg.: Encyclopaedia of Portal Technology and Applications, Idea Group Publishing, Hershey, PA, 2007.
- [BSM02] **Becks, A., Seeling, C., Minkenberg, R.,** *Benefits of Document Maps for Text Access in Knowledge Management: A Comparative Study,* Proceedings of ACM Symposium on Applied Computing (SAC'02), Madrid, Spain, pp.621-626, 2002.
- [Bü04] **Bürdek, B.E.,** *Clemens Lango: Visuos - a visuospatial operating software for knowledge work,* i-com Journal, 3/1, pp.39-40, 2004.
- [BW84] **Basili, V.R., Weiss, D.M.,** *A Methodology for Collecting Valid Software Engineering Data,* in: IEEE Transaction on Software Engineering, Vol.10, No.6, S.728-738, Juni 1984.
- [BWK00] **Baldonado, M.Q.W., Woodruff, A., Kuchinsky, A.:** *Guidelines for Using Multiple Views in Information Visualization,* Proceedings of the working conference on Advanced Visual Interfaces (AVI00), p.110-119, ACM Press New York, NY, USA, 2000.
- [CC05] **Craft, B., Cairns, P.,** *Beyond guidelines: what can we learn from the visual information seeking mantra?* In: Proceedings of the Ninth International Conference on Information Visualisation (IV'05), 2005.
- [CCS93] **Codd, E.F., Codd, S.B., Salley, C.T.,** *Providing OLAP to User-Analysts: An IT Mandate,* Codd & Associates, Ann Arbor/Michigan, 1993.
- [CH04] **Cruz, I.F., Huang, Y.F.,** *A Layered Architecture for the Exploration of Heterogeneous Information Using Coordinated Views,* in: Proceedings of the IEEE Symposium on Visual Languages - Human Centric Computing (VLHCC'04), Sep.26-29, 2004.
- [CM84] **Cleveland, W.S., McGill, R.,** *Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods,* Journal of the American Statistical Association, 79/387, pp.531-554, 1984.

- [CN02] **Card, S., Nation, D.**, *Degree-of-interest trees: A component of an attention-reactive user interface*, In Proceedings of the Working Conf. on Advanced Visual Interfaces (AVI'02), pp.231-246, 2002.
- [CK+02] **Cody, W.F., Kreulen, J.T., Krishna, V., Spangler, W.**, *The integration of business intelligence and knowledge management*, IBM Systems Journal, J.41, 4, pp.697-713, Oct. 2002.
- [CS94] **Chimera, R., Shneiderman, B.**, *An exploratory evaluation of three interfaces for browsing large hierarchical tables of contents*, ACM Transactions on Information Systems (ToIS), 12(4), pp. 383-406, 1994.
- [CS+97] **Carlis, J., Safonov, A., Perrin, D., Konstan, J.**, *The Neighborhood Viewer: A Paradigm for Exploring Image Databases*, In: Proceedings of Companion of ACM CHI'97 Conference - Human Factors in Computing Systems , pp. 299-300, 1997.
- [CZ03] **Computer Zeitung**, *Informationsinseln wachsen zusammen*, Nr.40, S.11, 2003.
- [DBW89] **Davis, F.D., Bagozzi, R.P., Warshaw, P.R.**: *User Acceptance of Computer Technology: A Comparison of Two Theoretical Models*, Management Science, Vol.35, No.8, pp.982-1003, JSTOR, 1989.
- [DF+04] **Dix, A., Finlay, J., Abowd, G., Beale, R.**: *Human-Computer Interaction*, 3. Auflage, Pearson Education Limited, 2004.
- [Do61] **Doyle, L.B.**, *Semantic Road Maps for Literature Searchers*, Journal of the ACM (JACM), Vol.8, No.4 , Oct. 1961.
- [Eg05] **Eggert, P.**, *Integration von Text Mining und strukturierten Datenbankanfragen für das Customer Relationship Management*, Diplomarbeit Informatik, RWTH Aachen, 2005.
- [EM02] **Eibl, M., Mandl, T.**, *Including User Strategies in the Evaluation of Graphic Design Interfaces for Browsing Documents*. In: Skala, Vaclav (Hrsg.): Journal of WSCG. Special issue. The 10th Int. Conf. on Computer Graphics and Visualization (WSCG'02), Pilsen, Tschechien, Vol.1, pp.163-169, Feb. 2002.
- [FF01] **Farbey, B., Finkelstein, A.**, *Evaluation in software engineering: ROI, but more than ROI.*, in: 3rd International Workshop on Economics-Driven Software Engineering Research (EDSER), Toronto, Canada, 2001.
- [FL04] **Fettke, P., Loos, P.**, *Referenzmodellierungsforschung*, in: Wirtschaftsinformatik 46/5, S. 331-340, 2004.

