

Der thematische Zugang zu Archiven und Dokumentbeständen - Grundprinzip und erweiterte Recherchemöglichkeiten

White Paper - TopicZoom GmbH

21. Oktober 2009

Zusammenfassung

Großarchive unterschiedlicher Einrichtungen ebenso wie Presse, Bibliotheken und Verlage stehen derzeit vor der Aufgabe, umfassende Dokumentbestände dem internen oder externen Publikum im Internet zugänglich zu machen. Gerade im Anblick einer ständig wachsenden Menge von Dokumenten, die in elektronischer textueller Form vorliegen, wird immer deutlicher, dass die traditionelle Volltextsuche keinen zufriedenstellenden Zugriff auf Dokumente ermöglicht, da sie kein Bild des Archivs und der auftretenden Themenbereich vermittelt und den Benutzer bei der Auswahl von Suchbegriffen alleine lässt. Als Gegenentwurf stellen wir den thematischen Zugriff auf Archive vor. Bei diesem Ansatz wird unter Einsatz einer Themenhierarchie nach einer speziellen Form der automatischen Indexierung dem Benutzer ein echter Überblick über das Archiv gegeben, in welchem allgemeine und spezielle Themen des Archivs mit ihrem Gewicht verzeichnet sind. Der Benutzer kann damit im Themenverzeichnis navigieren, um zu interessanten Dokumenten zu gelangen. Die Stärke des Ansatzes beruht auch auf einer ganzen Reihe weiterer interessanter Interaktionsformen, die dem Benutzer zur Verfügung stehen. In dieser Arbeit stellen wir die Grundprinzipien des thematischen Zugangs sowie nützliche erweiterte Recherchemöglichkeiten vor.

Inhaltsverzeichnis

1	Einleitung	2
2	Grenzen der traditionellen Volltextsuche	3
3	Der thematische Zugang - Grundprinzipien	5
3.1	Die Themenhierarchie als "thematische Landkarte"	5
3.2	Thematische Navigation - "Google Maps" auf dem Archivverzeichnis	7
4	Die TopicZoom Themenhierarchie	10
5	Erweiterte Funktionalitäten und Recherchemöglichkeiten	11
5.1	Vernetzung mit Wissensquellen	11
5.2	Vernetzung zwischen Texten und Themenhierarchie	12
5.3	Themenanalyse für Personen und Unternehmen	12
5.4	Vergleichende Recherchen	14
5.5	Facettierte thematische Recherche	14
6	Schlussbetrachtung	16

1 Einleitung

Im Pressebereich, bei Bibliotheken und Verlagen wächst derzeit der Druck, vorhandenes Archivmaterial nicht unerschlossen liegen zu lassen, sondern es internen und externen Benutzern im Internet zur Recherche zur Verfügung zu stellen. Selbst wenn das Material in elektronischer Textform vorliegt, fehlen dennoch auch heute in aller Regel übersichtliche Benutzerschnittstellen, die einen wirklich befriedigenden Zugriff auf das Archiv und seine Inhalte ermöglichen. Die heutige Standardsuchtechnologie, die auf dem Prinzip der Volltextsuche beruht, löst das Problem nicht wirklich. Ist der Benutzer noch nicht mit dem Archiv vertraut, ist es schwierig, gute Suchbegriffe zu finden. Auch nach vielen Anfragen bleibt die Vorstellung vom Archiv und seinen Inhalten oft sehr vage und wertvolle Inhalte bleiben ungenutzt.

Der hier vorgestellte *thematische Zugang* zu Archiven unterstützt gleichermaßen erfahrene wie unerfahrene Archivnutzer bei der Recherche, indem er am Bildschirm übersichtliche Formen der Navigation anbietet. Bei der vollautomatischen Indexierung der Dokumente wird eine umfangreiche Ontologie/Taxonomie von Themen eingesetzt. Als Ergebnis wird eine Art thematisches Inhaltsverzeichnis des Archivs berechnet, in welchem allgemeine und spezielle Themen und ihr Gewicht im Archiv verzeichnet sind. Bewegt sich der Benutzer in diesem Verzeichnis, werden für jedes Thema die wichtigsten Dokumente nebst den dort auftretenden themenspezifischen Schlüsselbegriffen angezeigt. Das Ausprobieren spezieller Suchbegriffe entfällt, statt dessen sind die Schnittstellen darauf

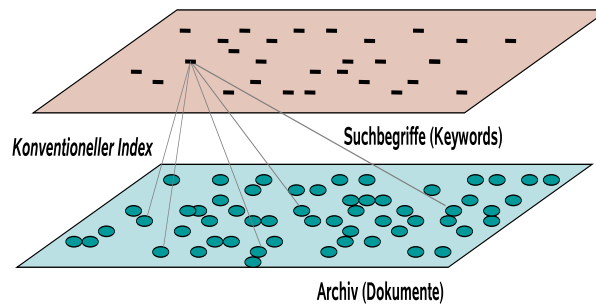


Abbildung 1: Architektur eines konventionellen Volltextindexes.

ausgerichtet, beim Informationssuchenden ein strukturiertes Bild des Archivs und seiner Inhalte entstehen zu lassen. Eine ganze Reihe von weiteren Interaktionsmöglichkeiten lässt sich zu unterschiedlichen Formen der Recherche nutzen oder schafft Vernetzungen zu anderen Wissensquellen.

Ziel dieses Berichts ist es, die Grundzüge des thematischen Zugangs sowie die sich ergebenden erweiterten Recherchemöglichkeiten allgemeinverständlich und ohne vertieften wissenschaftlichen Hintergrund darzustellen.

