



# Dynamische Menüs zur Recherche und Orientierung.

## Konzeption der thematischen Suche für den Informationsservice "WebSite 'Methodik"

Gerhard Knorz

Fachhochschule Darmstadt, Fachbereich IuD, Haardtring 100,  
D-64295 Darmstadt, Email: [knorz@iud.fh-darmstadt.de](mailto:knorz@iud.fh-darmstadt.de)

---

### Inhalt

- 1 Anwendungsfall
- 1.1 WebSite 'Methodik als zentrale Informationsressource für Lehren und Lernen
- 1.2 Architektur und Services von WebSite 'Methodik
- 2 Konzeption der thematischen Suche
- 2.1 Problemstellung
- 2.2 Lösungsmöglichkeiten
- 2.3 Vorgeschlagenes Konzept
- 2.4 Gestaltungsentwurf
- 3 Realisierung
- 3.1 Vokabular und Indexierung
- 3.2 Datenbank
- 3.3 Stand der Entwicklung

### Zusammenfassung

Das Konzept einer themenorientierten Suche wird vorgestellt, das gleichzeitig das Ziel verfolgt, dem Nutzer Orientierung über die inhaltliche Ausgestaltung eines Informationsbestandes zu geben. Es wurde für WebSite `Methodik entwickelt, einen WWW-Server für Distance Learning in dual mode, und ist übertragbar auf Dokumentensammlungen mittlerer Größe, die begriffsorientiert indexiert sind. Nutzer wählen aus einem kontrollierten Vokabular Themen aus und erhalten parallel zu den entsprechend indexierten Dokumenten alle weiteren Deskriptoren angezeigt, mit der sich die Suche vom Allgemeinen zum Speziellen fortsetzen läßt. Die Häufigkeit, mit der Deskriptoren gemeinsam indexiert sind, läßt in diesem Kontext die thematischen Schwerpunkte des Informationsservers erkennen. Die Suchfunktionalität kann in verschiedener Weise visualisiert werden. Konzeption und Aufbau der zugrundeliegenden Datenbank ist abgeschlossen, die Integration in WebSite `Methodik ist in Entwicklung.



## 1 Anwendungsfall

### 1.1 WebSite 'Methodik als zentrale Informationsressource für Lehren und Lernen

Hochschullehre ist aus vielfältigen Gründen in die Diskussion gekommen und in Bewegung geraten. Die Unterstützung der Lehre durch neue Medien und Informationstechnik ist nicht nur ein populäres Thema, sondern wird auch mit praktischen Konsequenzen von Lehrenden und Lernenden aufgegriffen (HIS 1998). Eine im Wintersemester 1997/98 durchgeführte Befragung bei Studierenden informationswissenschaftlicher Studiengänge im deutschsprachigen Raum hat ergeben, daß "Distance Learning" als Variante oder Komponente eines Studiums durchaus (wenngleich mit Einschränkungen) Akzeptanz findet (Beheim et al. 1998).

Der insgesamt positiven, gleichzeitig aber vorsichtigen Einstellung der Studierenden entsprechen Konzept und Praxis der Lehre im Fach Informationsmethodik im Fachbereich Information und Dokumentation der Fachhochschule Darmstadt. Der Ansatz ist als Distance Learning im Dual Mode zu klassifizieren: Zusätzlich zur Präsenzveranstaltung wird jeder Lehrveranstaltungstermin zunächst in Form einer Planung und abschließend in Form einer Dokumentation des tatsächlichen Verlaufs im Internet "gespiegelt". Unterlagen zu Lehrveranstaltungen werden vollständig im Netz bereitgehalten. Lehrveranstaltungsteilnehmer holen sich erforderliche Unterlagen vom Server selbst ab.

Die zentrale (technische) Komponente dieser Form der Lehre im Fach Informationsmethodik ist *WebSite 'Methodik*, dessen Wurzeln bis zum Wintersemester 1995/96 zurückgehen und die unter diesem Namen und in dieser Rolle seit Sommersemester 1996 besteht ([Knorz 1998](#)). Seit dieser Zeit dokumentiert der Server lückenlos alle im Bereich der Lehre relevanten Aktivitäten (auf der Ebene der einzelnen Termine und Inhalte) und etliches darüber hinaus. Gleichzeitig stellt er auch Kommunikationsmöglichkeiten bereit.

WebSite 'Methodik besteht mittlerweile aus annähernd 500 Dokumenten (Stand August 1998), die auf ca. 2500 Dateien in ca. 200 Verzeichnissen basieren (85 MB). Die Änderungsrate beträgt, bedingt durch die notwendigen Nacharbeitungen der Lehrveranstaltungsplanung, etwa bei 20 bis 30 Dateien/Woche. Das Volumen hat sich im Laufe des vergangenen Jahres nahezu verdoppelt.

Bei dem Aufwand, den Aufbau und Betrieb von WebSite 'Methodik für eine einzelne Lehrperson ohne weitere personelle Unterstützung erfordern, muß als Ergebnis mehr als "nur" ein Informationspool für Studierende resultieren. WebSite 'Methodik ist gleichberechtigt auch das Instrument, in dem sich alle planenden und vorbereitenden Tätigkeiten des Lehrenden abspielen, und das die Ergebnisse dieser Prozesse dokumentiert. Vorbereitung eines Vortrags oder eines Vorlesungstermins bedeutet zunächst Recherche nach einschlägigen Materialien auf dem Server, die anschließend strukturiert, aktualisiert, ergänzt und aufbereitet werden. Gleichzeitig ist der Informationsserver auch der Ort, an dem alle neuen Informationen zur späteren Verwendung abgelegt werden.

