

Institut für Medizinische Statistik und Epidemiologie
der Technischen Universität München
Klinikum rechts der Isar
(Direktor: Univ.-Prof. Dr. A. Neiß)

DATA MINING IN DER MEDIZINISCHEN
LITERATURDATENBANK
MEDLINE

Holger Tenner

Vollständiger Abdruck der von der Fakultät für Medizin der Technischen
Universität München zur Erlangung des akademischen Grades eines

Doktors der Medizin

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. D. Neumeier

Prüfer der Dissertation: 1. apl. Prof. Dr. G.R. Thurmayer
2. Univ.-Prof. Dr. A. Neiß
3. Univ.-Prof. Dr. A.J.W. Goldschmidt,
Universität Trier

Die Dissertation wurde am 26.06.2003 bei der Technischen Universität München
eingereicht und durch die Fakultät für Medizin am 05.05.2004 angenommen.

Gewidmet meinem Vater und der hilfsbereitesten Krankenschwester der Welt, Julia Eiber. Beide mußten jung sterben, weil ihnen die Medizin noch nicht helfen konnte.

INHALTSVERZEICHNIS

1	Einleitung.....	8
2	Problemstellung.....	10
2.1	PubMed als Präsentator der Suchergebnisse.....	11
2.1.1	Vorteil.....	11
2.1.2	Nachteil.....	11
2.2	Manuelle Nachbereitung der Daten.....	13
2.2.1	Vorteil.....	13
2.2.2	Nachteil.....	13
2.3	Fragestellung.....	14
3	Material.....	15
3.1	Definitionen.....	16
3.1.1	Data Mining.....	16
3.1.2	Suchmaße.....	16
3.1.3	Dienst.....	17
3.2	MEDLINE.....	18
3.2.1	Inhalt von MEDLINE.....	18
3.3	DIMDI's Free grips WebSearch.....	20
3.4	PubMed.....	22
3.4.1	Inhalt von PubMed.....	23
3.4.2	Präsentation.....	23
3.4.3	Merkmale der PubMed-Suche.....	24
3.5	MeSH.....	31
3.5.1	Inhalt des MeSH.....	32
3.5.2	Gliederung.....	33
3.6	Wichtige biomedizinische Datenbanken im Vergleich zu MEDLINE.....	36
3.6.1	MEDLINE.....	36
3.6.2	Cochrane Library.....	37
3.6.3	EMBASE.....	37
3.6.4	Current Contents.....	38
3.6.5	BIOSIS.....	38
3.6.6	Web of Science / Science Citation Index.....	38
4	Methodik.....	40
4.1	Einführung in Meva.....	41
4.2	Netz-Architektur.....	43
4.2.1	Modell.....	43
4.2.2	Propagierung im Internet.....	43
4.3	Server-Architektur.....	46
4.3.1	Portabilität.....	46
4.3.2	Performanz.....	46
4.4	Client-Architektur.....	48

4.4.1	Benutzeroberfläche.....	48
4.4.2	Benutzerfreundlichkeit.....	48
4.5	Entwicklungs-Architektur.....	55
4.5.1	Programmiersprache.....	55
4.5.2	Compiler.....	56
4.5.3	Entwicklungskonzept.....	56
4.5.4	Entwicklungsphasen.....	57
4.5.5	Verwendete Algorithmen.....	57
5	Ergebnisse.....	62
5.1	Meva-Eingabe.....	63
5.1.1	Eingabesektion von Mevas Konsultationsformular.....	63
5.1.2	Ausgabesektion von Mevas Konsultationsformular.....	66
5.1.3	Weitere Parameter.....	73
5.2	Meva-Ausgabe.....	74
5.2.1	Überschriftensektion.....	75
5.2.2	Meldungssektion.....	75
5.2.3	Parametersektion.....	76
5.2.4	Bilanzsektion.....	77
5.2.5	Histogrammsektion.....	78
5.2.6	Korrelationssektion.....	79
5.2.7	Detailsektion.....	81
5.2.8	MeSH-Sektion.....	83
5.2.9	Fußnotensektion.....	87
5.3	Fallbeispiele.....	88
5.3.1	Meva-Konsultation zum Profil von Filippi M.....	89
5.3.2	Meva-Konsultation zum Profil von Herfarth C.....	91
5.3.3	Meva-Konsultation zum Thema Lebertransplantation.....	93
5.3.4	Meva-Konsultation zur Gewinnung von Präsentationsmaterial.....	96
5.4	Technische Grenzwerte von Meva.....	98
5.4.1	Eingabemaxima.....	98
5.4.2	Ausgabemaxima.....	98
5.5	MePrep – Meva Preprocessor.....	99
5.5.1	Einführung.....	99
5.5.2	Installation.....	99
5.5.3	Bedienung.....	100
5.6	Administration von Meva.....	103
5.6.1	Installation.....	103
5.6.2	Wartung.....	105
5.7	Probleme und deren Lösungen in der Arbeit mit Meva.....	112
5.7.1	Falscher Rückgabedatentyp.....	112
5.7.2	Zu großes Resultat.....	112
5.7.3	MS-Access-Fehlermeldung.....	112
5.7.4	Fehlende Grafiken.....	113

5.7.5	Zerrissener Text im Ausdruck	113
5.7.6	Zerrupftes Tabellenformat im Textresultat	114
5.7.7	Falsche Sprache der Webpräsenz	114
5.7.8	Skriptmeldungen	116
5.7.9	Lange Sendezeit	117
5.7.10	Lange Reaktionszeit	117
6	Diskussion.....	119
6.1	Vorteile von Meva	120
6.1.1	Allgemeine Vorteile.....	120
6.1.2	Histogramme.....	120
6.1.3	Korrelationsmatrizen	121
6.1.4	Filter.....	121
6.1.5	Details.....	122
6.1.6	MeSH.....	122
6.2	Nachteile von Meva.....	123
6.3	Mevas Anwendbarkeit auf andere biomedizinische Datenbanken	124
7	Zusammenfassung	125
8	Stichwortverzeichnis	126
9	Literaturverzeichnis	131
10	Internet-Quellenverzeichnis.....	135
11	Abbildungsverzeichnis	137
12	Tabellenverzeichnis	138
13	Anhang A: Glossar	139
14	Anhang B: MEDLINE-Felder	146
15	Anhang C: DIMDI's Datenbankzugänge	149
16	Danksagung.....	150

Wenige Meter vor Alice hockte die Grinsekatz auf einem Baum.

„Grinsepussi!“ redete sie die Katze an, ziemlich unsicher, weil sie nicht wußte, ob ihr der Name behagte.

Indessen grinste die Katze noch breiter.

„Na, das scheint ihr zu gefallen“, dachte Alice und fuhr fort: „Würdest du mir bitte sagen, welchen Weg ich einschlagen muß?“

„Das hängt in beträchtlichem Maße davon ab, wohin du gehen willst“, antwortete die Katze.

„Oh, daß ist mir ziemlich gleichgültig“, sagte Alice.

„Dann ist es auch einerlei, welchen Weg du einschlägst“, meinte die Katze.

„Hauptsache, ich komme irgendwo hin“, ergänzte sich Alice.

„Das wirst du sicher, wenn du lange genug gehst“, sagte die Katze.

(Lewis Carroll, Alice im Wunderland)

1 EINLEITUNG

Mit Alice' Worten in der Präambel ist bereits viel über den Hintergrund der vorliegenden Arbeit berichtet: über die Schwierigkeit der menschlichen Suche nach Informationen. Diese Arbeit widmet sich speziell der Suche nach Informationen in medizinischen Datenbanken und der Frage, wie eine solche Recherche durch automatisierte Routinen erleichtert werden kann.

Die Welt um uns herum ändert sich. Es sind nicht nur politische Änderungen, die ins Auge stechen, sondern auch die im Bereich der Informatik und Kommunikation, im Informationssektor schlechthin. Die Menschheit erfindet immer neue und schnellere Formen des Informationsaustausches, die Komplexität der Rechner steigt alle anderthalb Jahre nach dem [Mooreschen Gesetz](#) um 100% und bietet damit ungeahnte Möglichkeiten zur Bewältigung der Aufgaben, die in der Forschung, Lehre, Bildung und Wirtschaft anstehen.

Parallel wird die Welt von der Tendenz zur Globalisierung ergriffen: Wirtschaftsmärkte werden zusammengelegt und neue Allianzen geschmiedet. Von der ökonomischen und politischen Globalisierung der Welt wird auch der Wissenschaftssektor ergriffen. Durch die Zusammenlegung von vormals singulären und schwer verfügbaren Datenquellen, durch die Schaffung neuer Informationskanäle und einer immer leichteren Verfügbarkeit von Daten stehen dem Menschen heute eine Vielzahl von Informationsmöglichkeiten zur Verfügung, von denen er früher nicht zu träumen wagte.

Als das sicher prominenteste Beispiel für ein neues Medium gilt das Internet mit seinen vielfältigen Recherchemöglichkeiten, speziell den Suchmaschinen. Nicht zuletzt dem exponentiellen Wachstum des Internets ist es zu verdanken, daß die maschinelle Suche nach Informationen in viele Bereiche des täglichen Lebens Einzug gehalten hat.

Doch jede neue Technologie hat ihren Preis: Suchen mit Suchmaschinen erfordert in der Regel Ausdauer, Frustrationstoleranz und die Anwendung kombinierter, heuristischer Suchstrategien, um zu passablen Ergebnissen zu kommen, denn eine undifferenzierte Suche mündet oft in zu viele Resultate. Zudem sind die meist linear angeordneten Resultate wenig aussagekräftig für das menschliche Auge.

Schon vor zehn Jahren wurde daher in der Managementberatung der Begriff des „*Just the right time learning*“ geboren; ein Begriff, der ausdrücken sollte, wie wichtig es ist, zum richtigen Zeitpunkt an die richtigen (hier

unternehmensrelevanten) Informationen zu gelangen. Das Gleiche gilt für die Suche nach wissenschaftlich fundierten, medizinischen Daten. Was wir früher zu wenig hatten, haben wir heute zuviel: Information.

Trotzdem die Medizin unter systemtheoretischen Gesichtspunkten ein konservatives¹ System ist, wird auch sie von der Dynamik der Entwicklung im Informationssektor erfaßt: Stichworte wie „Bayerisches Gesundheitsnetz“, DRG, die Möglichkeit für den niedergelassenen Arzt, Klinikabschlußberichte seiner Patienten on-line kurz nach der Entlassung zu lesen, Telematik und Hinzuziehung von externen Spezialisten bei schwierigen Operationen per *Video Conferencing*, Abrechnung bei den Kassen in elektronischer Form, die elektronische Patientenakte etc. zeugen davon. Durch die zunehmende Spezialisierung in den einzelnen medizinischen Fächern steigt zudem der Bedarf an Interkommunikation zwischen den Fachärzten.

Welche Möglichkeiten gibt es nun in der Medizin für den Arzt oder medizinisch Interessierten, schnelle an relevante Informationen zu kommen?

Weltweit gibt es unterdessen über hundert medizinische Datenbanken. Alleine in Deutschland bietet das Deutsche Institut für medizinische Dokumentation und Information (DIMDI) einen (zumeist kostenpflichtigen) Zugriff auf über 40 Datenbanken an. Um die Vielzahl medizinischer und medizinassoziierter aktueller Datenbanken anzudeuten, finden sich im [Anhang C](#) die Datenbanken, zu denen DIMDI einen Zugang gewährt.

Eine einfache Suche in *MEDLINE*, der bedeutendsten medizinischen Literaturdatenbank, mit *PubMed* im Internet zeitigt oft Zehntausende von Artikeln mit mehr oder minder großer Relevanz für den Anwender, die als unstrukturierte Liste zurückgeliefert werden.

Was kann der Anwender mit diesen Datenvolumina anfangen? Wie kann er diese Informationen sinnvoll für sich strukturieren und wichtige Informationen von unwichtigen trennen?

Diese und ähnliche Fragen will die vorliegende Arbeit für die Suche in *MEDLINE* mit *PubMed* beantworten.

¹ Das Attribut „konservativ“ bezeichnet hier keinen politischen Code, sondern entsprechend der kybernetischen Systemtheorie die Tendenz eines dissipativen Systems, die Parameter seines inneren Gleichgewichtszustandes konstant zu halten, bzw. nach der soziologischen Systemtheorie Niklas Luhmanns die Tendenz eines Sozialsystems, eingespielte Interaktionsbeziehungen faktisch normativ zu handhaben [Luhmann 1983, S. 42].

2 PROBLEMSTELLUNG

Das Kapitel Problemstellung erörtert die Probleme, denen ein Anwender ausgesetzt ist, der aus PubMeds Schnittstelle zu MEDLINE wichtige Erkenntnisse zu medizinischen Fragestellungen gewinnen will. Der Vorgang der Extraktion relevanter Information aus Massendaten wird nachfolgend auch *Data Mining* genannt.