-
- [Fr05] **Frese, D.**, *Annotation von Datenwürfeln mit heterogenen, externen Informationen*, Diplomarbeit Informatik, RWTH Aachen, 2005.
- [FSH02] **Fluit, C., Sabou, M., van Harmelen, F.**, *Ontology-based Information Visualization*, in: Geroimenko V., Chen C. (Eds.): *Visualizing the Semantic Web – XML-based Internet and Information Visualization*, Springer Verlag, London, pp.36-48, 2002.
- [Fu99] **Fugmann, R.**: *Inhaltserschließung durch Indexieren: Prinzipien und Praxis*, Frankfurt/Main: DGD, 1999.
- [Fu81] **Furnas, G.W.**, *The FISHEYE view: a new look at structured files*, in: *Readings in Information Visualization: Using Vision to Think*, S.K.Card, J.D.Mackinlay, B.Shneiderman, Eds., San Francisco: Morgan Kaufmann Publishers, Inc., pp. 312-330, 1981.
- [Fu97] **Furnas, G.W.**, *Effective View Navigation*. In: *Proc. of ACM Human Factors in Computing Systems (CHI'97)*, Atlanta GA, 367-374. ACM Press, New York, 1997.
- [GL+06] **Gresh, J., Lewis, B., Montemayor, J., Piatko, C., Turner, R.**, *Supporting insight-based information exploration in intelligence analysis*. *Communications of the ACM* 49/4, pp.63-68, 2006.
- [GHK01] **Gerstl, P., Hertweck, M., Kuhn, B.**, *Text Mining: Grundlagen, Verfahren, und Anwendungen*, HMD 222, Dezember 2001.
- [GR+00] **Gresh, D.L., Rogowitz, B.E., Winslow, R.L., Scollan, D.F., Yung, C.K.**, *WEAVE: A System for Visually Linking 3-D and Statistical Visualizations, Applied to Cardiac Simulation and Measurement Data*, in: *Proceedings of the Conference on Visualization 2000 (Vis00)*, pp.489-492. IEEE Computer Society Press, 2000.
- [Go05] **Goeken, M.**, *Wissensarten und Techniken im Anforderungsmanagement*, GI-Jahrestagung (2) 2005: S. 205-209, 2005.
- [GT95] **Goodhue, D.L.; Thompson, R.L.**: *Task-Technology Fit and Individual Performance*, *MIS Quarterly*, 19:2 (6/1995), pp.213-236, 1995.
- [GWG03] **Gross, T., Wirsam, W., Gräther, W.**, *AwarenessMaps: visualizing awareness in shared workspaces*, CHI03 Extended Abstracts, pp.784-785, 2003.
- [Ha92] **Harter, S.P.**, *Psychological Relevance and Information Science*, *Journal of the American Society for Information Science*, Vol.43, No.9, 1992.
- [Ha99] **Hackathorn, R.D.**, *Web Farming for the Data Warehouse*, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 1999.