## 1.2 Architektur und Services von WebSite 'Methodik

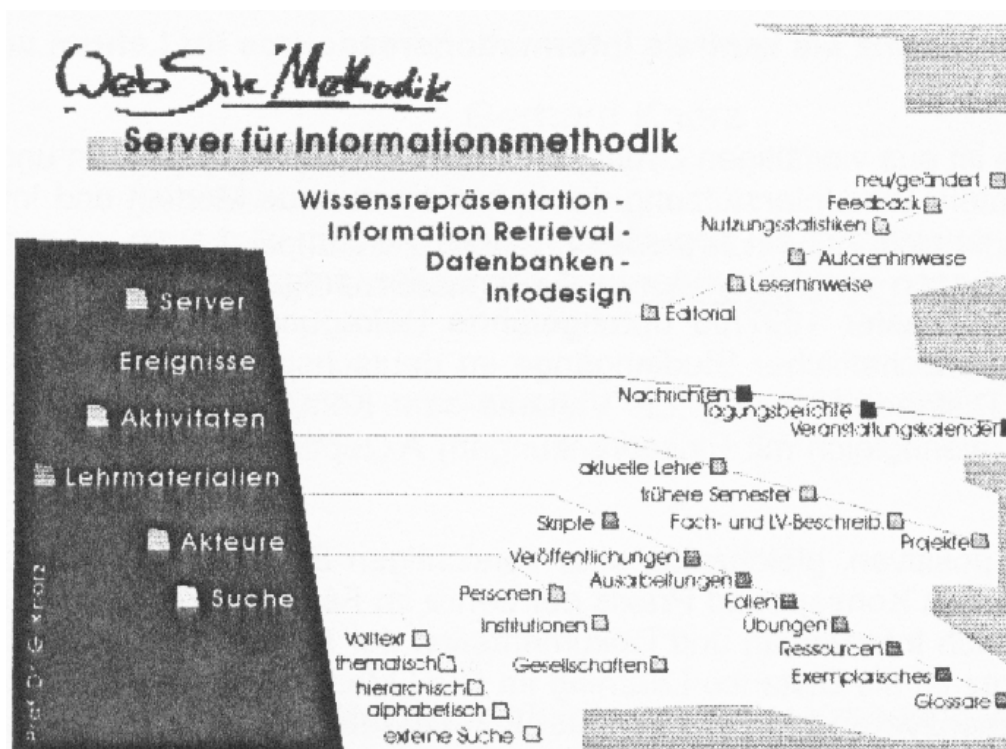


Abb. 1: Übersicht über die Rubriken von WebSite 'Methodik. Auf diese Site Map wird direkt von der Homepage aus verwiesen.

WebSite 'Methodik ist in 5 Rubriken (zuzüglich Suche) gegliedert, die weitgehend monohierarchisch, in Einzelfällen polyhierarchisch, weiterentwickelt werden (Abb. 1). Die Einordnung eines Dokumentes in die Rubrikstruktur ist zeitunabhängig und eindeutig. Die Strukturierung von WebSite 'Methodik ist bei quantitativem Wachstum und wachsenden Anforderungen stabil geblieben.

Folgende Prinzipien liegen WebSite 'Methodik zugrunde:

- **"Monotonie-Eigenschaft" des Informationsangebotes.** Es wird grundsätzlich auf Dokumentenebene nicht gelöscht: Ein gespeicherter Link auf ein Dokument bleibt valide, da dieses Dokument entweder zeitunabhängig von Interesse ist oder aber dem Ziel dient, einen in der Zeit ablaufenden Prozeß zu dokumentieren.
- **Vollständigkeit.** Ein zentrales Qualitätskriterium einer Datensammlung ist Vollständigkeit und Abdeckung. Wann immer eine Unterlage in der Lehre verwendet oder eine Veranstaltung gehalten wird, ist sie in WebSite 'Methodik zu finden.
- **Inhaltliche und formale Qualität.** In Stichworten: inhaltliche Bewertung, verschiedener Status von Dokumenten, vorgegebene reichhaltige formale Dokumentenstruktur, Meta Tags, kontrolliertes Indexierungsvokabular, durchgängige Namenskonventionen.
- **Navigation und Suche.** Jedes Dokument ist durch seine Indexierung thematisch eingeordnet und kann über diese Indexierung auch zielgerichtet

gefunden werden. Weiterhin wird von jedem Dokument aus eine Volltextsuche, ein alphabetisches und ein hierarchisches Register angeboten. Jedes Dokument listet darüber hinaus alle Standard-Pfade auf, über die dieses Dokument erreichbar ist. Ziel ist es, es dem Nutzer und seiner Zielsetzung zu überlassen, von welchen der angebotenen Navigationsmöglichkeiten er im einzelnen Gebrauch macht.

## **2 Konzeption der thematischen Suche**

### **2.1 Problemstellung**

Mit der Größe eines Informationsangebotes wächst der Stellenwert angemessener Navigationshilfen und Suchmöglichkeiten. In diesem Sinn hat *WebSite `Methodik* eine kritische Größe erreicht bzw. überschritten. Informationsangebote, die von gelegentlichen Nutzern inspiziert und punktuell genutzt werden, sind zusätzlich mit dem Problem konfrontiert, daß Nutzer erst nach und nach ein Bild davon bekommen können, zu welchen Themen Dokumente bereitgehalten werden. Außerdem können sich Themen wie "Gestaltung" oder "Retrieval" auf unterschiedlichen Informationsservern nicht nur quantitativ, sondern auch von der inhaltlichen Ausgestaltung her sehr unterschiedlich darstellen. Vielfach orientieren sich interessierte Nutzer zeitaufwendig per Navigation und recherchieren gezielt mittels Stichwortsuche.

Unter Nutzung der speziellen Voraussetzungen von *WebSite `Methodik* soll ein Konzept gefunden und vorgeschlagen werden, das eine gezielte themenorientierte Suche mit einer Orientierung über die inhaltliche Ausgestaltung des Informationsangebotes verbindet. Die Voraussetzungen betreffen

- die Größenordnung des Dokumentenbestandes (einige hundert bis einige tausend Dokumente)
- die Existenz einer Indexierung mit kontrolliertem Vokabular.

