

**Strategien und Tools zur Strukturierung und
Verwaltung heterogener Datenbestände**

Master Thesis

zur Erlangung des akademischen Grades

Master of Science, MSc

Universitätslehrgang Technische Kommunikation

eingereicht am 17.01.2005

Zentrum für Wissens- und Informationsmanagement
Abteilung Telekommunikation, Information und Medien
Donau-Universität Krems

von

Diplom-Ingenieur (FH), Diplomwirtschaftsingenieur (FH)
Leander Härter
Krems, 17.01.2005

Betreuer: Magister Franz Hable

Ich, Leander Härter, geboren am 26.08.1971 in Schweinfurt, erkläre,

- dass ich meine Master Thesis selbstständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und mich auch sonst keiner unerlaubten Hilfen bedient habe,
- dass ich meine Master Thesis bisher weder im In- noch im Ausland in irgendeiner Form als Prüfungsarbeit vorgelegt habe,
- dass ich, falls die Arbeit meine Schule/meine Unternehmung betrifft, meinen Arbeitgeber über Titel, Form und Inhalt der Master Thesis unterrichtet und sein Einverständnis eingeholt habe.

Ingolstadt, den 17.01.2005

Zusammenfassung

Informationen, die aus verschiedenen Quellen stammen und von unterschiedlichen Zielgruppen gelesen werden, sind schwierig zu organisieren. Unterschiedliche Anforderungen müssen berücksichtigt werden, um einen zielgruppengerechten Prozess aufzusetzen.

In dieser Master Thesis werden Strategien und Tools beschrieben, die zur Verwaltung heterogener Datenbestände dienen und die Arbeit mit einem System verbessern sollen. Es werden drei unterschiedliche Bereiche betrachtet, die den Zyklus der Informationsverarbeitung abbilden. Der Bereich „Publizieren“ beschäftigt sich mit dem Prozess, wie ein Dokument in eine Datenbank aufgenommen werden kann. Im Bereich „Suche und Navigation“ finden sich Hilfsmittel und Tools, um zu den gewünschten Dokumenten zu gelangen. Der Bereich „Maintenance“ schließlich beschreibt Wege, wie das vorhandene System den Anforderungen des Wachstums gerecht werden kann.

Abstract

Information created by several sources and target groups is complicated to organise. Different requirements need to be considered in order to accurately draw up a process for the target groups.

This Master Thesis is describing strategies and tools which serve the administration of the heterogeneous data set, and should improve the work with this system. Three different areas displaying the information assimilation cycle are being looked at: The area „Publish“ deals with the process of how information is stored in a database. In the area „Search and Navigate“ aids and tools can be found which help to arrive at the requested documents. Finally, the area „Maintenance“ describes ways on how the existing tools can be made suitable to deal with the requirements of a growing system.

Inhaltsverzeichnis

| | |
|--|-----------|
| Kurzzusammenfassung | 2 |
| Inhalt | 5 |
| 1 Themenbeschreibung | 7 |
| 1.1 Abgrenzung | 7 |
| 1.2 Problembeschreibung | 8 |
| 1.3 Zielbeschreibung | 9 |
| 1.4 Grundlegende Anforderungen | 9 |
| 1.5 Vorgehensweise | 9 |
| 2 Informationsarchitektur | 11 |
| 2.1 Definition der Ziele | 12 |
| 2.2 Definition der Zielgruppen | 12 |
| 2.3 Definition der Inhalte | 12 |
| 2.4 Definition einer Struktur | 13 |
| 3 Usecases | 14 |
| 3.1 Usecase Publizieren | 14 |
| 3.1.1 Usecase: Publizieren ohne Abstimmungsprozess | 15 |
| 3.1.2 Usecase: Publizieren mit Abstimmungsprozess | 16 |
| 3.2 Usecase Suchen | 21 |
| 3.2.1 Usecase: Suchen durch Navigieren | 21 |
| 3.2.2 Usecase: Navigieren zu einem bekannten Ort | 22 |
| 3.2.3 Usecase: Navigieren zu einem unbekanntem Ort | 22 |
| 3.2.4 Usecase: Suchen mit Hilfe des Suchdialogs | 24 |
| 3.3 Usecase Maintenance | 25 |
| 3.3.1 Usecase: Verbesserung der Struktur | 26 |
| 3.3.2 Usecase: Verbesserung des Inhalts | 27 |
| 3.3.3 Usecase: Verbesserung der Bedienbarkeit | 27 |
| 3.4 Zusammenfassung der Anforderungen | 27 |
| 4 Aufbau einer Struktur | 29 |
| 4.1 Methoden | 29 |
| 4.1.1 Card Sorting | 30 |
| 4.1.2 Logfileanalysen | 31 |
| 4.1.3 Mentales Modell | 34 |
| 4.2 Klassifizierungs- und Katalogsysteme | 35 |
| 4.2.1 Resource Description Framework (RDF) | 36 |
| 4.2.2 Open Directory Project | 37 |

| | | |
|----------|---|-----------|
| 4.2.3 | Association for Computing Machinery (ACM) Computing Classification System | 38 |
| 4.3 | Zentrale Strukturen | 40 |
| 4.4 | Dezentrale Strukturen | 40 |
| 4.5 | Mischformen | 41 |
| 5 | Strategien und Hilfsmittel zum Publizieren | 43 |
| 5.1 | Dokumenttypen | 43 |
| 5.2 | Metadaten | 44 |
| 5.2.1 | Metadaten-systematiken | 45 |
| 5.2.2 | Dublin Core | 47 |
| 5.2.3 | Anpassung des Dublin Core | 50 |
| 5.2.4 | Anpassungen am Publizieren-Dialog | 53 |
| 5.3 | Automatisches Publizieren | 53 |
| 5.4 | Workflow | 55 |
| 6 | Strategien und Hilfsmittel: Navigieren | 56 |
| 6.1 | Visualisierung | 56 |
| 6.2 | Hierarchische Visualisierung | 57 |
| 6.2.1 | Baumdarstellung | 57 |
| 6.2.2 | Treemap | 58 |
| 6.2.3 | Cone Trees und Cam Trees | 59 |
| 6.2.4 | Fractal Tree | 60 |
| 6.2.5 | Hyperbolic Tree | 62 |
| 6.3 | Netzartige Visualisierung | 63 |
| 6.3.1 | Hyperspace - Visualisierung von Hyperlinkstrukturen | 64 |
| 6.3.2 | Hotsauce - Webseitenbrowser | 65 |
| 6.3.3 | Infosky | 66 |
| 6.4 | Sitemap-Darstellungen | 67 |
| 6.4.1 | Listen | 68 |
| 6.4.2 | Dynamische Darstellungen | 69 |
| 6.5 | Alphabetisches Stichwortverzeichnis | 70 |
| 6.6 | Speedtyping | 70 |
| 7 | Strategien und Hilfsmittel: Suche | 72 |
| 7.1 | Indexbasiertes Suchen | 72 |
| 7.1.1 | Thesaurus | 73 |
| 7.1.2 | Synonymenliste | 74 |
| 7.2 | Musterbasiertes Suchen | 75 |
| 7.3 | Suchmaske | 76 |
| 7.4 | Trefferliste | 77 |

| | | |
|-----------|--|-----------|
| 7.4.1 | Ähnliche Dokumente | 78 |
| 7.4.2 | Unscharfe Suche | 78 |
| 7.5 | Stemming | 79 |
| 8 | Strategien und Hilfsmittel: Maintenance | 80 |
| 8.1 | Organisatorische Unterstützung | 80 |
| 8.1.1 | Pflege der Struktur | 80 |
| 8.1.2 | Pflege der Inhalte | 82 |
| 8.2 | Unterstützung durch Tools | 83 |
| 8.2.1 | Benachrichtigung | 83 |
| 8.2.2 | Doublettencheck | 84 |
| 9 | Zusammenfassung | 86 |
| 10 | Ausblick | 88 |
| | Abkürzungsverzeichnis | 89 |
| | Abbildungsverzeichnis | 91 |
| | Literaturverzeichnis | 92 |
| | Sachregister | 95 |

Danksagung

An dieser Stelle bedanke ich mich bei allen, die zum Gelingen dieser Master Thesis beigetragen haben. Mein Dank gilt insbesondere

- für die Unterstützung und die freundliche Aufnahme bei Hyperwave Research & Development in Graz Frau Astrid Jancke und Herrn Janez Hrastnik,
- Frau Dr. Stefanie Lindstaedt vom Know Center in Graz für die wissenschaftliche Beratung und überaus konstruktive Kritik,
- Herrn Hans Stecker für die germanistische Rundumbetreuung in allen Fragen,
- der AUDI AG und meinen Kollegen für die großzügige Unterstützung und dem Verständnis für meine Abwesenheit während des gesamten Studiums,
- Herrn Magister Franz Hable für die überaus geduldige Kritik und Hilfe zur Selbsthilfe.

In diesem Sinn vielen Dank und Glück auf!

1 Themenbeschreibung

Diese Arbeit beschäftigt sich mit der Problematik, wie ein komplexer Bestand an Dokumenten verwaltet werden kann, so dass zum einen der Informationslieferant die Dokumente einfach ablegen, zum anderen der Informationssuchende schnell und zielgerichtet zu den gewünschten Informationen gelangen kann. Mit der steigenden Anzahl von Dokumenten stellt sich die Frage, in welche Kategorien beziehungsweise Unterordner ein Dokument abgelegt werden soll.

Häufig werden Dokumente auf eigenen Laufwerken nach dem subjektiven Ermessen in einer beliebigen Struktur ablegt. Arbeiten mehrere Personen am selben Thema oder sollen Kollegen, Dienstleister oder andere Personen Zugang zu diesen Dokumenten haben, dann stellt sich die Frage, wie die Dokumente zu strukturieren sind.

Ziel dieser Arbeit ist es, Strategien und Hilfsmittel aufzuzeigen, mit deren Hilfe man auch in komplexen und heterogenen Datenbeständen effektiv arbeiten kann. Unter *komplex* und *heterogen* ist in diesem Zusammenhang zu verstehen, dass es eine Menge an Dokumenten gibt, die sich nach Herkunft, Inhalt und Format unterscheiden. Dabei soll es keine Rolle spielen, auf welchem Weg versucht wird, zu einem Dokument zu gelangen.

Es gibt inzwischen eine Reihe an Softwareprodukten, die es sich zum Ziel gesetzt haben, Wissen in Unternehmen zu verwalten.

In dieser Master Thesis wird als Umgebung zur Realisierung der Anforderungen die Software der Firma Hyperwave vorausgesetzt.¹

1.1 Abgrenzung

In der folgenden Arbeit werden Szenarien zugrunde gelegt, wie sie typischer Weise in mittelständischen Unternehmen der IT- und Telekommunikationsbranche mit circa 500 Mitarbeitern auftreten. Abhängig von der Branche und der Größe eines Unternehmens werden eventuell andere Methoden zum Zug kommen müssen, um zu einem befriedigenden Ergebnis zu gelangen.

Die Schwierigkeiten, die sich bei der Einführung oder Änderung von Prozessen ergeben, sind dem Verfasser bekannt. Dieses sogenannte „Changema-

¹Informationen zu Hyperwave unter <http://www.hyperwave.de>

nagement“ wird in dieser Arbeit nicht oder nur in Auszügen behandelt. Das soll aber nicht bedeuten, dass dieses Thema unterschätzt werden sollte. Vielmehr ist eine kontinuierliche Betreuung aller Beteiligten im gesamten Prozess der Einführung eines Wissensmanagementsystems von großer Bedeutung.

1.2 Problembeschreibung

Unternehmen sind ab einer gewissen Größe in Abteilungen beziehungsweise Divisionen organisiert. Die Unterteilungen entsprechen bestimmten Prozessen und Abläufen, aber auch Tätigkeiten, die zur Bereitstellung eines Produktes führen. Dabei gibt es zwar Unterschiede zwischen zum Beispiel Unternehmen des verarbeitenden Gewerbes und der Dienstleistungsbranche, ihnen ist aber im Zusammenhang mit der Themenstellung dieser Master Thesis eines gemeinsam: Informationen werden nicht nur von Kollegen innerhalb der eigenen Abteilung benötigt, sondern auch von Kollegen anderer Abteilungen, anderer Standorte oder auch von Partnern und Kunden.

Dies bezieht sich nicht nur auf Dokumente, die im Haus erstellt werden, wie Bedienungsanleitungen, Stücklisten, Broschüren oder Pressemitteilungen. Auch extern recherchierte Dokumente sind von großer Wichtigkeit. Sie liefern Hintergrundinformationen zum Entstehungsprozess, sie zeigen, mit welchen Themen sich die Entwickler beschäftigten, um sich selbst zu einer Thematik weiterzubilden oder in eine Technik einzuarbeiten.

Das Internet bietet inzwischen eine unüberschaubare Fülle an Informationen, Weiterbildungs- und Recherchemöglichkeiten, die gerade im IT-Bereich intensiv genutzt werden². Inzwischen gibt es zu vielen Themen so genannte eBooks, die ganze Bücher mit einer elektronischen Stichwortsuche zugänglich machen. Nicht zu vergessen die vielen Präsentationen, White Papers und so weiter, die man ohne große Mühe finden kann. Daneben hat auch das klassische Informationsmedium Buch nach wie vor seine Bedeutung für Informationssuchende.

An dieser kurzen Aufzählung wird deutlich, dass es diverse Medienformate gibt, die in Zukunft evtl. noch durch Audio- beziehungsweise Videoformate ergänzt werden, die aber an verschiedenen Stellen im Unternehmen verteilt sind und für die es keinen gemeinsamen Zugang gibt.

²Ein bekanntes Beispiel ist selfhtml, das als das zentrale deutschsprachige Nachschlagewerk zum Thema HTML gelten darf; siehe <http://teamone.selfhtml.de>

1.3 Zielbeschreibung

Ziel der Master Thesis ist es, zu überprüfen, welche Ansätze zur Strukturierung von Dokumenten möglich sind und diese zu bewerten. Dabei soll beachtet werden, dass der Prozess des Einstellens von Altdaten und von neuen Dokumenten möglichst einfach gestaltet werden soll. Parallel zur Betrachtung des Publizierens von Dokumenten gilt es zu untersuchen, welche Informationen zu einem Dokument (Metadaten) vorhanden sein müssen, um die Suche beziehungsweise Navigation zu unterstützen und eine Bewertung des Inhalts eines Dokumentes vornehmen zu können, ohne den kompletten Inhalt zu kennen.

1.4 Grundlegende Anforderungen

Zu Beginn der Arbeit wurden folgende grundlegende Anforderungen verbindlich definiert:

- Usability
Bei allen Aktionen, die von einem Benutzer im System ausgeführt werden, ist darauf zu achten, dass sie einfach und verständlich sind.
- Wartbarkeit
Erweiterungen oder Veränderungen des Standard-Hyperwave-Servers sollen einfach zu warten sein.
- Einschränkung der Komplexität
Durch neue Funktionen wird die Komplexität des bestehenden Systems erhöht. Deswegen sollen alle neuen Funktionen in sich geschlossen und überschaubar bleiben.

1.5 Vorgehensweise

Zur Bearbeitung der Fragestellung wurde folgende Vorgehensweise gewählt. Im ersten Schritt wurden anhand von Usecases die Anforderungen an das System ermittelt und abgegrenzt. Die Anforderungen entstanden durch die Diskussion mit den Betroffenen beziehungsweise mit den Administratoren des bestehenden Intranets bei Hyperwave. Als zweiter Schritt wurden verschiedene Ansätze recherchiert und beschrieben. Anhand der Anforderungen aus Schritt eins werden die Ansätze gegenübergestellt und bewertet, um so einen Vorschlag für die Realisierung zu erhalten. In einem iterativen Verfahren wurden anschließend die einzelnen Teilbereiche der Realisierung überprüft, um

in mehreren Wiederholungen das System möglichst optimal an die Benutzerbedürfnisse anzupassen.

Diese Vorgehensweise wurde deshalb gewählt, weil mit der Definition der Usecases verbindliche Kriterien zur Beurteilung der Vorschläge festgelegt werden konnten. Daher sind die Vorschläge dieser Master Thesis für andere Unternehmen auch nur insoweit anwendbar, wie die Anforderungen zumindest annähernd übereinstimmen. Die Kriterien entsprechen auch nicht den Wünschen aller Mitarbeiter bei Hyperwave. Sie sind vielmehr ein Konsens, der seine Bewährungsprobe erst im Laufe der Zeit bestehen muss.

2 Informationsarchitektur

Der Begriff Informationsarchitektur ist ein junger Begriff, der im Zusammenhang mit der Gestaltung von Websites geprägt wurde. Wirth [22] beschreibt die Informationsarchitektur einer Website als

„die Abfolge, Anordnung und Einteilung der Inhalte innerhalb der Hypertextstruktur.“

Der Begriff Informationsarchitektur wird in diesem Zusammenhang aber noch weiter verstanden. Er bezieht sich nicht nur auf die Gestaltung einer Website, die über das Internet zugänglich ist. Vielmehr ist er auch für die Arbeit mit Wissensmanagementsystemen von Bedeutung, weil zum Beispiel der Hyperwave Information Server vollständig auf Internettechnologie aufbaut und der Browser die Oberfläche zum Bedienen ist.

Zur Umsetzung der Informationsarchitektur sind eine ganze Reihe von Kenntnissen nötig, von denen hier die wichtigsten genannt seien:

- Grafikdesign bestimmt den formalen Aufbau eines Systems.
- Informationstheorie legt den inneren Aufbau des Systems fest.
- Usability-Tests helfen, den Bedürfnissen der Benutzer gerecht zu werden.

Das Thema Graphik Design wird weniger behandelt, weil als Grundsystem der Hyperwave Information Server IS/6.3 vorausgesetzt ist und hier möglichst wenig Änderungen vorgenommen werden sollen.

Deswegen wird im vorliegenden Zusammenhang unter Informationsarchitektur der Aufbau von Struktur(en) verstanden, die ein schnelles, intuitives Navigieren/Suchen und eine einfache, qualitativ hochwertige Administration der Dokumente sicherstellen.

Die Vorgehensweise, um alle Anforderungen aufzunehmen und umzusetzen, gliedert sich in mehrere Teilschritte:

- Definition der Ziele
- Definition der Zielgruppen
- Definition der Inhalte
- Entwicklung einer Struktur

2.1 Definition der Ziele

Ziel der im Rahmen dieser Master Thesis besprochenen Neustrukturierung ist es, ein System zu entwickeln, in dem man schnell, einfach und intuitiv extern recherchierte Dokumente finden kann. Ergebnis soll sein, dass ein beliebiger Mitarbeiter zunächst innerhalb des neuen Bereiches nach Dokumenten sucht, bevor eine Recherche im Internet oder in anderen externen Datenquellen (zum Beispiel Büchern) gestartet wird.

2.2 Definition der Zielgruppen

Zielgruppen sind in diesem beschriebenen Fall:

- Research und Development
Die Entwicklung bei Hyperwave ist in mehrere Abteilungen aufgeteilt, die den unterschiedlichen Schichten der Software und den notwendigen flankierenden Bereichen entspricht. Diese Gruppen haben zum Teil eine unterschiedliche Sichtweise auf Techniken und Methoden. Für jede Gruppe soll es den Anforderungen entsprechende Sichtweisen auf das vorhandene Wissen geben.
- Mitarbeiter in anderen Niederlassungen
Für die tägliche Arbeit beim Kunden und auch für das Verständnis von Zusammenhängen und Prozessen der Software, ist es für die Mitarbeiter in anderen Niederlassungen notwendig, auf das Wissen der Entwicklung zugreifen zu können. Dabei ist nicht nur das Wissen über die eigene Software, sondern auch über die Zusammenhänge, die Gründe und die technologische Art der Umsetzung wichtig.
- Partner und Kunden
Da auch Partner und Kunden am Wissen von Hyperwave interessiert sind, ist zu überlegen, ob eine einmal angelegte Struktur nicht nur intern, sondern auch für die externe Wissensvermittlung genutzt werden kann, um so ein einheitliches Gesamtbild der benutzten Dokumente zu erzeugen.

2.3 Definition der Inhalte

In dieser Master Thesis wird die Strukturierung von Dokumenten betrachtet, die aus externen Datenquellen stammen und im Haus inhaltlich nicht weiter verändert werden. Das hat zur Folge, dass über den Entstehungsprozess, die daran Beteiligten und die Hintergründe zu einem Dokument in der Regel nur

sehr wenig oder nichts bekannt ist. Umso wichtiger wird es sein, die bekannten Informationen zu einem Dokument zu erfassen.

2.4 Definition einer Struktur

Jede Zielgruppe hat unterschiedliche Anforderungen an die Struktur und unterschiedliche Perspektiven auf die Struktur. Dabei spielt es auch eine Rolle, ob man sich in der Rolle des Publizierenden befindet, der ein neues Dokument recherchiert hat und dieses schnell und für den richtigen Gebrauch zugänglich machen möchte. Oder ob man sich in der Rolle des Suchenden befindet, der zum einen Dokumente zu seinem speziellen Arbeitsgebiet sucht oder zum anderen Dokumente zu anderen Themen. Um die Anforderungen zu ermitteln, wurden Usecases definiert, die beschreiben, welche Prozesse ablaufen. Anhand dieser Prozesse werden dann Strukturvorschläge ermittelt, die mit den Beteiligten abgestimmt werden.

3 Usecases

Zur Bewertung der Vorschläge werden drei Usecases zu Grunde gelegt, die sich in mehrere Varianten aufteilen. Die Szenarien im Einzelnen sind:

1. Publizieren
2. Suchen
3. Maintenance

Die Usecases werden im Folgenden genau beschrieben, da die hier definierten Anforderungen die Grundlage für die Bewertung der Vorschläge bilden (Zur Vorgehensweise siehe Kapitel 1.5 Vorgehensweise auf Seite 9).

In Kapitel 3.4 Zusammenfassung der Anforderungen auf Seite 27 werden die Anforderungen der folgenden Usecases zusammengestellt.

3.1 Usecase Publizieren

Die Prozesse im Zusammenhang mit dem Publizieren von Dokumenten sind die entscheidenden Erfolgsfaktoren. Nur wenn das Publizieren einfach und neben den Dokumenten auch die zugehörigen Metadaten qualitativ hochwertig sind, wird ein System überhaupt zur Informationssuche verwendet werden. In diesem Umfeld sind deshalb folgende Rahmenbedingungen zu beachten:

- Usability
Das Einstellen von Dokumenten ist für einen Mitarbeiter Teil seiner täglichen Arbeit. Die Bereitschaft, einem Dokument Metadaten mitzugeben, Ordnerhierarchien oder Schlagwortlisten zu pflegen, ist sehr beschränkt. Darauf muss sich der Prozess des Publizierens einstellen. Wobei allerdings die Regel gilt, dass später nur mit den Daten gearbeitet werden kann, die das System kennt.
- Überblick und Konsistenz
Für einen Benutzer ist es wichtig, einen Überblick über den Aufbau eines Systems zu haben, und damit zu wissen, wo er sich gerade bewegt und wie er zu einem bereits gefundenen Dokument ein zweites Mal gelangen kann. Natürlich kann es mehrere Wege zu einem Dokument geben. Es muss aber sichergestellt sein, dass sich die Wege, zum Beispiel durch das Einstellen neuer Dokumente, nicht verändern.

Der Prozess des Publizierens besteht aus zwei großen Teilprozessen und mehreren Varianten. Die großen Teilprozesse sind zum einen die Recherche und

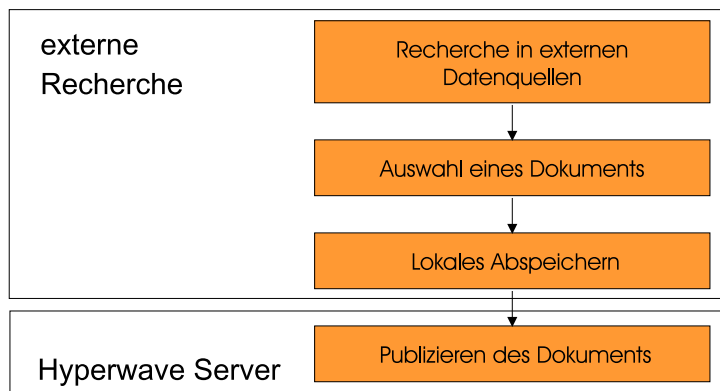


Abbildung 1: Usecase Publizieren
Quelle: eigene Darstellung

zum anderen das Publizieren auf dem Server selbst. Abbildung 1 verdeutlicht dies.

Im Verlauf der Arbeit hat sich herausgestellt, dass es im folgenden noch zwei Varianten des Publizierens gibt:

3.1.1 Usecase: Publizieren ohne Abstimmungsprozess

In diesem Szenario wird vorausgesetzt, dass jemand ein Dokument zu einem Thema abspeichert, das thematisch nur mit der Arbeit dieser einen Abteilung zu tun hat.

Das bedeutet, dass dieses Thema unter der Kontrolle einer abgegrenzten Gruppe liegt, die einen ausreichenden Überblick darüber hat und sich mit keiner anderen Gruppe die Recherche nach neuen Dokumenten teilt. Abbildung 2 zeigt den ablaufenden Prozess. Die Recherche und Bewertung der Wichtigkeit eines Dokumentes findet außerhalb von Hyperwave statt. Erst wenn sich jemand sicher ist, dass ein Dokument relevant für die Datenbank ist, wird es in Hyperwave abgespeichert. Zur Verwendung eines Workflows siehe Kapitel 5.4 Workflow auf Seite 55.