- [He95] **Hearst**, M., *TileBars: Visualization of Term Distribution Information in Full Text Information Access*, in Proc. of the ACM SIGCHI Conf. on Human Factors in Computing Systems (CHI'95), pp. 59-66, Denver, CO, May 1995.
- [He99] **Hearst**, M., *Untangling Text Data Mining*, in: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL'99), University of Maryland, 1999.
- [He06] **Hearst**, M., *Clustering versus Faceted Categories for Information Exploration*, Communications of the ACM, Vol.49, No.4, April 2006.
- [HH95] **Herget**, J., **Hensler**, S., *Online-Datenbanken in Wirtschaft und Wissenschaft: aktuelle Nachfragestrukturen und Nutzungstendenzen*, in: WIRTSCHAFTSINFORMATIK 37/2, S.129-138, 1995.
- [Hi06] **Hirschmeier**, S., *Exploratives Text Mining für den Zugriff auf Kundenmeldungen im Customer Support – eine nutzenorientierte Studie*, Diplomarbeit der Fachgruppe Informatik, RWTH Aachen, 2006.
- [Ho97] **Holten**, R., *Die drei Dimensionen des Inhaltsaspektes von Führungsinformationssystemen*, Arbeitsberichte d. Inst. für Wirtschaftsinformatik, Universität Münster, April 1997.
- [HT04] **Hetzler**, E., **Turner**, A., *Analysis experiences using information visualization*, IEEE Computer Graphics and Applications, Vol.24/5, pp.22-26, 2004.
- [HW+98] **Hetzler**, B., **Whitney**, P., **Martucci**, L., **Thomas**, J., *Multifaceted Insight through Interoperable Visual Information Analysis Paradigms*. In Graham Wills and John Dill, editors, In: Proceedings Information Visualization '98, pp.137-144, Los Alamitos, CA, 1998.
- [Ju04] **Judelman**, G.B., *Knowledge Visualization – Problems and Principles for Mapping the Knowledge Space*, MSc Thesis, University of Lübeck, Germany, 2004.
- [JS91] **Johnson**, B., **Shneiderman**, B., **Treemaps - A space filling approach to the visualization of hierarchical information structures**, Proceedings of IEEE Visualization 1991, IEEE CS Press, pp.275-282, 1991.
- [Ka03] **Kantardzic**, M., *Data Mining. Concepts, Models, Methods, and Algorithms*. Wiley&Sons, Piscataway, 2003.
- [KB+05] **Keller**, P., **Bange**, C., **Narr**, J., **Dahnken**, O., *OLAP & BI - 8 multidimensionale Datenbanken und 17 Reporting und Analyse-Werkzeuge im Vergleich*, Oxygon Verlag, München, ISBN: 3-93781-805-7, 2005.

-
- [KHA98] **Klement, T., Hemmje, M., Aberer, K.**, *Towards visual support for document navigation through multidimensional catalogues*. In Proc. of the Workshop on New Paradigms in Inf. Visualization and Manipulation (NPIV'98), Washington D.C., US, ACM Press, NY, pp.25-28, 1998.
- [KM78] **Keen, P.G.W., Morton, M.S.S.**: *Decision Support Systems: an organizational perspective*, Addison-Wesley, 1978.
- [KN98] **Kaplan, R. S., Norton, D. P.**: *Balanced Scorecard, Strategien erfolgreich umsetzen*, aus dem Amerikanischen von P. Hörvath, Stuttgart, 1997, Hörvath, P.: *Controlling*, Verlag Vahlen, München, 1998.
- [KK+00] **Kohonen, T., Kaski, S., Lagus, K., Salojarvi, J., Honkela, J.**, *Self Organization of a Massive Document Collection*, in: IEEE Transactions on Neural Networks, Vol.11, No.3, S.2003-2012, 2000.
- [KM+02] **Klein, P., Müller, F., Reiterer, H., Eibl, M.**, *Visual Information Retrieval with the SuperTable + Scatterplot*, 6th International Conference on Information Visualisation (IV'02), London, 10-12 July 2002.
- [KR98] **Königer, P., Reithmayer, W.**, *Management unstrukturierter Informationen: Wie Unternehmen die Informationsflut beherrschen können*, Frankfurt a.M.: Campus, 1998.
- [KR+03] **Klein, P., Reiterer, H., Müller, F., Limbach, T.**: *Metadata Visualisation with VisMeB*, IV'03, pp.600-605, 2003.
- [KSB06] **Kensche, D., Seeling, C., Becks, A.**: *MonA - An Extensible Framework for Web Document Monitoring*, Proc. of the ACM Symposium on Applied Computing (SAC'06), Dijon, France, 2006.
- [KSJ05] **Klamma, R., Spaniol, M., Jarke, M.**, *MECCA: Hypermedia Capturing of Collaborative Scientific Discourses about Movies*, Informing Science Journal, Vol.8, pp.3-38, 2005.
- [LN03] **Li, Q., North, C.**, *Empirical Experiment of Dynamic Query Sliders and Brushing Histograms*, Proc. of IEEE Inf. Vis. (IV03), Seattle, Washington, 2003.
- [LN05] **Lange, I., Nobs, A.**: *Was ist ein Bezugsrahmen?* Seminar für Doktorierende – Forschungsmethodik, Prof. Dr. R. Boutellier, ETH Zürich, WS 2004/2005.
- [Lo07] **Lohmann, S.**, *Die Beschaffung und die Integration externer Informationen zur situierten und rollenbasierten Entscheidungsunterstützung*, Dissertation, Universität Erlangen-Nürnberg, 2007.