Glaubt man einer Reichweitenstudie für medizinische Fachtitel [[↘ API-Studie](#)²], die der eingetragene Verein „LA-MED Kommunikationsforschung im Gesundheitswesen“ 2001 bei der NFO Infratest MediaResearch in Auftrag gegeben hat (ca. 1.000 Interviews mit niedergelassenen Allgemeinärzten, Praktikern und Internisten) und die am 17.08.2001 veröffentlicht wurde, nutzen bereits 37,2% der niedergelassenen Praktiker und Internisten unter 40 Jahren medizinische On-line-Datenbanken wie z.B. MEDLINE.

Entscheidet sich der Anwender für MEDLINE als Datenquelle, ist eine relativ hohe Nachweisquote und somit eine gute Repräsentation des medizinischen Wissens der Welt gesichert: MEDLINE ist die z.Z. größte medizinische Datenbank der Welt; die meisten Artikel aus medizinischen Zeitschriften sind in MEDLINE erfaßt.

Wählt der Anwender PubMed als Distributionsform, kommt er zudem in den Genuß der leichten und kostenlosen Verfügbarkeit der MEDLINE-Daten. Andere Vertriebsformen (*OVID* oder *WinSPIRS* – beide sind 2001 fusioniert) sind kostenpflichtig.

Wie werden die Suchergebnisse jedoch dem Anwender dargestellt? Welche Formen der Präsentation bietet PubMed dem Benutzer? Hier liegt das Problem. Der Anwender hat leider nur sehr begrenzte Möglichkeiten zur Auswahl.

Im folgenden werden die zwei derzeit möglichen Wege nebst ihren Problemen beschrieben, PubMeds Daten zu präsentieren: ein elementarer Ansatz und ein erweiterter, der zusätzliche Software benötigt.

² <http://www.dgn-service.de/Neues/api.html>; Alle Internetquellen sind im Unterschied zu Literaturquellen mit einem Pfeil (↘) markiert.

2.1 PubMed als Präsentator der Suchergebnisse

Die erste und einfachste Lösung ist, die Oberfläche von PubMed selbst zur Präsentation der Resultate zu nutzen. Dies hat zur Folge, daß dem Anwender alle gefundenen Artikel mit mehr oder minder großer Relevanz als simple Auflistung zurückgeliefert werden, portioniert auf einzelne, miteinander verknüpfte Ergebnisseiten. Der Anwender kann die Listen nach Autor, Journal oder Publikationsdatum sortieren. Verdichtungen, Häufigkeitsauswertungen, Korrelationen, Darstellung von Schlagworthierarchien o.ä. sind jedoch nicht möglich.

2.1.1 Vorteil

Der offensichtliche Vorteil dieser Lösung ist, daß keine zusätzliche Software benötigt wird. Ein Internetzugang und ein Browser ist alles, was der Nutzer benötigt. Durch die ubiquitäre Verbreitung der Browser und den Umstand, daß auch in vielen Kliniken heutzutage ein Internetanschluß zur Regel gehört, gestaltet sich die „Installation“ von PubMed sehr einfach: es gibt keine. Eine Schulung der Benutzer auf die Browser ist in der Regel nicht vonnöten. Diese Lösung ist somit universal.

Ein zweiter Vorteil ist die kostenlose Verfügbarkeit dieses Dienstes.

2.1.2 Nachteil

Der Nachteil der Auflistung in PubMed ist, daß die Daten dem Nutzer nicht verdichtet präsentiert werden. Die Extraktion wesentlicher Information wird somit dem Anwender erschwert.

Durch die enorme Fülle an zurückgelieferten Daten wird das Problem noch manifester: Wie kann der Benutzer in Tausenden Artikeln die für ihn wichtigen herausfiltern?

Die Suche nach Diabetes in PubMed zeitigte Stand März 2003 191.408 Artikelzitate. Aber selbst Krankheiten mit geringer Inzidenz in der Population sind in MEDLINE oft mit Hunderten von Beiträgen vertreten; als Beispiel nur sei die Suche nach der Ahornzuckerkrankheit (Maple Syrup Urine Disease) genannt, die 783 Artikelzitate Stand März 2003 in PubMed als einfache Liste zu Tage förderte.

Häufig gestellte klinische Fragen wie:

- Wer ist der Spezialist für ein bestimmtes Krankheitsbild?
- Wo ist der nächste Spezialist für ein bestimmtes Krankheitsbild?
- Welcher Therapieansatz zeitigte die besten therapeutischen Ergebnisse?
- Wer hat die beste klinische Erfahrung?
- Welche Begriffe gibt es in diesem Zusammenhang noch, nach denen ich suchen kann?
- Wie bekomme ich einen Überblick zu einem bestimmten Thema?

bzw. Fragestellungen, die bei statistischen Auswertungen auftreten wie:

- Welches sind die wichtigsten Schlagwörter oder Autoren?
- Welche zeitliche Verteilung von Autoren, Schlagwörtern oder Publikationstypen gibt es?
- Welche Institute oder Kliniken haben die meisten Artikel dazu publiziert?
- Welche räumliche Verteilung von Autoren gibt es?
- Wer hat wie oft über dieses Thema geschrieben?

lassen sich mit der Auflistung der Suchergebnisse in der Benutzeroberfläche von PubMed nicht oder nur schwer beantworten.

Auch ist eine iterative Verfeinerung des ursprünglichen Suchradius, um ein Thema einzukreisen, mit PubMed nur sehr schwer möglich.

2.2 Manuelle Nachbereitung der Daten

Neben der Nutzung von PubMed als Präsentator besteht eine zweite Lösung darin, sich die Ergebnisse von PubMed als Datei senden zu lassen. In diesem Fall kann der (sachkundige, motivierte und frustrationstolerante) Anwender selbsterstellte Skripte und lokale Bürosoftware nutzen, um PubMeds Resultate zu verdichten.

Unter ungünstigen Umständen muß jeder Arbeitsschritt händisch durchgeführt werden. Im besten Falle kann der Anwender durch die Nutzung von Programmier- oder Skriptsprachen, Stapelverarbeitung oder Makros die Arbeitsschritte zumindest semiautomatisieren.

Am Institut für Medizinische Statistik und Epidemiologie der Technischen Universität München existiert eine entsprechende, halbautomatische Lösung, die in mehreren Schritten mit Hilfe von Fortran-Programmen, MS-Word, MS-Access und MS-Powerpoint aufbereitete Folien generieren kann.

2.2.1 Vorteil

Der offensichtliche Vorteil ist der einer größeren Verdichtung der Daten. Die Ergebnisse können nach der Recherche professionell optisch aufbereitet werden.

2.2.2 Nachteil

Der Nachteil ist der, der jeder manuellen Lösung eignet: sie wird mit einem überproportional erhöhten Arbeitsaufwand erkaufte und eignet sich somit nur für spezielle und selten auftretende Fragestellungen. Außerdem hilft die Lösung in erster Linie dem, der sie erstellt: andere Benutzer mit vielleicht ähnlichen Fragestellungen können diese Lösung meist nicht für sich verwerten. Mit anderen Worten, die Lösung ist nicht universal.

2.3 Fragestellung

Beide vorgestellten Möglichkeiten – die direkte Nutzung von PubMed zur Präsentation als auch die manuelle Nachbereitung von Suchergebnissen – eignen sich nur bedingt für die Präsentation der MEDLINE-Suchergebnisse. Entweder ermangeln sie der Universalität oder der verdichtenden Aufbereitung der Daten.

Daher lautet die Fragestellung für diese Dissertation:

Wie kann man in einem *Data-Mining*-Ansatz zur interaktiven Analyse und Verdichtung von MEDLINE-Resultaten die Vorteile beider Lösungen – die fehlende Notwendigkeit einer Installation, die kostenlose Verfügbarkeit und die verdichtende Aufbereitung der Suchresultate – zusammenbringen, ohne deren Nachteile billigend in Kauf nehmen zu müssen?

Das nachfolgende Kapitel 3 (Material) zeigt die benötigten technischen Grundlagen für eine Lösung aus diesem Dilemma auf, während die darauf basierende Lösung im Kapitel 4 (Methodik) en detail vorgestellt wird.

3 MATERIAL

Das Kapitel Material definiert einige für die Arbeit mit PubMed grundlegende Begriffe und stellt MEDLINE, PubMed und mit ihnen verknüpfte Systeme detaillierter vor. Eine Gesamtübersicht verwendeter termini technici mit Kurzbeschreibungen definiert das Glossar im Anhang.

Wie schon in der Problemstellung erwähnt, behandelt die vorliegende Arbeit PubMed als Datenquelle, da die zugrundeliegende Datenbank – MEDLINE – die größte medizinische Datenbank der Welt ist, und PubMed im Gegensatz zu anderen Distributionen kostenlos verfügbar ist. Jedoch wird einführend eine kurze Gegenüberstellung von MEDLINE und anderen renommierten biomedizinischen Datenbanken wie *Current Contents*, *Science Citation Index*, *Excerpta medica*, *BIOSIS* und der *Cochrane Library* vorgenommen. Literaturdienste von Pharmaunternehmen werden nicht berücksichtigt, desgleichen nicht andere MEDLINE-Distributionen wie *OVID*, *Knowledge Finder* oder *WinSPIRS*, die einer anderen Suchsyntax gehorchen.

3.1 Definitionen

Um terminologische Unklarheiten beim Leser zu vermeiden, werden im folgenden die Begriffe *Data Mining*, die in der Informatik verwendete Suchmaße *Relevanz* und *Nachweisquote* sowie der Begriff des *Dienstes* resp. *Service* näher erläutert.

3.1.1 Data Mining

Der Vorgang der nichttrivialen und automatischen Suche nach Wissen in Massendaten wird im folgenden kurz *Data Mining* genannt [Lusti 2002, S. 279]. Diese Definition wird hier noch einmal angeführt, da in manchen Gebieten der Informatik, speziell auch in der Künstlichen Intelligenz, der Begriff enger gefaßt wird, wie z.B. bei Wiedmann et al. [Wiedmann et al. 2001, S. 21], die ihn konstitutiv mit dem Begriff des computergestützten Lernens verbinden.

Anzumerken ist, daß es keine kanonisierte Definition für den Begriff des *Data Mining* gibt. Zu einem entsprechenden Überblick über die Artenvielfalt rezenter Definitionen verhilft der Aufsatz von Wiedmann et al. [a.a.O., S. 19]. *Data Mining* i.e.S. ermöglichen vor allem die Statistik, die Visualisierung, die Regelinduktion und Neuronale Netzwerke. *Data Mining* i.w.S. umfaßt auch **OLAP** (*On-line Analytical Processing*) und konventionelle Abfrage- und Tabellenkalkulations-Funktionen.

Der Begriff des *Data Mining* wird in dieser Arbeit i.w.S. verwendet.

3.1.2 Suchmaße

Relevanz und *Nachweisquote* sind in der Informatik verbreitete Hilfsgrößen, um Sucherfolge zu quantifizieren. Zum Aufbau der Definitionen werden die folgenden Suchmengennotationen benötigt:

Gegeben sei eine Suche nach Artikeln in einem Artikelbestand und die damit verbundenen Artikelmenge wie folgt:

- A - „Treffer“ - relevante gefundene Artikel
- B - „Ballast“ - irrelevante gefundene Artikel
- C - „Vermißt“ - relevante nicht gefundene Artikel
- D - „Umgangen“ - irrelevante nicht gefundene Artikel.

Die Artikelmenge können als sich überschneidende Mengen dargestellt werden (Tab. 1):

	Gefunden	Nicht gefunden	Summe
Relevant	A	C	A + C
Nicht relevant	B	D	B + D
Summe	A + B	C + D	A + B + C + D

Tab. 1: Artikelmenge beim Suchen

Es lassen sich dann folgende Suchmaße (Wertebereich $Y = \{ y \in \mathbb{Q} \mid 0 \leq y \leq 1 \}$) aufstellen:

3.1.2.1 Relevanz

Die Relevanz (Genauigkeitsmaß, engl.: precision) ist eine Größe zur Abschätzung der Genauigkeit einer Suche.

Sie berechnet sich per definitionem wie folgt:

$$\text{Precision } P = \text{Relevante gefundene Artikel} / \text{Alle gefundene Artikel} = A / (A+B)$$

3.1.2.2 Nachweisquote

Die Nachweisquote (engl.: recall) ist eine Größe zur Messung der Vollständigkeit einer Suche.