Aus diesem Prozess ergeben sich folgende Anforderungen an die Struktur:

- Überblick
Speichert man Dokumente zum selben Thema ab, muss es sicher sein, dass der Prozess des Publizierens dies einfach zulässt. Das heißt, dass überschaubar sein muss, in welche Struktur ein Dokument einem Thema bisher zugeordnet war und dass gewährleistet ist, dass ein weiteres

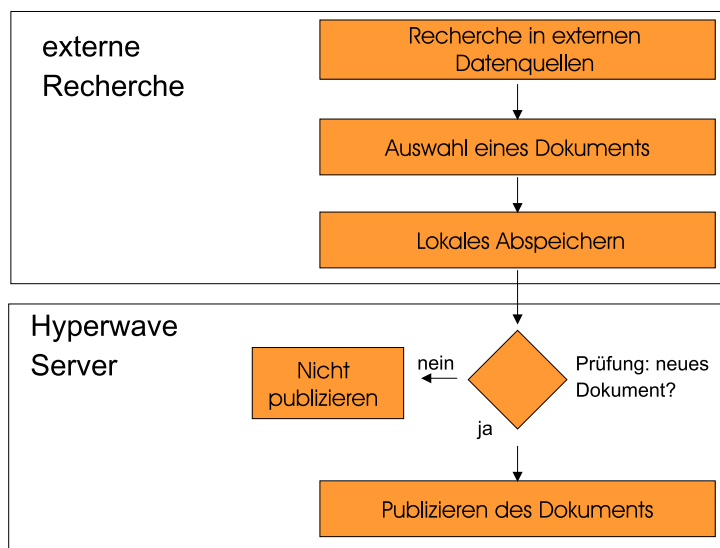


Abbildung 2: Publizieren ohne Abstimmungsprozess
Quelle: eigene Darstellung

Dokument auch diesem Thema zugeordnet wird. Es kann natürlich sein, dass ein Dokument auch weitere Themen behandelt, auch dies muss in der Struktur möglich sein.

- Usability
Die Bereitschaft, im Intranet ein Dokument zu publizieren, ist nur dann vorhanden, wenn es einfach ist und nicht zu viele Felder befüllt oder kontrolliert werden müssen. Dieses Thema wird in Zusammenhang mit Kapitel 5.2 Metadaten auf Seite 44 bearbeitet.
- Überprüfung des Datenbestandes
Wird ein neues Dokument publiziert, muss es die Möglichkeit geben, zu überprüfen, ob das Dokument in dieser Version nicht schon vorhanden ist. Nur wenn das Dokument einen neuen Inhalt hat oder geändert wurde, darf es publiziert werden.

3.1.2 Usecase: Publizieren mit Abstimmungsprozess

Es kann vorkommen, dass verschiedene Abteilungen Dokumente zu einem Thema publizieren. Schwierig wird dieser Fall, wenn Gruppen Dokumente zu ähnlichen oder dem selben Thema recherchieren. Dann kann es vorkommen, dass ein Dokument doppelt unter verschiedenen Namen und in verschiedenen

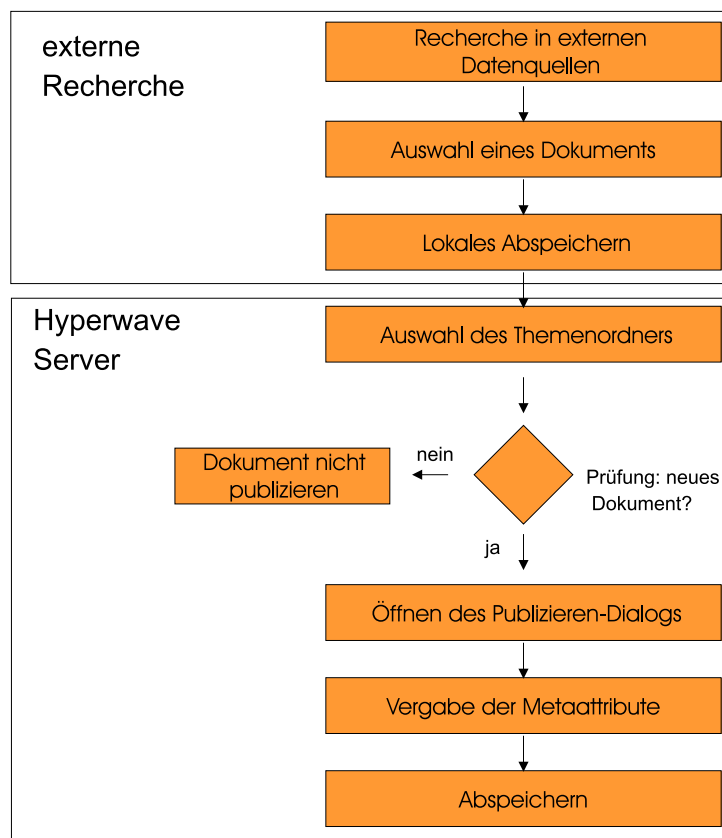


Abbildung 3: Publizieren Variante 1
Quelle: eigene Darstellung

Ordern abgelegt wird. Um dies zu vermeiden, muss es einen Abstimmungsprozess geben, der regelt, wie die Zuständigkeiten lauten. Dabei sind mehrere Varianten möglich.

- Variante 1: Gemeinsames Publizieren
In diesem Fall übernimmt derjenige, der publiziert, die Verantwortung dafür, dass es das neu einzustellende Dokument nicht bereits gibt. Das heißt, er überprüft den Inhalt der bestehenden Dokumente und vergleicht diesen mit dem neuen Dokument, um so sicherzustellen, dass es tatsächlich einen neuen Inhalt hat und nicht nur unter einem anderen Namen oder unter einer anderen Quelle recherchiert wurde. Der Vorteil dieser Variante liegt darin, dass keine umfangreichen Prozesse mit anderen Gruppenmitgliedern ablaufen müssen, die Verantwortlichen sind aber deswegen auch gezwungen, genau zu arbeiten.

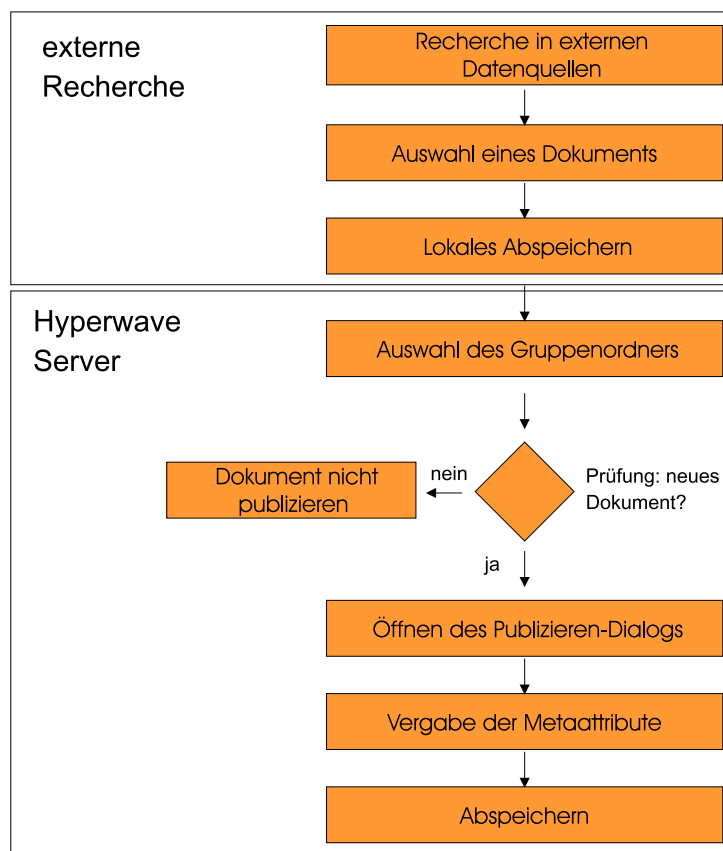


Abbildung 4: Publizieren Variante 2
Quelle: eigene Darstellung

- Variante 2: Subordner für die Gruppen unterhalb des Themenordners
Zu jedem Thema, das von mindestens zwei Gruppen bearbeitet wird, gibt es unterhalb des Themenordners noch weitere Ordner für jede Gruppe. In diesen Ordnern werden neue Dokumente gruppenweise publiziert. Das hat für jede Gruppe den Vorteil, dass sie einen eigenen Ordner hat, in der relevante Dokumente liegen und dass diese getrennt betrachtet werden können. Der Nachteil ist, dass so Dokumente mehrfach im System vorhanden sein können und möglicherweise mehrere Ordner durchsucht werden müssen.
- Variante 3: Publikation nur durch einen Verantwortlichen
Nach dem Recherchieren wird vom Benutzer geprüft, ob das Dokument zu einem Thema gehört, für das es einen Verantwortlichen gibt. Dieser bekommt das Dokument, zum Beispiel via E-Mail, geschickt und publiziert es dann im entsprechenden Themenordner. So bleibt die Qua-

lität des Datenbestandes gewährleistet und es ist immer klar, wer die Verantwortung für ein Dokument hat.

Die Anforderungen aus diesem Prozess decken sich zum Teil mit denen aus dem vorher genannten, es gibt aber auch neue:

- Überblick
Speichern mehrere Personen Dokumente zum selben Thema, ist es wichtig, dass sie auf das selbe mentale Modell zugreifen. Ansonsten wird ein neues Dokument nicht konsistent in das System eingefügt. Daraus resultiert das Problem, dass die Dokumente zu einem Thema alle über die Suche gefunden werden müssen.
- Überprüfung des Datenbestandes
Vor allem wenn mehrere Personen Dokumente zu einem Thema publizieren, muss es im Prozess des Publizierens schon am Anfang möglich sein, zu prüfen, ob das Dokument bereits vorhanden ist, oder ob es zum Beispiel eine ältere Version gibt.
- Usability
Diese Anforderungen decken sich mit dem Prozess ohne Abstimmungsprozess (siehe Kapitel 3.1.1 Usecase: Publizieren ohne Abstimmungsprozess auf Seite 15).

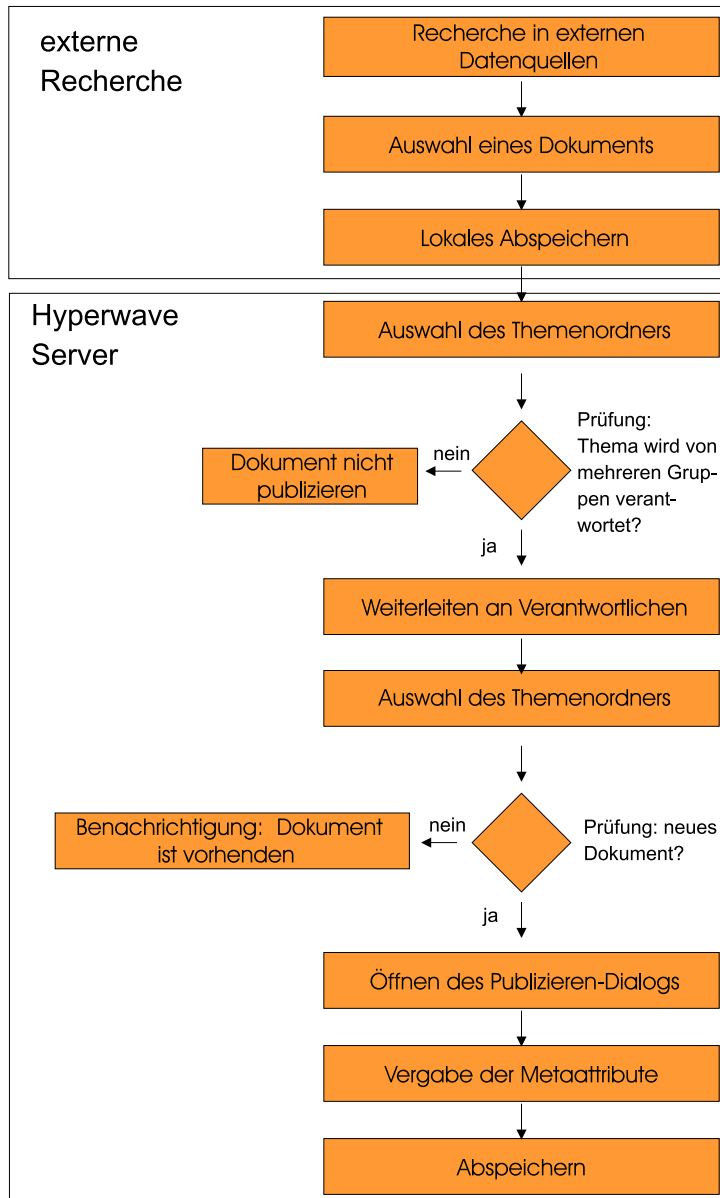


Abbildung 5: Publizieren Variante 3
Quelle: eigene Darstellung

3.2 Usecase Suchen

In diesem Kapitel wird beschrieben, welche grundsätzlichen Bedingungen für das Auffinden von Dokumenten definiert werden. Dabei wird zwischen zwei Zugriffswegen unterschieden:

- Suche durch Navigieren
Damit ist der Zugriff auf Dokumente gemeint, der durch ein Navigieren in einer bestehenden Struktur beziehungsweise Datenbank zustande kommt. Navigieren heißt dabei, so lange aus einem Bestand von Ordnungselementen Unterordnungselemente auszuwählen, bis man zur gewünschten Information gelangt ist. Damit sind nicht nur die klassischen Navigation mit Hilfe eines hierarchischen Baumes gemeint, sondern generell alle Interaktionen, mit deren Hilfe man sich in einem Ordnungsraum bewegt.
- Suche mit Hilfe eines Suchdialogs
Suche beschreibt in diesem Zusammenhang die Eingabe von Suchbegriffen, die als Ergebnis das Auffinden relevanter Dokumente zum Ziel hat. Dieses Verfahren ist durch so genannte Suchmaschinen aus der Internettechnologie bekannt und wird zum Beispiel von Google³ eingesetzt. In diesem Zusammenhang werden zwei unterschiedliche Ansätze betrachtet, die sich mit dem Hyperwave Information Server verarbeiten lassen. Dies ist zum einen der Einsatz der Suchtechnologie von Verity⁴, die auf einer indexbasierten Volltextsuche gründet, und zum anderen auf die Technologie von Autonomy⁵, die mit Algorithmen auf der Basis der Mustererkennung Informationen beurteilt.

3.2.1 Usecase: Suchen durch Navigieren

Bei diesem Usecase geht man davon aus, dass entweder bekannt ist, in welchem Bereich der Datenbank eine Information zu finden ist, oder dass der Aufbau der Datenbank so intuitiv ist, dass die Navigation in einem überschaubaren Zeitraum zum Erfolg führt. Dieser Zugang ist auch möglich, falls bekannt ist, zu welchem Dokument man gelangen will und man genau weiß, wo es liegt. Deswegen wird in diesem Zusammenhang noch zwischen zwei Varianten unterschieden. Variante eins beschreibt die Prozesse, falls man den Ort kennt, wo man ein Dokument finden kann. Varianten zwei, wenn der Ort unbekannt ist und durch Navigieren erst erschlossen werden muss.

³siehe dazu <http://www.google.de>

⁴siehe dazu <http://www.verity.com>

⁵siehe dazu <http://www.autonomy.com>

3.2.2 Usecase: Navigieren zu einem bekannten Ort

Benutzt man einen bestimmten Bereich einer Datenbank mehrmals, weil sich hier häufig verwendete Dokumente befinden, wird man sich den Aufbau der Datenbank schnell verinnerlichen und wissen, wo man ein bestimmtes Dokument zu suchen hat, beziehungsweise wo man Dokumente zu bestimmten Themen finden kann. Dies setzt voraus, dass man eine Datensammlung häufiger nutzt und sich mit dem Aufbau nach und nach vertraut macht.

Da in dieser Master Thesis der Aufbau einer internen Datenbank betrachtet wird, kann man voraussetzen, dass insbesondere die Dokumente, die von der eigenen Abteilung recherchiert sind und zunächst von ihr verwendet werden, in einer Struktur abgelegt sind, die bekannt ist.

Deshalb werden an diese Navigationsmöglichkeit folgende Anforderungen gestellt:

- Überblick
Der Aufbau der Struktur darf nicht zu weit verzweigen und muss in seinem Aufbau homogen sein. Das heißt, das Verhältnis von Breite zu Tiefe soll ausgeglichen sein, es darf nicht zu viele Einstiegspunkte geben, die einen ersten Überblick verhindern, um dann in einer einzigen Ebene zu enden. Es dürfen aber auch nicht zu wenige Einstiegspunkte vorhanden sein, die sich in eine zu tiefe Verästelung aufteilen, in der niemand den Überblick behalten kann.
- Beständigkeit
Um es vor allem den Mitarbeitern einer Abteilung zu ermöglichen, schnell und einfach zu ihren Dokumenten zu gelangen, ist es wichtig, dass sich die Struktur möglichst wenig ändert und stabil bleibt.
- Erweiterbarkeit
Neue Themen oder die Erweiterbarkeit bestehender Themen, zum Beispiel durch neue Versionen, müssen von der Struktur abgefangen werden können, ohne dass es zu großen Änderungen kommt. Generell gilt die Anforderung, dass Änderungen nur in tieferen Punkten der Hierarchien erfolgen sollen, um die Konsistenz und Nachvollziehbarkeit der Struktur zu bewahren.

3.2.3 Usecase: Navigieren zu einem unbekanntem Ort

Häufig befinden sich Dokumente an einem Ort oder in einer Struktur, die man nicht kennt. Um zu diesen zu gelangen, gibt es zwei Möglichkeiten:

- Navigieren
- Suchen

Im Folgenden wird beschrieben, welche Anforderungen ein System erfüllen muss, damit das Navigieren erfolgreich sein kann.

- Breite und Tiefe der Struktur
Hier gelten dieselben Bedingungen wie für das Publizieren von Dokumenten. Es muss ein ausgewogenes Verhältnis zwischen Anzahl der Einstiegspunkte (Breite) und Anzahl der Unterebenen (Tiefe) bestehen.
- Intuitiver Aufbau
Die Bezeichnung der Strukturen muss schlüssig sein und dem mentalen Modell der Firma entsprechen. Dies setzt dann zwar noch immer voraus, dass es sich mit dem des Mitarbeiters möglichst deckt. Da aber ein Intranet den Wissensstand eines Unternehmens und nicht den eines jeden einzelnen Mitarbeiters widerspiegelt, ist es wichtig, für alle eine gemeinsame Sprache zu definieren, die auch nach außen kommuniziert wird. Damit ist gemeint, dass zum Beispiel Produkte intern wie extern gleich, nach einem durchgehenden Schema bezeichnet werden und es zu keinen Inkonsistenzen kommt. Die Begriffe, die verwendet werden, sollen im Unternehmen geläufig sein.
- Fachtermini, Abkürzungen
Bei der Verwendung von Fachtermini und Abkürzungen ist darauf zu achten, dass diese bekannt sind und dass jeder Mitarbeiter die Begriffe zuordnen kann. Werden Abkürzungen aus dem Fachjargon verwendet, sollte die ausgeschriebene Bedeutung in der Beschreibung mit angegeben werden.
Für die Verwendung von Abkürzungen kann man weiter folgende Regeln verwenden ([7] Seite 44):
 - Die Anzahl muss begrenzt sein
 - Sie müssen sich untereinander deutlich unterscheiden
 - Sie müssen so häufig vorkommen, dass sie für den Benutzer zu festen, automatisierten, wortanlaogen Eintragungen in seiner mentalen Begriffswelt werden
- Überblick
Die Darstellung der Struktur muss so gewählt sein, dass ein schneller

Überblick über möglichst viele Ebenen gewährleistet ist, um Zusammenhänge zu erkennen und möglichst schnell einschätzen zu können, zu welchen Bereichen Dokumente vorhanden sind.

3.2.4 Usecase: Suchen mit Hilfe des Suchdialogs

Die Suche dient dazu, Dokumente zu einem bestimmten Thema zu finden, von denen man nicht weiß, wo sie zu finden sind. Natürlich ist es möglich, nur mit der Suche zu arbeiten, um zu einem Dokument zu gelangen. Dies hat aber folgende Nachteile:

- Der Überblick über die Anzahl der vorhandenen Dokumente geht verloren. Wird immer nur die Suche verwendet, findet man auch immer nur die Dokumente aus der Trefferliste. Alle anderen Dokumente, die zu einem Thema vielleicht auch interessant gewesen wären, den Suchbegriff aber nicht enthalten oder in einer anderen Kategorie sind, fallen aus der Trefferliste heraus. Dies ist auch das klassische Problem der Suchmaschinen im Internet, die alle nur einen mehr oder wenig großen Teil des Internets in ihren Indices erfassen.
- Der Suchbegriff muss bekannt sein. Eine Suche liefert immer nur Treffer zu einem eingegeben Begriff, nicht aber zu ähnlichen oder verwandten Begriffen. Um dies zu verbessern, muss man sich Hilfsmittel (Thesauri, Synonyme) bedienen. Es gibt auf dem Markt dafür Anbieter, die den allgemeinen Wortschatz einer Branche abdecken⁶. Aber diese Anbieter bieten keine individualisierte, auf eine Firma zugeschnittene Lösung, die auch die firmenspezifischen Begriffe abdeckt.

Aus diesen Gründen sollte für das Suchen folgende Empfehlung gelten:

1. Eingrenzung des Suchbereiches durch Navigieren durch die Struktur.
2. Suchen im eingegrenzten Bereich

An die Suche werden deshalb folgende Anforderungen gestellt:

- Aufrufbarkeit
Die Suche muss immer aufrufbar sein, das heißt, dass zu jedem Zeitpunkt eine Suche nach einem Begriff gestartet werden kann.

⁶siehe hierzu die Taxonomien der Firma Verity, die zum Beispiel Taxonomien für die pharmazeutische Industrie oder den Verteidigungssektor anbietet

- Einfache Suche
Mit der einfachen Suche soll die Anforderung nach einer schnellen und unkomplizierten Suche abgedeckt werden. Man gibt einen Suchbegriff ein, und eine Trefferliste erscheint. Diese Suche ist dazu gedacht, sich schnell einen Überblick über die Treffermenge und -liste zu machen.
- Erweiterte Suche
Ist die Treffermenge zu groß oder möchte man die Trefferliste auf bestimmte Bereiche beschränken, verwendet man die erweiterte Suche. Hier sollen Einschränkungen nach Verfasser, Autor, Erstellungsdatum, Dateiformat und Suchbereich möglich sein.
- Überblick
Die Trefferliste soll einen Überblick gewähren, wo in der Struktur Dokumente gefunden wurden. Bisher ist es üblich, Dokumente in der Reihenfolge der Relevanz anzuzeigen. Diese wird von den Suchmaschinen nach Algorithmen berechnet, die unter anderem die Häufigkeit des Auftretens des Suchbegriffes im Dokument und die Stellung innerhalb eines Dokumentes berücksichtigen. Da die Algorithmen nicht öffentlich zugänglich sind und deswegen im Detail nicht nachvollzogen werden kann, wie die Reihenfolge zustande kommt, ist diese Auflistung nicht optimal. Es ist besser, sich in der Trefferliste der Struktur zu bedienen und Dokumente zuerst der Struktur zuzuordnen und erst dann innerhalb der Struktur auf die Relevanz als Sortierkriterium zurückzugreifen. So wird auch deutlich, in welchen Bereichen der Struktur überhaupt Dokumente gefunden werden.
- Kontext
In der Trefferliste soll der Kontext des Suchbegriffs im Dokument dargestellt werden, um schneller entscheiden zu können, ob das Dokument relevant ist.
- Anzeige der Metadaten
Schon beim Publizieren werden Metadaten erfasst, um Dokumente genauer beschreiben zu können. Diese Metadaten sollen auch in der Trefferliste angezeigt werden.

3.3 Usecase Maintenance

Im Usecase Maintenance werden die Prozesse betrachtet, die im laufenden Betrieb des Intranets auftreten und die dafür sorgen sollen, dass der Datenbestand den Qualitätsstandards entspricht, der von den Benutzern erwartet

wird. Für die Nutzung eines Intranets ist die Qualität der Dokumente neben der Zugriffsgeschwindigkeit auf diese und deren einfache Bedienbarkeit von entscheidender Bedeutung. Der Aufbau eines Intranets ist mit großen Kosten verbunden. Dazu zählen nicht nur die Kosten für die Beschaffung der Hard- und Software, sondern auch die Projektkosten selbst. Dieser Kostenaufwand wird sich nur rechnen, wenn im anschließenden Betrieb die Qualitätsstandards der Startphase aufrecht erhalten bleiben und notwendige Verbesserungen zielgruppenorientiert umgesetzt werden. Daher gibt es für den Usecase Maintenance drei Aufgabenfelder:

- Verbesserungen der Struktur
- Verbesserungen des Inhalts
- Verbesserung der Bedienbarkeit

3.3.1 Usecase: Verbesserung der Struktur

Mit Verbesserung der Struktur ist nicht nur gemeint, weitere Anforderungen nach zusätzlichen Themen an die passende Stelle einbinden zu können. Es geht vielmehr auch darum, bestehende Themen weiterzuentwickeln und auszubauen. Von einem Benutzer kann im Laufe der Zeit nicht verlangt werden, dass weitere oder neue Themen zu häufigen Wechseln der Struktur führen. Der Aufbau muss vielmehr von Anfang an folgende Szenarien berücksichtigen:

- Wachstum
Gibt es zum Beispiel neue Versionen einer Software oder technologische Weiterentwicklungen zu bestimmten Themen (zum Beispiel JavaScript 2.0 statt JavaScript 1.1), dann muss es eine Möglichkeit geben, die neuen Punkte an der Stelle einzufügen, die die Benutzer auch auf Grund ihrer Vorkenntnisse erwarten, und zwar so, dass die bestehende Struktur nicht verändert, sondern lediglich ergänzt wird. Das bedeutet, dass die Struktur mit den bestehenden Themen wachsen kann.
- Erweiterbarkeit
Gerade im Bereich der Software gibt es häufig Veränderungen und Neuerungen, die zum Beispiel die technischen Grundlagen betreffen. Dies muss bereits beim Anlegen der Struktur berücksichtigt werden, so dass die Struktur einen breiten Rahmen bietet, um die neuen Themen aufzunehmen.