2 Grenzen der traditionellen Volltextsuche

Die traditionelle Volltextsuche beruht aus informationstechnischer Sicht auf der in Abbildung 1 dargestellten Architektur. Zu dem Archiv, das in der Regel eine sehr große Menge von Dokumenten umfasst (untere Ebene), wird ein konventioneller Index aufgebaut. Dieser Index beinhaltet eine umfangreiche Menge von *Schlüssel-* oder *Suchbegriffen* (obere Ebene). Zu jedem Schlüsselwort ist im Index gespeichert, in welchen Dokumenten (ggfs. an welchen Textpositionen) der Suchbegriff auftritt.

Abbildung 1 suggeriert allerdings eine Panorama-Vogelperspektive, die in der Realität deutlich anders aussieht. Vom eigentlichen Archiv, seiner Struktur und seinen Inhalten, sieht der Benutzer nichts. De facto besteht sein "Blick auf das Archiv" aus einem weißen Suchfenster, das auf seine Eingabe wartet. In der Regel hat der Benutzer nur eine sehr vage Vorstellung, welche Suchbegriffe potentiell auf interessante Dokumente führen (Abbildung 2). Tippt der Benutzer einen Suchbegriff ein, erhält er eine gerankte Trefferliste, in der die Titel der betreffenden Dokumente aufgelistet sind. Zu jedem Dokument werden zudem Textausschnitte präsentiert, in denen der Suchbegriff auftritt (Abbildung 3).

Diese Art des Zugriffs ist mit einer Reihe von Schwächen verbunden.

- Ist man mit den Inhalten des Archives noch nicht hinreichend vertraut, so ist es keinesfalls offenkundig, welche Suchbegriffe denn überhaupt Sinn machen.

Suchbegriffe (Keywords)

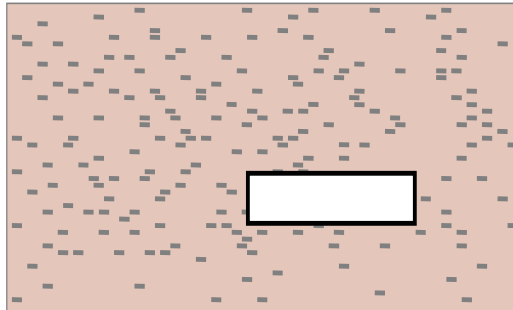


Abbildung 2: Volltextsuche - Sicht des Benutzers auf das Archiv.

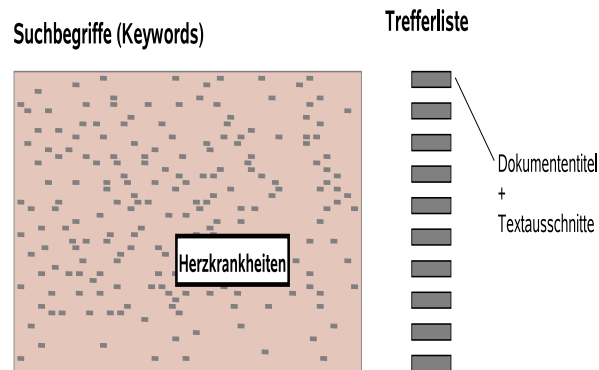


Abbildung 3: Standardinformation zu Treffern bei der Volltextsuche.

- Der Benutzer erhält keinen Überblick über das Archiv, die dort auftretenden Themen und Schwerpunkte.
- Benutzer, die an spezifischen Themen (Bayern, Nachkriegsdeutschland, Herzkrankheiten, ...) interessiert sind, werden nicht angesprochen. Tatsächlich sind spezifische Themen oft mit sehr vielen denkbaren Stichworten verbunden, so dass auch nach Eintippen einiger Suchbegriffe interessante Dokumente im Dunkeln bleiben.
- Das lästige Eintippen und Ausprobieren von Keywords ist die einzige Form der Interaktion. Das System bleibt passiv.
- Vergleichende Analysen, Statistiken zu Themenbereichen und andere weitergehende Funktionalitäten werden nicht unterstützt.

In Abwandlung einer bekannten Redeweise kann man zusammenfassend sagen, dass die Volltextrecherche dem Angebot gleicht, sich eine von zig- oder hunderten von Nadeln (sprich Suchbegriffen) auszusuchen (Abbildung 2). Hat man eine dieser Nadeln gewählt, bekommt man quadratzentimeterweise Ausschnitte des Heuschobers präsentiert, wo diese Art Nadel zu finden ist. Einen echten Überblick über den Heuschober und seine Inhalte, damit einen vernünftigen Zutritt, gewinnt man auf diese Weise schwerlich.

Suchen wir Nadeln? Web versus Archiv

Es ist an dieser Stelle anzumerken, dass die Frage nach dem Sinn des “Nadel-suchens” bei der Websuche ganz anders zu beantworten ist als bei der Suche in Archiven. Tatsächlich stellt ein großer Teil der Anfragen, die an Internetsuchmaschinen gestellt werden, eine Art Nadelsuche dar, etwa wenn nach der Adresse oder Telefonnummer einer Organisation oder nach der Homepage einer Person gesucht wird. Eine anderer erheblicher Teil der Anfragen im Internet betrifft bestimmte Formen von Transaktionen, wie Hotelbuchungen oder Buchbestellungen. Lediglich ein relativ kleiner Teil der Anfragen im Internet ist darauf ausgerichtet, interessante Dokumente zu einem bestimmten Thema zu finden. Ganz anders stellt sich hingegen der Zugriff auf Archive und Bibliotheksinhalte dar. Hier steht in aller Regel die Suche nach Informationen zu vorgegebenen Themen im Vordergrund. Die Andersartigkeit der Benutzerinteressen begründet, dass Designprinzipien von Websuchmaschinen keinesfalls direkt auf die Archiv- und Bibliothekssuche zu übertragen sind.