Sie berechnet sich per definitionem wie folgt:

$$\text{Recall } R = \text{Relevante gefundene Artikel} / \text{Alle relevante Artikel} = A / (A+C)$$

Anzumerken ist, daß Vollständigkeit nicht gemessen, sondern nur geschätzt werden kann, z.B. durch statistische Analysen (kleine Datenbestände ausgenommen). Vollständigkeit ist oft kein gutes Kriterium für Benutzerzufriedenheit, da die meisten Benutzer weniger Wert auf Vollständigkeit legen als auf Ergebnisse mit wenigen irrelevanten Artikeln, also auf eine hohe Relevanz eines Suchergebnisses. Ist das gesuchte Thema hingegen hochspezifisch, rücken Fragen der Art „Zeige mir alle Artikel zu diesem Thema“ in den Vordergrund und dementsprechend wird auch die Nachweisquote ein wichtiges Kriterium.

3.1.3 Dienst

Als letztes wird der Begriff des *Dienstes* definiert. Mit *Dienst* bzw. *Service* wird die Bereitstellung einer Funktionalität für Benutzer auf einem Server bezeichnet. Ein Dienst kann öffentlich verfügbar sein oder nur bestimmten Benutzern offenstehen. Er kann lokal oder über ein Netzwerk benutzt werden. Als ein Beispiel sei das Internet genannt, das seine Funktionalität über Dienste zur Verfügung stellt.

3.2 MEDLINE

MEDLINE® (*Medical Literature, Analysis, and Retrieval System Online*) ist die bedeutendste englische Literaturdatenbank für die gesamte Medizin. Sie ist seit 1971 verfügbar und wird durch die National Library of Medicine (NLM), USA, als elektronische Version des *Index Medicus* und anderer gedruckter Bibliographien (*Index to Dental Literature, International Nursing Index*) gepflegt.

Der *Index Medicus*® ist das durch die NLM seit über 100 Jahren gepflegte, gedruckte bibliographische Verzeichnis aller Artikel der meisten biomedizinischen Zeitschriften und Journale der Welt. Jährlich wird eine Liste von im *Index Medicus* indizierten Zeitschriften herausgegeben, die geänderte Zeitschriftennamen, Neuerscheinungen und eingestellte Zeitschriften berücksichtigt. Der *Index Medicus* wurde bis zum Jahre 2000 monatlich aktualisiert und jahrgangsweise in Form mehrerer Bände als *Cumulated Index Medicus* zusammengefaßt herausgegeben. Danach wurde diese Distribution von der NLM eingestellt, da immer weniger Kunden diesen Dienst abonnierten und statt dessen den elektronisch verfügbaren PubMed-Dienst für MEDLINE im WWW benutzten.

Die MEDLINE-Datenbank wird von verschiedenen Anbietern auf CD-ROM vertrieben, von kommerziellen Datenbank-Hosts on-line angeboten und nicht zuletzt auch von der medizinischen Nationalbibliothek der USA on-line und kostenlos unter PubMed zur Verfügung gestellt.

Anbieter, die MEDLINE auf CD-ROM mit unterschiedlichen Systemen anbieten, sind z.B.:

- *SilverPlatter* (System: *SPIRS*)
- *Aries/Nova Idea* (System: *Knowledge Finder*)
- *CD-Plus* (System: *OVID*).

Hosts, die analog zu PubMed MEDLINE on-line aufliegen haben, sind z.B.:

- *DIMDI*
- *STN*
- *DataStar*.

3.2.1 Inhalt von MEDLINE

MEDLINE enthält die bibliographischen Angaben sowie die *Abstracts* der wichtigsten internationalen medizinischen Zeitschriften (über 4.600) seit 1966 und

wird lt. [eigenen Angaben](#) (Stand 10.05.2003) von Dienstag bis Samstag, Januar bis Oktober, täglich auf den neuesten Stand gebracht. Durchschnittlich werden täglich über 2.000 Zitate hinzugefügt. Das derzeitige Volumen beträgt über 12 Mio. Artikel aus den USA und 70 anderen Staaten in 30 verschiedenen Sprachen. Etwa 52% aller Artikel wurden in den USA publiziert. Die bibliographischen Angaben können entweder typisiert über Deskriptoren oder als Freitext gesucht werden. Ca. 76% aller Zitate verfügen über englische Abstracts.

Sucht man auch Zitate aus dem Zeitraum vor 1966, kann man über das [Gateway](#) der NLM auch auf *OLDMEDLINE* zugreifen. Außerdem gibt es einen [patientenorientierten Zugang](#) zu MEDLINE via *MEDLINEplus*.

MEDLINE enthält Nachweise der internationalen Literatur aus allen Bereichen der Biomedizin und aus klinischen Gebieten, einschließlich der Zahn- und Veterinärmedizin, Psychologie, Krankenpflege, Pharmazie, vorklinischer Fächer und des öffentlichen Gesundheitswesens. Ab 2000 wurde das Spektrum der erfaßten Zeitschriften im Bereich *Life Science* stark erweitert. Ende 2001 wurden in MEDLINE viele (vormals separat bestehende) Spezialdatenbanken der NLM integriert.

MEDLINE eignet sich sehr gut, um sich einen Überblick über ein Thema zu verschaffen. Bei komplexen und umfangreichen Recherchen sollte man jedoch auch auf Spezialdatenbanken zurückgreifen.

Jedes Zitat eines Artikels aus einem Journal oder einer Zeitung wird in MEDLINE als ein Datensatz dargestellt, welcher aus mehreren Suchfeldern besteht. Als Suchfelder ausgelegt sind u.a.

- der Titel einer Arbeit
- der Autorenname (bis 1999 zu 25 Namen pro Zitat, ab 2000 unbegrenzt)
- Abstracts (nur bei ca. 80% der Datensätze verfügbar)
- von den Bibliothekaren der NLM vergebene Schlagwörter (*Medical Subject Headings*, weiter ausgeführt unter Punkt 3.5 - *MeSH*)
- die Artikelsprache (80% in Englisch)
- bibliographische Angaben (Journal, Jahrgang, Nummer, Paginierung etc.) und
- der Publikationstyp (Brief, Editorial, Kommentar, Review etc.).

3.3 DIMDI's Free grips WebSearch

DIMDI (Deutsches Institut für Medizinische Dokumentation und Information) propagiert einen kostenlosen On-line-Zugriff auf MEDLINE über *Free grips WebSearch*®. Beim Testen dieses Dienstes im Rahmen der vorliegenden Dissertation stellte sich aber heraus, daß die Benutzerführung abhängig vom Browser zuzeiten zu wünschen übrig ließ (Stand 6.4.2003):

DIMDI's Webseiten verlassen sich darauf, daß JavaScript bei den Benutzern eingeschaltet ist. Bei etlichen Benutzern ist dies jedoch nicht der Fall. Viele Anwender dürfen auch aus Sicherheitsaspekten (Firmen- bzw. Institutsnetzwerke) JavaScript nicht aktivieren. Benutzer ohne aktiviertes JavaScript scheitern somit in der Benutzeroberfläche von *Free grips WebSearch*. Es gibt seit Jahren professionelle serverseitige Techniken, die JavaScript entbehrlich werden lassen. Warum DIMDI an dieser Stelle nicht darauf zurückgreift, muß offen bleiben.

Benutzer, die nicht die letzten Versionen von Internet Explorer und Netscape besitzen, werden darauf hingewiesen, daß ihr Browser nicht paßt. Beim Autor erschien beim Test ein lapidares Fenster mit der Meldung, daß man Version 2.1 (von welchem Browser, wurde nicht erklärt) benutze. Diese Information war falsch. Die Einschränkung des Dienstes auf bestimmte Browser wirkt unprofessionell und steht im Widerspruch zu den Webentwicklungsrichtlinien des [W3C](#) (World Wide Web Consortium).

Beim Test scheiterte die freie Recherche über alle Datenbanken, die aufgerufene Webseite blieb im Initialisierungsstatus hängen trotz mehrfacher Versuche – zugegebenermaßen mit einem Browser, der von DIMDI nicht „erlaubt“ war.

Der Test der neuen Beta-version von DIMDI's *MEDLINE Direct* funktionierte immerhin halbwegs, wenn auch mit einigen Fehlern. Bei der Suche nach „Herzinsuffizienz“ wurden

FT=Herzinsuffizienz mit 39.085 Artikeln

und

ME0A; ML66 mit 12.250.234 Artikeln

gefunden. Was der letzte Posten bedeutet, wurde nicht erklärt. Anzunehmen ist, daß er die Gesamtzahl der in MEDLINE vorgehaltenen Artikel bezeichnet, die sich in dieser Größenordnung bewegt, jedoch ist dieser Suchposten weder verständlich noch sehr hilfreich und außerdem irritierend.

Beim Anklicken der Option „Versenden“ öffnete sich ein Registrierungs-fenster;

offensichtlich muß man sich bei DIMDI registrieren lassen, um sich Ergebnisse zusenden zu lassen. Ein Hinweis, was mit Versenden gemeint ist (E-Mail oder postalisch) fand sich in der Hilfe nicht.

Das Anklicken der Option „Ausgeben“ blieb schlußendlich völlig wirkungslos.

Beim zweiten Test mit einem von DIMDI „erlaubten“ Browser funktionierte dann die Suche sowohl in den freien Datenbanken als auch in *MEDLINE Direct*. Maximal kann man sich 100 Artikel ausgeben lassen - PubMed hingegen (s. Punkt [3.4](#)) erlaubt bis zu 10.000 Artikel.

Fazit: Da DIMDI die Artikel von MEDLINE auch mit deutschen Schlagworten versieht, können des Englischen nicht Kundige direkt nach deutschen Schlagworten suchen, vorausgesetzt, sie haben einen „DIMDI-zertifizierten“ Browser und außerdem JavaScript in ihrem Browser aktiviert. Wer keine Zeit hat, sich mit DIMDI's spezifischem Vokabular auseinanderzusetzen (ME0A, FT etc.) und das Suchen nach englischen Begriffen nicht scheut, kann direkt auf PubMed zurückgreifen, das in der folgenden Sektion erklärt wird.

3.4 PubMed

Im folgenden wird unter *PubMed*® der öffentlich über HTTP (*Hypertext Transfer Protocol*) zugängliche Dienst unter der [URI \(Uniform Resource Identifier\)](http://www.ncbi.nlm.nih.gov/PubMed/) <http://www.ncbi.nlm.nih.gov/PubMed/> verstanden, der vom 1988 gegründeten National Center for Biotechnology Information (NCBI) in Bethesda, Maryland, USA betreut wird. Das NCBI gehört mit der National Library of Medicine (NLM) zum National Institute of Health (NIH), dem größten staatlichen medizinischen Forschungsinstitut des amerikanischen Gesundheitsministeriums, dem 27 separate Institute und Zentren angehören.

Abb. 1 zeigt die Suchmaske, die den Zugriff auf den Dienst zur Verfügung stellt:

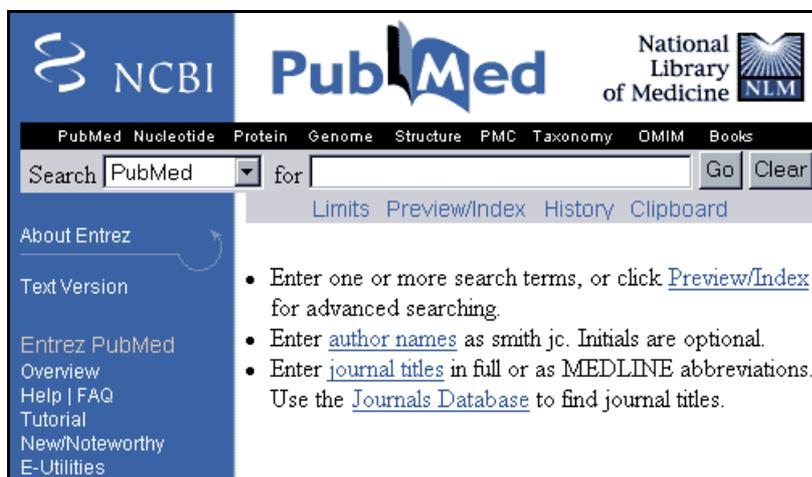


Abb. 1: Ausschnitt aus PubMeds Suchformular

Die medizinische Nationalbibliothek der USA hat PubMed 1997 als wissenschaftlichen und kostenlosen Suchdienst für MEDLINE und PreMEDLINE eingerichtet. Es dauerte nicht lange, daß auch viele andere Webpräsenzen einen (oft kostenpflichtigen) Zugang zu MEDLINE anboten. PubMeds Service zeichnet sich neben den weiter unten beschriebenen Merkmalen dadurch aus, daß er nicht nur authentifizierten medizinisch Ausgebildeten, sondern allen Interessierten kostenlos zur Verfügung steht. Als Benutzerinterface wird ein hypertextfähiger HTTP-Client (*Hypertext Transfer Protocol*) benötigt, der heutzutage ubiquitär in fast allen Bereichen von Arbeits- und Privatleben anzutreffen ist, eine besondere Installation ist somit auf den Benutzerrechnern nicht vonnöten.