- [LRP95] **Lamping, J., Rao, R., Pirolli, P.**, *A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies*, Xerox Palo Alto Research Center, 1995.
- [LM+02] **Limbach, T., Müller, F., Klein, P., Reiterer, H., Eibl, M.**, *Visualization of Metadata using the SuperTable + Scatterplot*, Proc. des 9. Int. Symposiums für Informationswissenschaft (ISI'02): 147-164, 2002.
- [LBS01] **Luzcak, H., Bullinger, H.-J., Schlick, C.**, *Unterstützung flexibler Kooperation durch Software. Methoden, Systeme, Beispiele*. Springer Verlag, ISBN 3-540-67500-0, 2001.
- [Ma92] **Marshall, C.C., Rogers, R.A.**, *Two Years Before the Mist: Experiences with AquaNet*. In Proceedings of the ACM Conf. on Hypertext (ECHT '92), Milano, Italy, pp.53-62, Dec. 1992.
- [Ma06] **Marchionini, G.**, *Exporatory Search: From Finding To Understanding*, Communications of the ACM, Vol.49, No.4, April 2006.
- [MC+03] **Mothe, J., Chrisment, C., Dousset, B., Alau, J.**, *DocCube: Multi-dimensional visualisation and exploration of large document sets*, JASIST 54(7), pp.650-659, 2003.
- [Me92] **Meadow, C.T.**, *Text information retrieval systems*, Academic Press, San Diego, California, 1992.
- [Me99] **Mertens, P.**, *Integration interner, externer, qualitativer und quantitativer Daten auf dem Weg zum Aktiven MIS*. in: Wirtschaftsinformatik 41(5): 405-415, 1999.
- [MG02] **Mertens, P., Griese, J.**, *Integrierte Informationsverarbeitung 2: Planungs- und Kontrollsysteme in der Industrie*, 9. Auflage, Gabler Verlag.
- [Mo97] **Modjeska, D.**, *Review of Research on Navigation in Electronic Worlds*. Technical Report, University of Toronto, 1997.
- [Mo99] **Morse, E.L.**, *Evaluation of Visual Information Browsing Displays*. PhD-Thesis, University of Pittsburgh, 1999.
- [NC+03] **North, C., Conklin, N., Indukuri, K., Saini, V.**, *Fusion: Interactive coordination of diverse data, visualizations, and mining algorithms*, ACM CHI'03, 2003.
- [No00] **North, C.**, *A user interface for coordinating visualizations based on relational schemata: SNAP-Together Visualization*, Dissertation, University Of Maryland, 2000.
- [NS97] **North, C., Shneiderman, B.**, *A taxonomy of, multiple window coordinations*, University of Maryland, Dept. of Computer Science Tech. Report #CS-TR-3854, 1997.

-
- [NS00] **North, C., Shneiderman, B.**, *Snap-Together Visualization: A User Interface for Coordinating Visualizations via Relational Schemata*, in: Proceedings of ACM Advanced Visual Interfaces (AVI'00), 2000.
- [NW05] **Novak, J., Wurst, M.**, *Collaborative Knowledge Visualization for Cross-Community Learning*, in: Knowledge and Information Visualization, pp.95-116, 2005.
- [Po98] **Pollitt, A.S.**, *The key role of classification and indexing in view-based searching*, Technical Report, University of Huddersfield, 1998.
- [PP01] **Pattison, T., Phillips, M.**, *View coordination architecture for information visualisation*. in: Proc. of Asia-Pacific Symposium on Information Visualisation, Sydney, Australia, pp.165-169, 2001.
- [PPS04] **Pankoke-Babatz, U., Prinz, W., Schäfer, L.**, *Stories about Asynchronous Awareness*, in Cooperative Systems Design - Scenario-Based Design of Collaborative Systems, F. Darses, R. Dieng, et al. (eds.), IOS Press, pp.23-38, 2004.
- [PR+06] **Pallot, M., Ruland, R., Traykov, S., Kristensen, K.**, *Integrating Shared Workspace, Wiki and Blog*, ICE 2006, 29th June, Milan, 2006.
- [Ra98] **Rahlenbeck, E.**: *AXA Colonia löst fast 90 Prozent der Kundenanfragen*, in: Computerwoche 48, 1998.
- [Ro00] **Robertson, G.**, *From hierarchies to polyarchies: visualizing multiple relationships*. In: Proc. of the Working Conference on Advanced Visual interfaces (AVI00), Palermo, Italy, ACM Press, New York, 2000.
- [Ro97] **Rosenberg, J.**, *The Interactive Diagram Sentence: Hypertext as Medium of Thought*. In Visible Language, 30(2), pp.103-116, 1997.
- [RW06] **Roberts, J.C., Wright, M.A.**, *Towards Ubiquitous Brushing for Information Visualization*, in: Proc. of the Conference on Information Visualization (IV'06), pp.151-156, 2006.
- [Sa00] **Sack, W.** *Conversation Map: A content-based Usenet newsgroup browser*, In: Proc. of the International Conference on Intelligent User Interfaces (IUI'00), ACM Press, pp.233-240, 2000.
- [Sa03] **Sahling, G.N.**, *Interactive 3D Scatterplots - From High Dimensional Data to Insight*, Master Thesis, Forschungszentrum VRVis, Wien, 2003.
- [Sc06] **Schuh, G.**, *Change Management — Prozesse strategiekonform gestalten*, Springer-Verlag Berlin Heidelberg, ISBN: 978-3-540-23657-3, 2006.