3 Der thematische Zugang - Grundprinzipien

Als Alternative zu einer konventionellen Volltextsuche bietet TopicZoom einen neuen Ansatz. Im Vordergrund steht das Anliegen, dem Benutzer einen echten Überblick über das Archiv zu vermitteln, der mit neuen Möglichkeiten zum “thematischen Navigieren” in der Dokumentensammlung verbunden ist. Wir stellen zunächst die Architektur und die sich hieraus ergebenden Grundfunktionalitäten beim Archivzugriff dar. In nachfolgenden Abschnitten gehen wir auf eine ganze Reihe zusätzlicher Möglichkeiten zur Recherche dar, die sich aus dem Ansatz in natürlicher Weise ergeben.

3.1 Die Themenhierarchie als “thematische Landkarte”

Die Grundarchitektur beim thematischen Zugriff geht zunächst von einem konventionellen Volltextindex aus und erweitert diesen (Abbildung 4). Über dem Raum der Suchbegriffe ist ein sehr umfangreicher *hierarchisch strukturierter Themenraum* aufgespannt, der von sehr allgemeinen Themen wie Politik, Wirtschaft, Sport, Religion und Weltanschauung bis zu sehr speziellen Themen wie Energiesparlampen, Seniorenheime, Kindergeld, Basketball-Bundesliga reicht. Abbildung 4 verdeutlicht das Prinzip, wobei die Hierarchie von allgemeinen Themen (oben dargestellt) zu spezielleren Themen (unten) führt.

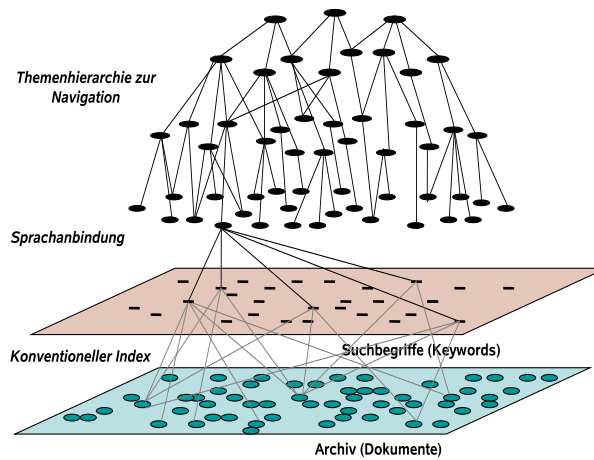


Abbildung 4: Architektur beim thematischen Zugang: Volltextindex, Themenhierarchie und Sprachanbindung.

Um eine sprachliche Anbindung zu erreichen, die eine automatische thematische Indexierung von Texten erlaubt, ist jedes Thema mit einer Liste “zum Thema gehöriger” Schlüsselbegriffe verknüpft.

Mit dieser Architektur erhalten wir eine bidirektionale Verknüpfung zwischen den Dokumenten des Archivs und den Themen der Hierarchie. Wie in Abbildung 5 angedeutet, können wir zu jedem Thema alle Dokumente des Archivs ermitteln, die Begriffe zu diesem Thema enthalten. Auch die *Anzahl* der themenrelevanten Begriffe in jedem Dokument wird erkannt. Damit können wir dem Benutzer die für das Thema wichtigsten Dokumente anzeigen.

Mindestens ebenso interessant ist auch der umgekehrte Blick vom Archiv (unten) zur Themenhierarchie (oben). Wie Abbildung 6 zu verdeutlichen sucht, stellen die sprachlichen Verbindungen eine Art Fluss dar, der zeigt, welche Themen im Archiv oft und welche selten auftreten. Hieraus ergeben sich eine Reihe wichtiger Möglichkeiten:

1. Die Themenhierarchie kann so eingeschränkt werden, dass wirklich nur die im Archiv auftretenden Themen integriert sind.
2. Themen können nach ihrer Wichtigkeit (Auftrittshäufigkeit) im Archiv geordnet und verglichen werden.
3. Mittels geeigneter Visualisierungen können dem Benutzer auf allen Ebenen der Hierarchie “thematische Landkarten” angeboten werden, welche die Teilgebiete und ihre Bedeutung im Archiv anzeigen.

Insgesamt erhält der Benutzer damit einen Überblick über die Inhalte des Archivs, der von allgemeinen bis speziellen Themen reicht. Abbildung 7 verdeutlicht diese Idee. Auf der obersten Ebene wird das Gewicht allgemeiner Themen

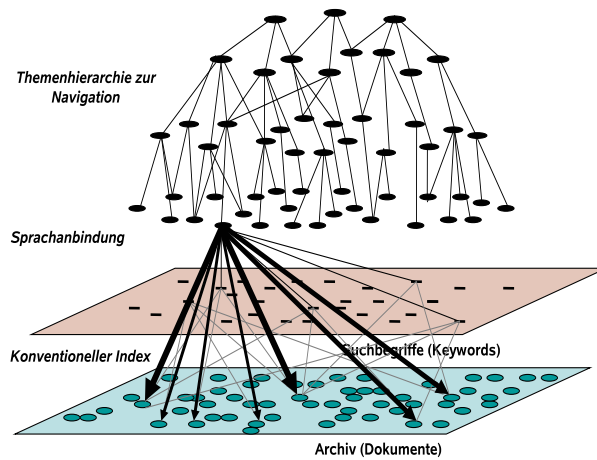


Abbildung 5: Der Fluss von oben nach unten zeigt die interessantesten Dokumente zu einem Thema der Hierarchie.

wie “Religion”, “Politik”, “Gesundheit”, “Sport” etc. angezeigt. Wandert der Benutzer zu spezielleren Themen wie politischen Parteien, sieht er auch hier das relative Gewicht der Teilthemen.