Gesucht ist also ein inhaltsorientiertes Such- und Navigationskonzept für gut erschlossene Informationsbestände, anwendbar auf klein- bis mittelgroße Dokumentensammlungen, umfangreiche Antwortmengen von Fachdatenbanken (zur Detailrecherche), oder von Subject Based Information Gateways (SBIGs) , wie sie im Web in verschiedenen Projekten zwischen Browsing- und Suchdiensten positioniert werden.

### **2.2 Lösungsmöglichkeiten**

Die Suche nach Lösungsansätzen für das gestellte Problem führt zu Arbeiten, die recht unterschiedlich motiviert sind:

**Web Mining:** Die Dynamik und spezifischen Anforderungen, Defizite und Möglichkeiten des WWW motivieren den Versuch, die Suche in einer mehr inhaltsorientierten Form zu unterstützen. Besonders bekannt ist die von Alta Vista angebotene Funktionalität, dynamisch Termcluster über eine statistische Kookkurrenzanalyse in Antwortdokumentmengen zu bilden um damit für die weitere Suche Anregungen für Inclusion oder Exclusion von Wörtern und „Begriffen“ (Clustern) zu

geben (Bourdoncle 1997). Weiter gehen Ansätze, in denen auch die Verweisstrukturen zwischen WWW-Dokumenten (und damit eine zwar unspezifische, aber dennoch intellektuell-basierte Entscheidung über inhaltliche Zusammenhänge) ausgewertet wird, um begriffliche Wortfelder (Stein 1997) oder Dokumentencluster (Kleinberg 1998) zu bilden.

**Visual Information Retrieval Interfaces:** Die graphische Darstellung und Aufbereitung von Informationsräumen hat große Potentiale, bekannte Ansätze des Information Retrieval (z.B. Clusteranalyse) mit eigenständigen Beiträgen so zu verbinden, daß die visuellen Fähigkeiten des Menschen für ein effektiveres Retrieval und ein Mehr an Orientierung genutzt werden (Rorvig 1998).

Dokumenten und Begriffs-Cluster sowie Dokument-Begriffs-Netzwerke werden als Scattermaps, als topographische Karten, als Gebirge, als Landschaften oder Räume (Däßler/Palm 1998) unter Verwendung verschiedenster Metaphern eingesetzt. So benutzt beispielsweise Lyberworld ein 3D-Raummodell auf der Basis einer Massenanziehungsmetapher (Hemmje 1993): Auf einer transparenten Kugel werden Anfrageterme interaktiv plaziert und bewegt, wobei deren „Anziehungswirkung“ auf die Dokumentenwolke im Innern der Kugel deren inhaltliche Struktur erkennbar werden läßt. Der Knowledge Browser (Däßler/Otto 1996) transformiert Antwortdokumentmengen in navigierbare 3D-Klötzchen-Landschaften.

**Graphische Umsetzung Boole'scher Suchanfragen:** Das gegenwärtig in der praktischen Anwendung immer noch vorherrschende Retrieval-Paradigma ist die logische Verknüpfung von Suchtermen durch Boole'sche Operatoren. Unabhängig von der Kontroverse über die prinzipielle Adäquatheit dieses Ansatzes gegenüber partial match-Verfahren (eine neuere Untersuchung dazu in (Paris/Tibbo 1998)) steht dieser Ansatz in einem paradoxem Spannungsfeld:

- Nutzer schätzen die Durchschaubarkeit und Nachvollziehbarkeit des Verfahrens (Frisch/Kluck 1997, S. 47), ...
- ... obwohl sie in der Anwendung der Operatoren objektiv Schwierigkeiten haben (Frisch/Kluck 1997, S. 42) und sich durchaus Retrievalergebnisse objektiv z.T. falsch erklären.

Es sind deshalb eine Reihe von Ansätzen entwickelt worden, um Boole'sche Suchen graphisch-interaktiv durchführen zu können, wobei einige davon Gewinne an Transparenz und Kontextinformation erreichen. Dazu gehören der InfoCrystal von (Spoerri 1994) sowie eine eigenständige Weiterentwicklung davon (Bürdek et al. 1998). Auch der Zugang zu intellektuell indexierten Dokumenten unter Anwendung von Methoden der formalen Begriffsanalyse, visualisiert durch Liniendiagramme, kann so aufgefaßt werden (Kollewe et al. 1994).

Der vorgeschlagene Ansatz muß das Problem der Data Mining-Ansätze nicht lösen, da bereits eine begriffsorientierte Indexierung vorliegt, die unmittelbar ausgewertet wird. Sie hat von ihrem gedanklichen Ausgangspunkt einen Bezug zu (Kollewe et al. 1994), läßt sich aber einfacher umsetzen. Methoden der Informationsvisualisierung lassen sich naheliegend bei der Schnittstellengestaltung mit einbringen (siehe Abschnitt 2.4).

### **2.3 Vorgeschlagenes Konzept**

Die Idee der thematischen Suche entwickelte sich aus der Nutzung des Schlagwortregisters in WebSite Methodik. Sie bietet an, daß der Nutzer iterativ vom Allgemeinen zum Speziellen hin sucht, und daß diese Suche gleichzeitig über das informiert, was gefunden werden kann. Die Suche liefert also Überblicksinformation über die quantitative und qualitative Ausgestaltung von Themen im vorgegebenen Dokumentenbestand.

**Start der Suche:** Die Suche startet entweder vom aktuellen Dokument aus, indem einer der indexierten Deskriptoren ausgewählt wird oder alternativ von einer speziellen Suchseite aus, in der die Liste aller Deskriptoren (Thesaurus) zur Selektion angeboten wird.