Durch den gleichzeitigen Zugriff auf MEDLINE und PreMEDLINE ist PubMed ein gut à jour gepflegter Suchdienst, da *PreMEDLINE* von Dienstag bis Samstag täglich mit den Zitaten neuer Artikel aktualisiert wird (die Übernahme eines Zitates aus dem Status *PreMEDLINE* in den Status *MEDLINE* hingegen, der die Verschlagwortung der Artikelzitate impliziert, kann einige Wochen dauern, unter Umständen ein Vierteljahr bis halbes Jahr).

In PubMeds Suchmaschine können Anwender eine Freitextsuche durchführen oder nach unterschiedlich rubrizierten Begriffen suchen (s.a. [Anhang B](#)), die in Suchfeldern erfaßt sind. Beispielhaft genannt werden soll hier die Suche nach Autoren, nach Schlagworten, welche die Bibliothekare der NLM den Artikeln zuordnen, oder nach anderen Feldern wie Titel, Abstract, Sprache u.a.m. Der Anwender kann außerdem den *MeSH* (*Medical Subject Headings*) benutzen, um sich die Stellung eines Schlagwortes im MeSH-Baum anzeigen zu lassen oder ähnliche Schlagwörter zu finden. In diesem MEDLINE-spezifischen Thesaurus sind alle Schlagworte hierarchisch aufgelistet, mit denen die Artikel indiziert wurden.

3.4.1 Inhalt von PubMed

PubMed enthält die Daten von MEDLINE sowie einige weitere Datenbestände:

- Nichtmedizinische Artikel aus MEDLINE-Journalen (z.B. zur Astrophysik oder Plattentektonik)
- Bibliographische Daten zu Artikeln aus MEDLINE-Journalen, die vor der ersten Erfassung dieser Journale in MEDLINE liegen (PreMEDLINE)
- Volltextversionen einiger Journale, deren Verlage sie elektronisch *PubMedCentral*TM zur Verfügung stellen und die durch die NLM qualitativ geprüft werden.

3.4.2 Präsentation

Die Ergebnisse werden dem Benutzer in der PubMed-Oberfläche als Listen präsentiert.

Der Anwender kann die Listen nach Autor, Journal oder Publikationsdatum sortieren. Häufigkeitsauswertungen, Korrelationsanalysen oder die Ausgabe gefundener Schlagwörter als MeSH-Baum sind jedoch nicht möglich. Durch die sequentielle Auflistung der gefundenen Artikel gestaltet sich die Erfassbarkeit wesentlicher Informationen für den Anwender zumeist sehr schwierig.

Abb. 2 zeigt einen Ausschnitt aus einem typischen PubMed-Resultat:

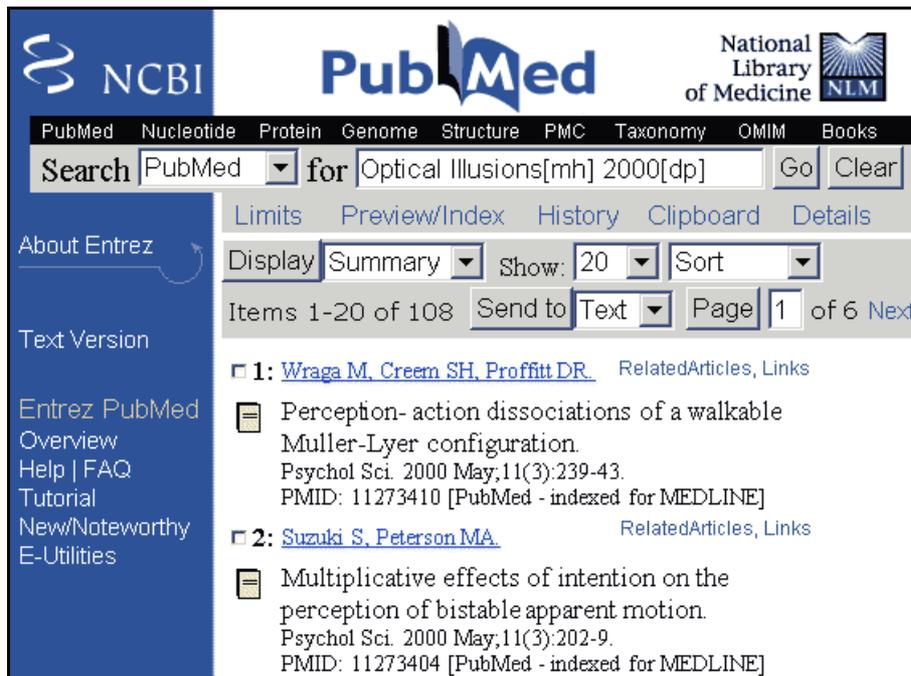


Abb. 2: Ausschnitt aus einem PubMed-Resultat

Die gefundenen Artikel können entweder direkt in der PubMed-Oberfläche angezeigt werden oder als Datei an den Benutzer geschickt werden. Maximal 10.000 Artikel sind zum Versand erlaubt.

Auch hier ist kritisch anzumerken, daß JavaScript aktiviert sein sollte, um bestimmte Funktionen von PubMed zu nutzen. Auch verstehen bestimmte Browser nicht die JavaScript-Funktionen, die PubMed verwendet, wie die Kundenbetreuung von PubMed in ihrer On-line-Hilfe einräumt.

3.4.3 Merkmale der PubMed-Suche

Die folgenden Sektionen beschreiben die einzelnen Merkmale (*Features*) genauer, die PubMed dem Benutzer anbietet, um seine Suche zu steuern. Wenn Artikelzahlen genannt werden, beziehen sie sich auf Stand Mai 2003.

3.4.3.1 Typisierte Feldsuche

Eine Recherche kann vom Benutzer auf bestimmte Felder eingeschränkt werden.

Die Suche nach dem Begriff `prostate[TI]` fände z.B. nur Artikel, die den Begriff im Titel führen, nicht jedoch solche, die ihn lediglich im Abstract aufweisen. Durch die Typisierung von Suchausdrücken kann die Relevanz eines Suchergebnisses gesteigert werden, was beim *Term Mapping* näher erläutert wird.

3.4.3.2 Boolesche Suche

Durch die Verwendung der logischen Operatoren *AND*, *OR* und *NOT* sowie runder Klammern kann eine Suche logisch strukturiert werden. Die Operatoren müssen großgeschrieben werden.

Eine Suche nach

```
tetrodotoxin AND voodoo
```

findet somit nur drei Zitate zu Artikeln, die beide Operanden enthalten, eine Suche nach

```
tetrodotoxin OR voodoo
```

zeitigt hingegen 9.568 Artikel.

Der an logischen Operatoren interessierte Leser sei an dieser Stelle auf die weiterführende Literatur verwiesen. Eine Einführung in die zweiwertige Aussagenlogik findet sich in den Lehrbüchern zur Informatik, als ein Bsp. nur [[Schöning 1992](#), S. 13 ff.].

3.4.3.3 Term Mapping

Um möglichst viele Artikel zu finden, verwendet PubMed ein sog. *Term Mapping*: eingegebene Begriffe werden auf Synonyme oder gleichlautende Begriffe im Schlagwortverzeichnis, dem Journal- und dem Autorenverzeichnis untersucht und der Suchbegriff entsprechend erweitert.

Eine Suchanfrage von

```
vitamin c common cold
```

würde von PubMed vor dem eigentlichen Suchvorgang intern zu

```
(ascorbic acid[mh] OR vitamin c[tw]) AND (common cold[mh] OR  
common cold[tw])
```

erweitert werden. Vitamin C erkennt PubMed als *Entry Term* für das Schlagwort *Ascorbic Acid* und *Common Cold* wird ebf. als Schlagwort erkannt. Danach

werden beide Schlagwörter von PubMed mit UND verknüpft. Bei dieser Suchanfrage gab PubMed 202 Treffer zurück.

Ohne ein *Term Mapping* würde lediglich nach

```
vitamin c[tw] AND common cold[tw]
```

gesucht. Dies ergäbe nur 131 Treffer, da Artikel, welche die Suchbegriffe nur in den Schlagworten tragen, entfielen.

Würde man hingegen eine typisierte Abfrage wie

```
vitamin c[mh] common cold[mh]
```

absetzen, erhielte man 195 Treffer. Durch die explizite Angabe eines Feldtyps, hier des MeSH-Feldes, wird das automatische *Term Mapping* abgeschaltet und keine Freitextsuche vorgenommen. Dies führt zu weniger, jedoch relevanteren Suchresultaten. Da in der Freitextsuche die Suchbegriffe auch in Abstracts von Artikeln enthalten sein können, die unwesentlich im Sinne der Suche sind, kann die Relevanz des Suchergebnisses durch die Verwendung von MeSH-Termen gesteigert werden [Coletti et al. 2001, S. 323]. Da jedoch viele Anwender mit der Schlagwortsuche nicht vertraut sind, nimmt PubMed ihnen mit dem automatischen *Term Mapping* dieses Nachdenken ab.

Fazit: Eine reine Freitextsuche birgt die Gefahr, viele Artikel zu übersehen. Deshalb verwendet PubMed automatisch ein *Term Mapping*, um die Nachweisquote durch Einbindung von typisierten Feldern zu erhöhen. Verwendet der Benutzer hingegen selbst typisierte Felder, kann zusätzlich die Relevanz des Suchergebnisses gesteigert werden.

3.4.3.4 Trunkierung

Des weiteren bietet PubMed die Möglichkeit zur Trunkierung von Suchbegriffen.

Eine Suche in PubMed nach

```
prost*
```

fand 89.268 Artikel mit sowohl *prostate* als auch *prostatectomy*, aber auch *prostaglandin* etc. Die Trunkierung schaltet das automatische *Term Mapping* ab.

3.4.3.5 Wörtliche Suche

Durch das Fassen eines Suchbegriffs in doppelte Hochkommata wird der Begriff als untrennbare Einheit betrachtet und wörtlich gesucht.

Die Suche nach

"diagnosis related groups"

sucht also nur nach Zitaten, welche die ganze Zeichenkette enthalten, und fand in PubMed 7.113 Treffer. Ein *Term Mapping* wird bei der wörtlichen Suche nicht vorgenommen.

Eine einfache Suche nach

diagnosis related groups

hingegen wurde von PubMed im Prozeß des *Term Mapping* in

"diagnosis-related groups"[mh] OR diagnosis related groups[tw]

umgeformt und ergab 7.223 Treffer.

3.4.3.6 Limits

PubMed bietet das *Limits*-Fenster an, um einige wenige ausgewählte Felder (Institution, Autor, Sprache, Volltext verfügbar, Beschränkung auf bestimmte Themata etc.) als einschränkendes Suchkriterium hinzuzufügen, ohne daß sich der Anwender mit PubMeds Notation der Feldtypen auseinandersetzen muß. Beherrscht man hingegen PubMeds Suchnotation, ist dieses Fenster obsolet.

3.4.3.7 Verknüpfte Artikel

Falls eine Suche nicht viel an Daten liefert, einer der gefundenen Artikel aber doch relevant ist, kann die Option *Related Articles* benutzt werden, um verwandte Artikel zu finden. Mit etwas Glück findet der zugrundeliegende Algorithmus weitere thematisch bezogene Artikel, darauf verlassen kann man sich jedoch nicht. Viele Anwender nutzen dieses Merkmal als bequeme Möglichkeit, um ihren Suchradius zu verändern, ohne viel nachdenken zu müssen [Shea 2002, S. 167].

3.4.3.8 Klinische Fragestellungen

PubMed erlaubt es, mehrere Filter einzusetzen, die auf klinische Fragestellungen Rücksicht nehmen. Diese Filterung wird in heutiger Zeit immer wichtiger, um die Spreu vom Weizen zu trennen. Nicht jeder Aufsatz besitzt klinische Relevanz nach EBM-Kriterien (*Evidence based Medicine*), um das darin enthaltene Wissen in der Praxis einzusetzen zu können.

Als Beispiel nur ein Zitat von Grandage, die über die jahrelange Verschreibung von Antiarrhythmica bei asymptomatischen ventrikulären Arrhythmien berichtet.

Diese Pharmaka wurden verschrieben, da man vermutete, daß die Reduktion der Arrhythmieanfälle das Risiko einer letal ausgehenden kardialen Dekompensation reduziere. Nach sechs Jahren entdeckte eine Studie, die untersuchen sollte, ob die Mortalität dieser so behandelten Patienten signifikant reduziert wurde, daß die Mortalität im Vergleich zu unbehandelten Patienten jedoch noch gesteigert wurde. Dieses erstaunliche Ergebnis konnte in mehreren nachfolgenden Studien verifiziert werden [Grandage et al. 2002, S. 300 ff.]. Dies ist nur ein Beispiel von vielen dafür, daß vorläufige Ergebnisse aus der wissenschaftlichen Forschung nicht immer im klinischen Alltag verwertbar sind.