Diese Punkte sind nicht nur für den Benutzer wichtig, sie sind auch die Grundlage für die Administration eines Systems. Eine Verzeichnisstruktur muss so aufgebaut sein, dass sie personenunabhängig verwaltet werden kann und dem mentalen Modell der Benutzer entspricht. Dieses Modell ändert sich selbstverständlich im Lauf der Zeit, da sich auch die Mitarbeiter verändern. Bei der Bearbeitung der Master Thesis hat sich gezeigt, dass es unterschiedliche Vorstellungen zu den Strukturen zwischen den Abteilungen gegeben hat. Dies ist auch nachvollziehbar, hat doch jede Abteilung eine, jeweils sich - wenn auch leicht - unterschiedliche Sichtweise. Beim Aufbau einer neuen Verzeichnisstruktur gibt es die Möglichkeit, die Differenzen zu überwinden, da es ein gemeinsames Ziel gibt. Im laufenden Betrieb darf es aber nicht wieder zur Verselbstständigung der Bereichsstrukturen kommen, da sonst das mentale Modell des Gesamtunternehmens in Teilmodelle aufgelöst wird.

3.3.2 Usecase: Verbesserung des Inhalts

Dieser Punkt ist eng mit dem Publizieren von Dokumenten verwandt, da es hier darum geht, die Qualität des Inhalts zu garantieren und zu verbessern. Es muss möglich sein, neue Dokumente einfach zu publizieren. Die Bedingungen, die sich daraus ergeben, sind eng mit dem Usecase Publizieren (siehe Kapitel 3.1) verbunden, gehen aber noch darüber hinaus, da es hier um ein kontrolliertes Wachstum geht. Die Tendenz, für immer neue Inhalte neue Ordner anzulegen, ist die Hauptschwierigkeit. Es muss einen abgestimmten Prozess geben, der definiert, wer die Erlaubnis hat, die Struktur zu erweitern. Nur so wird sichergestellt, dass es ein organisches und kein unkontrolliertes Wachstum der Struktur gibt.

3.3.3 Usecase: Verbesserung der Bedienbarkeit

Neue Systeme haben Grundannahmen als Basis für die Bedienbarkeit. Der Erfolg eines Systems hängt auch - wie oben erwähnt - von der Usability ab. Es ist sehr wichtig, auf die Bedürfnisse der Benutzer einzugehen. Vor allem nach dem Going live gilt es, berechnete Anforderungen schnell umzusetzen, um die Zufriedenheit mit dem System auf einem akzeptablen Niveau zu halten.

3.4 Zusammenfassung der Anforderungen

In den Usecases werden umfangreiche Anforderungen definiert, die sich auch zum Teil widersprechen. Es gibt aber einige Punkte, die durchgehend für alle Prozesse gelten. Dies sind:

- Überblick/Konsistenz
Egal, ob man Dokumente einstellt, sucht oder Inhalte pflegt, immer ist es wichtig, sich schnell und zuverlässig einen Überblick über den vorhandenen Datenbestand machen zu können.
- Beständigkeit
Innerhalb eines Datenbestandes darf es nicht zu viele Änderungen in einem überschaubaren Zeitraum geben. Die Struktur muss in sich konsequent und nachvollziehbar sein.
- Erweiterbarkeit
Auch wenn die Beständigkeit ein hohes Gut ist, muss es dennoch möglich sein, neue Themen an eine passende Stelle einzufügen, ohne dabei große Änderungen an der bestehenden Struktur vornehmen zu müssen.
- Usability
Bei der Auswahl der gewählten Softwareunterstützung muss darauf geachtet werden, dass sich die Prozesse im Sinn des mentalen Modells der Benutzer bewegen.

4 Aufbau einer Struktur

Der Aufbau einer Struktur dient der Zuordnung von Dokumenten zu Klassen, er stellt die Verbindung zwischen Dokumenten her und ermöglicht so deren Einordnung und Bewertung. Für Nielsen ([18], Seite 198) gibt es zwei wichtige Regeln für den Aufbau der Struktur einer Website:

- Das Vorhandensein einer Struktur überhaupt
- Die Erfüllung der Erwartungen der Benutzer hinsichtlich der Informationen und Dienste

Die Anforderungen an eine Website sind auch für diese Master Thesis zu übertragen. Mit der Möglichkeit, Dokumente zu strukturieren, werden Verbindungen und Zuordnungen zwischen den Dokumenten erst möglich. Diese müssen - und das gilt für jedes System - den Anforderungen und Erwartungen der Benutzer entsprechen.

Nach Meinung von Nielsen ([18], Seite 198) ist der Aufbau einer Struktur von den Aufgaben bestimmt, die die Benutzer an das System haben. Das kann unter anderem auch dazu führen, dass Informationen in einem Strukturpunkt untergebracht sind, die von mehreren Abteilungen benötigt werden.

Schwierig beim Aufbau von Klassen ist, dass es bei vielen Beteiligten unterschiedliche mentale Modelle gibt, die in unterschiedlichen Ansätzen deutlich werden. Kapitel 4.1.3 Mentales Modell beschreibt den Umgang mit den Vorstellungen der Mitarbeiter.

Der zentrale Ansatz (siehe Kapitel 4.3 Zentrale Strukturen auf Seite 40) wird im Folgenden dem dezentralen (Kapitel 4.4 auf Seite 40) gegenübergestellt und danach werden die Mischformen (Kapitel 4.5 Mischformen auf Seite 41) aus beiden diskutiert.

4.1 Methoden

Das Erstellen einer Struktur zur Ablage der Dokumente ist ein zentraler Baustein für ein Wissensmanagementsystem. Im folgenden werden zwei Methoden beschrieben, die im Vorfeld (Card Sorting) und im Nachhinein (Logfileanalyse) Hilfestellung leisten können, um diesen komplexen Vorgang zu unterstützen.

4.1.1 Card Sorting

Card Sorting ist eine Methode, um Strukturen und Klassifikationen zu entwickeln. Mit ihrer Hilfe wird festgestellt, wie Personen Begriffe zuordnen. Aus dem Vergleich der Zuordnungen lassen sich Rückschlüsse für eine zusammengeführte Struktur ziehen. Ziel dieses Verfahrens ist es,

„ein kundenfreundliches System zu entwickeln, das von allen Benutzern gleichermaßen akzeptiert wird, erstens, weil sie sich eingebunden fühlen, und zweitens, weil die Navigation großteils ihren eigenen Denkprozessen entspricht. ([10], Seite 2)“

Mit diesem Ziel wird schon deutlich, dass es wichtig ist, auf ein gemeinsames Verständnis des Inhalts eines Systems zu kommen, das in dieser Master Thesis als mentales Modell bezeichnet wird (siehe dazu Kapitel 4.1.3 Mentales Modell auf Seite 34).

Card Sorting wird laut Gaffney [11] aus folgenden Gründen eingesetzt:

- Card Sorting hilft die Wahrscheinlichkeit zu erhöhen, dass die Anwender Strukturelemente auf einer Website finden.
- Es ist einfach durchzuführen und kostengünstig umzusetzen.
- Es liefert ein Verständnis, wie Anwender Begriffe strukturieren.
- Es kristallisieren sich Begriffe heraus, die die Anwender schwer gruppieren können, beziehungsweise nicht zuordnen können.

Ein nicht zu unterschätzender Vorteil ist auch noch, dass dieses Verfahren ohne der Mithilfe eines Computers durchgeführt werden kann. Auch wird durch dieses Verfahren die Motivation der Teilnehmer deutlich erhöht, da sie „ihre“ eigene Struktur entwickeln können und diese dann auch in anderen Diskussionsrunden als eigenen Beitrag verteidigen können.

Die Durchführung von Card Sorting kann entweder in Einzelsitzungen oder in Gruppen erfolgen. Für Gruppen schlägt Gaffney ([11]) mindestens sechs Personen vor, mehr sind auch möglich. Bei den Gruppenübungen sollten aber auch nicht zu viele Anwender teilnehmen, da sonst das Finden der gemeinsamen Lösung erschwert wird.

Für die Durchführung benötigt man nur wenige Utensilien. Alle Begriffe, die strukturiert werden sollen, werden auf kleinen Karten notiert. Nach der Vorstellung des Verfahrens bekommt der Teilnehmer eine gewisse Zeit,

um sich einen Überblick über die Begriffe zu verschaffen. Gaffney [11] geht davon aus, dass man für das Ordnen von 50 Begriffen circa 30 Minuten benötigt. Jeder Teilnehmer hat nun die Aufgabe, die Begriffe einzelnen Stapeln zuzuordnen und für diese Stapel eigene Oberbegriffe zu finden. In der Nachbesprechung erläutert der Teilnehmer seine Erfahrungen und beschreibt seine persönlichen Eindrücke.

Bei der Auswertung werden die Ergebnisse der Teilnehmer miteinander verglichen. So lässt sich schnell herausfinden, welche Begriffe einheitlich zugeordnet und benannt wurden. Für die Begriffe, für die es keine gemeinsame Zuordnung gibt, bietet sich ein gemeinsamer Durchlauf des Verfahrens an, um zu einer einheitlichen Sichtweise zu gelangen.

Card Sorting ist nicht nur für den erstmaligen Entwurf einer Struktur ein gutes Hilfsmittel. Auch für die spätere Pflege und Weiterentwicklung ist dieses Verfahren sehr nützlich, weil es schnell durchzuführen und grafisch einfach zu handhaben ist.

4.1.2 Logfileanalysen

Die Logfileanalyse bietet die Möglichkeit, das Verhalten der Benutzer auf einer Website zum Teil nachzuvollziehen. Dazu nutzt man die Einträge des Webservers in den sogenannten Server Logfiles. Diese wurden ursprünglich entwickelt, um die Performance und Verfügbarkeit von Webservern zu protokollieren ([13] Seite 277). Die Logfiles werden als Textfiles geschrieben und sind mit jedem gängigen Editor zu betrachten. Nicht jeder Webserver protokolliert allerdings nach dem gleichen Format.

Der Hyperwave Information Server schreibt die Logfiles in verschiedenen Formaten, darunter auch dem Expanded Common Logfile Format, also einer Erweiterung des Common Logfile Formats. Tabelle 1 zeigt die Namen und Bedeutungen der Eintragungen in einem Common Logfile.

| Feldname | Bedeutung des Feldinhaltes |
|----------|---|
| Host | IP-Adresse des zugreifenden Servers |
| Ident | Identifikation (falls vorhanden, sonst Bindestrich) |
| Authuser | Benutzername (bei passwortgeschützten Webseiten, sonst Bindestrich) |
| Date | Datum und Uhrzeit des Zugriffs im Format dd/mmm/yyyy:hh:mm:ss |
| Timezone | Abweichung von der Greenwich Mean Time (GMT) in Stunden |

| Feldname | Bedeutung des Feldinhaltes |
|----------|--|
| Request | Methode, Dokument und Protokoll des Zugriffs (oft „GET/...“) |
| Status | Antwortstatus als Codenummer (z.B. 200 = „Seite erfolgreich übertragen“) |
| Bytes | Gesamtzahl der übertragenen Bytes |

Tabelle 1: Eintragungen im Common Logfile Format;
Quelle: [13] auf Seite 276

Im Expanded Common Logfile sind zusätzlich noch der Referrer (URL der Seite, die den Link zur angefragten Seite enthielt) und der Agent (Name und Versionsnummer des anfragenden Browsers) enthalten. Folgende Zeilen zeigen einen Logfileeintrag im Expanded Common Logfile Format, wie beispielhaft vom Hyperwave Information Server erzeugt wird:

```
123.456.78.9 - lhaerter - [01/Sep/2004:08:08:59 +0200] GET /Hyperwave/Dokumentation/Hilfe.doc HTTP/1.1 200 2161 http://Hyperwave/Dokumentation/Hilfe.doc;internal&Parameter=1094018926425 Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)
```

In diesem Eintrag steht, dass der Anwender `lhaerter` über die IP-Adresse `123.456.78.9` am `1. September 2004` um `8:08` das Dokument `Hilfe.doc` erfolgreich aufgerufen und dabei den `Internet Explorer 6.0` (abgeleitet aus der Abkürzung `MSIE 6.0`) mit dem Betriebssystem `Windows NT 5.1` benutzt hat.

Im Gegensatz zu den Logfiles von Webservern im Internet hat man im Intranet auch die Möglichkeit, festzustellen, von welchem User die Einträge stammen. Problematisch an dieser Auswertung ist allerdings, dass die Verarbeitung personenbezogener Daten in der Bundesrepublik Deutschland dem Bundesdatenschutzgesetz unterliegt. Der Zweck dieses Gesetzes ist

„den einzelnen davor zu schützen, dass er durch den Umgang mit seinen personenbezogenen Daten in seinem Persönlichkeitsrecht beeinträchtigt wird.“[4]

Eine mögliche Verfahrensweise, um Daten dennoch auswerten zu können, ist das Zusammenfassen von mehreren Benutzern zu Gruppen, wobei sich auch für diese Vorgehensweise eine Abstimmung mit dem Betriebsrat beziehungsweise der Arbeitnehmervertretung empfiehlt.

Da die Logfiles ursprünglich nicht für die statistische Erfassung des Benutzerverhaltens gedacht waren, ist es wichtig, sich zu Beginn zu überlegen, was man auswerten möchte und wie man etwas zählen will. Im Zusammenhang mit Auswertungen werden häufig die Begriffe Hits und Page views verwendet. Ein Hit entspricht dabei jedem Eintrag im Logfile. Das bedeutet, dass auch jedes aufgerufenen Grafikelement als Hit erscheint. Dies ist aber für eine Auswertung nicht gewollt. Deshalb sind die Page views aussagekräftiger, da sie die Anzahl der tatsächlich übertragenen Seite aufzeigen. Unwägbarkeiten treten aber beispielsweise dann auf, wenn .pdf-Dateien übertragen werden. Ist in diesem Format die sogenannte schnelle Webansicht aktiviert, dann werden mindestens drei Einträge geschrieben, die sich nur durch den Status unterscheiden:

```
123.456.78. - lhaerter [02/Sep/2004:14:59:20 +0200] GET /Hyperwave/
Dokumentation/Hilfe.pdf HTTP/1.1 200 194348 Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1)
123.456.78. - lhaerter [02/Sep/2004:14:59:26 +0200] GET /Hyperwave/
Dokumentation/Hilfe.pdf HTTP/1.1 200 194348 Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1)
123.456.78. - lhaerter [02/Sep/2004:14:59:27 +0200] GET /Hyperwave/
Dokumentation/Hilfe.pdf HTTP/1.1 206 119471 Mozilla/4.0
(compatible; MSIE 6.0; Windows NT 5.1)
```

Die ersten beiden Einträge beziehen sich unmittelbar auf das Aufrufen der Datei `Hilfe.pdf` und werden in der Statistik als zwei Page views gezählt. Der dritte Eintrag kommt durch das Nachladen des Acrobat Readers zustande, was am Status 206 erkennbar wird.

Ein weiterer Wert der Auswertungen sind die Visits. Sie zählen die IP-Adressen aufgerufener Clients. Ein Benutzer, der sich einmal angemeldet hat, erzeugt einen Visit. Meldet er sich erneut an, wird ein weiterer Visit erzeugt. Generell ist zu sagen, dass man die Logfileanalyse mit Vorsicht beurteilen sollte. Sie bieten sich aber dennoch an, um Tendenzen zu erkennen, in welchen Bereichen wichtige Informationen vorliegen. Daraus lassen sich Rückschlüsse ziehen, ob die Struktur in manchen Bereichen zu tief beziehungsweise zu breit ist.

Zur Analyse der Logfile wurde spezielle Software entwickelt. Mit dieser Software können die Logfiles, die zum Teil mehrere 100 Megabyte groß sein können, eingelesen und ausgewertet werden.. Das Programm `WebSuxess` der Firma `Exody` stellt zum Beispiel eine Reihe von voreingestellten Auswer-

tungen zur Verfügung, darunter auch die der Navigationsanalyse. Die Navigationsanalyse bietet eine Übersicht über das Besucherverhalten. Besucher kommen auf Ihre Web-Site, schauen sich dort eine oder mehrere Seiten an und verlassen sie wieder. Die Folge der Seiten, die von einem Besucher angesehen wurden, wird als Pfad bezeichnet. Jedem Visit entspricht also ein Besucherpfad. Werden die Pfade zu lang oder zeigen sie unterschiedliche Bereiche, ist dies ein Zeichen dafür, dass die Struktur an dieser Stelle verbessert werden sollte.

4.1.3 Mentales Modell

Der Begriff Mentales Modell wird im Zusammenhang mit der menschlichen Kognition verwendet ([7] Seite 10). Mentale Modelle sind notwendig zur internen Repräsentation von Wissen, das durch Erfahrungen gewonnen wurde. Diese Erfahrungen beziehen sich auf Sachverhalte und Funktionen, die über die Außenwelt aufgenommen wurden. Sie sind die Basis zur Planung und Steuerung von Handlungen([6] Seite 2).

Mentale Modelle stellen die Außenwelt in unterschiedlicher Form dar ([7] Seite 43). In der allgemeinen, semantischen Form zum Beispiel als „Autos haben Räder“, in der episodischen Form „Mein Auto hat 115 PS“. Bei der Verarbeitung von Information wird auf diese mentalen Modellen zurückgegriffen. Sie bilden die Grundkomponenten und Ableitungsregeln, auf deren Basis hin reagiert wird.

„Erinnerung und Wissensanwendung sind also stets Rekonstruktionen. Denken ist das bewußte Operieren mit mentalen Modellen“ ([7] Seite 43)

Ein Beispiel für die Umsetzung eines mentalen Modells in eine Software sind die Fenster des Betriebssystems Windows von Microsoft. Grundlage des mentalen Modells ist hier ein Schreibtisch mit vielen Papieren. Das gerade bearbeitete liegt oben. In der Umsetzung von Windows ist das aktuelle Fenster immer im Vordergrund, alle anderen Fenster sind im Hintergrund oder ausgeblendet. Auch das Löschen von Dokumenten bezieht ein mentales Modell zur Verarbeitung mit ein. Dokumente werden vor dem endgültigem Löschen in der Papierkorb verschoben und können von dort auch wieder rekonstruiert werden. Dies entspricht den Möglichkeiten des realen Papierkorbs.

Unter mentales Modell wird im Zusammenhang mit dieser Master Thesis das gemeinsame Verständnis einer Gruppe zu einem abgegrenzten Bestand an Begriffen verstanden. Je größer die Übereinstimmung innerhalb einer Gruppe

zur Zuordnung von Dokumenten in eine einzige Begriffswelt ist, umso einfacher ist es, den Begriffen Dokumente und Informationen beizugeben.

Menschen suchen in ihrer Umwelt Sinn, Ordnung und Zusammenhänge ([7] Seite 43). Auf diese Anforderung muss jede Software und Struktur innerhalb eines Datenbestandes reagieren. Strukturen, die nicht dem Modell der Benutzer entsprechen, machen die Arbeit mit einem System schwierig, führen zur Unzufriedenheit, werden nur ungern benutzt und führen deshalb letztlich nicht zum Erfolg. Dabei ist es nicht unbedingt nötig, das mentale Modell eines jeden Einzelnen exakt zu treffen, es genügen schon Andeutungen, um die mentalen Modelle treffend anzuwenden ([7] Seite 43).

4.2 Klassifizierungs- und Katalogsysteme

In Bibliotheken wird der Bestand an Büchern durch Kataloge erschlossen. Haller [12] beschreibt die Notwendigkeit, einen Katalog zu erstellen,

- wenn die Bücher ohne Systematik in Bezug auf den Inhalt aufgestellt sind,
- eine größere Anzahl erreicht ist,
- gezielte Suchen unabhängig vom Standort eines Buches gestellt werden, oder
- Informationen in verschiedenen Bibliotheken recherchiert werden sollen.

Diese Anforderungen treffen auch auf einen elektronischen Datenbestand zu, der in einer Datenbank gespeichert ist. Im Zusammenhang mit dieser Master Thesis wird der Aufbau eines Katalogs unter dem Kapitel 4 Aufbau einer Struktur auf Seite 29 beschrieben. Anwendungsgebiete finden sich in vielen Bereichen, auch in den Naturwissenschaften, zum Beispiel in der Biologie. Mit Hilfe eines Klassifizierungssystems werden Tiere und Pflanzen in Beziehung gebracht und ein Überblick hergestellt. Dazu erstellt man Klassen und Unterklassen, die das Gerüst bilden. In dieses Gerüst werden die Objekte eingefügt. Auch im Zusammenhang mit Bibliotheken entstanden Klassifizierungssysteme, weil hier der Bedarf besteht, große Bestände überschaubar und zugänglich zu halten.

Für die Arbeit in Bibliotheken sind mehrere Regelwerke entwickelt worden, so zum Beispiel im deutschsprachigen Raum die Berliner Anweisungen, die Preußischen Instruktionen und die Regeln für die alphabetische Katalogisierung (RAK). Mit Hilfe dieser Kataloge werden die Bücher sortiert und

wieder auffindbar gemacht. Die Regeln für die alphabetische Katalogisierung sind seit Anfang der 70er Jahre des 20. Jahrhunderts im Einsatz. Für den Einsatz elektronischen Medien wurden weitere Katalogisierungen weiterentwickelt. Dazu zählen zum Beispiel:

- Resource Description Framework (RDF)
- Open Directory Project
- ACM Computing Classification System

4.2.1 Resource Description Framework (RDF)

Das Resource Description Framework wurde vom W3C-Konsortium⁷ entwickelt, um einen einheitlichen Standard für die Auszeichnung von Objekten mit Metadaten zu definieren. Damit sollen laut Schön [19] mehrere Ziele erreicht werden:

- Die Metadaten sollen maschinell/elektronisch und von Menschen verarbeitet werden können.
- Die Metadaten sollen geeignet sein, um über Datennetze, zum Beispiel das Internet, geschickt werden zu können.
- Applikationen für verschiedene Zwecke (Datenbanken, Suchmaschinen, Softwareagenten, digitale Bibliotheken) sollen die Metadaten nutzen können.
- Die Verwaltung der Metadaten soll neutral sein, das heißt, sie darf nicht auf bestimmte Medienformate (Bibliothek, Internet, Multimedia) beschränkt sein. Mit dem RDF wird aber nur ein Rahmenwerk (Framework) definiert, das einen Mechanismus darstellt, mit dem Objekte beschrieben werden können.

RDF ist aus drei Bestandteilen aufgebaut:

- Der Uniform Resource Identifier (URI) identifiziert ein Objekt eindeutig. Es spielt dabei keine Rolle, ob das Objekt eine Website, ein Buch oder ein elektronisches Dokument ist.
- Mit Hilfe von Eigenschaften wird ein Objekt beschrieben. Zusammen mit dem Wert der Eigenschaft entsteht eine Aussage (Statement). Diese Darstellung in Form von Subjekt-Prädikat-Objekt ist auch die Grundlage für semantische Netze.

⁷zum W3C-Konsortium siehe <http://www.w3c.org>

- Bei der Entwicklung wurden verschiedene, bereits bestehende Metadatenmodelle berücksichtigt, allerdings nicht nur mit dem Ziel, diese zu vereinheitlichen, sondern um ein semantisches Netz im Internet aufzubauen. Dadurch soll erreicht werden, dass Verknüpfungen zwischen Dokumenten nach deren Inhalt geknüpft werden und die maschinell weiterverarbeitet werden können.

Der Begriff *Semantisches Netz* wurde Anfang der 1960er Jahre vom Sprachwissenschaftler M. Ross Quillian eingeführt, um die Bedeutung von Wörtern in verschiedenen Satz- und Verwendungskontexten zu bestimmen. In ihnen werden die semantische Verwandtschaft unterschiedlicher Begriffe durch verschiedenartige Relationen repräsentiert. In semantischen Netzen wird menschliches Wissen als ein Gefüge von Begriffen und assoziativen Beziehungen dargestellt.

Bei der Modellierung semantischer Netze bedient man sich bei Modellen, die die Arbeit des menschlichen Gedächtnisses nachbilden. In diesen Modellen gibt es assoziative Beziehungen, die Teilbereiche der realen Welt abbilden. Karagiannis und Telesko [16] beschreiben den Aufbau eines semantischen Netzes in Form von Knoten und Kanten. Die Knoten werden graphisch zumeist durch Rechtecke oder Kreise dargestellt und stellen Objekte, Situationen, Konzepte oder Begriffe graphisch dar.

Pfeile symbolisieren die Kanten und stellen Verweise zwischen den Wissensseinheiten dar. Gerade hierin ist ein Grund für Netzwerkmodelle zu suchen, indem nämlich die Organisation der Wissenstrukturen in den Vordergrund der Untersuchungen gestellt wird. Aus psychologischer Sicht betrachtet, sind Netzwerkmodelle als Gedächtnismodelle aufzufassen.

4.2.2 Open Directory Project

Das Open Directory Project hat zum Ziel, den Aufbau eines Webkataloges in die Hand vieler Internetuser zu legen. Mehrere tausend Internetuser navigieren im Internet, kategorisieren die Seiten und können auch teilweise das Ranking von Seiten beeinflussen. Diese Internetuser werden Editoren genannt.

Für jede Kategorie gibt es einen oder mehrere Editoren, die auch per E-Mail erreichbar sind. Es gibt inzwischen mehr als 590.000 Kategorien, die von mehr als 62.000 Editoren gepflegt werden.

Manche Suchmaschinen, wie zum Beispiel Google und AllTheWeb, stufen Websites, die durch das Open Directory Project erfasst werden, im Ranking höher ein.

Seit einiger Zeit wächst die Kritik am Open Directory Project, weil die Qualität und die Aktualität des Katalogs wegen der zum Teil langen Bearbeitungsdauer nicht Schritt halten können. Das Open Directory Project bietet aber die Möglichkeit, beim Aufbau eines eigenen Katalogs/Verzeichnisbaumes auf ein bestehendes System zugreifen zu können, das millionenfach getestet wurde.⁸

4.2.3 Association for Computing Machinery (ACM) Computing Classification System

Die Association for Computing Machinery wurde 1947 mit dem Ziel gegründet, als Austauschforum für Ideen, Entwicklungen und Informationen im Umfeld der elektronischen Datenverarbeitung zu dienen. Heutzutage gibt es Einrichtungen in mehr als 100 Ländern, die zu verschiedenen wissenschaftlichen Themenstellungen publizieren.