**Ergebnis der Suche:** In beiden Fällen führt die Auswahl eines Deskriptors D parallel zu zwei Ergebnissen:

- Liste aller Dokumente, die mit dem Deskriptor D indexiert sind (Abb. 2b).
- Liste aller Deskriptoren, die zusammen mit D in der Indexierung von Dokumenten vorkommen, ergänzt um die Angabe der Anzahl der Kookkurrenzen.

Tatsächlich werden zwei Deskriptorenlisten erzeugt, indem Deskriptoren, die formale Aspekte beschreiben von solchen, die inhaltliche Aspekte benennen unterschieden werden (Abb. 2a).

**Iteration:** Der Suchprozess kann fortgeführt werden:

- Die Selektion eines Dokumententitels führt direkt zu dem entsprechenden Dokument und beendet die Iteration.
- Die Selektion eines der Deskriptoren (z.B. von D2) führt erneut zu einem Suchergebnis, bei dem D durch (D UND D2) ersetzt wird.

**Interpretation:** Mit etwas Pragmatismus können wir die Dokumentenliste als die Extension des gewählten Begriffs D auf dem Server ansehen: Das Thema wird repräsentiert durch die Menge aller Dokumente, in denen es behandelt wird. Die Liste der gewichteten Deskriptoren können wir als Approximation der Intention des Begriffs D auffassen: Als Intension gilt die Menge aller Merkmale, die die Dokumente der Extension gemeinsam haben (Kollewe et al. 1994). Als Merkmale interpretieren wir die inhaltlichen Kennzeichnungen, also die zugeteilten Deskriptoren. In dieser strengen Form werden außerordentlich hohe Anforderungen an die Differenziertheit, Konsistenz und Vollständigkeit der Indexierung gestellt. Praktisch handhabbarer wollen wir demnach als Intension des Themas die Menge aller Deskriptoren ansehen, die solchen Dokumenten zugeteilt sind, die das selektierte Thema behandeln. Diese Liste von Deskriptoren, nach Häufigkeit sortiert, gibt in der Regel einen deutlich schnelleren Überblick darüber, in welcher Hinsicht das gewählte Thema ausgearbeitet ist als die Menge der Dokumententitel (Extension). Abb. 2a und 2b sollen dies am Beispiel des Themas „Gestaltung“ verdeutlichen.

Bezeichnung		Notation
Gestaltung	19	Sac
Benutzungsschnittstelle	13	Sac
Informatik	3	Sab
Informationswissenschaft	2	Saa
Hypermedia	2	Saab
Bildbearbeitung	2	Sac
Lehrinhalt	2	Sba
Didaktik	2	Sba
Informationslinguistik	1	Saaca
Klassifikationssystem	1	Saacbb
Retrieval	1	Saacbc
Evaluierung	1	Sba

Bezeichnung	Anz	Notation
Overhead-Folie	14	Fa
Beispiel	10	Fb
Publikation	2	Fa
publiziertes Dokument	2	Fa
Lehrveranstaltung	2	Fb
Definition	2	Fcbb
Hilfe	1	Fb
Statistik	1	Fb
Richtlinie	1	Fb
WebSite Methodik	1	Fcaa

*Abb. 2a: Es gibt 19 Dokumente, denen „Gestaltung“ zugeordnet ist. Dieses Thema ist eindeutig in Richtung „Benutzungsschnittstelle“ ausgestaltet (13 Dokumente). Alle anderen Begleitthemen sind randständig. Was die Form betrifft, so handelt es sich in erster Linie um Overheadfolien und Beispiele (schließt sich nicht aus!). Die dritte Spalte dient (nur) der Sortierung innerhalb eines Ranges*

Titel		Ext
„Media System Design“, ein neuer Studiengang an der Fachhochschule Darmstadt (Info 7, Knorz/Seeger)	infosieben1	htm
LDV-Forum, Jg. 1998, Heft 1: Rezension R. Köhler - Schiltz: Der Dialektometrische Atlas von Südwest-Baden (G. Schiltz)	kohler2	htm
Multimedia Ausbildung an der Fachhochschule Darmstadt. (NfD, Knorz/Seeger)	nfd1	htm
Overhead-Folie: Benutzungsoberfläche - Fallbeispiel Intranet-	boffi1	htm
Overhead-Folie: Design ist Interface-Design	diid1	htm
Overhead-Folie: GUI - Minimalistische Gestaltung von	mavoa1	htm
Overhead-Folie: Interaktionsparadigma: Direkte Manipulation	iodm1	htm
Overhead-Folie: Multimedia - Manipulationsmöglichkeit durch	ddmfbm1	htm
Overhead-Folie: Optische Wahrnehmung - Farben im Kontext	owfik1	htm
Overhead-Folie: Optische Wahrnehmung -	ownap1	htm
Overhead-Folie: Optische Wahrnehmung - Linien und	ownluf1	htm
Overhead-Folie: Visualisierung hierarchischer Strukturen	vhs1	htm
Overhead-Folie: Visualisierung von Dokumentmengen	vvdm1	htm
Overhead-Folie: Visuelle Gestaltung - 7 Dimensionen	sddva1	htm
Overhead-Folie: Visuelle Metapher in 3D-Welten	vmi3dw1	htm
Overhead-Folie: Visuelle Metaphern - Klassifikation	kvm1	htm
Overhead-Folie: Zeichnen planarer Graphen: Verschiedene	zpaav1	htm
Rezension zu: Bonsiepe. Interface. Design neu beaufreien.	bonsieb1	htm
WebSite Methodik - Vorlagen und Richtlinien für den Autor	autor2	htm

Abb. 2b: Bei nur 19 Dokumenten kann man sich noch vergleichsweise leicht einen inhaltlichen Überblick verschaffen, aber dennoch zeigt sich bereits der zusätzliche Orientierungseffekt durch die verdichtete deskriptorbasierte Darstellung in Abb. 2a.