3.2 Thematische Navigation - “Google Maps” auf dem Archivverzeichnis

Die einfachste Form des Benutzerzugriffs besteht in der Navigation im Themenverzeichnis. Diese kann als eine Art “Google Maps” auf der “Weltkarte der Archivthemen” betrachtet werden. Der Benutzer sieht zu Beginn die allgemeinsten Themen und ihr relatives Gewicht. Wählt er nun etwa das Thema “Gesundheit” aus, so zeigt eine Trefferliste die interessanten Dokumente zum Bereich Gesundheit. Jedes dieser Dokumente ist auf eine Weise portraitiert, die sofort einen detaillierteren Eindruck vermittelt: das Portrait zeigt die im Dokument auftretenden *Begriffe zum Bereich Gesundheit* an sowie eine Auswahl weiterer Schlüsselbegriffe.

Das derart vermittelte nähere Bild sowie der Blick auf die Teilthemen des Bereichs Gesundheit im Themenverzeichnis können nun den Benutzer veranlassen, zum Teilthema “Krankheiten” und damit zur nächstfeineren “Landkarte” zu zoomen. Wieder werden in der Trefferliste die interessantesten Dokumente zum fokussierten Thema angezeigt, wobei in den Dokumentenportraits nun die Schlüsselbegriffe zum Thema “Krankheiten” hervorgehoben sind. In einer von vielen denkbaren Fortsetzungen wandert der Benutzer aufgrund des sich ergebenden Eindrucks vom Archiv zum Teilthema “Herzkrankheiten” und weiter zu “Vorhofflimmern”, wo er nun die wichtigsten Dokumente explizit inspiziert.

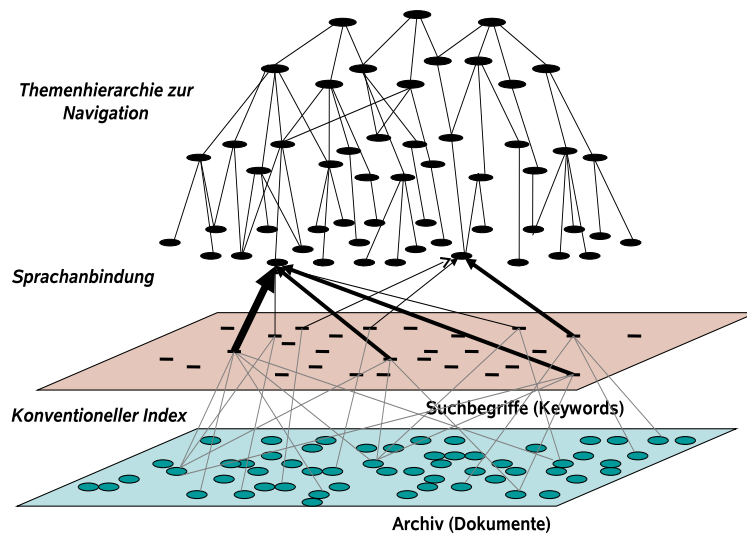


Abbildung 6: Der Fluss von unten nach oben zeigt die Bedeutung der Themen im Archiv.

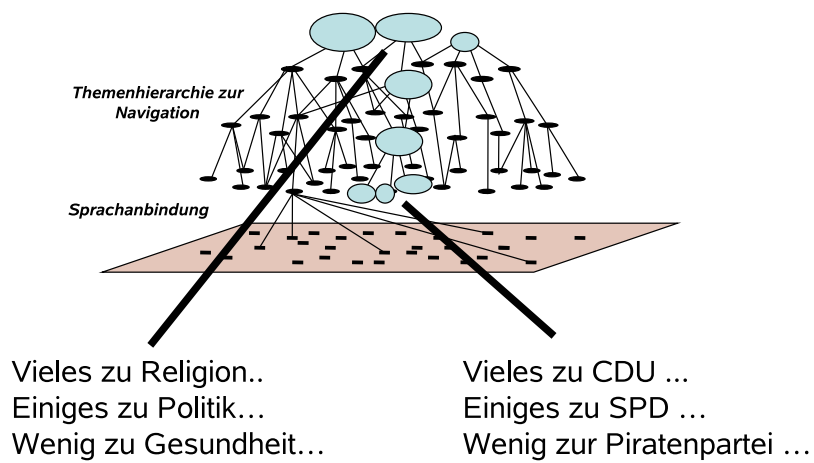


Abbildung 7: Hieraus abgeleitete "thematische Landkarten" dienen als Abbild des Archivs zur Überblicksvermittlung.