So begann man 1982 mit dem Aufbau eines Klassifikationssystems für den Bereich Computer mit dem Namen Computing Classification System (CCS). Das System wurde mehrfach erweitert und liegt seit 1998 in seiner aktuellen Form vor. Es wird als Basis für den ACM Guide to Computing Literature (Guide) verwendet, der jährlich um circa 20.000 neue Einträge erweitert wird und der inzwischen mehr als 250.000 Einträge umfasst. Grundlagen für die Entwicklung waren folgende Punkte [1]:

- Herzstück des CCS ist ein Baum, der die einfachste Darstellung einer hierarchischen Struktur in einem linearen Format ist.
- Jeder Eintrag im Klassifikationsbaum in den ersten drei Ebenen ist mit einer Kombination aus Buchstaben und Zahlen versehen. Die vierte Ebene ist ohne eigene Bezeichnung.
- Neue oder geänderte Einträge werden gekennzeichnet. Alte Einträge bleiben in der Klassifikation erhalten, werden aber gesondert ausgewiesen.

Der Baum besteht derzeit aus elf so genannten first-level nodes, denen die Buchstaben A-K zugeordnet sind. In der Ebene darunter wird jedes Element mit einer Ziffer versehen, wobei die Ziffer 0 immer fest für den Eintrag *General*

⁸zum Open Directory Projekt siehe auch <http://dmoz.org/about.html>

reserviert ist. Neben General gibt es stets den festen Eintrag *Miscellaneous*, der mit dem Buchstaben *m* gekennzeichnet ist. Alle weiteren Ebenen werden mit einem Punkt und einer Ziffer versehen. Das CCS wird laufend bearbeitet und ergänzt und umfasst aktuell mehr als 1400 Einträge. Das Computing

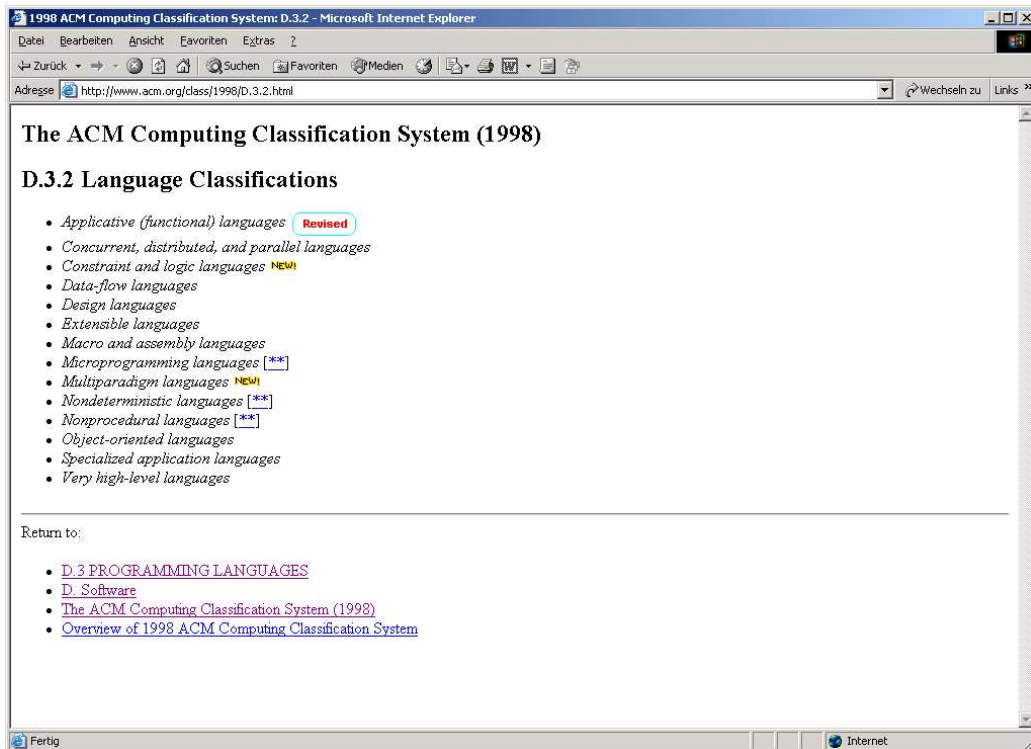


Abbildung 6: Auszug aus dem Computing Classification System
Quelle: <http://www.acm.org/class/1998/D.3.2.html>

Classification System ist sehr gut geeignet, die eigene Struktur hinsichtlich einer durchgängigen Bezeichnung zu überprüfen. Da sich vor allem im Bereich der Hard- und Software Dinge sehr schnell ändern und neue Begriffe hinzukommen, erscheint es sinnvoll, bestehende Standards so weit wie möglich zu übernehmen, um nach innen und außen mit einer einheitlichen Sprache aufzutreten. Selbstverständlich gilt es, auch die eigene Unternehmenssprache sowie Marketingaspekte zu berücksichtigen. Ein System dagegen, das auf das Auffinden von Informationen ausgelegt ist, die häufig im Internet recherchiert wurden, darf nicht zu sehr differenzieren. Das Computing Classification System kann als Referenz hierfür große Dienste leisten. Aus diesen Gründen wurden die bestehenden Strukturvorschläge mit dem Computing Classification System verglichen und soweit wie möglich der Referenz angepasst.

4.3 Zentrale Strukturen

Zentrale Strukturen eines Wissensbestandes bilden wichtige Sichtweisen, um den Datenbestand in seiner Fülle zu überblicken. Dies setzt allerdings voraus, dass mit Hilfsmitteln der Visualisierung auch tatsächlich ein Überblick möglich ist (siehe dazu Kapitel 5 Strategien und Hilfsmittel zum Publizieren).

Die Vorteile einer zentralen Struktur sind:

- Ein Überblick über die gesamte Breite der Informationen ist gewährleistet. Alle Informationen sind in einer Struktur untergebracht.
- Die Pflege der Struktur ist einfach, solange jeder Punkt nur in einem Kontext vorkommt und eine Hierarchie erzeugt werden kann, die ein Verlinken von Inhalten/Strukturen erlaubt.

Daneben gibt es aber auch Nachteile der zentralen Struktur:

- Ein spezifischer Blick auf einen Teilbereich ist schwer möglich, wenn dieser sich aus verschiedenen Strukturpunkten zusammensetzt. Gibt es Informationen zu einem Thema an mehreren Stellen in der Struktur, wird die Wartung sehr schwierig.
- Die Pflege der Dokumente setzt unter Umständen voraus, dass ein Mitarbeiter in unterschiedlichen Bereichen publizieren muss. Das macht die Pflege umständlich und aufwendig, weil man so einen großen Bereich der Struktur kennen muss. Vor allem für die Akzeptanz eines Systems ist dieser Punkt sehr kritisch.

4.4 Dezentrale Strukturen

Dezentrale Strukturen bilden keine gesamte Ansicht auf einen Dokumentbestand ab, sondern sind nach Kriterien strukturiert (zum Beispiel Abteilungen), die den Gesamtblick außer Acht lassen und nur eine spezielle Sichtweise repräsentieren. Dies hat mehrere Vorteile:

- Einzelne Abteilungen können ihre eigene Sichtweise auf ihren Dokumentenbestand selbstständig verwalten. Die Akzeptanz des Systems ist groß, weil nur die für die eigene Arbeit interessanten Dokumente mit wenigen Klicks verfügbar sind.
- Die Struktur ist schnell aufgebaut, da keine Rücksicht auf andere Bereiche genommen werden muss und man sich in einer kleinen Gruppe zügig auf eine gemeinsame Struktur einigen kann.

Diesen Vorteilen stehen aber auch Nachteile gegenüber:

- Der Gesamtüberblick über die vorhandenen Informationen in einem Unternehmen geht verloren. Themen, die von verschiedenen Abteilungen behandelt werden, sind in der Struktur mehrfach vorhanden, ohne dass es der Suchende erkennen kann.
- Dokumente zu einem Thema können von verschiedenen Abteilungen mit ganz unterschiedlichen Titeln abgelegt werden. Es ist für den Suchenden nicht klar, in welchem Bereich er sich bewegen soll und ob er tatsächlich alle Dokumente zu einem Thema gefunden hat. Auch die Volltextsuchmaschine wird dies nicht beheben können. Vor allem wenn ein populäres Thema gesucht wird, tritt das Problem der langen Trefferliste auf.

4.5 Mischformen

Um die Vor- und Nachteile der zentralen und dezentralen Sichtweisen auszugleichen, sind Mischformen möglich. Zum Beispiel, dass beide Sichtweisen parallel angeboten werden und der Benutzer sich aussuchen kann, in welchem Bereich er nach Informationen sucht. Dies setzt aber einen großen Wartungsaufwand und eine sehr intelligente Verlinkung der Ordnerstrukturen voraus.

Voraussetzung für eine gute Struktur sind strikte Regeln und ein immer wiederkehrender Aufbau innerhalb der Struktur. Dies kann entweder durch verbindliche Regeln oder durch eine kleine Anzahl von Redakteuren sichergestellt werden. Die Erfahrung zeigt, dass strikte Regeln häufig zu Ablehnung oder Umgehung des Systems führen und in Nischen Parallelstrukturen entstehen. Ein bekanntes Beispiel hierfür sind die Gruppenlaufwerke.

Ein kleines Redaktionsteam, das die Struktur - in Zusammenarbeit mit den Anwendern - pflegt, Änderungen vornimmt und eventuell auch selbst Dokumente publiziert, ist im Zusammenhang mit den hier dargestellten Use-cases anderen Lösungen vorzuziehen.

Im betrachteten Fall wurde das Problem so gelöst, dass nur ein Mitarbeiter pro Abteilung das Recht hat, die Struktur zu ändern und dieser für die Koordination innerhalb seines Bereiches zuständig ist.

Es gibt eine zentrale Ansicht, die sich aus den Strukturen der einzelnen Bereiche zusammensetzt. Damit es nicht zu Überschneidungen kommt, wird

beim Anlegen eines neuen Ordners überprüft, ob es diesen schon gibt. Falls ja, bekommt der Benutzer einen Hinweis auf das Vorhandensein, kann den Ordner aber dennoch anlegen. Um zu verhindern, dass ein Dokument mehrfach publiziert wird, wird beim Einstellen eine Checksumme gebildet (siehe dazu Kapitel 8.2.2 Doublettencheck auf Seite 84). Diese Checksumme ist für jedes Dokument eindeutig. So kann festgestellt werden, ob ein Dokument bereits existiert. In diesem Fall gibt es wieder einen Hinweis mit der Angabe, wo das Dokument bereits abgespeichert ist. Man kann es aber dennoch publizieren.

5 Strategien und Hilfsmittel zum Publizieren

Der Schwerpunkt der Verbesserungen zum Thema Publizieren liegt auf der Anreicherung der Informationen. Die Strategie, alleine mit einer Volltextsuche einen Datenbestand zu verwalten, muss scheitern, weil Begriffe an sich nichts über den Kontext aussagen und erst der Kontext eine schnelle Einschätzung möglich macht, ob ein Dokument von Relevanz ist.

Deshalb wurde versucht, jedem Dokument schon beim Publizieren wichtige Zusatzinformationen zuzuordnen, die für das spätere Navigieren und Suchen eine Hilfestellung geben können. Die Zusatzinformationen lassen sich unterscheiden in Informationen über das Dokument und zum Dokument.

Informationen zu einem Dokument sind wichtig, weil hier festgelegt wird, welcher Typ das Dokument ist (siehe folgendes Kapitel). Bei den Informationen zu einem Dokument wurden einige Standards entwickelt, die ab Kapitel 5.2 Metadaten ab Seite 44 beschrieben werden.

Den Abschluss des Kapitels bildet eine kurze Untersuchung, ob das Publizieren eines Dokumentes vollständig von einer Software vorgenommen werden kann, die mit vordefinierten Beispielen - so genannten Regelsets - Inhalte erkennen und einer Struktur zuordnen kann. (Kapitel 5.3 Automatisches Publizieren Seite 53).

5.1 Dokumenttypen

Bei einer Analyse der Arten bestehender Dokumente hat sich gezeigt, dass es verschiedene Dokumente gibt, die sich durch ihr Format oder ihre Verwendung unterscheiden. Diese Arten werden im Folgenden Dokumenttypen genannt.

Durch diese Einteilung lassen sich die Dokumente später einfacher klassifizieren. Dies ist hilfreich, um in der Struktur schnell zu erkennen, um welchen Typ Information es sich handelt. Auch in der Trefferliste der Suche kann man darüber eine bessere Sortierung erreichen.

Entsprechend der Anforderung nach einem guten Überblick/Konsistenz aus Kapitel 3.4 Zusammenfassung der Anforderungen auf Seite 27 wurden deshalb folgende Dokumenttypen definiert:

- Dokumente

Unter Dokumenten versteht man alle recherchierten Informationen. Dazu zählen beispielsweise Guidelines, Referenzen, eBooks etc.

- Links
Hierunter fallen alle Verweise zu externen Seiten im Internet.
- Bilder
Hierzu gehören alle Bildformate, die nicht durchsucht werden können.
- Bücher
Bücher, die in der hauseigenen Bibliothek vorhanden sind.

Es wäre auch denkbar gewesen, noch weitaus mehr Dokumenttypen zu bestimmen, bzw. genauer zu differenzieren. Im Sinn der Usability, der Akzeptanz und der Nachvollziehbarkeit beim Einstellen wurde darauf verzichtet.

5.2 Metadaten

Mit Metadaten werden Dokumente beziehungsweise Objekte zusätzlich ausgezeichnet, um sie zum einen genauer zu beschreiben und zum anderen in einen Kontext zu stellen. Der Brockhaus ([8], Seite 278) beschreibt Metadaten als

„Daten, die andere Daten beschreiben, z.B. Datenbankbeschreibung in einer Datenbank (beschreibt die Bedeutung und Struktur der Primärdatensätze, eine Indextabelle (gibt die Speicheradresse von Daten an) oder ein Metadatei.“

In unserem Fall ist mit Attributen die genauere Beschreibung von Eigenschaften eines Dokumentes gemeint, wie es zum Beispiel die Herkunft, der Autor oder das Erstellungsdatum sein können.

Es gibt verschiedene Möglichkeiten, einem Objekt Metadaten zuzuordnen. Bei HTML-Dokumenten werden in so genannten Meta-Tags Informationen im Objekt direkt abgelegt. Ein Beispiel dafür ist die Homepage der Donau Universität Krems, die unter anderem folgende Metadaten direkt mitliefert:⁹

```
<meta name="keywords" content="krems, donau, universität, university, master, llm, mba, mas, msc, ma, mes, weiterbildung, ausbildung, fortbildung, postgradual, postgraduate Donau, Universität, Krems, Donauuni, Uni, Studium, Studien, Weiterbildung, postgraduate, Forschung, Master, MBA, MSc, LLm, Lehrgänge, Seminare, Wissenschaft">
```

⁹Auszug aus dem Quellcode der Homepage der Donau Universität Krems; <http://www.donau-uni.ac.at>

```

<meta name="description" content="Donau-Universität Krems, Europas
einzige staatliche Weiterbildungsuniversität Die Donau-Universität
Krems ist Europas einzige staatliche Weiterbildungsuniversität mit
exklusiven Aufbaustudien in den Bereichen Wirtschafts- und Manage-
mentwissenschaften, Telekommunikation, Information und Medien, Eu-
ropäische Integration, Kulturwissenschaften sowie Umwelt- und Medi-
zinische Wissenschaften.">
<meta name="author" content="(c) 2003 Donau-Universität Krems, by
NETural Communication GmbH, Linz/Austria, office@netural.com. All
rights reserved.">
<meta name="publisher" content="Donau-Universität Krems">

```

In diesem Beispiel wird sehr schön deutlich, wie mit Metadaten komplexe Inhalte beschrieben werden, aber auch, welche Problemstellungen dadurch auftreten. Obwohl der Eintrag unter "content" schon sehr umfangreich ist, werden immer noch nicht alle Inhalte, die zum Angebot der Donau-Universität gehören, erfasst. Es erfolgt eine Beschreibung des Inhalts auf einer aggregierten Ebene. Das Problem dabei ist, diese Ebene treffend zu wählen, da sonst die Beschreibung im einen Extrem zu ungenau und im anderen Extrem zu feingliedrig ist.

Eine andere Form der Darstellung von Metadaten sind Nachschlagewerke, wie sie zum Beispiel in gedruckter Form vorliegen.¹⁰ Auch im Internet gibt es inzwischen sehr umfangreiche Nachschlagewerke, die zum Teil auch kostenfrei zur Verfügung stehen.¹¹ Diesen Nachschlagewerken ist gemeinsam, dass sie Verweise zu anderen Stichpunkten enthalten, und man sich so durch ein komplexes Wissensgebiet bewegen kann.

Bei Computerdateien besteht zudem die Möglichkeit, Metadaten der Datei als Anhang mitzugeben. Der Hyperwave Information Server verwaltet zum Beispiel den Titel und die Beschreibung, aber auch alle Zugriffsrechte und systembedingte Attribute, die indizierbar sind und damit auch durchsucht werden können.

5.2.1 Metadatensystematiken

Für verschiedene Anwendungsszenarien wurden inzwischen standardisierte Metadatensystem entwickelt, die eine einheitliche Auszeichnung von Infor-

¹⁰Zu den bekanntesten Nachschlagewerken im deutschsprachigen Raum gehört zum Beispiel Brockhaus

¹¹siehe hierzu zum Beispiel <http://www.wikipedia.de>

mationen gewährleisten sollen.¹² Vom Max-Planck-Institut für Bildungsforschung in Berlin wurde eine Website *Clearingstelle für Metadaten*¹³ eingerichtet. Wie schwierig der Umgang damit ist, zeigt aber, dass sie zuletzt 1999 aktualisiert wurde.

Die Anzahl der Arbeiten und Initiativen zu diesem Thema ist schier unüberschaubar. Eine kleine Auswahl zeigt folgende Tabelle[20].

- Dublin Core
Der Dublin Core ist der „Kern“ der inhaltlichen und formalen Erschließungsmerkmale, die sonst für die bibliothekarische und inhaltliche Erschließung benutzt worden sind. Ursprüngliches Ziel der Anstrengungen einer Vielzahl von Beteiligten (Bibliothekaren, Informationswissenschaftlern, Informatikern und Systemspezialisten) war es, einen Minimalsatz von Erschließungselementen zu definieren. Sie sollten helfen, die Präzision und Retrievalfähigkeit digitaler Dokumente bei Recherchen im Internet, zum Beispiel mittels Suchmaschinen, zu verbessern. Durch „Metatags“ im Header der HTML-Dokumente beziehungsweise der dokumentenähnlichen Objekte (DLOs) sollten Suchmaschinen in der Lage sein, Dokumente einer Suchanfrage zuzuordnen.
- Text Encoding Initiative
Basierend auf der Auszeichnungssprache SGML beziehungsweise XML wurde von der Text Encording Initiative (TEI) ein Dokumentformat zur Kodierung und zum Austausch von Texten entwickelt. Es legt in einer Dokumenttypdefinition (DTD) fest, aus welchen Bestandteilen die Beschreibung eines Dokumentes zusammengesetzt ist. Innerhalb der Geisteswissenschaften hat es sich zu einem de-facto-Standard entwickelt und wird zum Beispiel zur Auszeichnung von sprachlichen Informationen in der Linguistik und zur Kodierung von gedruckten Werken verwendet [20].
- Encoding Archive Description
Auch diese Auszeichnung beruht auf SGML und wird als Suchkriterium für Archiv-Daten verwendet. Es ist umfangreicher als der Dublin Core und es lassen sich viele Elemente des Dublin Core hier wieder finden.
- USMARC
USMARC wurde vom Network Office der Library of Congress erstellt.

¹²Eine Übersicht über gebräuchliche Metadaten-systeme ist zu finden unter <http://de.wikipedia.org/wiki/Metadaten>

¹³Link zur homepage: <http://www.mpib-berlin.mpg.de/dok/metadata/gmr/gmr1d.htm>

Es ist ein Austauschformat von Metadaten speziell zur maschinellen Verarbeitung bibliographischer Daten. Das Format ist sehr umfangreich.

- MAB2

In Deutschland wurde das USMARC angepasst und etwas vereinfacht. Es hat nicht mehr so viele Felder und die Unterteilung in Unterfelder ist wesentlich einfacher. Es wird in Nordrhein-Westfalen für die Katalogisierung von Internet-Ressourcen benutzt. Inzwischen gibt es auch einen Abgleich zwischen MAB2 und dem Dublin Core.

5.2.2 Dublin Core

Für die Arbeit mit Metadaten im Internet wurde der Dublin Core entwickelt.¹⁴ Dieser Standard umfasst in seiner Basisversion 15 Attribute.

| Element | Name | Beschreibung |
|--------------------|---|---|
| title | Titel | Der vom Verfasser, Urheber oder Verleger vergebene Name (=value) der Ressource. |
| creator | Autor | Die Person(en) oder Organisation(en), die den intellektuellen Inhalt verantworten. |
| subject or keyword | Schlagwörter aus dem kontrollierten Vokabular | Das Thema der Ressource beziehungsweise Schlagwörter oder Phrasen, die das Thema oder den Inhalt des Dokuments beschreiben. |
| description | Inhaltliche Beschreibung | Eine textuelle Beschreibung des Ressourceninhalts inklusive eines Referats (Abstract) bei dokumentenähnlichen Ressourcen oder Inhaltsbeschreibungen bei graphischen Ressourcen. |
| publisher | Verleger/Herausgeber | Einrichtung, die verantwortet, dass diese Ressource in dieser Form zur Verfügung steht, wie zum Beispiel ein Verleger, ein Herausgeber, eine Universität oder ein korporatives Unternehmen. |

¹⁴Homepage des Dublin Core: <http://dublincore.org/>

| Element | Name | Beschreibung |
|---------------------|----------------------------------|--|
| other contributors | sonstige Beteiligte | Zusätzlich(e) Person(en) und Organisation(en) zu jenen, die im Element 2 (creator) genannt wurden, die einen bedeutsamen intellektuellen Beitrag zur Ressource geleistet haben, deren Beitrag aber sekundär im Verhältnis zu denen im Element 2 (creator) zu betrachten ist. (zum Beispiel Herausgeber, Übersetzer, Illustratoren, auch Konferenzleiter, Moderatoren). |
| date | Datum in normierter Schreibweise | Das Datum, an dem die Ressource in der gegenwärtigen Form zugänglich gemacht wurde. |
| resource type | Ressourcenart | Art der Publikation wie zum Beispiel Dissertation, Diplomarbeit, Gedicht, Homepage, Essay und so weiter aus einem kontrollierten Vokabular. Verwendet wird ein kontrolliertes Vokabular aus einer Liste zugelassener Bezeichnungen. |
| format | Format | Datentechnisches Format der Ressource, zum Beispiel: Text/HTML, ASCII, Postscript-Datei, ausführende Anwendung, JPEG-Bilddatei und so weiter (Die Angabe für dieses Element muss einem kontrollierten Vokabular entnommen werden, zum Beispiel den angemeldeten Internet Media Types (MIME Types)(siehe auch RFV2046)). |
| resource identifier | Ressourcenidentifikation | Zeichenkette oder Zahl, die eine eindeutige Identifikation des Dokuments ermöglicht (zum Beispiel ISBN, ISSN). Bei vernetzten Ressourcen sind URLs und URNs vorgesehen. |

| Element | Name | Beschreibung |
|-------------------|----------------------------------|--|
| source | Quelle | SWB-ID-Nr. der Titelaufnahme des Dokuments in der SWB-Verbunddatenbank und in einem 2. Meta-Element (falls nötig) der Eintrag des Werks (gedruckt oder elektronisch), aus dem diese Ressource stammt. |
| language | Sprache | Sprache(n) des intellektuellen Inhalts der Ressource als Sprachcode, in der der Inhalt des Dokuments vorliegt. (Darstellung 3-stellig aus dem normierten Sprachcode NISO Z39,50) |
| relation | Verhältnis zu anderen Ressourcen | Eine formale Spezifizierung für dieses Element ist in Arbeit |
| coverage | Abdeckungsspektrum | Abdeckungsaspekte (zeitliche, örtliche, flächenhafte etc. Aspekte, die zur Charakterisierung des Objekts sinnvolle Ergänzungen geben, zum Beispiel bei Karten, Bildern etc.) <i>Eine formale Spezifizierung für dieses Element ist in Arbeit.</i> |
| rights management | Rechtliche Bedingungen | Vorgesehen für den Inhalt dieses Elements ist ein Link (URL oder andere passende URI wie zutreffend) zu einem Urhebervermerk, ein „Rights-Management“-Vermerk über die rechtliche Bedingung oder ggf. zu einem Server, der solche Informationen dynamisch erzeugt. |

Tabelle 2: Elemente des Dublin Core; Quelle: Kreieren eines Dublin Core Metadaten-Eintrags zur Übernahme in eine HTML-Seite in <http://www.bsz-bw.de/diglib/medserv/konvent/metadat/dcmake2b.html>

Von manchen Sites im Internet wird inzwischen eine Auszeichnung mit Hilfe des Dublin Core vorgenommen. Das folgende Beispiel zeigt, wie dies aussehen kann¹⁵:

<head>

¹⁵Beispiel entnommen aus <http://selfhtml.teamone.de/html/kopfdaten/meta.htm.dublin.core>

```

<meta name="DC.Title" content="SELFHTML/Meta-Angaben">
<meta name="DC.Creator" content="Stefan Münz">
<meta name="DC.Subject" content="Meta-Angaben">
<meta name="DC.Description" content="Heute bekannte Meta-Angaben in HTML">
<meta name="DC.Publisher" content="TeamOne">
<meta name="DC.Contributor" content="Wolfgang Nefzger">
<meta name="DC.Date" content="'2001-12-15">
<meta name="DC.Type" content="Text">
<meta name="DC.Format" content="text/html">
<meta name="DC.Identifizier" content="http://selfhtml.teamone.de/">
<meta name="DC.Source" content="HTML-Referenz">
<meta name="DC.Language" content="de">
<meta name="DC.Relation" content="Unterkapitel">
<meta name="DC.Coverage" content="Munich">
<meta name="DC.Rights" content="Alle Rechte liegen beim Autor">
  <!-- ... andere Angaben im Dateikopf ... -->
</head>

```

5.2.3 Anpassung des Dublin Core

Für die Auszeichnung der Dokumente mit Metadaten wurde der Dublin Core gewählt, weil er mittlererweile im Internet große Verbreitung gefunden hat und so ein Abgleich mit anderen Datenquellen möglich ist. Sein Vorteil ist auch, dass Anpassungen möglich sind, die im Folgenden näher beschrieben werden.