3.4.3.8.1 Clinical Queries

Die Suche mit Hilfe von *Clinical Queries* erlaubt es, mehrere Filter zu verwenden, welche auf Fragen aus dem klinischen Alltag Bezug nehmen. Ein Filter ist hier eine Kombination fixer Ausdrücke, die – intern von PubMed hinzugefügt zum Suchbegriff des Benutzers – seine Suche auf bestimmte Bereiche einschränkt.

In PubMed kann der Anwender nach Eingabe der Suchbegriffe über die *Clinical Queries* angeben, ob es sich um eine therapeutische, diagnostische, ätiologische oder prognostische Fragestellung handelt, und ob er Sensitivität oder Spezifität höher wichtet. Damit lassen sich schnell klinisch relevante Arbeiten auffinden.

Der interne Filter, den PubMed hierbei verwendet, ist eine Kombination von Schlag- und Textwörtern basierend auf der Arbeit von Haynes et al. [Haynes et al. 1994, S. 447 - 458]. Haynes empfahl diese Suchterme bei der Recherche nach klinischen Studien für den Abfragezeitraum von 1991 bis heute, PubMed verwendet sie der Einfachheit halber jedoch für jeden abgefragten Zeitraum. Mit diesen spezifischen Termkombinationen kann die Sensitivität der Suche nach klinischen Studien – verglichen mit einer einfachen Suche – um mehr als 30%, allerdings auf Kosten der Spezifität, erhöht werden.

Tab. 2 veranschaulicht die erreichten Sensitivitäten nach Haynes für ausgewählte Studien aus dem Jahre 1991:

Klinische Kategorie	Spitzensensitivität [%]
Ätiologie	82
Prognose	92
Diagnose	92
Therapie	99

Tab. 2: Sensitivität kombinierter Suchbegriffe für klinische Studien 1991 nach Haynes et al.

Wurde der Fokus auf eine hohe Spezifität gelegt, erreichten die entsprechenden Termkombinationen eine Spezifität von über 93% über alle klinischen Kategorien

für ausgewählte Studien aus den Jahren 1986 und 1991.

Das Risiko, auch einige relevante Artikel zu verpassen, ist allerdings nicht zu unterschätzen. Bei eng umgrenzten Recherchen mit wenigen Fundstellen ist es sinnvoll, diese Option auszulassen.

Nach der Implementierung des klinischen Filters in PubMed sind wiederholt Versuche unternommen worden, seine Leistung zu verbessern. Als ein Beispiel nur seien Bachmann et al. von der Zürcher Universität [[Bachmann et al. 2002](#), S. 653 - 658] genannt, die berichten, PubMeds Filterleistung für diagnostische Studien mit einer eigenen Suchtermkombination verbessert zu haben, die bei hoher Sensitivität eine bessere Suchpräzision aufweise. Dazu ist kritisch anzumerken, daß es immer möglich ist, mit einer sorgfältig ausgesuchten Studienmenge einen generischen Algorithmus zu schlagen, der für alle Zeiträume und Studientypen gelten soll.

Fazit: Durch die Fülle der dargebotenen Informationen in PubMed bei breitflächigen Recherchen ist PubMeds Filter für den Kliniker ein unschätzbares Hilfsmittel zur Reduktion der Datenflut in der täglichen Arbeit. Speziell Fragestellungen wie „Welche Therapie ist für den Patienten in dieser Situation am besten?“, in der angloamerikanischen Literatur in Anlehnung an die *EBM POEM* (*Patient-oriented evidence that matters*) genannt [[Shaughnessy et al. 1994](#), S. 489 - 499], lassen sich so gezielt ausfiltern.

3.4.3.8.2 Systematic Reviews

Eine Alternative zu den *Clinical Queries* stellt PubMeds Filter *Systematic Reviews* dar, der auf EBM-Kriterien fokussiert. Er basiert zum großen Teil auf einer Arbeit von Shojania und Bero [[Shojania et al. 2001](#), S. 157 - 162].

Dieser Filter kombiniert den Suchterm des Anwenders mit Suchtermen, die nach Systematischen Reviews, Metaanalysen, Reviews klinischer Studien, EBM, Konsensuskonferenzen oder Richtlinien suchen. Zitate von Journalen, die sich auf klinische Reviews spezialisiert haben, werden ebf. berücksichtigt.

Als Fazit ist der Schluß zu ziehen, daß bei breitflächigen Recherchen mit klinischem Fokus *Clinical Queries* und *Systematic Reviews* eine wertvolle Hilfe für den Anwender darstellen.

Als parallele Ansätze zur Erhöhung der Relevanz in der Suche nach klinischen Studien oder Reviews sei hier unter anderen noch die *Cochrane Library* genannt, deren Reviews nur randomisierte kontrollierte Studien bewerten und deren

Reviews qualitativ höherwertig sind als andere Reviews [[Olsen et al. 2001](#), S. 829 - 832], sowie das Programm *Info Retriever*, dessen Aufgabe darin besteht, den Arzt parallel zur Arbeit mit dem Patienten (und nicht danach) schnell mit klinisch hochwertigen Informationen zu versorgen. Dazu sucht *Info Retriever* nach EBM-Kriterien in festgelegter Reihenfolge die klinisch relevantesten und validesten Datenquellen zuerst ab, ehe das Programm auch weniger relevante Quellen berücksichtigt [[Grandage et al. 2002](#), S. 298 - 304].

3.5 MeSH

MeSH® (*Medical Subject Headings*) ist die Kurzbezeichnung für den Thesaurus der Datenbank MEDLINE. Er ist gedruckt und als Datenträgerversion erhältlich (jeweils in verschiedenen Sprachen), aber auch [on-line](#) abrufbar. Der MeSH wird von der NLM jährlich aktualisiert.

Alle Zitate in MEDLINE werden mit Hilfe von genormten Deskriptoren (*Subject Headings, Controlled Terms, CT*) durch die NLM indiziert. Über ein Schlagwort kann man somit die bibliographischen Daten aller zugehörigen Artikel finden. Die Deskriptoren sind hierarchisch zusammengefaßt im *MeSH, den Medical Subject Headings (Controlled Vocabulary)*.

Die Verwendung von MeSH-Termen statt einer Freitextsuche erhöht die Relevanz des Suchergebnisses beträchtlich. Manche Autoren stehen sogar auf dem Standpunkt, daß man jede Suche in PubMed – sofern möglich – mit dem MeSH durchführen sollte [[Coletti et al. 2001](#), S. 323].

Die Artikel von PubMed sind nach dem Prinzip *Main heading/Subheading* verschlagwortet. Die *Main headings* kennzeichnen ein Gebiet, auf welches die *Subheadings* gezielter eingehen. Die *Subheadings* erlauben somit eine Feinabstufung der Schlagwörter für den betreffenden Artikel, wie z.B. die Einstufung, ob ein Schlagwort dieses Artikels im Zusammenhang mit Ätiologie, Diagnose, Therapie o.ä. genannt wurde, z.B. `multiple sclerosis/etiology`.

Um gezielt nach Zitaten zu bestimmten Themengebieten zu recherchieren, verwendet man den MeSH-Browser (Abb. 3). In die Eingabezeile gibt man Suchbegriffe zum interessierenden Thema ein.

The screenshot shows the MeSH search interface on the NCBI website. The search bar contains 'vitamin c' and the results list 'Ascorbic Acid' and 'Ascorbic Acid Deficiency'. The interface includes a navigation menu with options like PubMed, Nucleotide, Protein, Genome, Structure, PMC, and Taxonomy. The search results are displayed in a table with columns for Limits, Preview/Index, History, Clipboard, and Details. The results are sorted by relevance and show the first two items.

Abb. 3: Ausschnitt aus einem MeSH-Resultat

Ist der Begriff kein im MeSH festgelegtes Schlagwort, bietet PubMed eine Reihe ähnlicher Terme an, aus denen man das gewünschte Schlagwort auswählen kann.

Zur sachlichen Erschließung eines Themas reicht der MeSH jedoch nicht immer aus: Oftmals erscheinen Publikationen zu einem neuen Thema, bevor ein entsprechendes Schlagwort vergeben wurde. Diese Publikationen entgehen einem bei der Schlagwortsuche, da keine nachträgliche Anpassung erfolgt. *Genetic Privacy* z.B. findet sich erst ab 2002 im MeSH. In so einem Fall kann nur eine Freitextsuche weiterhelfen.

DIMDI übersetzt seit 1996 den MeSH jedes Jahr ins Deutsche und bietet ihn elektronisch oder als Druckversion an – im Gegensatz zur *National Library of Medicine* jedoch kostenpflichtig. Die Preise bewegen sich für die nicht-kommerzielle Nutzung lt. [DIMDI's Angaben](#) zwischen 180 € und 1.330 € (Stand 06.04.2003); für die kommerzielle Nutzung sind keine Preise angegeben. Abonnenten zahlen etwas weniger.

3.5.1 Inhalt des MeSH

Der englische MeSH enthält lt. [NLM](#) (Stand 06.04.2003):

- 21.973 Schlagwörter (*Descriptors*)
- 125.858 Querverweise (*Entry Terms*)
- 132.123 chemische Schlagwörter (*Supplementary Concept Records*)

DIMDI's deutscher MeSH enthält lt. [DIMDI](#) (Stand 06.04.2003):

- 20.600 Schlagwörter
- 22.000 deutsch-englische Querverweise
- 5.500 zusätzliche deutsche Bezeichnungen (ZT)
- 212 nicht übersetzbare englische Querverweise

Die englische Version kennt im Gegensatz zur deutschen sehr viele Querverweise (*Entry Terms*). Sie dienen als Synonyme für die Schlagwörter. Für *Ascorbic Acid* listet der amerikanische MeSH z.B. die folgenden 13 Querverweise:

Acid, Ascorbic
L-Ascorbic Acid
Acid, L-Ascorbic
L Ascorbic Acid
Vitamin C
Sodium Ascorbate
Ascorbate, Sodium
Ascorbic Acid, Monosodium Salt
Ferrous Ascorbate
Ascorbate, Ferrous
Hybrin
Magnesium Ascorbicum
Magnorbin

3.5.2 Gliederung

Der MeSH ist polyhierarchisch strukturiert und besteht aus vier Teilen: dem hierarchischen, dem alphabetischen, dem permutierten und dem chemischen Teil. Alle vier Teile werden im folgenden ausführlich beschrieben.

3.5.2.1 Hierarchischer Teil

Der hierarchische Teil (*MeSH Tree Structures*) besteht aus den 15 folgenden Kategorien und wird auch als systematischer Teil bezeichnet. Jede Kategorie umfaßt spezifischere Unterkategorien, die hier aus Platzgründen jedoch nicht aufgelistet werden können:

A: Anatomy
B: Organisms
C: Diseases
D: Chemicals and Drugs
E: Analytical, Diagnostic and Therapeutic Techniques and Equipment
F: Psychiatry and Psychology
G: Biological Sciences
H: Physical Sciences
I: Anthropology, Education, Sociology and Social Phenomena
J: Technology and Food and Beverages
K: Humanities
L: Information Science
M: Persons
N: Health Care
Z: Geographic Locations

3.5.2.2 Alphabetischer Teil

Im *MeSH Annotated Alphabetic List* sind alle Deskriptoren alphabetisch gelistet:

- Einführung, von DIMDI nicht übersetzt
- Schlagwörter (*Main headings*), z.B. *ascorbic acid*
- Zusatzbezeichnungen (*Subheadings*), z.B. *ascorbic acid/blood*
- Querverweise (*Entry Terms*), z.B. *vitamin c see ascorbic acid*
- Anmerkungen

3.5.2.3 Permutierter Teil

Das permutierte Verzeichnis des MeSH (*Permuted MeSH*) entlastet das alphabetische Verzeichnis bei Mehrwortbezeichnungen durch die Auflistung der zu einem signifikanten Teilbegriff vorhandenen Deskriptoren. Dies ist eine wesentliche Hilfe, um eine schnelle elektronische Suche gewähren zu können.

Beispiel „Index“:

Consumer Price Index

→ See Economics

Cornell Medical Index

Dental Plaque Index

DMF Index

Gingival Index

→ See Periodontal Index

Health Status Index

→ See Health Status Indicators

Index Medicus

→ See MEDLARS

...

3.5.2.4 Chemischer Teil

Im chemischen Teil, den *Supplementary Concepts Records*, sind die Bezeichnungen chemischer Substanzen aufgelistet. Diese werden von DIMDI nicht übersetzt. Beispiel:

4-Benzoxazolecarboxylic acid, 5-(methylamino)-2-((3,9,11-trimethyl-8-(1-methyl-2-oxo-2-(1H-pyrrol-2-yl)ethyl)-1,7-dioxaspiro(5.5)undec-2-yl)methyl)-, (6S-(6alpha(2S*,3S*),8beta(R*),9beta,11alpha))-

3.6 Wichtige biomedizinische Datenbanken im Vergleich zu MEDLINE

Im folgenden wird ein Vergleich einschlägiger biomedizinischer Datenbanken mit MEDLINE bzw. PubMed vorgenommen, indem wichtige Vor- und Nachteile der einzelnen Datenbanken, speziell in den Punkten Aktualität und Umfang, beleuchtet werden.