Eine andere Interpretation faßt die generierten Deskriptorenlisten als Repräsentation einer dynamisch und nutzerangepaßte Hierarchie (Menüauswahl) auf. Jeder Deskriptor kann vom Nutzer an die Spitze der Hierarchie gewählt werden. Generiert wird die jeweils nächste Hierarchieebene, also die Menge der jeweils möglichen erfolgreichen Fortsetzungen einer einschränkenden Suche. Die Auswahlentscheidungen des Nutzers bestimmen also die spezifische, ihm angebotene Themenhierarchie vom Allgemeinen zum Speziellen.

**Angestrebte Vorteile:** Der wesentliche Vorteil der dargestellten Suchstrategie soll sein, daß eine hierarchische, zielgerichtete Suche unterstützt wird, ohne daß vordefinierte klassifikatorische Strukturen erforderlich sind, die leicht an der aktuellen Perspektive des Nutzers vorbeigehen können. Und außerdem soll der Nutzer, als Nebeneffekt der Suche, darüber informiert werden, was sonst er hätte



alles finden können - und was nicht. Der Nutzer soll also eine Vorstellung davon aufbauen, inwieweit Themen (repräsentiert als Deskriptoren bzw. UND-Verknüpfung von Deskriptoren) zentral oder peripher abgedeckt werden.

## 2.4 Gestaltungsentwurf

Keywords: WebSite Methodik; Index; Suche; dynamisches\_Dokument.

### Das Thema "Retrieval" auf dem Server WebSite 'Methodik'

41 Dokumente zum Thema "Retrieval"

1. Beispiel (Klassifikationssystem): "Predicts Wörterbuch und Thesaurus"
2. Beispiel (Klassifikationssystem): "RIS Klassifikation" (D. Seergel Okt. 1992)
3. Beispiel (Thesaurus): "GID-12-Thesaurus (1985)"
4. Beispiel (Thesaurus): Alcohol and Other Drug (1993)
5. Beispiel (Thesaurus): INFODATA-Thesaurus /Inhalt (1994)
6. Forschungsprojekt SIFT: Information Retrieval und Computerlinguistik
7. Internet-Quellen zum Thema "Information Retrieval Engines"
8. LDV-Forum 97/1: "Inhaltsbasiertes Information Retrieval: Die TestMing-Technologie" (S. Özgeçer)
9. Lexikon-Schlüssel: Information Retrieval-Anwendungen (C3)
10. Overhead-Folie: Das Ranking-Prinzip
11. Overhead-Folie: Dokumente entstehen...
12. Overhead-Folie: Dokumentenretrieval via Faktetretrieval
13. Overhead-Folie: Ein Dokument versteckt sich...
14. Overhead-Folie: Entwicklung von Informationssystemen Overhead-Folie: Faktetretrieval "Holen statt suchen"
15. Overhead-Folie: Faktetretrieval: Archetypische Fragestellung
16. Overhead-Folie: Offener Wettbewerb für Retrievalsysteme: TRBC
17. Overhead-Folie: Qualitative Veränderung des Retrievalproblems ...
18. Overhead-Folie: Relevance Feedback
19. Overhead-Folie: Retrieval von Bildern - Aufgabeklassifikation
20. Overhead-Folie: Retrieval von Bildern - Formähnlichkeiten

Dokumente über "Retrieval" ...

... thematisch unterteilt:		... nach Form unterteilt:	
Anz. von Dokumenten	über Retrieval UND ...	Anz. von Dokumenten	über Retrieval als ...
11	Inhaltserschließung	21	Overhead-Folie
9	Automatische Indizierung	14	Beispiel
5	Evaluierung	10	publiziertes_Dokument (Dokumentenstatus)
	Graphik	6	Script
	Hypermedia	3	Übung
	Informationenlinguistik	2	Faksimile
	Informationswissenschaft		Photo
	Thesaurus		
4	Dokumenten-Management		
3	Veranstaltung		
2	Externe Institution		
	Klassifikationssystem		
	Mediendokumentation		
	Ordnungssystem		
1	Datenbank		
	Forschungsprojekt		
	Grammatik		
	Internet		
	Publikation		
	Statistik		
	Wirtschaftsinformation		
	Wörterbuchentwicklung		

Abb. 3: Exemplarische Darstellung eines Suchergebnisses am Beispiel "Retrieval". Die s/w- Wiedergabe läßt die unterschiedliche Farbgestaltung von linker und rechter Bildschirmhälfte nur ansatzweise erkennen. Die Dokumentenliste zeigt, daß eine systematische Benennung der Dokumente den Überblick darüber unterstützt, welche Arten von Dokumente (Übung, Beispiel, Folie, Skript, etc.) vorliegen.

Es ist klar erkennbar, daß Retrieval auf diesem Server in erster Linie zusammen mit Inhaltserschließung/automatischer Indexierung behandelt wird.

Eine einfache Umsetzung des vorgestellten Konzeptes illustriert Abb. 3 an einem Beispiel. Ein in seiner Grundstruktur zweigeteilter Bildschirm bietet als Ergebnis der Auswahl des Deskriptors „Retrieval“ alle Dokumente zum Thema an, deren Durchsicht aber bereits längeres Scrollen erfordert. Parallel werden die zusammen mit „Retrieval“ indexierten Deskriptoren, nach inhaltlich und formal orientierten Deskriptoren unterschieden, aufgelistet. Die einzelnen Häufigkeitsränge werden durch Farbabstufung (in Gelb) so abgesetzt, daß jeder Rang (z.B. Rang 1 mit Inhaltserschließung, 11 mal gemeinsam mit Retrieval indexiert) gegenüber den nachfolgenden hervorgehoben erscheint.