Abhängig vom gewählten Dokumenttyp (siehe Kapitel 5.1 Dokumenttypen auf Seite 43) werden die zugehörigen Metadaten definiert. Da die Dokumenttypen sich durch ihr Format beziehungsweise ihre Herkunft unterscheiden, dennoch aber eine einheitliche Vergabe der Metadaten zwingend notwendig ist, wurden zwei unterschiedliche Typen von Metadaten definiert:

- Standardattribute
- dokumenttypen-spezifische Attribute

Zu den Standardattributen gehören Titel, Dokumenttyp, Verfasser, Besitzer, Erstellungsdatum und Quelle. Hyperwave kennt im Standard eine ganze Reihe von Attributen. Aus der folgenden Übersicht geht hervor, wie die Übereinstimmung mit dem Dublin Core und die Abweichungen von ihm zu verstehen sind.

| Dublin Core | Hyperwave | Bemerkung |
|---------------------|------------------|--|
| title | Titel | Der Titel wird bei Hyperwave automatisch übernommen, wenn ein bestehendes Dokument in der Datenbank gespeichert wird. |
| creator | Dokument-Autor | Dieses Dokument ist bei Hyperwave optional, wird aber im Zusammenhang mit der Master Thesis als Pflichtfeld definiert. Daneben gibt es noch das Attribut Besitzer, das unter anderem für die Steuerung der Zugriffsrechte wichtig ist. |
| subject | Schlüsselwort | Im Dublin Core wird festgelegt, dass die Schlüsselwörter aus einem kontrollierten Vokabular stammen müssen. Davon wird hier abgewichen. |
| description | Beschreibung | Optionales Feld zur umfassenden Beschreibung des Titels. |
| publisher | Besitzer | Wird bei Hyperwave standardmäßig aus den Attributen des Benutzerobjektes generiert. |
| other contributors | nicht definiert | Da nur auf externe Dokumente zugegriffen wird und diese Angaben dem Publizierenden in der Regel nicht zur Verfügung stehen, wird auf dieses Attribut verzichtet. |
| date | Erstellungsdatum | Wird standardmäßig mit dem aktuellen Datum gefüllt, eine Änderung ist möglich. |
| resource type | Dokumenttyp | Werte werden aus einer definierten Liste ausgewählt, die editierbar ist (siehe Kapitel 5.1). |
| format | MIME Type | Wird von Hyperwave beim Upload automatisch erkannt. |
| resource identifier | Objektkennung | Wird von Hyperwave automatisch vergeben und besteht aus zwei hexadezimalen Zeichenketten, die in der Datenbank eindeutig sind. |
| source | Quelle | Dieses Attribut existiert noch nicht, es wird für alle Objekte als neuer Standard definiert. |
| language | Sprache | Bei der Vergabe des Dokumenttitels wird die Sprache des Dokumentes angegeben. |
| relation | - | Da diese Angabe dem Publizierenden in der Regel nicht bekannt ist, wird sie nicht erfasst. |

| Dublin Core | Hyperwave | Bemerkung |
|-------------------|-----------|--|
| coverage | - | Da diese Angabe dem Publizierenden in der Regel nicht bekannt ist, wird sie nicht erfasst. |
| rights-management | - | Da diese Angabe dem Publizierenden in der Regel nicht bekannt ist, wird sie nicht erfasst. Dieses Attribut darf nicht mit dem Rechteattribut von Hyperwave verwechselt werden, über das alle Zugriffsrechte zu einem Objekt in der Datenbank gesteuert werden. |

Tabelle 3: Übersicht Dublin Core und Hyperwave Standard Attribute; Quelle: eigene Darstellung

Von den 15 Attributen werden folgende Attribute standardmäßig für jedes Dokument vergeben:

- title
- creator
- publisher
- date
- resource type
- format
- resource identifier
- source
- language

Folgende Attribute können optional angegeben werden:

- subject
- description

5.2.4 Anpassungen am Publizieren-Dialog

Bei der Standardauslieferung des Hyperwave Information Servers IS/6 wird mit dem Publizieren-Dialog ein neues Dokument auf dem Server abgespeichert. In diesem Dialog können alle Standardattribute verwaltet und beliebig neue Attribute hinzugefügt werden. Dieser große Funktionsumfang ist in der Regel nicht nötig, da einige Attribute ohnehin vom System befüllt werden und nur ein bestimmter Ausschnitt tatsächlich relevante Informationen für den Benutzer enthält. Deswegen wurde dieser Dialog angepasst, auch um die unterschiedlichen Dokumenttypen getrennt voneinander publizieren zu können. Der neue Dialog berücksichtigt auch die Anpassungen, die sich aus der Übernahme des Dublin Cores ergeben.

5.3 Automatisches Publizieren

Der Hyperwave Information Server IS/6 bietet die Möglichkeit, Software der Firma Autonomy einzubinden. Diese Software hat die Funktion, Dokumente nach vorher festgelegten Regeln automatisch in der Struktur an die passende Stelle abzulegen. So erspart man sich die manuelle Suche des Platzes in der Struktur und vermeidet - im Idealfall - Inkonsistenzen, da die Software ja immer nach den gleichen Regeln arbeitet.

In Abbildung wird der Prozess des automatischen Publizierens deutlich. Grundlage für das automatische Publizieren ist das Erstellen sogenannter Regelsets. Regelsets bestehen aus Ordnern mit vorbildhaften Dokumenten zu einem Thema. Sie geben vor, mit welchem Inhalt ein Thema definiert wird. Zur Abgrenzung wird eine weitere Regel festgelegt, die das Verhalten für die Ausnahmen steuert. Ist dies geschehen, erstellt Autonomy einen Kategorisierer, der das Zuordnen der Dokumente in die in dem Regelset definierten Ordner übernimmt. Es ist möglich, für verschiedenste Themen Regelsets beziehungsweise Kategorisierer anzulegen.

Bei verschiedenen Tests hat sich allerdings gezeigt, dass das automatische Publizieren für diese Anwendung nicht geeignet ist. Der Aufwand, die Regelsets zu erstellen, ist sehr hoch, dadurch wird auch die Wartung schwierig. Ein weiteres Problem besteht in der Definition der Ausnahmeregeln. Dokumente, die in diese Kategorie fallen, müssen per Hand der Struktur zugewiesen werden. Das setzt eine Kenntnis der Struktur voraus.

Grundvoraussetzung für die Anwendung einer automatischen Kategorisierung ist die permanente Überwachung der Inhalte, um schnell auf neue

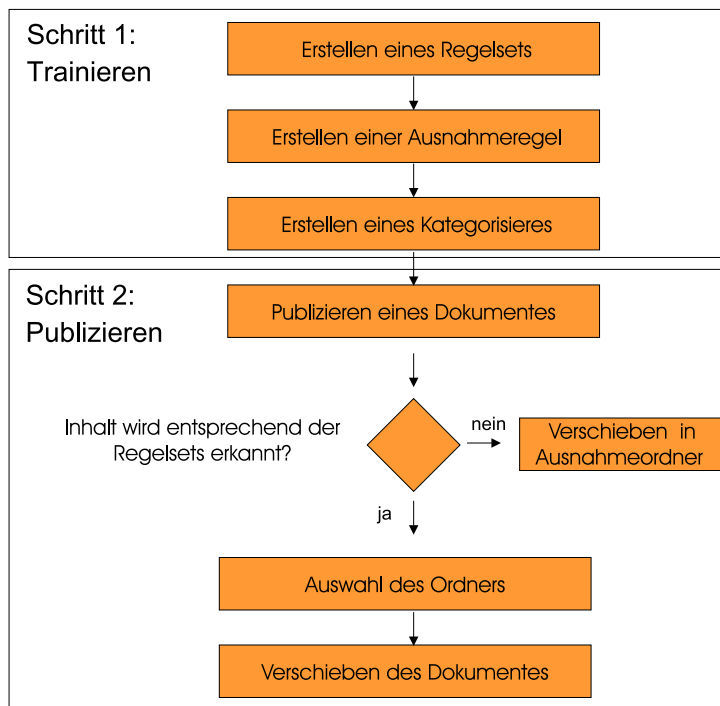


Abbildung 7: Prozess des automatischen Publizierens eines Dokumentes
Quelle: eigene Darstellung

Themen reagieren und neue Regelsets anlegen zu können. Dies war im Umfeld dieser Master Thesis nicht möglich.

Dass das automatische Publizieren aber durchaus eine große Hilfe sein kann, zeigt das Beispiel des Dokumentations- und Informationszentrums München GmbH (DIZ)¹⁶. Hier wird das Archiv der Süddeutschen Zeitung verwaltet. Mehrere tausend Artikel müssen täglich in ein Hyperwave System eingepflegt werden. Unterstützt werden die Redakteure durch ein Tool, das eine Vorabqualifizierung der neuen Artikel vornimmt und dem Redakteur Vorschläge unterbreitet. Das Funktionieren dieses Tools ist aber nicht nur von der unterstützenden Software abhängig, sondern auch ganz wesentlich vom organisatorischen Aufbau.

Die Redakteure sind ein relativ kleines Team, das mit dem Inhalt und dem Aufbau vertraut ist. So ist eine hohe Datenqualität durch kontinuierliche Arbeit gewährleistet. Im Zusammenhang mit dieser Master Thesis sollte aber der Ansatz verfolgt werden, dass viele unterschiedliche Mitarbeiter auf

¹⁶Siehe hierzu <http://www.diz-muenchen.de/html/wir.html>

möglichst einfache Art Dokumente publizieren. Um dies zu erreichen, wurde auf das automatische Publizieren verzichtet.

5.4 Workflow

Zur Abstimmung von inhaltlichen und gestalterischen Fragen sind in vielen Content Management Systemen Workflows etabliert. Diese haben den Zweck, dass ein Dokument vor seiner Veröffentlichung von den entscheidenden Beteiligten auf seine Richtigkeit geprüft werden kann. Durch den standardisierten Prozess, der innerhalb eines Workflows abläuft, kann eine hohe Qualität erreicht werden.

Auch mit dem Hyperwave Information Server ist das Erstellen von Workflows möglich. Im vorliegenden Fall wurden allerdings aus folgenden Gründen keine Workflows erstellt:

1. An den recherchierten Dokumenten werden keine Änderungen vorgenommen.
Alle Dokumente, die in die Struktur eingepflegt werden sollen, werden extern verfasst. Das bedeutet, dass es im nachhinein zu keinen Änderungen am Inhalt, der Gestaltung oder anderen Bestandteilen des Dokumentes kommt.
2. Das Hauptkommunikationsmittel ist der E-Mail-Versand
Gibt es Unklarheiten bezüglich eines Dokumentes, so kann auf die Standardkommunikationswege innerhalb eines Unternehmens zugegriffen werden. Im Sinne einer Vereinfachung erscheint es sinnvoll, Technologie nur dann einzusetzen, wenn der tatsächliche Nutzen dem Aufwand in einem angemessenen Mass gegenübersteht. Hat ein Mitarbeiter Klärungsbedarf, kann er jederzeit das Dokument an einen Kollegen schicken und um dessen Meinung fragen. In einem mittelgroßen Unternehmen sind die entsprechenden Kontaktpersonen bekannt. Der Aufbau umfangreicher, dauernd zu pflegender Prozess- und Workflowfunktionalitäten ist vor dem Hintergrund des tatsächlichen Nutzens deswegen kritisch zu sehen.
3. Mit Hilfe von organisatorischen Abstimmungen und Tools ist die Qualität hoch.
In Kapitel 8 Strategien und Hilfsmittel: Maintenance auf Seite 80 sind die Möglichkeiten beschrieben, die die Qualität innerhalb der Struktur und des Inhalts auf einem hohen Niveau zu halten.

6 Strategien und Hilfsmittel: Navigieren

6.1 Visualisierung

Visualisierung wird ganz allgemein definiert als

„Bezeichnung für bildliche Formulierung und Kommunikation, das heißt für die Aufbereitung von Information mit vor allem bildlichen Mitteln wie auch für die visuelle Wahrnehmung.“(Brockhaus [9], 23. Band, Seite 380)

Unter Visualisierung wird im Zusammenhang mit dieser Master Thesis die grafische Darstellung von Strukturen und Informationen verstanden. Es geht darum, dem Benutzer eine Oberfläche anzubieten, die intuitiv schnell erfassbar, in sich geschlossen ist und ein mentales Modell abbildet, das dem des Benutzers entspricht.

Im Laufe der Zeit wurden diverse Systeme entwickelt, um große Dokumentenmengen darzustellen.¹⁷ Unter Informationsvisualisierung versteht Däßler ([5], Seite 3) dabei

„alle Konzepte, Methoden und Tools zur visuellen Darstellung von Informationen aus Datenbanken, digitalen Bibliotheken oder anderen Dokumentsammlungen[...].“

Weiter lassen sich nach Däßler ([5], Seite 3) dabei

„vier elementare Aufgaben der Visualisierung formulieren:

- Das Auffinden spezifischer Informationsdomänen in großen Datenbeständen.
- Das Erkennen von Relationen, Strukturen oder Trends in unstrukturiert erscheinenden Informationsmengen.
- Multiple Views, das heißt verschiedene Sichten auf identische Datenbestände
- Die Darstellung von Informationen im Kontext zu anderen Informationen“

¹⁷Einen kleinen Überblick bietet <http://www.cg.tuwien.ac.at/courses/InfoVis/Links> und <http://www.cybergeography.org/atlas/>

Bei der Visualisierung von Daten ist es wichtig, zu unterscheiden, aus welchem Bereich die Daten stammen. Geht es darum, zum Beispiel wissenschaftliche Daten aus der Geographie oder Medizin zu visualisieren, liegen gebräuchliche Metaphern vor, die identifiziert werden können. Bei Dokumenten ist dies schwieriger, weil die Information abstrakt, das heißt in Textform, vorliegt.

Weiter schreibt Däßler ([5], Seite 7):

„Um abstrakte Informationen, das heißt Dokumente, Begriffe oder andere textbasierte Informationen aus Datenbanken, Retrievalsystemen oder im Web zu visualisieren, ist es zunächst notwendig, einen Informationsraum zu definieren. Der Informationsraum besitzt in der Regel eine spezifische Raumstruktur, in die Objekte und Objektbeziehungen kartographiert (engl. Mapping) werden. Das Mapping von abstrakten Informationen kann wesentlich komplizierter sein als bei der wissenschaftlichen Visualisierung, wo durch Simulationsparameter oder die Realität bereits räumliche Dimensionen vorgegeben sind.“

6.2 Hierarchische Visualisierung

In dieser Master Thesis wird besonders die Darstellung hierarchischer Strukturen betrachtet, die in einer Datenbank abgelegt sind. Für hierarchische Daten gibt es mehrere Möglichkeiten der Visualisierung:

- Baumdarstellung
- Treemap
- Cone Trees und Cam Trees
- Fractal Tree
- Hyperbolic Tree

6.2.1 Baumdarstellung

Baumdarstellungen entsprechen den Vorstellungen, die ein Benutzer an ein herkömmliches Ordnersystem hat. Ausgehend von einem Startverzeichnis zweigen beliebig viele Knoten ab und es entstehen Subverzeichnisse. Ein Subverzeichnis nennt man auch Sohn oder Nachfahre. Es entsteht so ein hierarchisches System von Verzeichnissen, in dem Relationen und Ähnlichkeiten

dargestellt werden. Je tiefer die Navigation fortschreitet, umso ähnlicher werden sich die Objekte.

An die Grenzen der Leistungsfähigkeit stößt diese Darstellung, wenn sehr viele Informationen angezeigt werden sollen und es entweder zu flache oder zu tiefe Strukturen gibt. Bei zu flachen Strukturen fällt es schwer, einen Überblick zu gewinnen und einen Einstiegspunkt zu finden. Gibt es eine zu große Anzahl von Elementen in der Tiefe, dauert es zu lange, bis man zu einem Thema gelangt.

Der große Vorteil dieser Darstellung ist allerdings, dass wohl jeder Benutzer schon einmal mit dieser Technik gearbeitet hat, man die Darstellung deswegen gewohnt ist und damit umgehen kann. Bei allen anderen Darstellungen, die in diesem Kapitel besprochen werden, handelt es sich um „Exoten“, die keine größere Verbreitung nachweisen können und deswegen mit Akzeptanzproblemen zu kämpfen haben. Deshalb wurde auch die bestehende Darstellung beibehalten, zumindest um die normale Navigation abzubilden.

Das bekannteste Beispiel für die Darstellung eines hierarchischen Verzeichnisbaumes ist der Windows Explorer.

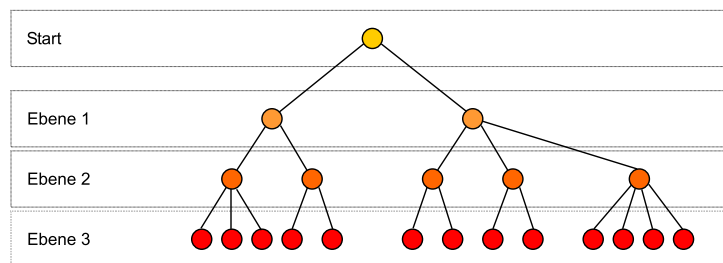


Abbildung 8: Schema einer Baumdarstellung
Quelle: eigene Darstellung

6.2.2 Treemap

Bei der Darstellung einer Treemap benutzt man die Eigenschaft des Menschen, Bilder schnell zu erfassen und zu verstehen. In einer Treemap werden die Informationen als Rechtecke dargestellt, die ineinander verschachtelt sind. Die Größen der Rechtecke sind unterschiedlich, je nach dem, wie viele und wie große Dokumente sich in einem Cluster befinden. Durch Farben werden

die Dateitypen kenntlich gemacht. Treemaps eignen sich gut dafür, hierarchische Verzeichnisse zu visualisieren, so dass man zum einen einen schnellen Überblick bekommt und zum anderen ohne viel Klicken zum gewünschten Objekt gelangt.

Probleme bei der Treemap sind laut Däßler [5], dass ein Navigieren aufgrund der fehlenden expliziten Strukturdarstellung zum Kontextverlust und zur Desorientierung führen kann. Die Darstellung wirkt unter Umständen infolge der Vernetzung und Verzerrung überladen. Darüber hinaus mag es schwierig sein, die Hierarchieebenen der Regionen auseinanderzuhalten.

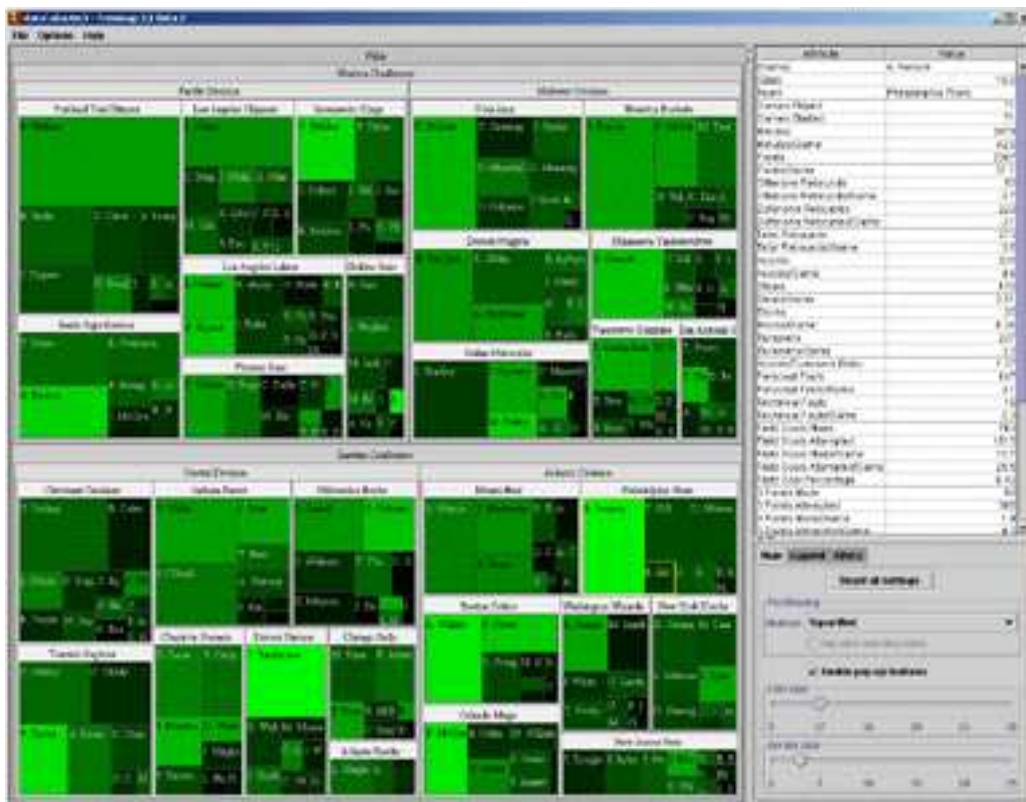


Abbildung 9: Darstellung von Verzeichnissen einer Treemap
 Quelle: <http://www.cs.umd.edu/hcil/treemap/#publications>

6.2.3 Cone Trees und Cam Trees

Das Konzept der Cone beziehungsweise Cam Trees wurde 1991 von der User Interface Group des Xerox Palo Alto Research Center (PARC) veröffentlicht.

Im Gegensatz zu den bisher vorgestellten Methoden wird hier eine dritte Dimension der Visualisierung eingeführt. Die dritte Dimension wird ausgenutzt, indem die Verzeichnishierarchie kegelförmig aufgespannt ist. Sind die Kegel horizontal angeordnet, spricht man von Cam Trees, vertikal heißen sie Cone Trees. Auf jedem Kegel gibt es Objekte, die im Vordergrund und solche, die im Hintergrund sind.

Diese Darstellung stößt an ihre Grenzen, wenn mehr als 1000 Knoten oder 30 Verzweigungen pro Ebene dargestellt werden sollen oder es mehr als zehn Ebenen gibt (Däßler in [5]).

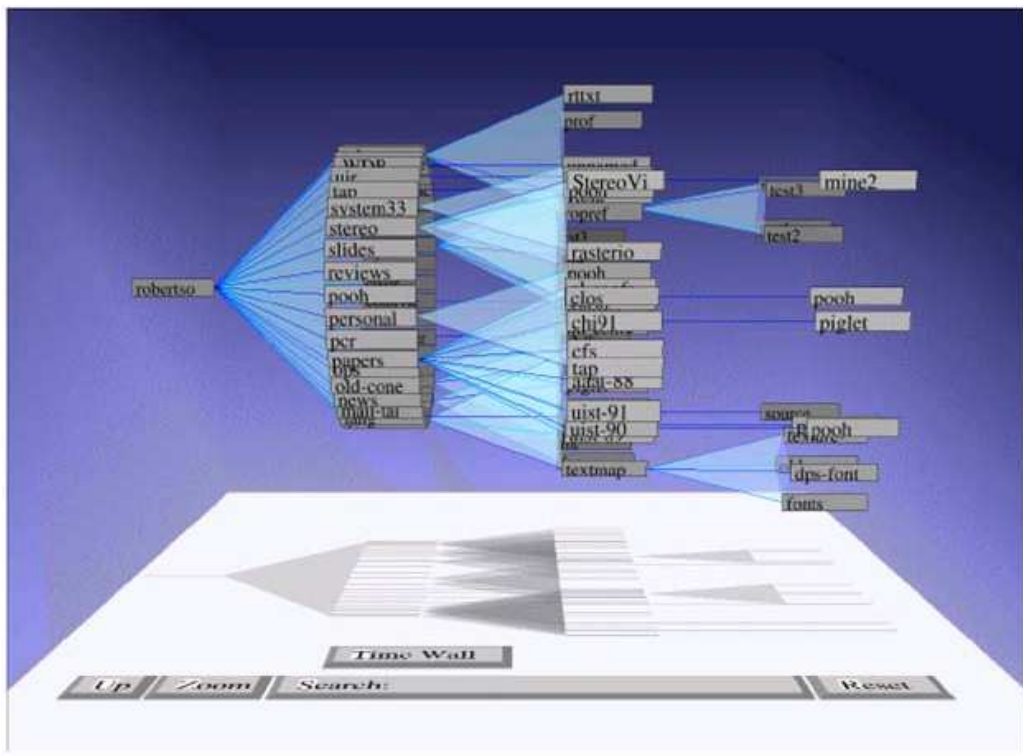


Abbildung 10: Darstellung eines Cam Trees

Quelle:

<http://www.ifs.tuwien.ac.at/mlanzenberger/psinfovis/ss03/img/cam.jpg>

6.2.4 Fractal Tree

Dem Problem der begrenzten Darstellungsfähigkeit der Cone Trees beziehungsweise Cam Trees bei großen Strukturen begegnet das Konzept des Frac-

tal Trees. Es wurde unter anderem von Hideki Koike und Hirotaka Yoshihara entwickelt. Laut Koike/Yoshihara [17] hilft die fraktale Darstellung dabei, die Kontrolle über eine große Anzahl von Dokumenten zu erhalten. Vor allem deswegen, weil mit dem Fractal Tree auch tiefere Verzeichnisstrukturen abgebildet werden können.