- [Se01] **Seeling, C.**, *Eine Vergleichende Studie zur Evaluierung der Aufgabenangemessenheit von Dokumentenlandkarten im Wissensmanagement*, Diplomarbeit, RWTH Aachen, 2001.
- [SB04] **Seeling, C., Becks, A.**, *Analysing Associations of Textual and Relational Data with a Multiple Views System*, in: Proc. of the 2nd Int. Conf. on Coordinated and Multiple Views in Exploratory Visualization (CMV'04), IEEE Computer Society Press, London, England, pp.61-70, 2004.
- [SBB96] **Spenske, M., Beilken, C., Berlage, T.**, *FOCUS: The interactive table for product comparison and selection*, In: Proceedings of the ACM Symposium on User Interface Software and Technology (UIST'96), Seattle, pp.41-50, Nov. 1996.
- [SF+99] **Shneiderman, B., Feldman, D., Rose, A., Grau, F.X.**, *Visualizing Digital Library Search Results with Categorical and Hierarchical Axes*, Technical Report CS-TR-3992, Department of Computer Science, University of Maryland, 1999.
- [Sh92] **Shneiderman, B.**, *Tree visualization with treemaps: a 2-d space-filling approach*, ACM Transactions on Graphics, vol.11/1, p.92-99, 1992.
- [Sh96] **Shneiderman, B.**, *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization*, Technical Report 96-66, Institute for Systems Research, University of Maryland, 1996.
- [Si01] **Simon, B.**, *Wissensmedien im Bildungssektor – Eine Akzeptanzuntersuchung an Hochschulen*, Dissertation, Wirtschaftsuniversität, Wien, 2001.
- [Sm02] **Smith, B.**, *Aristoteles, Kant und die Quantenphysik*, in: Philosophie und Wissenschaft, Würzburg: Königshausen und Neumann, p.79-97, 2002.
- [SM87] **Salton, G., McGill, M.J.**: *Information Retrieval*, MacGraw-Hill, 1987.
- [SM99] **Shipman, F.M., Marshall, C.C.**, *Spatial Hypertext: An Alternative to Navigational and Semantic Links*. ACM Computing Surveys 31(4), December 1999.
- [SP04] **Scholtz, J., Potts Steves, M.**, *A framework for real-world software system evaluations*, in: Proceedings of CSCW04, pp.600-603, 2004.
- [SPB07] **Seeling, C., Prinz, W., Becks, A.**, *Providing Awareness in Complex Cooperation Networks*, Proc. of IADIS Int. Conference on Web Based Communities (WBC07), IADIS Press, Salamanca, Spain, Feb.18-20, 2007.
- [SP+07] **Seeling, C., Prinz, W., Ruland, R., Becks, A.**, *Kooperatives Arbeiten unter der Lupe*, in: Tagungsband der 7. Konferenz Mensch & Computer, Weimar, 2007.