3.6.1 MEDLINE

Nach einer Studie der Münsteraner Zweigbibliothek für Medizin im August 2000 [Obst 2000, S. 34], in der die Prävalenz aktueller Artikel ausgewählter medizinischer Periodika in Datenbanken untersucht wurde, schnitt PubMed als die mit Abstand aktuellste Datenbank ab (durchschnittlich eins bis vier Wochen Vorsprung), verglichen mit *Current Contents*, *Web of Science*, *Embase Alert* und den Homepages der Periodika – keine Spur von „current“ oder „alert“ bei *Current Contents* und *Embase Alert*. Einen deutlichen Schwerpunkt setzt PubMed naturgemäß auf die angloamerikanische Literatur.

Journale und Institutionen, die mit PubMed kooperieren, versenden ihre Artikelzitate an PubMed bereits vor oder gleichzeitig mit der Publizierung der entsprechenden Artikel. Zumeist findet sich in PubMed dann auch ein Link auf den Volltext, der vom entsprechenden Provider vorgehalten wird. Stand Mai 2003 fanden sich bereits 973 mit PubMed kooperierende Provider. Dies mag auch ein Grund für die Aktualität des vorgehaltenen Datenpools von PubMed sein.

MEDLINE, auf das PubMed aufsetzt, gilt als eine der bedeutendsten Literaturdatenbanken der Welt, so wie die NLM als umfassendste medizinische Bibliothek der Welt gilt. Jedoch ist anzumerken, daß MEDLINE für umfassende Recherchen alleine nicht ausreicht: In diesem Fall empfiehlt sich zusätzlich eine Suche in *EMBASE*, *Current Contents*, *BIOSIS*, *Science Citation Index/Web of Science* o.a. Datenbanken.

MEDLINE verfügt z.B. nicht über umfassende Referenzen auf dem Gebiet der Psychologie, Medizinischen Soziologie und nichtklinischen Pharmakologie [Greenhalgh 1997, S. 180 - 183]. Des weiteren werden europäische Journale, wie schon erwähnt, weniger berücksichtigt als beispielsweise bei *EMBASE*.

Topfer et al. [Topfer et al. 1999, S. 297 - 303] kommen zu dem Schluß, daß man relevante Artikel verlieren würde, verließ man sich bei der Suche nur auf MEDLINE, und schlagen daher eine kombinierte Suche in MEDLINE und *EMBASE* vor.

Zu demselben Schluß kommen auch Suarez-Almazor et al. [Suarez-Almazor et al. 2000, S. 476 - 487], wenn sie in einer Studie zum Vergleich von MEDLINE mit EMBASE feststellen, daß für eine umfassende Suche nach kontrollierten klinischen Studien die Benutzung von MEDLINE alleine nicht ausreicht; vielmehr sei die Benutzung von zwei oder mehr Datenbanken sowie die manuelle Einsichtnahme in ausgewählte Journale vonnöten.

3.6.2 Cochrane Library

Für die umfassende Suche nach kontrollierten klinischen Therapiestudien eignet sich die *Cochrane Library*. Interessant ist hier die Komponente *Cochrane Controlled Trials Register (CTR)*: sie enthält alle Publikationen zu kontrollierten klinischen Studien aus den genannten Datenbanken und zusätzliche Artikel, die in keiner Datenbank erfaßt sind. Systematische Reviews sind in den Komponenten *DARE* und *CDSR* erfaßt. Mit der Schaffung des *Cochrane Consumer Network*³ in Form einer Webpräsenz hat die *Cochrane Collaboration* einen ersten Schritt in die Richtung getan, auch medizinischen Laien Zugang zum medizinischen Wissen zu gewähren [White 2002, S. 218 ff.].

Die Cochrane-Datenbank wird quartalsweise aktualisiert.

Ein kontrolliertes Vokabular gibt es nicht, MEDLINE-Artikel sind jedoch mit ihren Schlagwörtern (MeSH) erfaßt.

3.6.3 EMBASE

EMBASE wird vom niederländischen Verlag Elsevier Science Publishers hergestellt und auf CD-ROM oder on-line bei DIMDI u.a. Datenbankservern angeboten. Schwerpunkt sind Medikamente und Pharmakologie, die Datenbank enthält die Inhalte der Referatzeitschriften *Excerpta Medica*. Der Datenbestand ist dem von MEDLINE sehr ähnlich (60 – 80%), *EMBASE* verfügt jedoch über mehr Artikel aus Europa. Ca. 4.000 medizinische Zeitschriften werden von *EMBASE* erfaßt, zusätzlich aber auch Kongreßberichte. Ca. neun Mio. Datensätze sind seit 1970 erfaßt mit ca. 450.000 neuen Zitaten und Abstracts pro Jahr.

Der Datenbestand wird monatlich aktualisiert. Kritische Worte zur realen Aktualität der Datenbank verliert jedoch Obst [Obst 2000, S. 34], wie schon oben erwähnt.

Bemerkenswert ist ferner, daß idente Suchanfragen an *EMBASE* unter unterschiedlichen Anbietern (DataStar, Dialog, DIMDI und STN) „beträchtliche“

³ <http://www.cochraneconsumer.com/>

Diskrepanzen im Resultat zeitigen können [[Gretz et al. 1996](#), S. 494 - 506].

Ein kontrolliertes Vokabular liegt vor (*EMTree*, nicht ident zu MEDLINEs MeSH).

3.6.4 Current Contents

Current Contents ist eine fachübergreifende Inhalts- und Abstract-Datenbank von Artikeln, Tagungsberichten und Editorials aus über 7000 Zeitschriften (mit Link vom Literaturzitat zum Inhaltsverzeichnis der Zeitschrift).

Die Datenbank wird seit 1990 gepflegt und wöchentlich aktualisiert, teilweise werden Artikel schon vor oder mit ihrem Erscheinen indiziert. Zur realen Aktualität der Datenbank vgl. jedoch noch einmal Obst [[Obst 2000](#), S. 34], der in seiner Studie ausführt, daß trotz des Anspruches von *CCMed* (*Current Contents Medicine* der ZB Medizin Köln), in MEDLINE oder EMBASE indizierte deutsche Zeitschriften schneller als dort nachzuweisen, im beobachteten Zeitraum z.B. die drei letzten Hefte des Deutschen Ärzteblatts in der Datenbank fehlten.

Ein kontrolliertes Vokabular gibt es nicht, statt dessen finden sich unkontrollierte Autorenschlagwörter und automatisch generierte Schlüsselwörter.

3.6.5 BIOSIS

BIOSIS Previews ist die wichtigste Datenbank auf dem Gebiet der Biologie und wird von Biosis, Philadelphia, USA angeboten. Die Datenbank ist außerdem online u.a. abrufbar bei DIMDI und DataStar. Mehr als 50% der Artikel stammen aus der Human- und Tiermedizin, Biochemie und -physik, Pharmakologie, Umweltforschung sowie Land- und Forstwirtschaft. Für den Mediziner ist sie besonders interessant für Recherchen zu Mikrobiologie und Hygiene, aber auch für molekularbiologische Fragestellungen, z.B. für die Suche nach einer bestimmten Methode. *BIOSIS* enthält ca. 9.300 Zeitschriften (davon 80% selektiv), sowie Konferenzberichte und US-Patente bis 1989. Die Hälfte der Zitate verfügt über Abstracts.

Lt. Angaben des Herstellers wächst der Datenbestand täglich um ca. 1.500 Dokumente.

BIOSIS verfügt über zwei kontrollierte Vokabularien (nicht ident zu MEDLINEs MeSH).

3.6.6 Web of Science / Science Citation Index

Der fachübergreifende *Science Citation Index*, der vom Institute for Scientific Information (ISI), Philadelphia, USA hergestellt wird, umfaßt Artikel aus dem

Zeitraum 1991 – 1996. Ab 1997 gibt es das *Web of Science (SCI - Science Citation Index expanded)*. Es umfaßt Literaturzitate aus ca. 5.300 bedeutenden Journalen mit Angabe und Suchmöglichkeiten der zitierten Literaturstellen (*Citation indexing*). Letzteres ist bemerkenswert, da Fragen der Art „Wie oft wurde die Publikation eines Autors zitiert“, die als Qualitätsindikator für wissenschaftliche Arbeiten gelten, mit Hilfe der Datenbank schnell beantwortet werden können. Die Datenbank ist auf CD-ROM erhältlich sowie u.a. bei DIMDI und DataStar on-line abrufbar.

Ein Vokabular existiert in der Form von Autorenschlagwörtern.

4 METHODIK

Eine Lösung für die Problemstellung – eine interaktive Analyse und Verdichtung von MEDLINE-Resultaten – sollte universal sein, möglichst kostengünstig wie in PubMed und ohne viel Aufwand beim Benutzer zu installieren und von ihm interaktiv zu bedienen sein. Außerdem wurde bei der Entwicklung der Lösung ein Augenmerk gelegt auf die Möglichkeit zur

- Verdichtung von bibliographischen Feldern
- Darstellung von Häufigkeiten von bibliographischen Feldern
- Darstellung von Korrelationen von bibliographischen Feldern
- Darstellung der zugehörigen Schlagworthierarchie und
- bequemen iterativen Verfeinerung des ursprünglichen Suchradius und damit einer Erhöhung der Relevanz bzw. der Nachweisquote.

Diese Lösung mit dem Namen *Meva* (*MEDLINE Evaluator*) wurde als [Webdienst](#) implementiert.

Das Kapitel Methodik stellt zuerst *Meva* vor und gibt anschließend anhand der Netz-, Server-, Client- und Entwicklungsarchitektur einen Überblick über Aufbau und Funktionsweise der zugrundeliegenden Systeme.

4.1 Einführung in Meva

Zur Lösung der in der Problemstellung genannten Problematik – eine interaktive Analyse und Verdichtung von MEDLINE-Resultaten – wurde ein Webdienst namens *Meva* (*MEDLINE Evaluator*) entwickelt.

Meva ist ein MEDLINE-Postprozessor: Es setzt auf einem Suchresultat auf, das der Anwender über eine Suche in PubMed gewonnen hat und anschließend an Meva zur Auswertung geschickt hat. Mit Meva können Anzahl, Vorkommen und Beziehungen bibliographischer Felder in einem PubMed-Suchresultat analysiert und graphisch verdichtet dargestellt werden.

Der Ablauf einer Meva-Konsultation gestaltet sich wie folgt:

1. Der Benutzer ruft das PubMed-Suchformular auf und führt darüber eine Suche in MEDLINE durch.
2. Er speichert die Suchergebnisse als Datei im MEDLINE-Format auf seinen Rechner ab.
3. Anschließend ruft der Benutzer das Meva-Formular auf, trägt den Pfad zur abgespeicherten PubMed-Datei ein, wählt die zu analysierenden Suchfelder aus und setzt optional weitere Parameter zur Steuerung der Analyse: so kann er u.a. Suchfilter oder Minimalhäufigkeiten spezifizieren, um den Suchradius und damit Mevas Analyseresultat weiter einzuschränken. Auch eine Beschränkung der Analyse auf Erstautoren ist möglich. Falls gewünscht, können zusätzlich Codes zu MeSH-Termen und weiterführende Suchverknüpfungen zu PubMed ausgegeben werden.
4. PubMeds Suchergebnisse stellt Meva dar als Histogramme und Korrelationstabellen der bis zu 50 häufigsten Ausprägungen, als sortierte Detailliste vergesellschafteter Felder und als MeSH-Baum. Der Benutzer kann sich die Ergebnisse wahlweise im Hypertextformat oder im datenbanktauglichen Textformat liefern lassen, wenn er eine Weiterverarbeitung der Daten wünscht. Wenn die Felder des Meva-Resultates verlinkt sind, löst er durch ihr Anklicken eine weitere Suche in PubMed aus und kann dann entweder mit Punkt 2 fortfahren (erneute Analyse durch Meva) oder sich die Abstracts bzw. den Volltext der gefundenen Artikel in PubMed ausgeben lassen.

Des weiteren können zusätzliche Einstellungen durch den Nutzer im Meva-Formular getätigt werden, die im Kapitel 5 (Ergebnisse) detailliert werden.