Dies wird dadurch erreicht, dass jede neue Ebene um einen konstanten Fak-

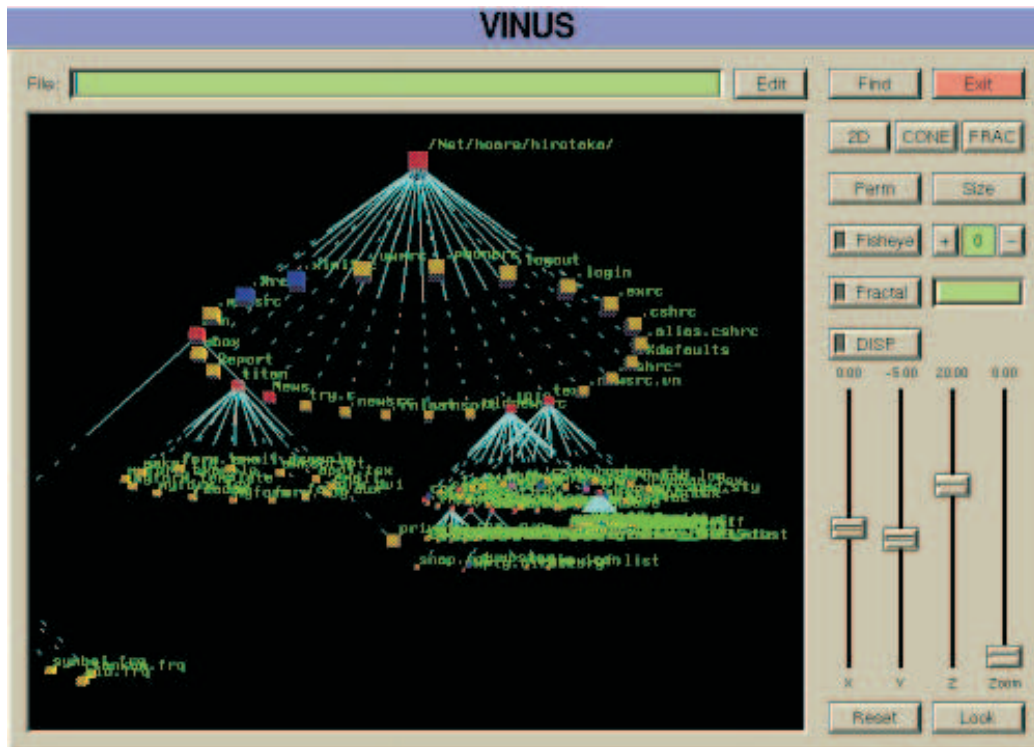


Abbildung 11: Darstellung eines Fractal Trees

Quelle: <http://www.vogue.is.uec.ac.jp/~koike/vinus/vinus.html>

tor kleiner dargestellt wird als die vorhergehende. Das Problem dabei ist, dass sich die Darstellung bei einem tief verzweigten Baum exponentiell verkleinert. Erst durch Klicken auf einen Baum wird dieser auf die volle Bildschirmbreite vergrößert. Dadurch geht aber die Übersichtlichkeit verloren.

Wie auf der Abbildung zu erkennen ist, kann die Darstellungsgenauigkeit mit verschiedenen Schieberegler gesteuert werden. Die Handhabung der Regler erfordert einiges Fingerspitzengefühl.¹⁸

¹⁸zu testen unter <http://id.mind.net/~zona/mmts/geometrySection/fractals/tree/treeFractal.html>

6.2.5 Hyperbolic Tree

Der Vorteil des Hyperbolic Trees liegt in der Möglichkeit, sehr große Verzeichnisstrukturen darstellen zu können. Dies erreicht man dadurch, dass ein räumlich verzerrter Blick auf die Struktur berechnet wird. Objekte, die im Vordergrund sind, werden größer angezeigt, Objekte, die in der nächsten Ebene stehen, kleiner und so weiter. Man kann sich den hyperbolischen Raum vereinfacht als eine Kugel vorstellen, die gedreht werden kann. Zu Beginn ist der Baum auf den Startpunkt fokussiert. Zieht man ein Objekt aus dem Umkreis ins Zentrum, wird dieses immer größer dargestellt und mit ihm auch die zusammenhängenden Strukturen. Der große Vorteil des Hyperbolic Trees ist die schnelle Navigationsmöglichkeit in großen Verzeichnisstrukturen.

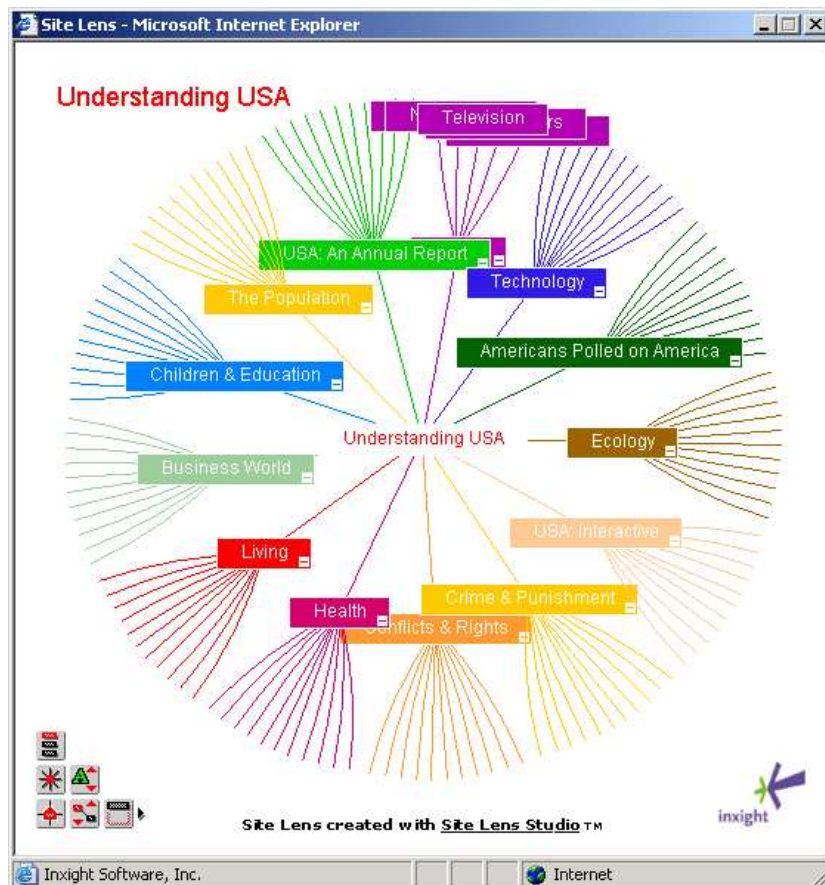


Abbildung 12: Darstellung eines Hyperbolic Trees
Quelle: <http://www.inight.com/products/sdks/st/features.php>

Problem dieser Darstellung ist, dass sie für den Benutzer eine sehr unge-

wöhnliche Darstellung ist, die sich dynamisch verändert (siehe hierzu auch Kapitel 6.4.2 Dynamische Darstellung einer Sitemap auf Seite 69).

Weder das Zentrum/Einstiegspunkt bleiben als Konstante erhalten, noch ist ein ganzheitlicher Überblick möglich. Vielmehr muss man sich immer wieder in ganz unterschiedlichen Bereichen innerhalb der Darstellung bewegen und verliert so leicht den Gesamtüberblick.

Dennoch ist diese Darstellung sehr interessant, vor allem für User, die sich in der Struktur auskennen und schnell in tiefer verzweigte Regionen vorstoßen möchten.

6.3 Netzartige Visualisierung

Die hierarchische Darstellung von Verzeichnissen stößt dort an ihre Grenze, wo die Dokumente untereinander in Beziehung stehen. Mögen die Beziehungen der Sites einer Homepage vielleicht noch darzustellen sein, die Beziehung zwischen mehreren Websites ist so nicht mehr zu visualisieren, weil die Beziehungen netzartig sind und es deswegen keinen zentralen Ausgangspunkt mehr gibt, von dem aus man einsteigen könnte. Durch die Verlinkung gibt es viele gleichberechtigte Einstiegspunkte, die miteinander in Kontakt stehen; sei es durch eine direkte Verbindung oder durch einen oder mehrere Zwischenschritte. Zur Visualisierung dieser Beziehungen wurden verschiedene Systeme entwickelt, von denen sich allerdings bisher kein System am Markt durchsetzen konnte. Um die Ansätze zu zeigen, seien hier folgende Anwendungen genannt:

- Hyperspace - Visualisierung von Hyperlinkstrukturen
- Hotsauce - Webseitenbrowser
- Infosky - Visualisierung von vielen Dateien mit der Metapher einer Galaxis

6.3.1 Hyperspace - Visualisierung von Hyperlinkstrukturen

Hyperspace wurde 1995 an der University of Birmingham entwickelt, mit dem Ziel, die Hyperlinkstrukturen im Web zu visualisieren. Die Darstellung von Verzeichnissen erfolgt mit Hilfe von Kugeln. Über Hyperlinks verknüpfte Dokumente werden nahe beieinander angezeigt, andere Dokumente ohne Links sind weiter voneinander entfernt. Die Objekte selbst sind im Raum frei verteilt.

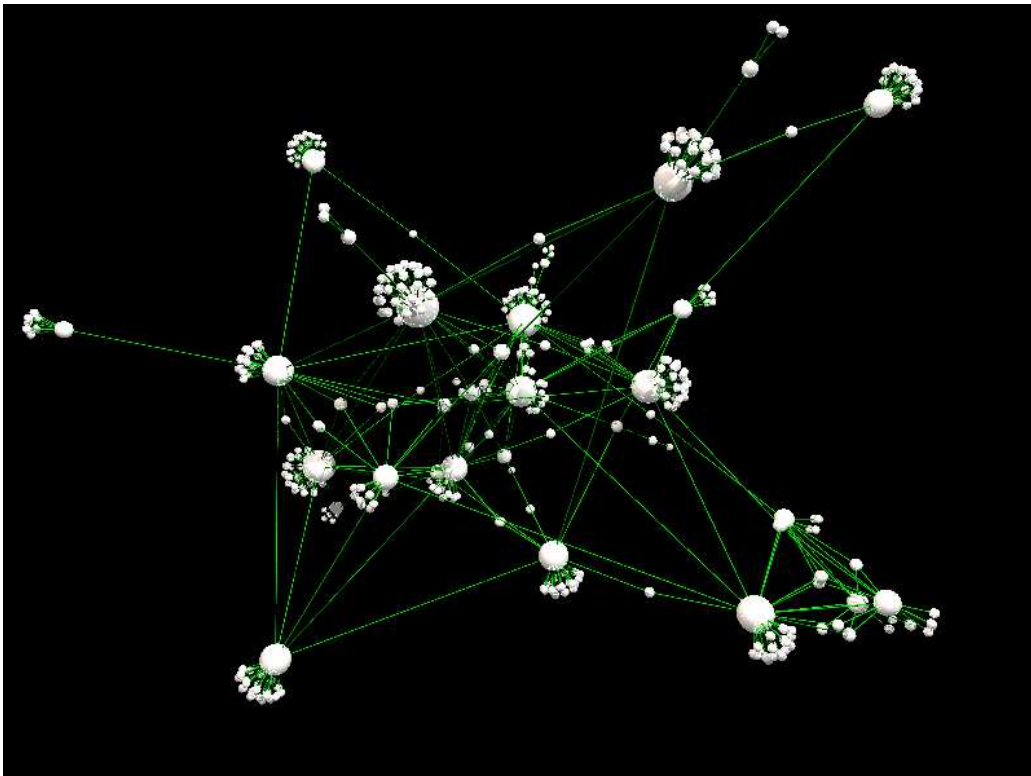


Abbildung 13: Darstellung einer Website mit Hyperspace

Quelle:

http://www.igd.fhg.de/archive/1995_www95/proceedings/posters/35/

6.3.2 Hotsauce - Webseitenbrowser

Hotsauce war ein Projekt von Apple,¹⁹ bei dem es darum ging, Verzeichnisstrukturen dreidimensional darzustellen. Dazu wurde ein eigenes Dateiformat definiert und über ein PlugIn war es möglich, mit einem Internetbrowser zu navigieren. Es war geplant, das Dateiformat frei zugänglich zu machen, so dass Websites ihren Inhalt für das PlugIn aufbereiten hätten können. Abbildung 14 zeigt den Einstiegspunkt zur Navigation. Man kann hier schon erkennen, wie schwierig es ist, sich in einem dreidimensionalen Raum zu bewegen, da durch die Überlappungen nur wenige Ebenen hintereinander sichtbar sind. Wenn man nicht weiß, wo man suchen soll, fällt es schwer, zu Informationen zu gelangen. Wie schon die vielen „toten“ Links in den Websites der Fußnoten zeigen, war Hotsauce kein Erfolg vergönnt.

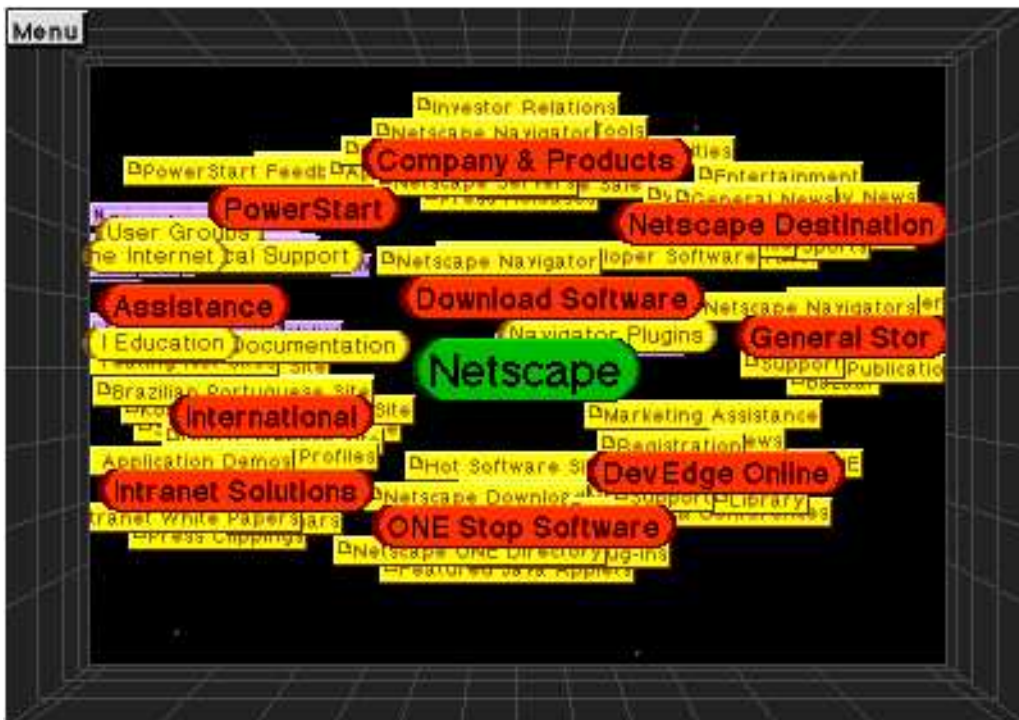


Abbildung 14: Hotsauce - dreidimensionales Navigieren
Quelle: <http://www.alvitec.ch/hotsauce/img/projekt.gif>

¹⁹Siehe hierzu die Websites <http://www.intern.de/97/05/06.htm>, sowie auch <http://www.xspace.net/hotsauce/>

6.3.3 Infosky

Infosky wurde gemeinsam von Hyperwave und dem KnowCenter der TU Graz entwickelt. Ziel war es, ein System zu entwickeln, das „die effiziente Erforschung großer, hierarchisch strukturierter Dokumentablagesysteme ermöglicht.“[15]

Infosky soll laut Kappe et al. [15] folgende Anforderungen erfüllen:

- Skalierbarkeit:
Visualisierung großer, hierarchisch strukturierter Dokument-Ablagen (Millionen von Objekten)
- Hierarchie plus Ähnlichkeit:
Repräsentation sowohl der Hierarchie als auch der Dokument-Ähnlichkeit in einer einzigen, konsistenten Visualisierung
- Fokus plus Kontext:
Integration von lokaler und globaler Sicht des Informationsraumes in einer einzigen Visualisierung
- Stabilität:
Verwendung einer Metapher, die die Wiedererkennung von „Plätzen“ im Informationsraum unterstützt. Insbesondere soll die Visualisierung auf globaler Ebene relativ robust gegen Änderungen auf lokaler Ebene sein, so dass dieselben Objekte mit hoher Wahrscheinlichkeit an denselben Orten bleiben.
- Einheitliche Sicht:
Unterstützung einer einheitlichen, konsistenten Sicht auf den Informationsraum für alle Benutzer, unabhängig von deren Zugriffsrechten, so dass eine Verständigung darüber zwischen den Benutzern möglich wird.
- Exploration:
Einfache, intuitive Navigation und Suche im Informationsraum. Im Idealfall ist die Visualisierung in der Lage, eine möglichst große Zahl an Dokument-Eigenschaften (zum Beispiel: Alter des Dokuments) und Beziehungen zwischen den Dokumenten darzustellen.

Als Metapher zur Visualisierung der Hierarchie wurde der Sternenhimmel gewählt, der aus Sternen, Galaxien und Clustern besteht. In diesem kann beliebig navigiert und gezoomt werden. Dabei entsprechen Sterne einem Dokument, Galaxien repräsentieren Ordner. Mehrere Ordner bilden ein Cluster.

Infosky erzeugt selbst keine Hierarchie, sondern verwendet die vorgegebene Ordnerstruktur. Voraussetzung für die Visualisierung ist ein signifikanter Text-Inhalt der Dokumente, der ausgewertet werden kann. Ausgehend vom Inhalt werden Ähnlichkeiten bestimmt, die mit der oben genannten Metapher dargestellt werden können. Dokumente mit ähnlichem Inhalt liegen in einer Galaxie nahe beieinander.

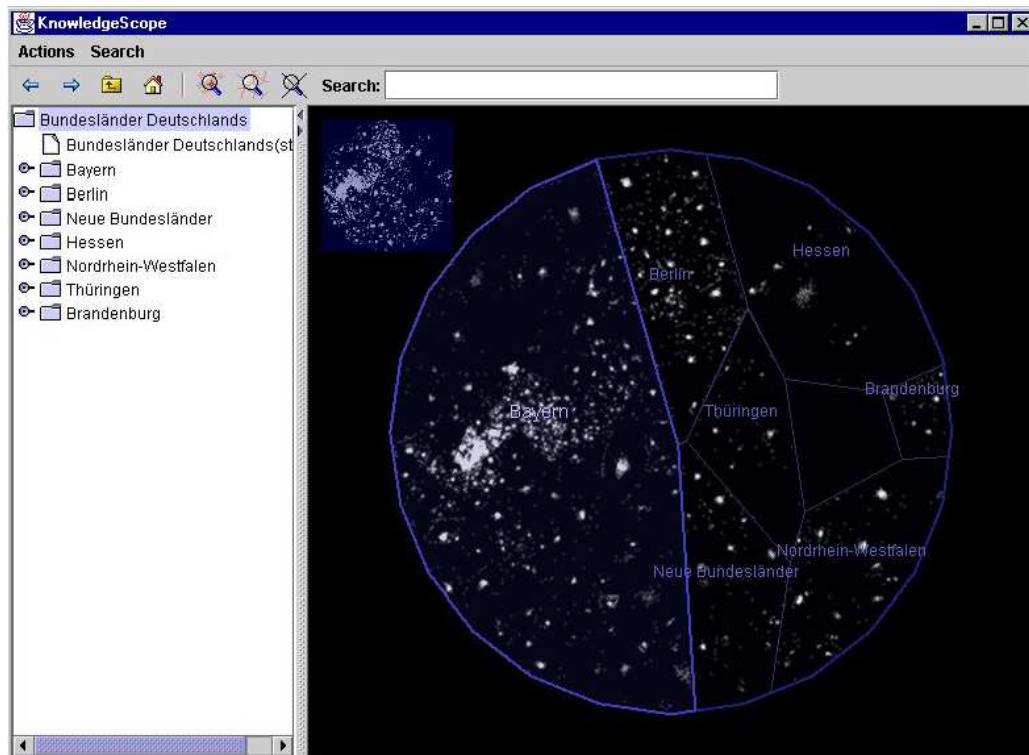


Abbildung 15: Benutzerinterface von Infosky

Quelle: The InfoSky Visual Explorer: Exploiting Hierarchical Structure and Document Similarities; Proc. of HCI International 2003 (Computers and Advanced Technology in Education), Volume 3, 1268-1272, Lawrence Erlbaum Associates

6.4 Sitemap-Darstellungen

Sitemap-Darstellungen sind auf Websites ein beliebtes Medium, um einen schnellen Überblick über den Aufbau und die Struktur der Website zu geben. Es gibt verschiedenen Arten von Sitemaps, von denen einige im Folgenden kurz erläutert werden.

6.4.1 Listen

In diesem Fall werden alle Strukturpunkte einer Sitemap in Form von Listeneinträgen dargestellt. Je nach gewählter Bildschirmauflösung gibt es unterschiedlich viele Spalten. Innerhalb eines Eintrages kann es mehrere Unterpunkte geben, die die Hierarchie darstellen. Diese Art der Darstellung ist im Web häufig anzutreffen, weil sie einfach zu pflegen ist. Das Problem vieler Sitemaps ist, dass sie nur schwer zu finden sind. Die Suche auf der Homepage von Audi (<http://www.audi.de>) listet die Sitemap zum Beispiel unter dem Titel *Inhalt* auf.

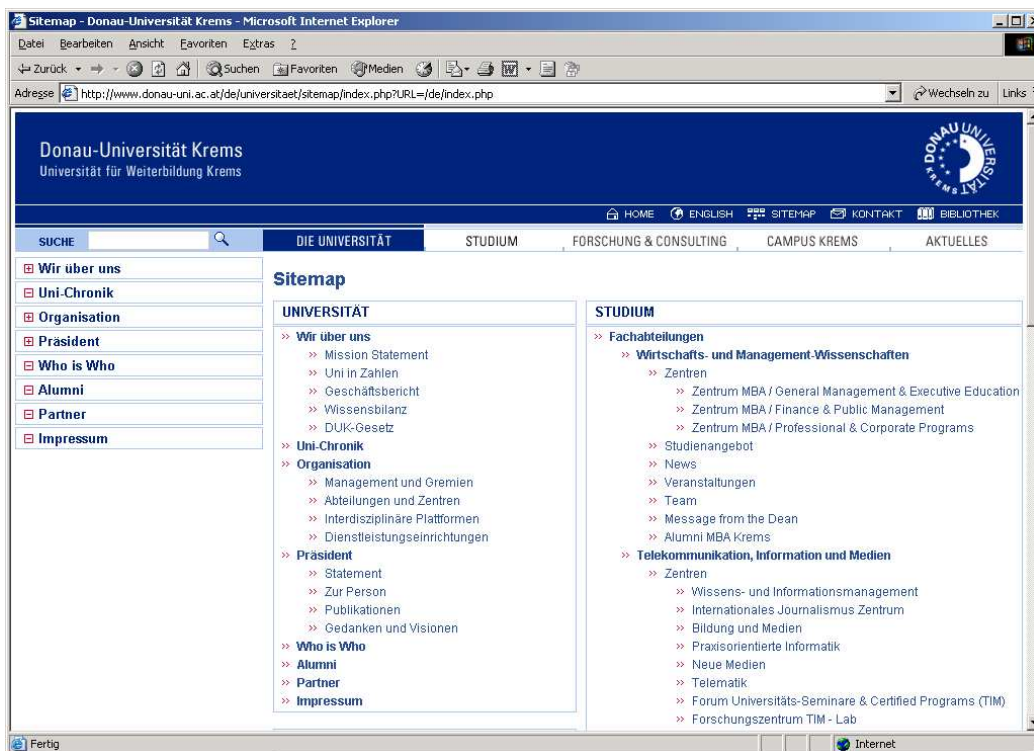


Abbildung 16: Sitemap-Darstellung in Form einer Liste
Quelle: <http://www.donau-uni.ac.at/en/universitaet/sitemap/index>

6.5 Alphabetisches Stichwortverzeichnis

Im Unterschied zur Sitemap werden im alphabetischen Stichwortverzeichnis nicht die Struktur, sondern alle Begriffe, die in den Titeln der Struktur vorkommen, abgebildet. Die Einträge sind mit Links versehen, so dass man durch Klicken darauf zu dem entsprechenden Strukturpunkt springen kann.

Diese Darstellungsform ist als schnelle Navigationsmöglichkeit gedacht, für den Fall, dass man einen Begriff bereits kennt, aber nicht genau weiß, wo man ihn suchen soll. Natürlich gibt es die Möglichkeit, über die Suche an die selbe Stelle zu gelangen. Das setzt allerdings voraus, dass man den Begriff exakt kennt. Es kann aber vorkommen, dass man nur eine ungefähre Vorstellung hat und in der Umgebung des Begriffes in der Tabelle noch einen besseren Stichpunkt findet.

6.6 Speedtyping

Speedtyping ist eine schnelle Möglichkeit, um in einem Verzeichnis zu navigieren. Es wird zum Beispiel in der Hilfe von Microsoft Word im Bereich Index verwendet. Durch Eintippen einer Buchstabenkombination schränkt man den übereinstimmenden Bereich immer weiter ein. Bereits nach wenigen Buchstaben hat man so die Trefferliste stark verkürzt und kann durch Klicken auf den Begriff zum gewünschten Ziel springen.