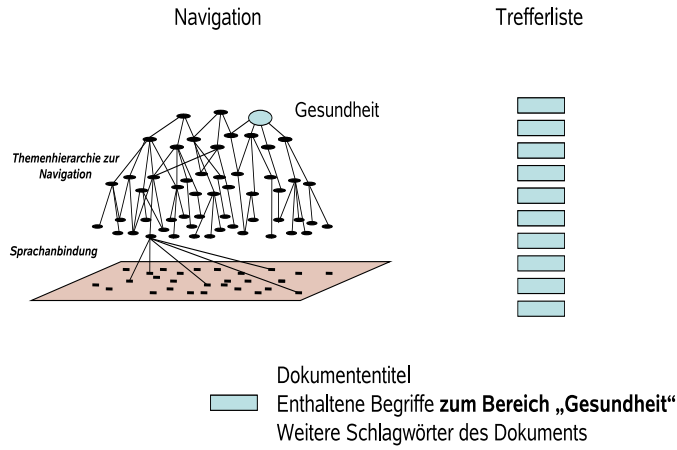


Abbildung 8: Thematische Navigation - Ausgangspunkt beim Thema “Gesundheit”. Die Trefferliste zeigt die interessantesten Dokumente. Zu jedem Dokumenten werden die dort auftretenden Begriffe des Themenbereichs “Gesundheit” sowie weitere Schlüsselbegriffe angezeigt.

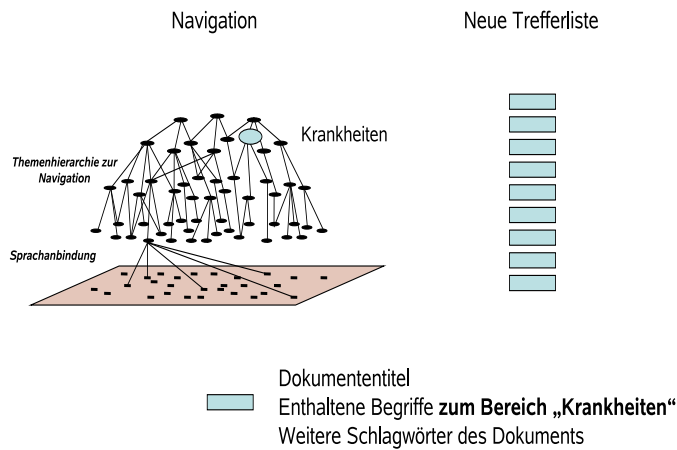


Abbildung 9: Erster Zooming-Schritt zum Thema “Krankheiten”. In den Dokumentenportraits sind nun Begriffe zum Thema “Krankheiten” hervorgehoben.

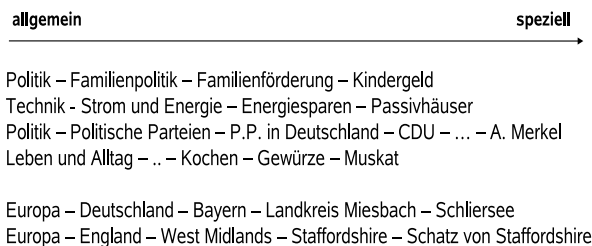


Abbildung 10: Beispiele für Themen und Verfeinerungen in der TopicZoom Hierarchie.

4 Die TopicZoom Themenhierarchie

Der themengesteuerte Zugriff auf das Archiv ist umso wertvoller, je vollständiger und besser ausgearbeitet die zugrundegelegte Themenhierarchie ist, wobei die sprachliche Anbindung eine Schlüsselrolle spielt.

Die TopicZoom Themenhierarchie enthält nicht weniger als 80.000 Themen, sie wird ständig ergänzt und erweitert. Damit ähnelt sie in ihrer Abdeckungsbreite und -Tiefe einer umfangreichen Enzyklopädie. Wichtige Bestandteile sind

1. eine geographische Teilhierarchie mit Kontinenten, Staaten, administrativen Gliederungen und Städten,
2. eine temporale Teilhierarchie mit Jahren, Jahrzehnten, Jahrhunderten und wichtigen zeitlichen Epochen,
3. eine Vielzahl thematisch erfasster Entitäten von Personen über Organisationen und Ereignisse bis hin zu Kunstwerken oder politischen Verträgen,
4. eine deutsche und englische sprachliche Anbindung, die eine automatische Indexierung deutscher und englischer Texte ermöglicht.

Die geographische Teilhierarchie erlaubt es, als Spezialfall der thematischen Navigation eine geographische Navigation zu wählen. Hierbei navigiert der Benutzer etwa von “Europa” über “Deutschland” zu “Baden-Württemberg”. Bei jedem Schritt umfasst die Trefferliste Dokumente, die Orte oder andere “lokalisierbare” Begriffe des gewählten geographischen Raums enthalten.

Die temporale Teilhierarchie bietet die Möglichkeit, auf Dokumente mit Inhalten zu speziellen Jahrhunderten und Zeitperioden (Mittelalter, Nachkriegszeit, Renaissance, ...) zuzugreifen.

Die in der TopicZoom Hierarchie kodierten Informationen umfassen ca. 6 Millionen Zuordnungen der Form “Begriff X gehört zu Thema Y”. Abbildung 10 stellt exemplarisch einige Beispiele für Themen und Verfeinerungen in der Hierarchie dar.