Durch Auswahl eines der dargebotenen Deskriptoren (hier: Inhaltserschließung) wird die Suche fortgesetzt, was zu einem strukturell gleichartigen Ergebnis für das Thema „Retrieval und Inhaltserschließung“ führt (Abb. 4).

The screenshot shows a web page titled "Das Thema 'Retrieval' und 'Inhaltserschließung' auf dem Server WebSite 'Methodik'". The page includes a search bar on the left and a main content area with two columns of results. The top right corner displays keywords: "WebSite Methodik, Index, Suche, dynamisches Dokument".

**11 Dokumente zum Thema "Retrieval" und "Inhaltserschließung"**

1. Beispiel (Klassifikationssystem): "Predicative Wörterbuch und Thesaurus"
2. Beispiel (Klassifikationssystem): "RIS Classification" (D. Soergel, Okt. 1997)
3. Beispiel (Thesaurus): "GJD JZ Thesaurus (1993)"
4. Beispiel (Thesaurus): "Alcohol and Other Drug (1993)"
5. Beispiel (Thesaurus): "INFODATA-Thesaurus Inhalt (1994)"
6. LDV Forum 97/1: "Inhaltsbasiertes Information Retrieval: Die TextMining-Technologie" (S. Gessert)
7. Overhead-Folienfolge: "Indexierung komplexer Dokumente"
8. Skript: "Indexieren... Kap. 2: Extraktionsmethode"
9. Skript: "Indexieren... Kap. 7: Qualität von ..."

**Dokumente über "Retrieval" und "Inhaltserschließung"...**

... thematisch unterteilt:	... nach Form unterteilt:
Anz. von Dokumenten über "Retrieval" und "Inhaltserschließung" UND ...	Anz. von Dokumenten über "Retrieval" und "Inhaltserschließung" als ...
4 Thesaurus	7 Beispiel
3 Informationslinguistik Informationswissenschaft	4 publiziertes Dokument
2 Automatische Indexierung Evaluierung Externe Institution Klassifikationssystem	3 Skript
1 Publikation Veranstaltung Wirtschaftsinformation	2 Exsimpla
	1 Overhead-Folie

Abb. 4: Zur Fortsetzung der Suche (Abb. 3) wurde der Deskriptor Inhaltserschließung ausgewählt.

**Ranking versus Ortscodierung:** Der bisherige Vorschlag ist dadurch gekennzeichnet, daß er die häufig vorkommenden Themen an die ersten Positionen der Liste setzt (Ranking). Nur innerhalb eines Ranges kann etwa über eine Notation thematisch sortiert werden (was nur bei den kleinen Häufigkeiten einen Effekt hat, und auch dieser Effekt ist gering. Siehe dazu Abb. 2a).

Alternativ könnte auch eine Ortskodierung gewählt werden, bei der ein Deskriptor absolut oder relativ zu anderen immer an der gleichen Position einer Liste auftaucht.

**Text versus Graphik:** Alternativ zu der listenorientierten Oberfläche könnte (aufwendiger) auch eine graphische Darstellung etwa in Form eines Ausschnittes eines Begriffsnetzes gewählt werden oder als Ebene, aufgespannt vom Klassifikationsschema und der Kookkurrenzhäufigkeit, in der die einzelnen Deskriptoren eingetragen werden.

**Einfachheit versus Funktionsvielfalt:** Alternativ zur Beschränkung der Suchmethode auf eine einschränkende Suche (mittels UND-Operator) kann damit experimentiert werden, inwieweit auch Möglichkeiten anderer Suchstrategien (ODER-Operator, Einbeziehung begrifflicher Relationen) angeboten werden können, ohne das Interface und dessen Bedienung zu überfrachten.

### 3 Realisierung

#### 3.1 Vokabular und Indexierung

Grundlage der Umsetzung des vorgestellten Konzeptes der thematischen Suche

ist die Dokumentenbasis von WebSite `Methodik, ein Vokabular von ca. 100 Deskriptoren sowie eine entsprechende Dokument-Indexierung.

Das Vokabular hat die Schwäche, daß es begleitend zum Aufbau des Informationsservers entstand. Die Entscheidung über das generische Niveau des Deskriptoren fiel auf der Basis der prognostizierten quantitativen Verhältnisse und unter dem Aspekt, Anpassungen bereits indexierter Dokumente zu vermeiden. In diesem Sinne sind einige (zentrale) Deskriptoren zu allgemein gewählt und die Unterbegriffe sind Nichtdeskriptoren. Es hat andererseits den Vorteil, daß die Deskriptoren sehr zielgenau der Logik des Faches und seiner Lehrveranstaltung entsprechen, so daß in diesem Zusammenhang sehr effektiv Dokumente aufgesucht werden können.

Jeder Deskriptor ist in ein Klassifikationssystem eingebunden, um fachsystematische Sortierungen von Deskriptormengen zu ermöglichen.