Speedtyping wurde in dem Rahmen dieser Arbeit nicht implementiert, da

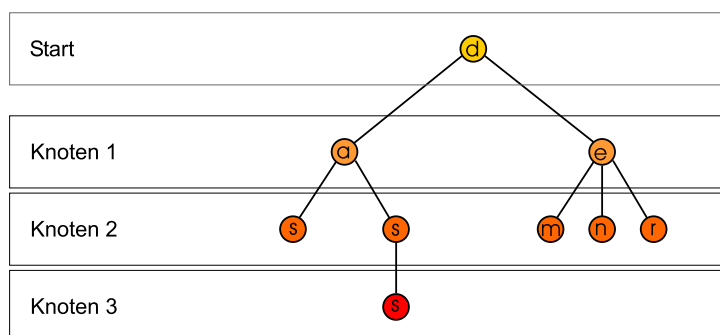


Abbildung 18: Aufbau eines Index

Quelle: eigene Darstellung

der technische Aufwand zur Umsetzung zu groß ist. Unter anderem ist das Rechte- und Rollenkonzept zu berücksichtigen, das erlaubt, nur die Objekte darzustellen, die gelesen werden dürfen. Das bedeutet, dass eventuell für jeden Benutzer eigene Listen mit den lesbaren Objekten erzeugt werden, auf

die mit Speedtyping zugegriffen wird, oder dass während des Eintippens die Rechte kontrolliert werden. Bei der zweiten Variante ist mit Performance-schwierigkeiten zu rechnen, die erste Variante erzeugt eine große Anzahl an Listen, die verwaltet und stets aktualisiert werden müssen.

7 Strategien und Hilfsmittel: Suche

Die Verbesserungsmöglichkeiten im Zusammenhang mit dem Einsatz der Suche betreffen zwei Punkte:

- Eingabe des Suchbegriffs
Hier gibt es einige wichtige Punkte, die beachtet werden müssen, um eine Suche einfach ausführen zu können.
- Trefferliste
Die Trefferliste ist eine dynamische Darstellung der Inhalte einer Struktur, die sich auch dynamisch verändern lassen muss. Dazu gibt es eine Reihe von Hilfsmitteln (Thesaurus, Kapitel 7.1.1, Seite 73 und ähnliche Dokumente, Kapitel 7.4.1 auf Seite 78), die hilfreich zum Einsatz kommen können.

Mit dem Hyperwave Information Server ist es möglich, zwei verschiedene Suchmaschinen anzubinden:

- Die Suchmaschine von Verity verfolgt den Ansatz einer indexbasierten Suche. Das bedeutet, dass der Inhalt aller Dokumente beim Publizieren erfasst und in einen Index aufgenommen wird.
- Die Suchmaschine von Autonomy hingegen erfasst den Inhalt eines Dokumentes nicht über den Inhalt, sondern sie sucht Muster und vergleicht diese mit einander.

Im Folgenden werden beide Suchansätze noch genauer beschrieben.

7.1 Indexbasiertes Suchen

Bei der indexbasierten Suche wird jedes Dokument beim Abspeichern erfasst und sein Inhalt in einem Index abgelegt. Dieser Index umfasst den Begriff und die zugehörigen Informationen, in welchen Dokumenten der Begriff vorkommt.

Den Indizierungsprogrammen ist es möglich, alle gängigen Dateiformate zu durchsuchen und die einzelnen Begriffe zu extrahieren. Mit Hilfe einer Stopwortliste werden zu häufig vorkommende Worte erst gar nicht in den Index übernommen. So führt zum Beispiel eine Suche nach „der“, „die“ oder „das“ in der Regel zu keinem Treffer. Dadurch vermeidet man eine zu große

Anzahl von Treffern bei Kombinationen von Suchbegriffen mit dem Booleschen Operator „or“ und mehreren Begriffen wie zum Beispiel „Das Buch der Bücher“.

In diesem Index werden alle Begriffe erfasst, unabhängig davon, ob sie richtig geschrieben sind. Lässt man sich einen Gesamtwortschatz ausgeben, kann man sehr einfach auch die unterschiedlichen Schreibweisen von Begriffen sehen, was zur Korrektur der Unternehmenssprache durchaus hilfreich sein kann.

Da das indexbasierte Suchen alleine auf die im Index erfassten Begriffe zugreifen kann, gibt es Hilfsmittel, die die Qualität der Suchergebnisse verbessern. Einige werden im Folgenden erläutert.

7.1.1 Thesaurus

Der Thesaurus ist laut Brockhaus ([9], Band 22, Seite 99) die

„alphabetisch und systematisch geordnete Sammlung aller sprachl. und sonstigen Beziehungen eines bestimmten Anwendungsbereichs (zum Beispiel einer Fachsprache) nach den semant. Beziehungen in einem System syntagmat. und paradigm. Querverweise. Die T. sind als Funktionsträger im Rahmen eines Dokumentationssystems grundlegende Hilfsmittel zur Wiederauffindung und inhaltl. Erschließung von Dokumenten und zur Wiedergewinnung von Information über jedes gewünschte Element des erfassten Bereichs.“

Abbildung 19 zeigt einen Ausschnitt aus einem Thesaurus,²⁰ der die Einordnung des Begriffes „Euro“ in ein Schema darstellt. Ein Thesaurus bietet die Möglichkeit, die Trefferliste einer Suchmaschine zu beeinflussen. Der Hyperwave Information Server wird in der Regel zusammen mit der Volltextsuchmaschine von Verity ausgeliefert. Diese Software lässt die Verwendung eines Thesaurus zu. In der Verity-Version des Hyperwave-Servers ist bereits ein kleiner Standard-Thesaurus integriert, der aber für die Anforderungen hier nicht ausreichend ist.

Verity bietet die nötige Software an, um den Thesaurus anzupassen. Auf diese Arbeit wurde aber verzichtet, da die Erstellung - unabhängig von der Software - als zu aufwendig angesehen wurde.

²⁰siehe zum Beispiel http://www.gbi.de/_de/thesaurus/

| | |
|--------------------|--|
| Oberbegriff | Wahrung |
| Synonym | Weltwahrung |
| Unterbegriff | DM |
| | Schilling |
| | Euro |
| Verwandte Begriffe | Devisen |
| | Geld |
| Zuordnung | V.05 Geld und Finanzmarkte |
| | V.07.03.01 Internationales Wahrungssystem |

Abbildung 19: Auszug aus einem Thesaurus
Quelle: eigene Darstellung aus
http://www.gbi.de/_de/thesaurus/ ubernommen

7.1.2 Synonymenliste

Ganz allgemein versteht man unter einem Synonym laut Brockhaus ([9] 21.Band, Seite 535) ein

„Wort, das mit einem anderen Wort oder einer Folge von Wortern derselben Sprache bedeutungsgleich, bedeutungsahnlich oder sinnverwandt ist.“

Fur das Wort „hoflich“ gibt es zum Beispiel die Synonyme galant, artig, nett, zuvorkommend, hilfsbereit.²¹ Die Schwierigkeit bei der Verwendung einer Synonymenliste besteht in der Pflege. Beim Erstellen einer neuen Datenbank sind die Ressourcen fur den Aufbau eventuell noch vorhanden. Bei der taglichen Arbeit mit einem System wird man schnell an die Grenzen des Machbaren stoen.

Ein weiterer Grund dafur war der unklare Prozess, wie die Pflege vorgenommen werden soll. In der Regel besitzt kein Mitarbeiter das Wissen, um den Wortschatz eines Unternehmens auch nur annahernd in einer Liste

²¹entnommen aus der Synonymenliste von Microsoft Word

abzubilden. Des Weiteren stellt sich die Frage, wer für die weitere Pflege verantwortlich ist und wer entscheidet, wann Begriffe synonym zueinander sind.

Bei der Verwendung von Synonymen sollte weiter beachtet werden, dass dadurch die Anzahl der Treffer anwächst. Dies kann auch zu Unzufriedenheit und Unverständnis bei den Anwendern führen. Deswegen sollte darauf geachtet werden, dass diese Funktionalität vom Benutzer selbst auswählbar ist. Auch sollte bei der Schulung auf die Unterschiede bei der Arbeitsweise der Suchmaschine hingewiesen werden.

7.2 Musterbasiertes Suchen

Basis für das Erkennen von Mustern der Software von Autonomy sind die Arbeiten der Mathematiker Thomas Bayes und Claude Shannon([2] Seite 8-9). Autonomy²² versucht, die Kernaussagen von Dokumenten - so genannte „Konzepte“ - mit mathematischen Verfahren aus einem Text zu extrahieren. Dabei spielt es dann keine Rolle mehr, in welcher Sprache oder welchem Alphabet ein Dokument verfasst ist. Das wichtigste Tool dafür ist die „Dynamic Reasoning Machine“. Diese Software hat vier Aufgaben([2] Seite 2):

- Vergleichen der Inhalte
Ein Quelltext wird analysiert und eine Liste mit Verweisen auf andere Dokumente, die ein hohes Maß an inhaltlicher Übereinstimmung aufweisen, wird übergeben.
- Erstellen von Agenten
Ein Text wird analysiert und als Ergebnis ein „Konzept-Agent“ generiert, der die Kernaussage des Textes in kodierter Form zusammen mit dem zugrunde liegenden Muster ausgibt.
- Trainieren von Agenten
Durch weitere Texte werden die Agenten trainiert und die Abstimmung der Muster immer weiter verbessert.
- Textsuche
Die Suchabfrage, in natürlicher Sprache eingegeben und/oder mit den booleschen Operatoren kombiniert, wird entgegengenommen und eine Trefferliste entsprechend der Relevanz angezeigt.

²²zum Thema Autonomy siehe auch <http://www.autonomy.com>

7.3 Suchmaske

Die Suchmaske darf den User durch Eingabefelder nicht überfordern. Der Erfolg von Google ist auch darin begründet, dass die Suchmaske sehr einfach gehalten ist und intuitiv schnell erlernt werden kann. Aus den Anforderungen aus Kapitel 3.2 Usecase Suchen ergeben sich folgende Umsetzungsvarianten:

- Aufrufbarkeit
Die Suche ist zu jedem Zeitpunkt aufrufbar, das heißt sie ist nicht in den Contentframe integriert, sondern immer im statischem Bereich.
- Einfache Suche
Dieser Teil der Suche entspricht der einfachen Suche, wie man sie auch von der Website von Google kennt. Das bedeutet, dass man einen Suchbegriff eingibt, die Suche startet und dann der gesamte Inhalt durchsucht wird. Es können ein oder mehrere Suchbegriffe verwendet werden.
- Erweiterte Suche
Die erweiterte Suche ist auch jederzeit aufrufbar. In ihr können einige zusätzliche Einstellungen zur besseren Einschränkung der Trefferliste vorgenommen werden. Die Einschränkungen sind möglich, weil die entsprechenden Attribute bereits beim Einstellen erfasst (siehe Kapitel 5) und anschließend indiziert werden.

Diese Ansätze beheben grundsätzliche Schwierigkeiten bei der Verwendung der Suche an sich. Das Problem für viele User ist aber der Suchbegriff selbst. Wie in Kapitel 3.2 Usecase Suchen auf Seite 21 beschrieben, ist häufig nicht klar, welchen Suchbegriff man verwenden soll beziehungsweise überhaupt kann. Hilfsmittel wie zum Beispiel ein Thesaurus (siehe Kapitel 7.1.1 auf Seite 73) und eine Synonymenliste (siehe Kapitel 7.1.2 auf Seite 74), helfen in diesem Zusammenhang zwar, es gibt aber noch weitere Möglichkeiten.

Verwendet man eine indexbasierte Suchmaschine wie zum Beispiel Verity (siehe Kapitel 7.1 auf Seite 72), so wird ein eigener Gesamtwortschatz erstellt. In ihm sind alle indizierten Worte erfasst. Diesen Gesamtwortschatz kann man als Grundlage für das Speedtyping (siehe Kapitel 6.6 Speedtyping auf Seite 70) bei der Eingabe des Suchbegriffes verwenden. Gibt ein Benutzer einen Suchbegriff ein, der in der Liste des Gesamtwortschatzes nicht auftaucht und so über das Speedtyping nicht gefunden wird, gibt es auch keine Treffer. So lässt sich schon vor dem Starten der Suchabfrage klären, ob es überhaupt Treffer geben wird. Schwieriger wird dieses Verfahren bei der Eingabe mehrerer Suchbegriffe, aber zumindest für alle Suchen zu einem Suchbegriff ist eine Hilfestellung gegeben.

7.4 Trefferliste

Bisher wird die Trefferliste nach einem internen Algorithmus der Suchmaschine generiert, der nicht offengelegt ist und auf den man auch kaum Einfluss hat. Neben den speziellen Sortierkriterien der verschiedenen Suchmaschinen gibt es aber eine Reihe von Kriterien, die wohl von den meisten Suchmaschinen - in unterschiedlicher Gewichtung - zur Erzeugung der Trefferliste herangezogen werden. Im Rahmen dieser Master Thesis wird davon ausgegangen, dass diese Kriterien eine gewichtige Rolle bei der Erzeugung der Reihenfolge der Trefferliste spielen. Die Kriterien können laut [3] zum Beispiel sein:

- Übereinstimmung mit den Metadaten
Zu manchen Dokumenten werden die Metadaten erfasst und können so ausgewertet werden (siehe dazu Kapitel 5.2.1 auf Seite 45).
- Die Anzahl der gefundenen Wörter
Die Häufigkeit, mit der ein Suchbegriff im Dokument vorkommt, ist ebenso ausschlaggebend. Bei der Eingabe von zum Beispiel fünf Suchbegriffen ist das Dokument relevanter, das statt zwei Begriffen vier gesuchte Begriffe enthält.
- Position der Wörter
Je nach der Position des Wortes innerhalb eines Satzes oder einer Überschrift, wird das Wort unterschiedlich hoch gewichtet. Wichtig kann in diesem Zusammenhang auch sein, ob das Wort zum Beispiel einen Link enthält oder fett gedruckt ist.
- Der Abstand der Suchbegriffe im Dokument
Bei der Suche nach mehreren Begriffen ist der Abstand zwischen ihnen von Bedeutung. Folgen die Suchbegriffe im Dokument unmittelbar hintereinander, wird angenommen, dass dieses Dokument von besonderer Relevanz ist.
- Die Häufigkeit von Suchbegriffen innerhalb von Dokumenten
Je häufiger ein Suchbegriff in einem Dokument auftaucht, um so höher wird das Dokument gewichtet.

Bei der Darstellung der Trefferliste werden vom Hyperwave Information Server die Logarithmen der Suchmaschinen von Verity und Autonomy bereits berücksichtigt.

Neben dem Titel werden in der Standard-Hyperwave-Trefferliste noch diverse andere Attribute (zum Beispiel Autor, Datum der Erstellung, Datum

der Änderung, Dateigröße) angezeigt, darunter auch der Pfad, in dem das Dokument steht. Von der Trefferliste aus ist es möglich, über den Pfad direkt zum Dokument zu springen.

Wie in den Kapiteln 4.3 Zentrale Strukturen, 4.4 Dezentrale Strukturen und 4.5 Mischformen beschrieben, ist es aber außerordentlich wichtig, mit intuitiven, immer wiederkehrenden Strukturen zu arbeiten. Eine Trefferliste, die darauf keine Rücksicht nimmt und ausschließlich mit einem statistischem Modell die Relevanz berücksichtigt, ist wenig hilfreich. Vielmehr sollte die Trefferliste so gestaltet sein, dass die Treffer in ihrem Verhältnis zur Struktur dargestellt werden und erst innerhalb der Struktur die Relevanz eine Rolle spielt.

7.4.1 Ähnliche Dokumente

Neben der Verwendung des Thesaurus ist es mit Verity auch möglich, eine Suche nach ähnlichen Dokumenten durchzuführen. Dazu ist es nicht nötig, den Suchbegriff ein weiteres mal einzutippen, die Suche kann vielmehr über einen zusätzlichen Button gestartet werden.

Für die ähnliche Suche können bei Verity sowohl ganze Dokumente als auch Textpassagen verwendet werden. Der Befehl zum Aufrufen dieser Funktion lautet `LIKE`. Die Ähnlichkeit von Dokumenten untereinander wird durch eine statistische Bewertung von Begriffen ermittelt, die miteinander verglichen werden. Verity ist in der Lage, die wichtigsten Begriffe eines Textes zu extrahieren und kann deswegen Texte auch miteinander vergleichen. Aus diesen Vergleichen lässt sich auch auf eine Ähnlichkeit schließen.

7.4.2 Unscharfe Suche

Die unscharfe Suche berücksichtigt unterschiedliche Schreibweisen beziehungsweise Rechtschreibfehler bei der Eingabe eines Suchbegriffes. Ein anderer Begriff anstelle von unscharfer Suche lautet fehlertolerante Suche. Mit diesem Begriff wird noch deutlicher, für welchen Bereich dieses Tool geeignet ist. Bei der Internetsuchmaschine Google ist dieses Konzept umgesetzt. Hat man hier einen Suchbegriff falsch eingegeben, zum Beispiel „Suchmaschiene“, so zeigt Google in der Trefferliste ganz oben einen Alternativbegriff mit der wahrscheinlichen Schreibweise an. Abbildung 20 zeigt dies. Um dieses Verhalten mit Verity abzubilden, gibt es zwei Aufrufe.

Mit dem Aufruf `<SOUNDEX>` wird eine Suche nach ähnlich klingenden Begriff-

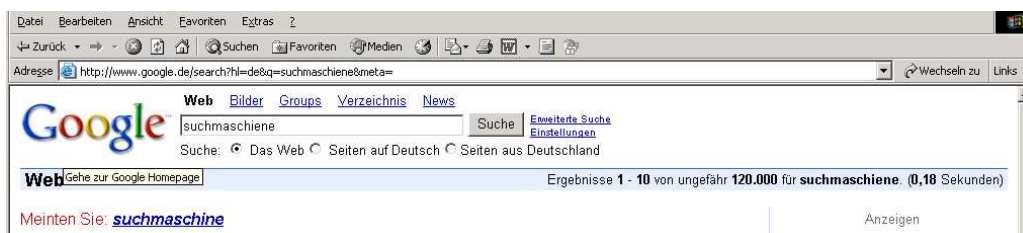


Abbildung 20: Unscharfe Suche bei Google

Quelle: <http://www.google.de>

fen verbunden. Dabei ist es lediglich notwendig, dass die Begriffe mit dem selben Anfangsbuchstaben beginnen. Verwendet man zum Beispiel den Aufruf `<SOUNDEX> sale`, so werden auch die Begriffe „sale“, „sell“, „seal“, „shell“, und „soul“ mitgesucht.

Der Aufruf `TYPO/N` verändert Buchstaben innerhalb eines Begriffes. Abhängig vom Wert `N` werden Buchstaben verändert, die weiter weg sind. Liegt der Wert zum Beispiel bei 1, so wird ein Buchstabe nur in der nächsten Umgebung verändert. Ein Beispiel veranschaulicht dies ([21] Seite 3-9):

mouse, house (Aus dem „m“ wurde ein „h“)
 agreed, greed (Das „a“ wurde gelöscht)
 cat, coat (Das „o“ wurde hinzugefügt)

Diese Funktion kann man sehr gut für die Weiterverarbeitung von Dokumenten verwenden, die mittels Texterkennung (OCR) erfasst wurden. Bei der Erkennung kann es zu Fehlern kommen, da Buchstaben manchmal nicht eindeutig ermittelt werden. Der Aufruf `TYPO/N` hilft, die falsch erkannten Begriffe zu berichtigen.

7.5 Stemming

Neben der Korrektur von Fehlern werden auch unterschiedliche Schreibweisen von Begriffen berücksichtigt. Dazu zählen zum Beispiel das Verwenden des Plurals oder der unterschiedlichen Kasus im Deutschen. Die Methode, um Worte auf ihren Stamm zurückzuführen, nennt man „stemming“.

Verity unterstützt auch diese Funktion. Mittels eines eigenen Aufrufs `<STEM>` werden verschiedene Wortstämme berücksichtigt. So findet zum Beispiel der Aufruf `<STEM> film` alle Dokumente, in denen auch die Begriffe „films“, „filmed“, und „filming“ vorkommen([21] Seite 3-8).

8 Strategien und Hilfsmittel: Maintenance

Bei der Unterstützung der Prozesse im Zusammenhang mit dem Usecase Maintenance ist zwischen zwei großen Bereichen zu unterscheiden:

- Organisatorische Unterstützung
- Unterstützung durch Tools

Die organisatorische Unterstützung beschreibt die Aufgabenverteilung innerhalb eines Unternehmens und soll sicherstellen, dass das mentale Modell innerhalb der Struktur gewahrt bleibt.

Bei der Unterstützung durch Tools geht es darum, wie die eigentliche Pflege des Systems unterstützt werden kann.

8.1 Organisatorische Unterstützung

Die organisatorische Unterstützung beginnt mit der Festlegung der Verantwortlichkeiten. Diese unterteilen sich in die Bereiche

- Pflege der Struktur
- Pflege der Inhalte

8.1.1 Pflege der Struktur

Für die Pflege der Struktur sind zwei Varianten intensiv diskutiert worden. Die Variante 1 ist ein System eines kleinen Redaktionsteams von zwei oder drei Personen, die für alle Abteilungen das Anlegen und Pflegen der Struktur übernehmen. Da diese sich sehr häufig in der Struktur bewegen und sich schnell absprechen können, ist eine hohe Qualität der Struktur gewährleistet. Abbildung 21 veranschaulicht die Bedeutung eines zentralen Redaktionsteams.

Dieses Team ist auch Ansprechpartner für alle Fragen und Anregungen, sowie verantwortlich dafür, die Anforderungen der User umzusetzen. Sobald ein neues Thema aufkommt und entweder vom Redaktionsteam selbst erkannt oder von den Abteilungen vorgeschlagen wird, entscheidet das Team, wo das Thema in der Struktur aufgenommen wird und legt dazu den Strukturpunkt an. Dies muss in enger Abstimmung mit der zuständigen Abteilung geschehen.

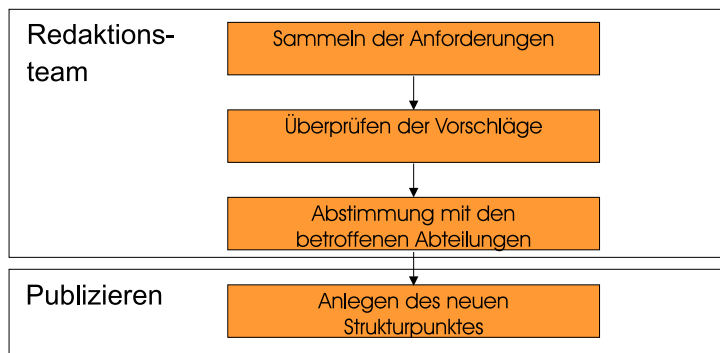


Abbildung 21: Prozess der Pflege der Struktur bei einem zentralen Redaktionsteam

Quelle: eigene Darstellung

Dem großen Vorteil der hohen Qualität der Struktur steht der Nachteil gegenüber, dass für diese Arbeit keine Ressourcen zur Verfügung standen. Der zentrale Ansatz eines Redaktionsteams wird aber immer stärker von Gewicht sein, je größer ein Unternehmen wird und je häufiger auf diese Daten zugegriffen wird. Nur in einem kleinen Team sind schnelle Entscheidungen zu treffen und durchzusetzen. Da in keinem Fall für die Abteilungen die Anforderungen komplett erfüllt werden können, ist es nicht notwendig, Diskussionen über Gebühr zu führen. Selbstverständlich müssen die Betroffenen ausreichend die Möglichkeit haben, sich in den Prozess einzuklinken. Letztlich muss aber das gemeinsame mentale Modell über Partikularinteressen gestellt werden, um eine Zerstückelung zu verhindern, die keinem mehr nutzt.

Variante 2 verteilt die Verantwortlichkeiten an die Abteilungen. Dabei gilt es die oben genannten Schwächen der Verteilung zu minimieren. Dies geschieht dadurch, dass für die Pflege der Struktur nur ein Mitarbeiter pro Abteilung zuständig ist. Dieser Mitarbeiter ist der Redakteur der Abteilung, der nicht nur die Struktur betreut, sondern auch inhaltlich den Überblick behält. Alle Redakteure bilden gemeinsam das Redaktionsteam. Bei dieser Variante ist besonders darauf zu achten, dass die Abstimmungsprozesse der Redakteure geordnet ablaufen.

Großer Vorteil dieses Verfahrens ist, dass die Redakteure über ein sehr gutes Fachwissen in ihren Bereichen verfügen und so die Struktur hochwertig fortentwickeln können. Es besteht allerdings die Gefahr, dass es zu Inkonsistenzen bei abteilungsüberschreitenden Themen kommt. Den Ablauf der Prozesse zeigt Abbildung 22. In dieser Abbildung wird deut-

lich, dass es zu parallelen Prozessen kommen kann, die zu parallelen Strukturpunkten führen können. Im Fall dieser Master Thesis wurde Variante 2



Abbildung 22: Pflege der Struktur mit einem dezentralen Redaktionsteam
Quelle: eigene Darstellung

gewählt. Zur Eingrenzung dieser Entwicklungen siehe Kapitel 8.2 Unterstützung durch Tools auf Seite 83.

8.1.2 Pflege der Inhalte

Bei der Pflege der Inhalte treten vergleichbare Probleme auf. Nur kommt hier noch erschwerend hinzu, dass alle Mitarbeiter Dokumente recherchieren können und den Bedarf haben, diese zu publizieren und ihren Kollegen zugänglich zu machen.

Deshalb wäre es auch in diesem Fall sinnvoll, alle Dokumente von einem zentralen Redaktionsteam dahingehend bewerten zu lassen, ob der Inhalt relevant für die Datenbank ist und anschließend das Dokument zu verorten.