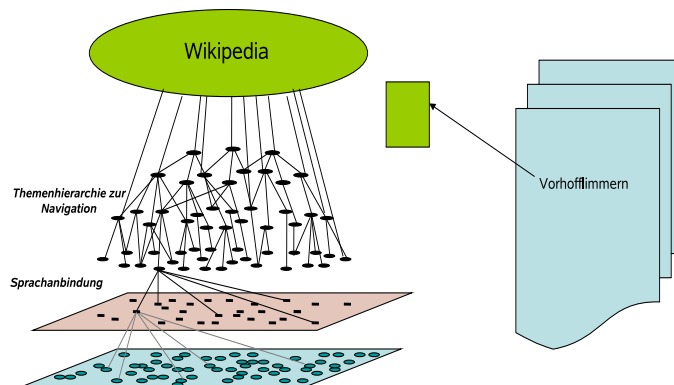


Abbildung 11: Vernetzung mit der Wikipedia. Zum Thema “Vorhofflimmern” ist in der Hierarchie der Link auf die entsprechende Wikipediaseite abgespeichert. Der Text wird bei der maschinellen Bearbeitung automatisch verlinkt. Klickt der Benutzer danach angereicherte Begriffe im Text an, wird er auf die entsprechenden Wikipediaseiten geführt.

Im Hinblick auf Anwendungen im Presse- und Zeitungsbereich wurde die sogenannte IPTC-Klassifikation in der Themenhierarchie miteinbezogen.

5 Erweiterte Funktionalitäten und Recherchemöglichkeiten

Im nachfolgenden stellen wir einige höhere Funktionalitäten vor, die sich beim thematischen Zugang auf Grundlage der in Abbildung 4 dargestellten Architektur leicht realisieren lassen und die zu zusätzlichen interessanten Recherche- und Interaktionsmöglichkeiten der Benutzer führen.

5.1 Vernetzung mit Wissensquellen

Die Konzepte einer Themenhierarchie können leicht mit externen Wissensquellen, wie zum Beispiel Wikipedia, Wörterbüchern, Artikeln aus anderen Kollektionen oder diversen Online-Angeboten etc. vernetzt werden. Bei der in Abbildung 11 dargestellten Erweiterung wurden zu den Themen in der Hierarchie entsprechende Seiten der Wikipedia mit abgespeichert. Bei der Bearbeitung der Texte können nun Begriffe im Text, die Themen der Hierarchie darstellen, automatisch mit einem Link zu Wikipedia versehen werden. Klickt der Benutzer daraufhin ein Vorkommen des Wortes “Vorhofflimmern” im Text an, so wird er direkt auf die Wikipediaseite zum Begriff “Vorhofflimmern” geführt.

5.2 Vernetzung zwischen Texten und Themenhierarchie

In einer anderen Variante werden die Begriffe der Texte direkt mit den Konzepten in der Themenhierarchie verbunden. Das Prinzip ist in Abbildung 12 dargestellt, wo der Benutzer Dokumente zum Thema "Vorhofflimmern" inspiziert und in einem der Dokumente auf den Begriff "Kardiologie" stößt. Durch ein Anklicken der betreffenden Textpassage springt man in der Hierarchie direkt zum Thema "Kardiologie". Als Treffer werden nun demgemäß die wichtigsten Dokumente zum Thema "Kardiologie" angezeigt.

Diese Art der Vernetzung von Textbausteinen mit der Themenhierarchie lädt den Benutzer auf besondere Weise zum "Stöbern" im Archiv ein. Neue Begriffe in den inspizierten Texten führen direkt in verwandte thematische Bereiche des Archivs und damit auf neue Trefferlisten. Diese enthalten wiederum interessante Begriffe, wodurch sich der Kreis schließt (vgl. Abbildung 13). Im Information Retrieval wird diese Form der Benutzerinteraktion manchmal durch das Bild des "Beerenpflückens" verdeutlicht. Auch beim Beerenpflücken (oder Pilzsammeln) im Wald verfolgt man keine gerade Linie, sondern lässt sich durch die Funde treiben.

5.3 Themenanalyse für Personen und Unternehmen

Im Presseumfeld, speziell beim Pressemonitoring, ist es ein häufiges Anliegen, die Medienresonanz von unterschiedlichen Personen oder Organisationen zu untersuchen. Ein um sein Renomee besorgtes Unternehmen bzw. eine in der Öffentlichkeit stehende Person mag sich zum Beispiel fragen, in welchen thematischen Kontexten die Firma (Person) erwähnt ist.

Für Pressearchive ergeben sich hier durch den thematischen Zugriff optimale Möglichkeiten. Es muss lediglich in einem ersten Schritt die Menge der Dokumente, die die Firma (Person) erwähnen, als "Teilarchiv" ausgesondert werden. Daraufhin können alle Trefferzahlen auf dieses Teilarchiv eingeschränkt werden, wodurch sich der erwünschte Überblick ergibt.

Erweitert man diese Idee, kommt man zu einem allgemeineren Begriff der "Queranalyse": hierbei werden die Dokumente zu einem beliebigen Thema unter einem zweiten Gesichtspunkt thematisch analysiert. Einige Beispiele mögen dies verdeutlichen:

- Wir analysieren die Dokumente zum Thema "Doping" und fragen, welche Sportarten (oder Staaten,...) dort erwähnt sind.
- Wir analysieren die Dokumente zum Thema "Ernährung" und fragen, welche Krankheiten oder Gesundheitsthemen dort erwähnt sind.
- Wir analysieren die Dokumente zum Thema "Religion" und fragen, welche Zeiträume und Epochen dort angesprochen werden.