### 3.2 Datenbank

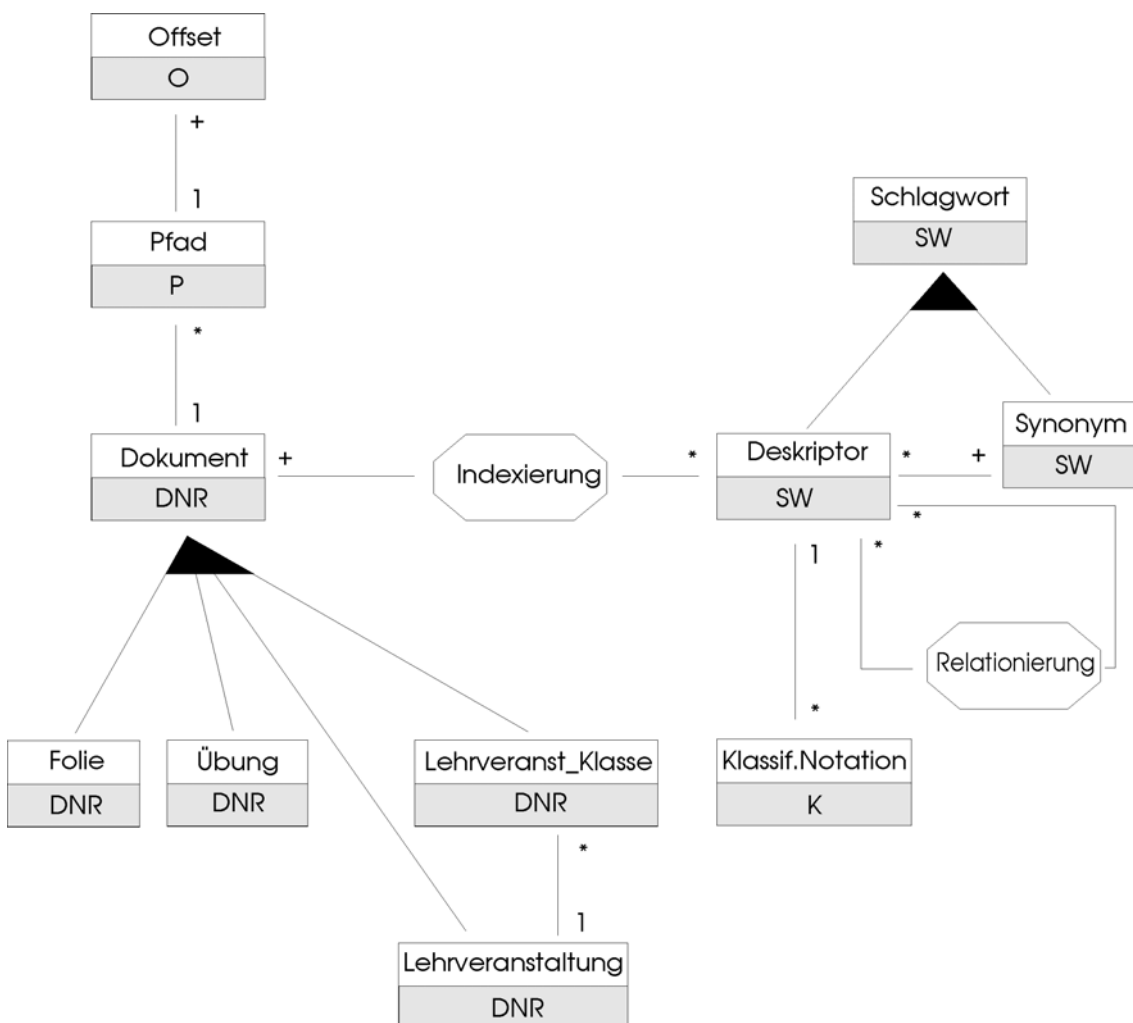


Abb. 5: Konzeptioneller Datenbankentwurf. Im Zentrum steht die Beziehung „Indexierung“ zwischen Dokumenten und Deskriptoren.

Die Datenbank stellt den Kern einer Datenhaltung dar, die in weiterer Zukunft immer mehr Aufgaben übernehmen soll. Ziel ist es, den Anteil von Informationen in statischen HTML-Dokumenten zurückzudrängen, um die entsprechenden Seiten dynamisch aus der Datenbank generieren zu können.

Den konzeptionellen Datenbankentwurf in einer Notation nach (Dennert 1991) zeigt Abb. 5. Die Umsetzung in Datenbankrelationen (unter Verwendung von ACCESS) erfolgt weitestgehend schematisch (Knorz 1996) und wird an dem SQL-Statement konkret, das die Tabelle in Abb. 2a erzeugt. Man findet in diesem Statement die Objektklassen des ER-Modells als Namen der Datenbankrelationen mit den entscheidenden Attributen wieder.

Obwohl der SELECT-Ausdruck 5 Relationen (die Relation Indexierung zwei mal) in einem Join verbindet und eine Schachtelung enthält, ergeben sich keinerlei Effizienzprobleme.

```
SELECT Schlagwort.Bezeichnung,
       count(Schlagwort.Bezeichnung) AS Anz, Klassifikation.Notation
FROM Schlagwort, Indexierung AS I1, Indexierung AS I2, Deskriptor,
     Klassifikation
WHERE I1.SW = (SELECT Schlagwort.SW
              FROM Schlagwort
              WHERE
Schlagwort.Bezeichnung="Gestaltung")
  AND I1.DNR=I2.DNR
  AND I2.SW=Schlagwort.SW
  AND I2.SW=Deskriptor.SW
  AND Deskriptor.K=Klassifikation.K
  AND Klassifikation.Notation like 'S*'

GROUP BY Klassifikation.Notation, Schlagwort.Bezeichnung
ORDER BY 2 DESC;
```

*Abb. 6: SQL-Statement zur Erzeugung der Tabelle in Abb.2. Die letzte Selektion (like 'S\*') ist dafür zuständig, das Ergebnis anhand der Klassifikationsnotation auf inhaltlich orientierte Deskriptoren einzuschränken.*

### **3.3 Stand der Entwicklung**

Im Rahmen einer Diplomarbeit am Fachbereich Information und Dokumentation ist das Konzept der thematischen Suche prototypisch realisiert worden (Windows NT Internet Server, Internet Database Connection und Access-Datenbank). Allerdings waren Datenbankstruktur und –zugriff für eine Verwendung im laufenden Betrieb nicht geeignet.

Die Abbildungen 3 und 4 sind nicht originale Bildschirmabzüge der vollzogenen Systemumstellung, sondern eine von mehreren Gestaltungsstudien, die mit realen Daten das Konzept illustrieren sollen und Vorgaben für die Implementierung machen.