Es wurde eine sehr freie Art des Publizieren gewählt, die es jedem Mit-

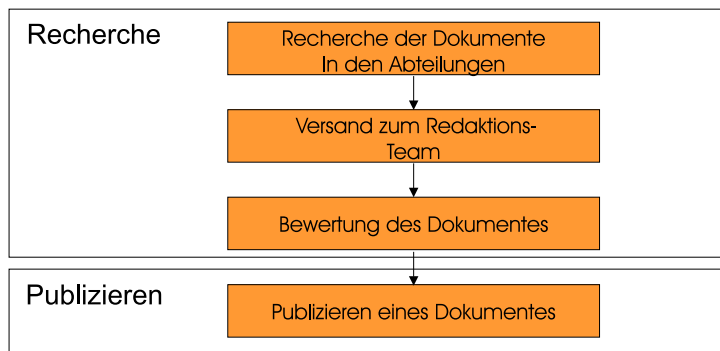


Abbildung 23: Variante 1: Zentrales Redaktionsteam
Quelle: eigene Darstellung

arbeiter ermöglicht, innerhalb seiner Struktur ein beliebiges Dokument zu publizieren. Dies hat auch einige nicht zu unterschätzende Vorteile:

- **Qualität**
Die Qualität der Dokumente innerhalb einer Bereichsstruktur ist hoch, da die Fachleute dieses Themengebietes sehr genau Bescheid wissen, welche Dokumente sie benötigen und an welcher Stelle diese abzulegen sind.
- **Zufriedenheit**
Die Möglichkeit, selbst Dokumente zu publizieren, macht aus Betroffenen Beteiligte, weil Verantwortlichkeiten an einzelnen Personen direkt festgemacht werden können. Es gibt keine Diskussionen mit einer übergeordneten Stelle, bürokratischer Aufwand entfällt.

Dennoch ist diese Variante nur dann sinnvoll umzusetzen, wenn mit entsprechenden Tools Unterstützung angeboten wird.

8.2 Unterstützung durch Tools

Für die gewählte organisatorische Einbettung der Pflege der Struktur und der Inhalte ist es wichtig, unterstützende Tools zu verwenden, damit die Qualität der Daten gewährleistet werden kann. Dazu wurden zwei Hilfsmittel gewählt:

- **Benachrichtigung**
Veränderungen an der Struktur sollen an einen fest definierten Benutzer/Administratorkreis gemeldet werden.
- **Doublettencheck**
Bereits vorhandene Strukturpunkte beziehungsweise Dokumente sollen beim Erstellen erkannt werden und nur nach einer Sicherheitsabfrage publiziert werden können.

8.2.1 Benachrichtigung

Weil die Variante des verteilten Redaktionsteams gewählt wurde, ist es wichtig, dass Veränderungen in der Struktur überwacht werden. Dies geschieht dadurch, dass beim Anlegen eines neuen Ordners eine E-Mail an eine fest definierte Person geschickt wird, die anschließend überprüft, ob der Strukturpunkt notwendig ist und falls ja, gegebenenfalls die Bezeichnung und den gewählten Ort prüft. Ist dies nicht der Fall, muss der Strukturpunkt - nach Absprache selbstverständlich - wieder gelöscht werden. Brisant wird diese Abstimmung bei abteilungsübergreifenden Themen. In diesem Fall ist eine

Redaktionssitzung notwendig, um die strittigen Punkte zu klären.

Auch beim Publizieren eines neuen Dokumentes wird eine E-Mail an den Abteilungsredakteur geschickt und dieser auf das neue Dokument aufmerksam gemacht. Dieser muss dann im Einzelfall entscheiden, was mit dem Dokument zu tun ist.

Die Gefahr bei dieser Vorgehensweise ist, dass die Beobachtung durch die Abteilungsredakteure mit der Zeit nicht nachlassen darf. Dies erfordert ein hohes Maß an Zuverlässigkeit und auch Durchsetzungsvermögen.

8.2.2 Doublettencheck

Beim Anlegen eines neuen Ordners wird geprüft, ob es diesen Ordner an einem anderen Ort bereits gibt. Dabei spielt es keine Rolle, an welcher Stelle innerhalb der Struktur der Ordner schon angelegt wurde. Ist dies nicht der Fall, kann der Ordner einfach angelegt werden.

Falls es schon einen Ordner mit der selben Bezeichnung schon gibt, wird angezeigt, an welcher Stelle der Ordner schon existiert. Nach einer Abfrage ist es möglich, dennoch einen Ordner gleichen Namens ein zweites Mal anzulegen.

In beiden Fällen wird eine E-Mail an den Verantwortlichen geschickt.

Mit Hilfe der Checksumme ist es möglich, festzustellen, ob es ein Doku-

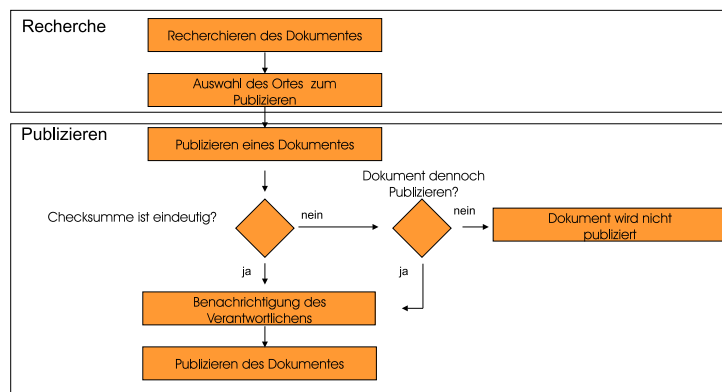


Abbildung 24: Toolunterstützung beim Publizieren

ment bereits in der Datenbank gibt. Die Checksumme ist ein statistisches Verfahren, das den Inhalt eines Dokumentes in Zahlen umsetzt und für jedes Dokument eine eindeutige Summe ermittelt. Nur wenn zwei Dokumente

exakt inhaltsgleich sind (unabhängig vom Titel) ergibt sich die selbe Checksumme. Beim Abspeichern eines Dokumentes wird die Checksumme unmittelbar ermittelt und mit den bestehenden Checksummen verglichen. Bei einer Gleichheit wird der Pfad des bestehenden Dokumentes angezeigt, und erst nach einer Abfrage kann es publiziert werden.

9 Zusammenfassung

Grundlage für ein funktionierendes Ablagesystem ist ein gemeinsames mentales Modell für alle Beteiligten. Dies setzt voraus, dass es ein gemeinsames Verständnis zwischen den Personen, die Dokumente in eine Struktur einpflegen, und den Personen, die auf diese Dokumente zugreifen wollen, gibt.

Die Benennung der Dokumente wie der Struktur muss einer übergreifenden Logik folgen, die sich aus den speziellen Anforderungen einer Firma und dem vorhandenen Weltwissen der Mitarbeiter aufbaut. Da man davon ausgehen kann, dass der erste Entwurf einer Struktur nur sehr selten den späteren Anforderungen entsprechen wird und es zu Veränderungen kommen wird, ist es sinnvoll, zuerst mit den einfachen, unstrittigen Punkten zu beginnen und nicht zu Anfang den Nutzen eines Systems durch lange Grundsatzdiskussionen zu hinterfragen.

Für den Aufbau eines Informationsverzeichnisses ist nicht die Vollständigkeit wichtigstes Kriterium, sondern die einfache Bedienbarkeit und die schnellen Erfolgserlebnisse für den Benutzer.

Vielmehr ist es sinnvoll, flexibel auf Wünsche zu reagieren, ohne dabei das große Ganze außer Acht zu lassen. Der Aufbau einer Struktur für eine Datenbank unterscheidet sich nicht wesentlich vom Aufbau einer Systematik für eine Bibliothek. Eine Struktur, die sich außerhalb des Computers als kompliziert und unübersichtlich erweist, wird auch mit Hilfe von Software nicht besser. Die Vorteile einer elektronischen Verwaltung treten erst dann im vollen Umfang zu Tage, wenn sie den Benutzer in seiner Arbeit unterstützen, so dass er sie im besten Fall überhaupt nicht bemerkt. Dazu zählt die Fähigkeit, des gezielten Umstellens einer Struktur, deren einfacher Ausbau und die Möglichkeit, mit Verknüpfungen zu arbeiten.

Bei dieser Aufgabe darf man allerdings nicht verkennen, dass man nicht allen Benutzern ein System auf den Leib schneiden kann. Gerade beim Erstellen der Struktur ist es wichtig, mit Personen zu sprechen, die über ein fundiertes Wissen des Themengebietes verfügen. Dies ist eine Arbeit, die nicht auf Abteilungsleiterenebene, sondern auf operativer Ebene erledigt werden sollte. Aber auch hier gilt es zu beachten, zu einem in sich geschlossenem logischen Ergebnis zu kommen, das mehr als nur den Anforderungen einer Abteilung entspricht. Dazu gehört auch Durchsetzungsvermögen, Standhaftigkeit und Fingerspitzengefühl.

Die Strukturierung und Verwaltung heterogener Datenbestände ist möglich, allerdings muss es eine realistische Erwartungshaltung an ein System geben. Die Eingabe eines Suchbegriffes liefert nicht die Antwort auf eine Frage, sondern verweist lediglich auf Dokumente, in denen der Suchbegriff in irgendeiner Weise vorkommt. Die intellektuelle Aufbereitung bleibt nach wie vor dem Benutzer überlassen. Bei einer wohldosierten Unterstützung zur Pflege eines Datenbestandes mit verschiedenen Tools ist es möglich, ein einfaches System zu schaffen, das den Erfordernissen der Benutzer gerecht wird. Oberstes Ziel muss es sein, einfach und schnell zu sein.

10 Ausblick

Nicht alle in dieser Master Thesis gemachten Vorschläge konnten umgesetzt werden. Die Ressourcen standen nicht zur Verfügung. Viele Vorschläge stammen aus anderen Bereichen, in denen sie bereits umgesetzt sind. Das Speed-typing zum Beispiel wird unter Windows für die Suche im Hilfesystem verwendet. Dies auf die Navigation in einer Struktur zu übertragen, erscheint sinnvoll, da es auch hier darum geht, möglichst schnell zu einem Strukturpunkt zu gelangen.

Zum Teil stehen den Vorschlägen auch Hindernisse entgegen. Beim Erstellen einer Sitemap zum Beispiel muss das Rechte- und Rollensystem berücksichtigt werden. Das bedeutet, dass in einer Sitemap auch nur die Ordner dargestellt werden dürfen, auf die der Benutzer Zugriff hat. Im Extremfall hat das zur Folge, dass für jeden Benutzer individuell eine Sitemap erstellt werden muss, da sich die Struktur permanent ändern kann, bei jedem Aufruf zusätzlich noch von Neuem. Dies setzt eine hohe Performance und eine sehr flexible Darstellung voraus, die derzeit noch nicht realisiert ist.

Im Zusammenhang mit der Suche ist es denkbar, mehr mit den Auswertungen des bisherigen Verhaltens eines Benutzers beziehungsweise einer Benutzergruppe zu arbeiten. Ausgehend von den bereits getätigten Suchanfragen lassen sich eventuell Bewertungen der Nützlichkeit eines Dokumentes hinsichtlich eines bestimmten Suchbegriffes treffen. Da die Auswertung personenbezogener Daten aber zum Teil schon sehr restriktiv gesehen wird, ist es fraglich, ob sich etwas in dieser Art durchsetzen lassen wird.

Bei vielen Suchmaschinen ist es heute schon möglich, komplette Sätze als Suchbegriff einzugeben. Abhängig von der Formulierungskunst des Anwenders ist das sicherlich auch ein Weg, um zu relevanteren Trefferlisten zu kommen.

Unabhängig von aller zur Verfügung stehender Software wird der Nutzen und die Zufriedenheit mit einem heterogenen Datenbestand aber immer von einem abhängen:

Die Qualität des Inhalts ist von entscheidender Bedeutung!

Viele der recherchierten Dokumente werden in Form von Word, Powerpoint oder Adobe Acrobat Formaten vorliegen. Diese sind für das Austauschen von Dokumenten zum Teil sehr gut geeignet, bei der Recherche in

einem lokalen Netzwerk können diese aber Schwierigkeiten bereiten, das zum Beispiel immer die komplette Datei übertragen werden muss. Bei kleiner ausgelegten Netzwerken und grossen Dateien kann es zu Wartezeiten kommen. Deswegen kann es sinnvoll sein, über Konvertierungsprogramme diese grossen Dokumente in ein für das Internet besser geeignetes Format zu überführen. Das kann zum Beispiel ein html-Format sein. Berücksichtigt man noch, dass die Dateiformate auch über einen längeren Zeitraum und über mehrere Versionswechsel und auf unterschiedlichen Plattformen (Linux oder Unix) lesbar sein sollen, dann ist vielleicht auch der Aufwand gerechtfertigt, die Dokumente in ein xml-Format zu überführen.

Nur wenn die Erwartungen des Benutzers an den Inhalt in den meisten Fällen erfüllt werden können und die Aufbereitung der Inhalte seinen Wünschen entspricht, wird er ein System freiwillig mehrfach nutzen und für seine Arbeit Entlastung finden. Genau dies sind die Kriterien, die den Einsatz von Ressourcen und Geld für ein System rechtfertigen.

Abkürzungsverzeichnis

| | |
|---------|--|
| ACM | Association for Computing Machinery |
| ASCII | American Standard Code for Information Interchange |
| CCS | Computing Classification System |
| DIZ | Dokumentations- und Informationszentrum München GmbH |
| DLO | Dokumentähnliches Objekt |
| DTP | Dokumenttypdefinition |
| GMT | Greenwich Mean Time |
| HTML | Hypertext Markup Language |
| IP | Internet Protocol |
| IS/6.3 | Hyperwave Information Server 6.3 |
| ISBN | International Standard Book Number |
| ISSN | International Standard Serial Number |
| IT | Informationstechnologie |
| JPEG | Joint Photographic Experts Group |
| MAB2 | Maschinelles Austauschformat für Bibliotheken |
| MIME | Multipurpose Internet Mail Extensions |
| NISO | National Information Standards Organization |
| OCR | Optical Character Reader |
| PARC | Palo Alto Research Center |
| pdf | Portable Document Format |
| RAK | Regeln für alphabetische Katalogisierung |
| RDF | Resource Definition Framework |
| SGML | Standard Generalized Markup Language |
| TEI | Text Encoding Initiative |
| TU Graz | Technische Universität Graz |
| USMARC | United States Machine-Readable Cataloging records |
| URI | Uniform Resource Identifier |
| URL | Uniform Resource Locator |
| URN | Uniform Resource Name |
| W3C | World Wide Web Consortium |
| XML | Extensible Markup Language |

Abbildungsverzeichnis

| | | |
|----|---|----|
| 1 | Usecase Publizieren | 15 |
| 2 | Publizieren ohne Abstimmungsprozess | 16 |
| 3 | Publizieren Variante 1 | 17 |
| 4 | Publizieren Variante 2 | 18 |
| 5 | Publizieren Variante 3 | 20 |
| 6 | Auszug aus dem Computing Classification System | 39 |
| 7 | Prozess des automatischen Publizierens eines Dokumentes | 54 |
| 8 | Schema einer Baumdarstellung | 58 |
| 9 | Darstellung von Verzeichnissen einer Treemap | 59 |
| 10 | Darstellung eines Cam Trees | 60 |
| 11 | Darstellung eines Fractal Trees | 61 |
| 12 | Darstellung eines Hyperbolic Trees | 62 |
| 13 | Darstellung einer Website mit Hyperspace | 64 |
| 14 | Hotsauce - dreidimensionales Navigieren | 65 |
| 15 | Benutzerinterface von Infosky | 67 |
| 16 | Sitemap-Darstellung in Form einer Liste | 68 |
| 17 | Darstellung einer Sitemap mit einem Hyperbolic Tree | 69 |
| 18 | Aufbau eines Index | 70 |
| 19 | Auszug aus einem Thesaurus | 74 |
| 20 | Unschärfe Suche bei Google | 79 |
| 21 | Prozess der Pflege der Struktur bei einem zentralen Redak- tionsteam | 81 |
| 22 | Pflege der Struktur mit einem dezentralen Redaktionsteam | 82 |
| 23 | Variante 1: Zentrales Redaktionsteam | 82 |
| 24 | Toolunterstützung beim Publizieren | 84 |

Literatur

- [1] Text entnommen aus: <http://www.acm.org/class/1998/ccs98-intro.html> (06.11.2004)
- [2] Autonomy; Technology White Paper; Autonomy Inc; 2000; im Web verfügbar unter <http://www.autonomy.com/content/downloads/>
- [3] Babiak, Ulrich; Effektive Suche im Internet; O´Reillys Internetbibliothek; 4. Auflage 2001; Seite 107-108
- [4] Bundesdatenschutzgesetzes (BDSG) vom 20.12.1990 (BGBl. I S. 2954), zuletzt geändert durch Gesetz vom 21. August 2002 (BGBl. I S. 3322), § 1 Zweck und Anwendungsbereich des Gesetzes, aus <http://www.datenschutz-berlin.de/recht/de/bdsg/bdsg01.htm> (06.11.2004)
- [5] Däßler, Rolf; Informationsvisualisierung, Stand, Kritik und Perspektiven; Fachhochschule Potsdam, Fachbereich Archiv/Bibliothek/Dokumentation, Projektgruppe InfoViz; o.J.; <http://fabdp.fh-potsdam.de/daessler/paper/InfoVis99.pdf> (06.11.2004)
- [6] Dutke, Stephan; Mentale Modelle: Konstrukte des Wissens und Verstehens; Kognitionspsychologische Grundlagen für die Software-Ergonomie; Verlag für Angewandte Psychologie, Göttingen 1994
- [7] Eberleh, Edmund, Oberquelle, Horst, Oppermann, Reinhard; Einführung in die Software-Ergonomie, Gestaltung graphischer-interaktiver Systeme: Prinzipien, Werkzeuge, Lösungen; 2. Auflage; de Gruyter Verlag; Berlin 1994
- [8] F. A. Brockhaus; Der Brockhaus Computer und Informationstechnologie; Hardware, Software, Multimedia, Internet, Telekommunikation; Mannheim, Leipzig, 2003
- [9] F.A. Brockhaus GmbH; Brockhaus Enzyklopädie in vierundzwanzig Bänden; Mannheim, Leipzig, 1993, 1994

- [10] Friedl, Silvia; Dokumentation-Beschreibung Usability Tests; FH Joanneum, Informationsmanagement; http://dmt.fh-joaanneum.at/projects/ebus2/seite.php?name=Card_Sorting (06.11.2004)
- [11] Gaffney, Gerry; What is Card Sorting?; Usability Techniques series; 2000, in http://www.infodesign.com.au/usabilityresources/design/card_sorting.asp (06.11.2004)
- [12] Haller, Klaus; Katalogkunde, Eine Einführung in die Formal- und Sacherschließung; 3. erweiterte Auflage; K. G. Saur Verlag; München 1998
- [13] Hippner, Hajo, Wilde, Klaus D. (Hrsg.); IT-Systeme im CRM, Aufbau und Potenziale; Betriebswirtschaftlicher Verlag Dr. Th. Gabler/GWV Fachverlage GmbH; Wiesbaden, 2004
- [14] Kahn, Paul, Lenk, Krzysztof; Websites visualisieren; Rowohlt Verlag; Berlin 2001
- [15] Kappe, Dr. Frank, Droschl, Dr. Georg, Kienreich, Wolfgang et al; o.J.; InfoSky: Eine neue Technologie zur Erforschung großer, hierarchischer Wissensräume, veröffentlicht zur Knowtech 2002; http://www.knowtech2002.de/Kappe_Hyperwave_Graz.pdf; für weitere Informationen zu Infosky siehe: Andrews, Keith; Visualising Information Structures - Aspects of Information Visualisation; Professional Thesis; 2002
- [16] Karaguannis, Dimitris, Telesko, Rainer; Wissensmanagement: Konzepte der künstlichen Intelligenz und des Softcomputing; Oldenbourg, 2001
- [17] Koike, Hideki, Yoshihara, Hirotaka; Fractal Approaches for Visualizing Huge Hierarchies; Department of Communications and Systems; University of Electro-Communications Chofu; Tokyo 182; Japan; o.J.; <http://www.vogue.is.uec.ac.jp/~koike/papers/vl93/vl93.pdf> (06.11.2004)
- [18] Nielsen, Jakob; Erfolg des Einfachen; Markt+Technik Verlag, München 2004

- [19] Schön, Eckhardt; Das Resource Description Framework (RDF) - ein neuer Weg zur Verwaltung von Metadaten im Netz; Technische Universität Ilmenau, Fakultät für Informatik und Automatisierung; o.J.; <http://www.netobjectdays.org/mirrors/stja.cd/Beitraege/mik98/eschoen.pdf> (06.11.2004)
- [20] Rusch-Feja, Deann; Mehr Qualität im Internet - Entwicklung und Implementierung von Metadaten; Online-Tagung der DGD. Berlin: Deutsche Gesellschaft für Dokumentation; 1997; Seite 113-130
- [21] Verity, Incorporated; Verity Query Language Reference Guide V2.6; 01.06.2000
- [22] Wirth Thomas; Missing Links, Über gutes Webdesign; Carl Hanser Verlag; München, Wien, 2004

Sachregister

- Ähnliche Dokumente, 78
- Überblick, 14, 15

- Abkürzung, 23
- ACM - Association for Computing Machinery, 38
- AllTheWeb, 38
- Alphabetisches Stichwortverzeichnis, 70

- Anforderungen, 9
- Aufrufbarkeit, 24
- Ausnahmeregel, 53
- Automatisches Publizieren, 53
- Autonomy, 21, 53, 72, 75
- Autor, 25

- Baumdarstellung, 57
- Bayes, 75
- Begriffswelt, 35
- Benachrichtigung, 83
- Benutzerverhalten, 33
- Berliner Anweisungen, 35
- Beständigkeit, 28
- Betriebsrat, 32
- Breite, 22, 23
- Bundesdatenschutzgesetz, 32

- Cam Tree, 57, 59
- Card Sorting, 30
- CCS - Computing Classification System, 38
- Checksumme, 42, 84
- Common Logfile Format, 31
- Cone Tree, 57, 59
- Content Management System, 55

- Dateiformat, 25
- Datenbestand, 16
- Definition der Inhalte, 12
- Definition der Struktur, 13
- Definition der Ziele, 12
- Definition der Zielgruppen, 12
- Dezentrale Strukturen, 40
- Dokumenttyp, 43, 53
- Doublettencheck, 83, 84
- Dublin Core, 46, 47, 50
- Dynamische Darstellung, 69

- E-Mail, 18, 55, 83
- eBook, 8
- Einfache Suche, 25, 76
- Einstiegspunkt, 22, 23
- Encoding Archive Description, 46
- Erstellungsdatum, 25
- Erweiterbarkeit, 26, 28
- Erweiterte Suche, 25, 76
- Expanded Common Logfile Format, 31

- Fachjargon, 23
- Fractal Tree, 57, 60

- Google, 21, 38, 76, 78
- Gruppenlaufwerk, 41

- Hierarchische Visualisierung, 57
- Hit, 33
- Hotsauce, 63, 65
- Hyperbolic Tree, 57, 62
- Hyperspace, 63, 64
- Hyperwave, 7
- Hyperwave Information Server, 21, 31, 45, 51, 53, 55

- Indexsuchen, 72
- Informationsarchitektur, 11
- Informationsraum, 66
- Informationsvisualisierung, 56

Infosky, 63, 66
 IP-Adresse, 33

 Kategorisierer, 53
 KnowCenter, 66
 Komplexität, 9
 Konsistenz, 14, 28
 Kontext, 25

 LexisNexis, 69
 Listen, 68
 Logfileanalyse, 31

 MAB2, 47
 Maintenance, 80
 Mapping, 57
 Max-Planck-Institut für Bildungsforschung, 46
 Mentales Modell, 23, 27, 34, 81
 Meta-Tag, 44
 Metadaten, 14, 25, 36, 44, 45, 50, 77
 Metadatenmodell, 37
 Metadaten-systematiken, 45
 Metapher, 57
 Microsoft, 34
 Microsoft Word, 70
 Mischformen, 41
 Musterbasiertes Suchen, 75

 Navigationsanalyse, 34
 Navigieren, 21, 56
 Netzartige Visualisierung, 63

 Open Directory Project, 37

 Page view, 33
 Palo Alto Research Center (PARC), 59
 Personenbezogene Daten, 32
 Pflege, 82
 Preußische Instruktionen, 35
 Publizieren, 43

 Publizieren mit Abstimmungsprozess, 16
 Publizieren-Dialog, 53

 Redaktionsteam, 41, 80
 Regeln für die alphabetische Katalogisierung, 35
 Regelset, 53
 Relevanz, 25, 43, 77, 78
 Resource Description Framework, 36

 Schlagwortliste, 14
 Schreibweise, 78
 Semantisches Netz, 37
 Shannon, 75
 Sitemap, 67
 Sortierkriterium, 25
 Speedtyping, 70, 76
 Standardattribut, 50
 Stemming, 79
 Stopwortliste, 72
 Struktur, 13
 Strukturierung, 12
 Suchbereich, 25
 Suchdialog, 21
 Suchmaske, 76
 Synonym, 24
 Synonymenliste, 74

 Taxonomie, 24
 Text Encoding Initiative, 46
 Thesaurus, 24, 73
 Tiefe, 22, 23
 Treemap, 57, 58
 Trefferliste, 24, 25, 72, 77
 TU Graz, 66

 Unscharfe Suche, 78
 Usability, 9, 14, 16
 Usecase, 10, 14
 Usecase Publizieren, 14
 Usecase Suchen, 14

Usecase Verbesserung am Inhalt, 27
Usecase Verbesserung der Bedienbarkeit, 27
USMARC, 46

Verfasser, 25
Verity, 21, 72, 73, 76, 78
Verzeichnisstruktur, 27
Visit, 33
Visualisierung, 56
Volltextsuche, 41
Vorgehensweise, 9

W3C, 36
Wachstum, 26, 27
Wartbarkeit, 9
Webserver, 31
Websuxess, 33
Windows, 34
Windows Explorer, 58
Workflow, 15, 55
Wortschatz, 74

Zentrale Strukturen, 40
Zugriffsgeschwindigkeit, 26