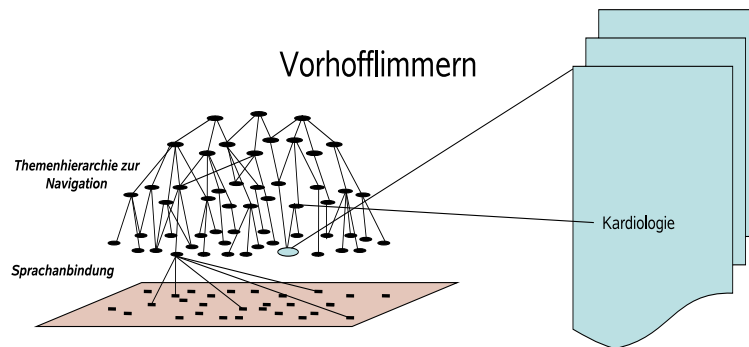


Abbildung 12: Vernetzung von Textbegriffen mit der Themenhierarchie. Das Anklicken von "Kardiologie" im Text führt zum Thema "Kardiologie" der Hierarchie und zu einer neuen Trefferliste.

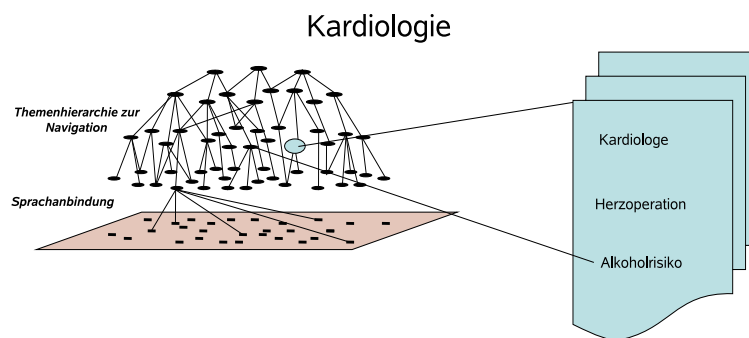


Abbildung 13: Begriffe in den neuen Trefferlisten führen auf neue Themen.

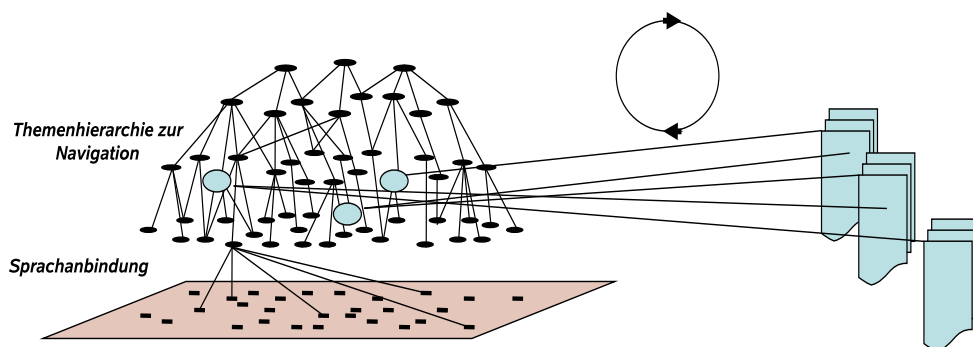
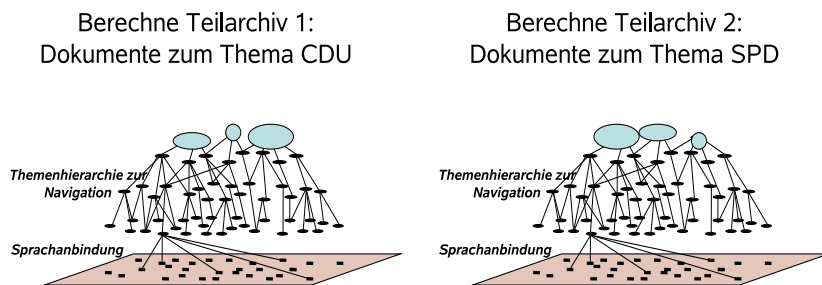


Abbildung 14: Bild vom "Beerenpflücken": Interessante Begriffe in Dokumenten führen auf neue Themen, deren Trefferlisten enthalten neue interessante Begriffe...



Wie unterscheidet sich die thematische Struktur? Welche Themen sind beim „CDU-Teilarchiv“ häufiger als im „SPD-Teilarchiv“ und umgekehrt?

Abbildung 15: Vergleich der thematischen Struktur der Dokumente zum Thema „CDU“ mit der thematischen Struktur der Dokumente zum Thema „SPD“.

5.4 Vergleichende Recherchen

Eine natürliche Fortsetzung der im vorherigen Abschnitt genannten Techniken besteht darin, die thematische Struktur unterschiedlicher Teilarchive miteinander zu vergleichen. So mag man sich nach einem Wahlkampf fragen, inwiefern bestimmte Parteien bevorzugt mit ganz bestimmten Themen in Verbindung gebracht wurden.

Die in Abbildung 4 dargestellte Architektur beim thematischen Zugang ermöglicht es, solche Vergleiche vollautomatisch zu machen (vgl. Abbildung 15). Wollen wir in einem Pressearchiv zum Beispiel die mit den Parteien CDU und SPD verbundenen Themen in Presseberichten vergleichen, so berechnen wir zunächst die

- Kollektion aller Dokumente, die das Thema „CDU“ erwähnen, und die
- Kollektion aller Dokumente, die das Thema „SPD“ erwähnen.

Für beide Teilarchive berechnen wir nun die thematische Struktur mit den auftretenden Themen und ihrem Gewicht. Wichtige Unterschiede, zum Beispiel in den Themengebieten „Wirtschaft“ oder „Soziales“, können sofort festgehalten werden.