Abb. 4 illustriert die interagierenden Module und die beteiligten Datenflüsse, die einer Meva-Konsultation zugrunde liegen:

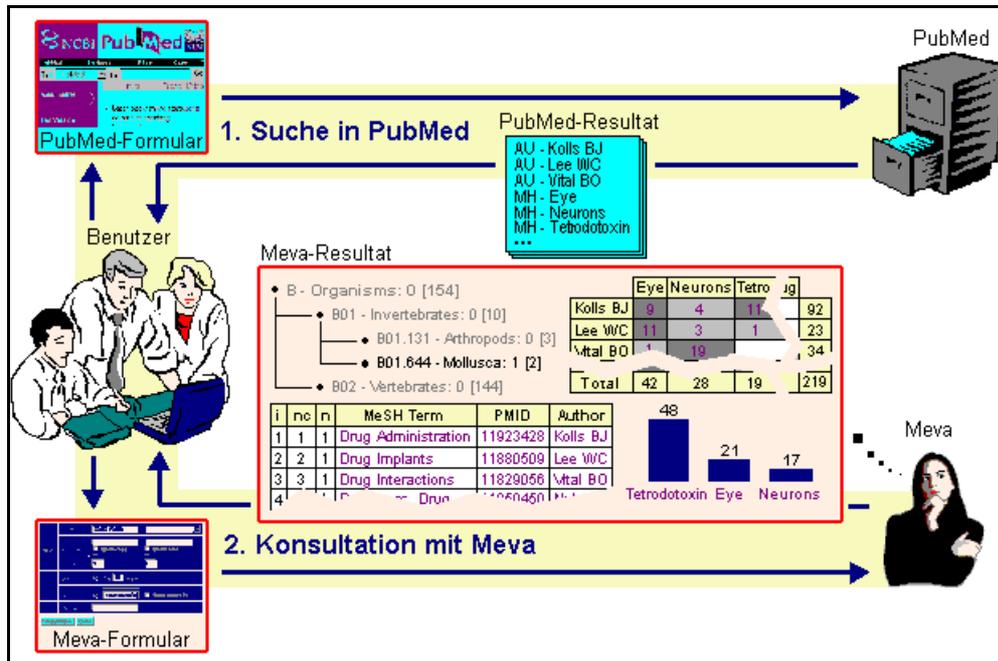


Abb. 4: Operative Architektur von Meva

Die Architektur der beteiligten Systeme soll in den nächsten Sektionen näher erläutert werden.

4.2 Netz-Architektur

4.2.1 Modell

Da Meva auf einem Webserver und nicht lokal auf den Rechnern der Benutzer läuft, wird im folgenden die Bereitstellung von Mevas Funktionalität auch als Dienst oder [Service](#) bezeichnet.

Meva beruht auf dem [Client-Server-Modell](#), bei dem die Funktionalität auf Arbeitsplatz und Server aufgeteilt ist. Dies bringt entscheidende Vorteile mit sich:

- Es wird eine Lasttrennung der Ressourcen des Arbeitsplatzrechners und der des Servers ermöglicht.
- Durch die Vorhaltung des Dienstes auf einem Webserver kann Mevas Funktionsweise zentral und einheitlich administrativ parametrisiert werden.
- Die Installation von Hunderten Meva-Instanzen für die Benutzer entfällt – es ist nur ein [Thin Client](#) auf den Arbeitsplatzrechnern erforderlich, in diesem Fall ein Webbrowser. Da Browser in der heutigen Zeit schon einen obligaten Bestandteil von PC-Betriebssystemen bilden und somit ubiquitär verbreitet sind, reduziert sich die Einführung des Dienstes auf die Installation der Server-Komponente.

4.2.2 Propagierung im Internet

Durch die Bereitstellung über das HTTP-Protokoll wird der Dienst im WWW (World Wide Web) verfügbar. Neben der technischen Bereitstellung ist es jedoch auch nötig, den Dienst im WWW zu propagieren, damit er überhaupt gefunden wird. Ist die URL der Webpräsenz noch keinem potentiellen Benutzer bekannt, kann sie naturgemäß nicht durch Eintippen in die Adreßzeile eines Browsers gefunden werden.

Anfang der 90er Jahre verwendeten Webbenutzer noch oft Webverzeichnisse, um an die gewünschten Informationen zu gelangen. Heute finden die meisten Benutzer Informationen, indem sie Suchmaschinen verwenden. Daher wurden zwei Strategien bei der Propagierung verfolgt:

4.2.2.1 Eintragung in Webkataloge

Der Dienst wurde unter der entsprechenden URL bei vier großen Webkatalogen angemeldet (Open Directory Project, Yahoo, Lycos, Witch), teilweise für Deutsch und Englisch separat, sofern entsprechende Kategorien verfügbar waren.

4.2.2.2 Propagierung bei Suchmaschinen

Der Dienst wurde unter der entsprechenden URL bei drei großen Suchmaschinen angemeldet (Google, Altavista, Lycos).

Als Nebeneffekt ist zu beachten, daß die o.g. Eintragung in Kataloge eine indirekte Propagierung bei vielen assoziierten Suchmaschinen vornimmt, die erfolgreicher ist als eine Direktbewerbung: so wertet Google z.B. die Eintragung einer Webpräsenz in einen Katalog wie dem Open Directory Project entsprechend hoch, so daß sich das Page Ranking dieser Präsenz in Google deutlich erhöht. Dies wird verständlich, wenn man bedenkt, daß Katalogeinträge immer mit Hilfe menschlicher Bewertung vorgenommen werden, Anmeldungen bei Suchmaschinen hingegen automatisch erfolgen.

Um die Bewertung der Präsenz durch die Suchmaschinen zu verbessern und damit die Auffindbarkeit im Web, wurden der Erstellung der Webpräsenz einige Kriterien [↘ [Schallhorn](#)⁴] zugrundegelegt, die nachfolgend nur auszugsweise beschrieben werden:

- Auf den Einsatz visueller HTML-Editoren wurde verzichtet, da diese bestenfalls als drittclassig einzustufenden HTML-Quelltext generieren und Suchmaschinen somit oft keine relevanten Texte extrahieren können. Die manuelle oder semimanuelle Erstellung wirkt sich hingegen positiv auf das *Page Ranking* der Suchmaschinen aus.
- Es wurde fast nur *HTML 4.01 Strict* [↘ [HTML 4.01](#)⁵], *Cascading Style Sheets 1* [↘ [CSS 1](#)⁶] und *2* [↘ [CSS 2](#)⁷] bei der Webseitenerstellung verwendet. Durch die Vermeidung von physischen HTML-Formatbefehlen und blinden Tabellen konnte eine Entkoppelung von Inhalt und Form erreicht werden. Viele Suchmaschinen honorieren diese klare Trennung mit einem verbesserten *Page Ranking*. Zudem sind so gestaltete HTML-Seiten auch für Behinderte besser verständlich.
- Auf Frames als Seitengestaltungsmittel wurde aus denselben Gründen verzichtet.
- HTML- und CSS-Syntax der Dateien wurden mit den W3C-Validatoren verifiziert.
- Webseiten-Dubletten wurden entfernt, da Suchmaschinen diese häufig als

⁴ <http://www.kso.co.uk/>

⁵ <http://www.w3.org/TR/html401/>

⁶ <http://www.w3.org/TR/1999/REC-CSS1-19990111>

⁷ <http://www.w3.org/TR/REC-CSS2/>

Indiz für Spam ansehen.

- Die Wortdichte der Schlüsselwörter wurde überprüft, da Dichtewerte über 12% von vielen Suchmaschinen als verdächtig (Spam) eingestuft werden.
- Schlüsselwörter wurden im Text logisch ausgezeichnet.
- Die wichtigsten Schlüsselwörter wurden in den ersten drei Sätzen jeder Datei abgelegt.
- Navigationssektionen wurden physisch an das Dateieneende gesetzt.
- Dateinamen wurden aus maschinell extrahierbaren Suchbegriffen aufgebaut, die Bestandteile zusammengesetzter Dateinamen wurden dementsprechend mit Bindestrichen gekoppelt.
- Wo sinnvoll, wurden die Verweistexte ident zu ihren URIs gehalten.
- Die Verschachtelungstiefe der Dateien wurde klein gehalten.
- Eine Seitenbeschreibung und eine Schlüsselwortliste (nicht mehr als 12 Schlüsselwörter) pro HTML-Datei wurden erstellt und in deren Kopfbereich eingetragen (viele Suchmaschinen lesen diese Informationen heute jedoch nicht mehr aus).
- Der Titel der Dateien wurde nachträglich noch einmal sorgfältig dem Inhalt angepaßt, da er von Suchmaschinen hoch bewertet wird.

4.3 Server-Architektur

Meva ist nach den Spezifikationen des [Common Gateway Interface](#) Version 1.1 [[↗ CGI 1.1](#)⁸] geschrieben. Der Webserver behandelt jeden *Client Request*, welcher einen URI (*Uniform Resource Identifier*) auf ein CGI-Verzeichnis anfordert, als ausführbares Programm. Meva wird als *Parsed Header CGI* aufgerufen, d.h. es gibt dem Webserver nur die minimal benötigten Kopfzeilen zurück, die der Webserver dann entsprechend mit eigenen Kopfzeilen ergänzt. Dieser Vorgang passiert transparent für den Benutzer; es gibt für den Benutzer keinen Unterschied, ob er mit der Anforderung eines URIs die Lieferung einer statischen HTML-Seite oder einen Programmablauf (der ebf. eine HTML-Resultat generieren kann) anstößt.

4.3.1 Portabilität

Als Betriebssystem kommt prinzipiell jedes System in Frage, auf dem ein ANSI-C-Compiler und ein Webserver lauffähig sind. Getestet wurden Windows NT, Windows 2000 und Linux. Da bei der Programmierung darauf Wert gelegt wurde, vom zugrundeliegenden Betriebssystem zu abstrahieren und systemspezifische Gegebenheiten wie Bytefolge (*Little Endian*, *Big Endian*) oder Datengröße nicht berücksichtigt wurden, sollte das Programm – evt. mit einigen Modifikationen – auf jedem der o.g. Systeme lauffähig sein.

Durch den Einsatz von [Assembler](#)-Code könnte zwar durch Einsparung von Prozessortaktzyklen eine Laufzeitoptimierung von ca. 30% - 50% erreicht werden [[Monadjemi 1991](#), S. 758], dies würde aber ebf. zum Verlust der Portabilität auf Quellcode-Ebene führen, da der erzeugte Code maschinenabhängig wäre.

Als Webserver kann jedes Produkt verwendet werden, das den CGI-Standard Version 1.1 und den MIME-Typ (*Multimedia Internet Mail Extensions*) *multipart/form-data* versteht [[Kientzle 1996](#), S. 291]. Jeder voll implementierte HTTP/1.1-Server beherrscht dies. Ältere Webserver (HTTP/1.0 oder HTTP/0.9) oder nur minimal implementierte Webserver eignen sich hingegen nicht für Meva. Getestet wurden die Webserver Apache und OmniHTTPD.

4.3.2 Performanz

Prinzipiell gilt: über je mehr Hauptspeicher und über je mehr ausgefeilte *Caching*-, *Multitasking*- und Speicherauslagerungs-Strategien, wie z.B. *Paging* [[Blieberger et al. 1992](#), S. 251] oder *Swapping* [a.a.O., S. 254], ein System verfügt, desto schneller laufen seine Applikationen. Meva profitiert davon gleich mehrfach: einerseits müssen große Datenfelder im Speicher aufgebaut werden (abhängig vom

⁸ <http://www.w3.org/CGI/>

Volumen der zu verarbeitenden Daten bis zu 15 MB), andererseits greift das Programm auch mehrfach auf die lokale Platte zu, um Index- oder Konfigurationsdateien zu lesen. Meva unterstützt allerdings kein *Multithreading*.

Dennoch verläßt sich Meva nicht vollständig auf eine optimale Umgebung: da Betriebssystemaufrufe und Dateizugriffe je nach System zeitwierig ausfallen können, puffert Meva einerseits Dateizugriffe und bündelt andererseits externe Systemaufrufe für die Speicherverwaltung, um Zeit zu sparen (*Granulated heap allocation*). Außerdem werden nur hochperformante Routinen (s.a. Such- und Sortierverfahren) eingesetzt, deren Komplexität unter quadratischen Werten liegt. Somit sollte die Antwortzeit auch auf langsamen Servern oder unter Vollast unter einer Minute liegen.

Die Performanz von Meva hängt allerdings auch von der Performanz des Webbrowsers und von der Anbindung des *Clients* an das Netz ab: Verfügt der Benutzer nur über ein langsames System oder fällt die Netzanbindung des Benutzers niedrig aus, wie z.B. bei einem langsamen Einwählzugang, transportiert der Webserver die Daten entsprechend auch nur langsam zu Meva.

Eine Entlastung des Datenaufkommens (und damit des Netzes) beim Senden der Resultate konnte durch den konsequenten Einsatz von CSS1 und CSS2 (*Cascading Style Sheets*) erreicht werden [[Nielsen](#)⁹].

Summa summarum bestimmen somit drei Faktoren Mevas Laufzeiten:

- Performanz des *Servers*
- Performanz des *Clients*
- Transportgeschwindigkeit des *Netzes*

Das Gesamtsystem ist dabei nur so schnell wie das langsamste Glied in der Kette.