Mittlerweile ist die Datenbank gemäß Abb. 5 vollständig aufgebaut. Alle Daten, soweit benötigt, stehen darin zur Verfügung, genauso wie die erforderlichen SQL-Abfragen. Es wird nun begonnen, das Konzept in seiner einfachsten Form (entsprechend Abb. 3 und 4) auf der Basis von ASP (Active Server Pages) für Windows NT Internet Server umzusetzen und zur Verfügung zu stellen. Direkt anschließend ist eine Evaluierung geplant.

## Literatur

[Bourdoncle 1997]

Bourdoncle, F.: *LiveTopics – recherche visuelle d'information sur l'internet*. In: Dossiers de l'Audiovisuel, La Documentation Française, (1997),74, S. 36-38

[Beheim et al. 1998]

Beheim et al.: *Distance Learning for Students via Internet - a new Catchword for Studying?* Erscheint in: Shaping the Knowledge Society. Proceedings of BOBCATSSS 1998, Budapest, 1998.

Preprint: <http://www.iud.fh->

[darmstadt.de/iud/wwwmeth/paper/bobcat97/paper/vortrag/DL-Proceeding-HTML/DL-Proceeding-Final.html](http://www.iud.fh-darmstadt.de/iud/wwwmeth/paper/bobcat97/paper/vortrag/DL-Proceeding-HTML/DL-Proceeding-Final.html)

[Bürdek et al. 1998]

Bürdek, B.E.; M. Eibl; J. Krause: Visualisation in Document Retrieval: An example of the integration of software ergonomics and an aesthetic quality in design. <http://www.bonn.iz-soz.de/research/information/vri98/vri98eng.htm> (vorläufige Fassung; erscheint in den Proc. zu: Visual Representations and Interpretations 98, Liverpool, September 1998), 1998.

[Däßler/Otto 1996]

Däßler, R.; A. Otto: *Knowledge Browser - ein VRML-basiertes Navigationstool für Information Retrieval Systeme im World Wide Web*. Internationales Symposium für Informationswissenschaft (Proc. ISI'96, Berlin), Konstanz: Universitätsverlag Konstanz, 1996, S. ????

[Däßler/Palm 1998]

Däßler, R.; H. Palm: *Virtuelle Informationsräume mit VRML . Information recherchieren und präsentieren in 3D*. Heidelberg: dpunkt, 1998

[Dennert 1991]

Denert, E.: *Software Engineering*. Berlin, Heidelberg et al.: Springer, 1991.

[Frisch/Kluck 1997]

Frisch, E.; M. Kluck: *Pretest zum Projekt German Indexing and Retrieval Testdatabase (GIRT) unter Anwendung der Retrievalsysteme Messenger und freeWAISsf*. Informationszentrum Sozialwissenschaften Bonn, IZ-Arbeitsbericht Nr. 10, 1997

[Hemmje 1993]

Hemmje, M.: *Eine inhaltsorientierte, intuitive 3D-Benutzerschnittstelle für Information-Retrieval-Systeme*. In: Knorz, G., Kraus, J. & Womser-Hacker, C. (Hr.): *Information Retrieval '93. Von der Modellierung zur Anwendung (Proc. IR'93, Regensburg)*, Konstanz: Universitätsverlag Konstanz, 1993, S. 82-99

[HIS 1998]

HIS: Dokumentation medienunterstützten Lehrens und Lernens.  
HIS: Hochschul-Informations-System-GmbH. (Recherchierbare WWW-Adressen-Sammlung zum Einsatz elektronischer Medien an Hochschulen).  
<http://www.his.de/abt3/proj/676/index.html> (20.8.1998).

[Kleinberg 1998]

Kleinberg, J.: *Authorative Sources in a Hyperlinked Environment*. Proceedings of the 9<sup>th</sup> ACM-SIAM Symposium on Discrete Algorithms, 1998

[Knorz 1996]

Knorz, G.: *Datenbank-Entwurfsmethoden*. In: Buder, M; Rehfeld, W.; Seeger, Th. (Hr.): *Grundlagen der praktischen Information und Dokumentation*. Bd. 2., 4. Ausgabe. München, London, New York, Paris: K.G. Saur, 1996, S. 664-687

[Knorz 1998]

Knorz, G.: *WebSite 'Methodik als zentrale Informationsressource für Lehren und Lernen*. LDV-Forum 15(1998), 1, S. 45-58  
Preprint: <http://www.iud.fh-darmstadt.de/iud/wwwmeth/paper/ldvf97b/paper1.htm>

[Kollewe et al. 1994]

Kollewe, W.; M. Skorsky; F. Vogt und R. Wille: *TOSCANA - ein Werkzeug zur begrifflichen Analyse und Erkundung von Daten*. In: R. Wille; M. Zickwolf (Hr.), *Begriffliche Wissensverarbeitung - Grundfragen und Aufgaben*, B.-I. Wissenschaftsverlag, Mannheim-Leipzig-Wien-Zürich, 1994, S. 267-288.

[Paris/Tibbo 1998]

Paris, L.A.H.; H.R. Tibbo: *Freestyle vs. Boolean: A Comparison of Partial and Exact Match Retrieval Systems*. *Information Processing and Management*, 34(1998),2,3, S. 175-190

[Rorvig 1998]

Rorvig, M.: *Scaled Structure in Visualized TREC Data and Query Feedback*. *Information Processing and Management*, 34(1998),2,3: S. 151-160

[Spoerri 1994]

Spoerri, A.: *InfoCrystal: A Visual Tool for Information Retrieval & Management*. In: M. Galer; S. Harker; J. Ziegler (Hr.): *Conference on Human Factors in Computing Systems (Proc. ACM CHI'94)*, 1994, S. 11-12

[Stein 1997]

Stein, P.: *Verbesserung von Suchindizes im World Wide Web*. Ludwig-Maximilians-Universität München, CIS-Bericht, 1997, S. 97 - 105