5.5 Facettierte thematische Recherche

Als letzte Erweiterungsmöglichkeit stellen wir die themenbasierte facettierte Recherche vor. Hierbei erfolgt die Navigation mittel mehrerer orthogonaler Richtungen, mit denen die Menge der relevanten Dokumente gleichzeitig unter meh-

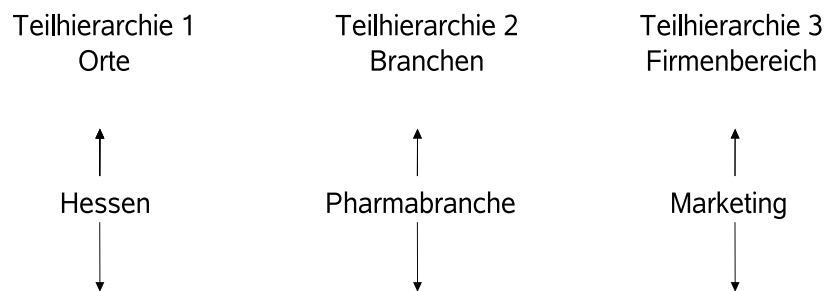


Abbildung 16: Facettierte Recherche.

renen unterschiedlichen Aspekten eingeschränkt wird. Für ein einfaches Beispiel stellen wir uns ein Szenario vor, wo eine arbeitsvermittelnde Stelle

- ein umfangreiches Archiv von Lebensläufen oder Bewerbungen zu verwalten hat und
- einen Bewerber für eine bestimmte Stelle sucht.

Die Stellenanforderungen beziehen sich in der Regel auf unterschiedliche Kriterien wie z.B.

1. auf Erfahrungen in einem bestimmten Firmenbereich (Marketing, Produktentwicklung,..),
2. auf Erfahrungen in einer bestimmten Branche,
3. auf geographische Merkmale des Bewerbers.

Weitere Richtungen können hinzukommen. Zunächst ist festzuhalten, dass der thematische Zugang in einem solchen Szenario große Vorteile hat. So wird ein Lebenslauf auch dann mit "Marketing" assoziiert, wenn dieser Begriff selbst gar nicht auftritt, solange andere Schlüsselbegriffe zum Thema "Marketing" auftreten. Ein Bewerber aus München wird bei einer Suche im Bereich Bayern erfasst.

Unter Verwendung von Themenhierarchien lassen sich Bedingungen in unterschiedlichen Richtungen besonders leicht feintunen. Stellen wir uns als Beispiel ein Szenario vor, wo wir Lebensläufe mit Texttreffern zum Firmenbereich "Marketing", zur "Pharmabranche" und zum Bereich "Hessen" aus der Sammlung von Lebensläufen herausgefiltert haben (Abbildung 16). Sollte die Treffermenge nun sehr groß sein, so kann es Sinn machen, auf einer der drei Achsen in Abbildung 16 den Fokus noch weiter einzuschränken. Führt jedoch die obige Menge von Bedingungen auf eine zu kleine Menge interessanter Lebensläufe, so können wir etwa "Pharmabranche" zu "Gesundheitsbranche" verallgemeinern.

Diese und vergleichbare Formen einer facettierte Suche werden beim thematischen Zugriff direkt unterstützt.

6 Schlussbetrachtung

Der hier vorgestellte thematische Zugang zu Archiv- und Bibliotheksinhalten versucht, die allgegenwärtige “Stichwortsuche” beim rechnergestützten Archivzugriff durch eine neue themengesteuerte Form der Interaktion zu ersetzen bzw. zu ergänzen. Hierzu ist anzumerken, dass natürlich seit vielen Jahrhunderten Bibliotheksinhalte themenbasiert geordnet und recherchierbar gemacht werden. Als zentrale Hilfsmittel seien die vorhandenen Systematiken für Bibliotheksbestände und/oder Klassifikationsschemata (z.B. universelle Dezimalklassifikation) genannt. Allerdings erfolgt die Auszeichnung des Bestands in den uns bekannten Fällen stets in manueller Weise. Dies ist zum einen sehr kostspielig, hat aber auch prinzipielle Nachteile. So werden Dokumente oft nur einer oder wenigen Kategorien zugeordnet, selbst wenn eigentlich Verbindungen zu vielen Themen existieren.

Der thematische Zugriff bei TopicZoom basiert auf einer vollautomatischen Indexierung des Dokumentbestands. Ordnen und Einteilen sind Sache des Rechners. In diesem Sinn werden beim thematischen Zugang traditionell bewährte Zugriffsmechanismen, die bislang nur manuell genutzt werden konnten, auf das Niveau der automatischen Informationsverarbeitung geliftet. Bereits vorhandene Ordnungsschemata können leicht integriert werden. Der große Vorteil liegt im Kosten-Nutzen Verhältnis. Extrem umfangreiche Dokumentsammlungen können mit TopicZoom Technologie in kürzester Zeit und kostengünstig zugänglich gemacht werden. Auch ein sich täglich verändernder dynamischer Dokumentbestand mit vielen Neuzugängen wird problemlos verarbeitet. Bei der Navigation im Themenverzeichnis bleiben prinzipiell auch Dokumente sichtbar, die nur wenige themenspezifische Begriffe enthalten. Es ist nicht die Rolle des Systems, derartige Dokumente dem Benutzer prinzipiell vorzuenthalten. Stattdessen wird angezeigt, wie viele und welche thematisch einschlägige Begriffe in einem Dokument auftreten. Es ist dann Sache des Benutzers zu entscheiden, was er näher betrachten möchte.