-
- [SS98] **Stauss, B., Seidel, W.**, *Beschwerdemanagement*, Carl Hanser Verlag, München, 1998.
- [SSS04] **Schlechtweg, S., Schulze-Wollgast, P., Schumann, H.**, *Interactive Treemaps With Detail on Demand to Support Information Search in Documents*, in: Proc. of IEEE Symposium on Visualization, 2004.
- [St95] **Stürmer, U.**, *Vergleichende Evaluierung verschiedener Interaktionsparadigmen für Information-Retrieval*, Dissertation, GMD-Studien Nr.277, Nov. 1995.
- [Su01] **Sullivan, D.**, *Document Warehousing and Text Mining: Techniques for Improving Business Operations*, John Wiley & Sons, Inc., New York, NY, USA, pp.275-287, 2001.
- [Su03] **Sullivan, D.**, *Enterprise Content Management: Defining a Meta Data Standard for Enterprise Content Management*, DMReview online, <http://www.dmreview.com>, 2003.
- [SWY75] **Salton, G; Wong, A.; Yang, C.S.**, *A Vector Space Model for Information Retrieval*. Communications of the ACM, 18(11): 613-620, 1975.
- [SW+04] **Stuckenschmidt, H., de Waard, A., Bhogal, R., Fluit, C., Kampman, A., van Buel, J., van Mulligen, E.M., Broekstra, J., Crowlesmith, I., van Harmelen, F., Scerri, T.**: *A Topic-Based Browser for Large Online Resources*, EKAW 2004: 433-448, 2004.
- [Te01] **Tewis, H.**: *Analysemethoden im Spannungsfeld von Softwareneuentwicklung und Einführung von Standardsoftware*, Diplomarbeit, Universität Hamburg, 2001.
- [TS06] **Thomas, O., Scheer, A.-W.**, *Business Engineering mit Referenzmodellen – Konzeption und informationstechnische Umsetzung*, Information Management & Consulting, 21/1, 2006.
- [Va01] **Valente, A., Housel, T.**: *A Framework to Analyze and Compare Knowledge Management Tools*, In Proc. of the Knowledge-Based Intelligent Information Engineering Systems and Technologies (KES01), IOS Press, Ohmsha, Japan, 2001.
- [Va05] **Vara, V.**, *Tweaking Google*, The Wall Street Journal Online, 3rd April 2005.
- [WL90] **Wehrend, S., Lewis, C.**, *A problem-oriented classification of visualization techniques*. Proceedings of IEEE Visualization, pp. 139-143, IEEE Computer Society Press, 1990.
- [WT+95] **Wise, J.A., Thomas, J.J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., Crow, V.**, *Visualizing the non-visual: spatial analysis and interaction with information from text documents*, in: Proc. Information Visualization, Oct. 1995, Atlanta, Georgia, USA, IEEE Computer Society Press, Los Alamitos, California, 1995.

- [WMM06] **White**, R.W., **Muresan**, G., **Marchionini**, G., *Evaluating Exploratory Search Systems*, Preface, in: Proc. of ACM SIGIR Workshop on Evaluating Exploratory Search Systems (EESS'06), Seattle, Washington, USA, 2006.
- [Yo07] **Yost**, B.A., *The Visual Scalability of Integrated and Multiple View Visualizations for Large, High Resolution Displays*, Dissertation, Virginia Polytechnic Institute and State University, 2007.

Lebenslauf

Name: Christian Reinhard Johannes Seeling

Geburtsdatum: 1. April 1974

Geburtsort: Aachen

Wohnort: Starenweg 22
52146 Würselen
Tel.: 02405 – 46 24 00
E-Mail: christian.seeling@gmx.de

Familienstand: Verheiratet (2 Kinder)

Schulbildung: 1985-1993 Kaiser-Karls-Gymnasium in Aachen (Abitur)

Zivildienst: 1993-1994 Diakonisches Werk in Aachen

Studium: RWTH Aachen: Informatik mit Nebenfach BWL (Schwerpunkt Operations Research). Abschluss mit Diplom. Gesamtnote „sehr gut“.
Diplomarbeit: „Vergleichende Studie zur Evaluierung der Aufgabenangemessenheit von Dokumentenlandkarten im Wissensmanagement“.

Berufliche Tätigkeit: Ende 2001 bis Ende 2002: Wissenschaftlicher Angestellter am Lehrstuhl Informationssysteme. Bis Mai 2007: Forschung und Entwicklung am Fraunhofer-Institut für Angewandte Informationstechnik FIT, St. Augustin in den Bereichen Suchtechnologien, Text Mining und Informationsvisualisierung. Zahlreiche internationale Veröffentlichungen in Zeitschriften und Konferenzbeiträgen. Seit Juni 2007 angestellt bei Fa. Aspera OHG, dem führenden Anbieter für Softwarelizenzmanagement in internationalen Konzernstrukturen (Innovationsmanagement & strategische Produktentwicklung).