4.3.2.1 Weitere Verbesserungen

Eine Möglichkeit, serverseitig die Performanz noch weiter zu erhöhen, besteht zweifelsohne im Aufruf von Meva als *Non parsed header CGI (NPH)*. Hierbei interagiert das Programm ohne Zwischenschaltung des Servers direkt mit dem Client, d.h. es muß auch den vollen Satz an Kopfzeilen zurückliefern. Durch das Auslassen des Parsens durch den Webserver erniedrigt sich die Reaktionszeit. Da aber einige Webserver NPH nicht unterstützen, wurde diese Möglichkeit nicht umgesetzt.

⁹ <http://www.w3.org/Protocols/HTTP/Performance/Pipeline>

4.4 Client-Architektur

4.4.1 Benutzeroberfläche

Der Benutzer interagiert mit Meva über ein Webformular, das vom Server über einen Webbrowser abgerufen werden kann. Hier können alle Parameter, die für eine korrekte Verarbeitung benötigt werden, vom Benutzer eingestellt werden. Beim Absenden des Formulars werden Parameterwerte nebst zu analysierenden Daten an den Webserver geschickt, der sie an Meva weiterleitet. Bei der Erstellung des Formulars wurde darauf geachtet, gängige Forderungen der Ergonomie (Intuitivität, Ermüdungsfreiheit der Oberfläche, Funktionalität, Benutzerführung, kontextsensitive Hilfe zu Arbeitsschritten) einzuhalten.

Neben dem Formular – der eigentlichen Interaktionsfläche mit Meva – wurde eine komplette *Web Site* erstellt, in die sich das Formular nahtlos einbettet. Für den Benutzer stehen Hilfeseiten, FAQs (Antworten zu häufigen Fragen) und Hintergrundinformationen zum Service und zur Entwicklungsgeschichte zur Verfügung.

4.4.2 Benutzerfreundlichkeit

Die nachfolgenden Punkte illustrieren nur einige der angelegten und überprüften Kriterien für Benutzerfreundlichkeit von Webseiten:

4.4.2.1 Direktes Editieren des HTML-Quelltextes

Wie schon vorher bei der Diskussion der Eignung für Suchmaschinen besprochen, wurde auf den Einsatz visueller HTML-Editoren verzichtet. Dadurch wurden die HTML-Dateien um ein Vielfaches kleiner, als sie visuelle Editoren erzeugen. Die Folge sind verkürzte Ladezeiten, Entlastung des Netzwerkes und eine bessere Verfügbarkeit.

4.4.2.2 Entkoppelung von Inhalt und Form

Durch den ausschließlichen Einsatz von *HTML 4 Strict*, *CSS 1* und *CSS 2* bei der Webseitenerstellung konnte eine Entkoppelung von Inhalt und Form erreicht werden. Das Design wurde ausschließlich mit CSS erstellt, in HTML wurde lediglich logisch ausgezeichnet. Dies bewirkte eine einheitlichere Darstellung in den Browsern und eine verbesserte Ergonomie und Anpaßbarkeit des Designs durch den Benutzer mit benutzerdefinierten *Style Sheets*. Letzteres ist enorm wichtig für Farbschwache oder Farbblinde (s.a. Punkt [4.4.2.11](#) - Eignung für behinderte Benutzer), aber auch für Benutzer, die das Web gerne in ihren „eigenen“ Farben explorieren wollen.

4.4.2.3 Verifikation der Syntax

Die Syntax der Webseiten wurde mit entsprechenden Validatoren überprüft, die das W3C zur Verfügung stellt [[↘ HTML-Validator¹⁰](#); [↘ CSS-Validator¹¹](#)]. Ein korrekter Quellcode führt zu einer fehlerfreien Darstellung beim Benutzer. Bei den meisten Seiten im World Wide Web ist der Quellcode leider nicht korrekt.

4.4.2.4 Ausschluß von Werbung

Der Einsatz von Werbebannern verbietet sich selbstredend (und war auch nie intendiert). Im Gegensatz zu den Aussagen einiger selbsternannter „Marketingstrategen“ dulden Benutzer Werbung nur ungern und nur, wenn sie nicht anders können, z.B. weil sie nach einem Produkt auf einer Seite mit eingblendeter Werbung suchen. Besteht jedoch Wahlfreiheit, verlassen die meisten Benutzer werbeanimierte Seiten frühzeitig wieder, viele schon nach zwei bis drei Sekunden, wie die Auswertung der Logdateien der Webserver belegen.

4.4.2.5 Orthographie und Syntax

Orthographie und Syntax aller Texte wurden händisch und maschinell überprüft. Auch hier gilt, daß die meisten Webpräsenzen dies versäumen.

4.4.2.6 Lesbarkeit

Die Startseite besitzt eine überragende Bedeutung für Benutzer. Häufig wird die Startseite als erste Seite dem Benutzer präsentiert. Laut den Statistiken der Webserver entscheiden sich die meisten Benutzer relativ rasch, oft schon in zwei bis fünf Sekunden, ob sie auf einer Seite bleiben oder nicht. Als Blickfänger für potentielle neue Benutzer verdient daher die Startseite besondere Beachtung bei der Erstellung einer Webpräsenz.

Daher wurde speziell die Lesbarkeit der Startseite überprüft und optimiert. Die Bewertung erfolgte mit dem Lesbarkeitsgrad nach Flesch [[Klare 1974](#), S. 62 - 102] entsprechend Tab. 3 anhand der durchschnittlichen Anzahl von Silben pro Wort sowie Wörtern pro Satz.

¹⁰ <http://validator.w3.org/>

¹¹ <http://jigsaw.w3.org/css-validator/>

Quote	Lesbarkeit
90 – 100	Sehr einfach
80 – 90	Einfach
70 – 80	Relativ einfach
60 – 70	Standard
50 – 60	Relativ schwierig
30 – 50	Schwierig
0 – 30	Sehr schwierig

Tab. 3: Lesbarkeitsgrad nach Flesch

Die Formel für den Lesbarkeitsgrad nach Flesch lautet

$$206,835 - (1,015 \times \text{DSL}) - (84,6 \times \text{DSW})$$

mit DSL als durchschnittliche Satzlänge (die Anzahl der Wörter geteilt durch die Anzahl der Sätze) und DSW als durchschnittliche Anzahl der Silben pro Wort (die Anzahl der Silben geteilt durch die Anzahl der Wörter).

Durch die Optimierung konnte der Lesbarkeitsgrad der Startseite von initial drei auf 21 erhöht werden.

Vielleicht ist der Leser nun irritiert, zumindest der Autor war es, scheint die Lesbarkeit doch mit 21 lt. obiger Tabelle immer noch sehr gering zu sein. Daher ist anzumerken, daß die Formel nach Flesch einerseits sehr einfach gestrickt ist, berücksichtigt sie doch nur Silben- und Wortzahlen, und andererseits wissenschaftliche Texte naturgemäß komplexerer Natur sind. Zudem zog die gezielte Positionierung von Schlüsselwörtern für Suchmaschinen in den einleitenden Sätzen eine weitere Verschlechterung der Lesbarkeit nach sich.

Um die Relativität von Fleschs Formel aufzuzeigen, erscheinen nachfolgend vier Varianten einer Textpassage mit ihren unterschiedlichen Lesbarkeitsgraden. Der Autor hat sich bemüht, die inhaltliche Deckungsgleichheit der Passagen beizubehalten:

Originale Textpassage aus der Startseite. Flesch = 0 „sehr schwierig“:

Mit Meva können Anzahl, Vorkommen und Beziehungen der bibliographischen Felder im PubMed-Resultat in einer graphisch verdichteten Form dargestellt werden, optional können Schlagwörter hierarchisch als Baum (MeSH Tree) ausgegeben werden. Sofern gewünscht, kann sich der Benutzer alternativ Teile der Meva-Analyse in textueller Form liefern lassen, die für eine Weiterverarbeitung in einer Datenbank geeignet ist.

Aktuelle, optimierte Textpassage aus der Startseite. Passiva entfernt, Text simplifiziert, Thema optisch hervorgehoben. Flesch = 36 „schwierig“:

Was gibt Meva aus? Meva stellt Anzahl und Beziehungen der Felder graphisch verdichtet dar. Schlagwörter kann es hierarchisch als Baum (MeSH Tree) ausgeben. Alternativ liefert es Teile der Analyse in Textform, die der Benutzer in eine Datenbank zur Weiterverarbeitung importieren kann.

Dieselbe Textpassage aus der Startseite, stärker „optimiert“. Flesch = 56 „relativ schwierig“:

Was gibt Meva aus? Anzahl und Beziehungen der Felder werden geprüft. Meva erzeugt daraus Graphiken. Schlagwörter kann es als Baum zeichnen. Es kann auch einen Teil als Text liefern. Den kann der Benutzer in eine Datenbank einlesen. Dort kann er ihn weiterverarbeiten.

Dieselbe Textpassage aus der Startseite, zusätzlich „optimiert“. Flesch = 86 „einfach“. Inhaltliche Entstellungen waren leider nicht mehr zu vermeiden:

Was macht Meva? Es prüft viel. Dann macht es Bilder daraus. Es kann einen Baum aus Wörtern zeichnen. Oder es liefert einen Teil als Text. Den kann man in eine Bank für Daten stecken. Dort kann man mit ihm danach arbeiten.

Ein Beispiel für Flesch „Sehr einfach“ bleibt dem Leser an dieser Stelle erspart.

Wie man sehen kann, sollte man nur sehr kurze Sätze bilden und tunlichst jedem mehrsilbigen Wort aus dem Wege gehen, um gemäß Flesch als gut lesbar zu gelten. Der Absatz beispielsweise, den Sie gerade durchlesen, hat einen Flesch-Lesbarkeitsgrad von 44 und gilt somit als „schwierig“. Der Autor hofft dennoch auf Ihr wohlwollendes Verständnis.

Fazit: Für Fachthemen kann man Fleschs Formel nicht anwenden, da das entsprechende Fachvokabular eine *Conditio sine qua non* darstellt. Zudem steht zu befürchten, daß man mit „einfachen“ Texten wie im letzten Kasten wahrscheinlich eine andere Klientel anspricht als man sich erhofft hatte. Fleschs Formel ist hingegen als Mittel zur Abschätzung der Verbesserung bzw. Verschlechterung der Lesbarkeit eines Textes gegenüber seinem Urtext bedingt verwendbar.

4.4.2.7 Einsatz von Druckformaten

Durch den Einsatz von CSS konnten Druckformate implementiert werden, welche die Benutzerfreundlichkeit deutlich erhöhen halfen. Statt eines oft auf Webseiten zu findenden Hinweises „Hier finden Sie die Druckversion dieser Seite...“ und zusätzlichen Ladens der druckoptimierten Seitenversion reicht nun ein simpler Ausdruck derselben Seite. Zusätzlich werden somit im Ausdruck Seitenelemente ausgeblendet, die nur für das Betrachten der Seite am Bildschirm notwendig sind, nicht aber für den Ausdruck, wie z.B. Navigationselemente.

4.4.2.8 Reduktion der Speichergröße

Alle verwendeten Bilder und HTML-Seiten wurden auf eine geringe Speichergröße optimiert und die Ladezeiten dadurch drastisch reduziert. Dies ist ein wesentliches Kriterium für Benutzerfreundlichkeit in Netzwerken.

4.4.2.9 Konsistente Navigation

Da nur die wenigsten Webbenutzer erfolgreich durch Seiten mit einer komplexen netzartigen Verlinkung navigieren können und sich schnell das frustrane Gefühl des *Lost in Hyperspace* einstellt, wurde Wert auf eine konsistente und lineare Menüführung in Form einer einfachen Menüliste am Anfang jeder Seite gelegt. Die maximale Verschachtelungstiefe beträgt somit eins. Alle Seiten wurden zudem klar in Sektionen gegliedert.

4.4.2.10 Verzicht auf dynamisches HTML

Auf den Einsatz von dynamischen HTML wurde komplett verzichtet (Flash, JavaScript, Java). Zum einem deshalb, da diese Bestandteile nur unnötigen Ballast mit sich bringen, aber keine oder kaum Funktionalität; zum anderen, da ca. 30% der Webnutzer lt. Statistiken der Webserver ohnehin diese Technologien in ihrem Browser zum Zwecke eines schnelleren und ungestörteren Surfens ausgeschaltet haben.

JavaScript wird lediglich als zusätzliche Möglichkeit der Formularprüfung angeboten. Dies ist sinnvoll, da es das Netzwerkaufkommen reduzieren hilft. Das Formular funktioniert jedoch auch ohne JavaScript.

4.4.2.11 Eignung für behinderte Benutzer

Es wurde darauf geachtet, die Webpräsenz auch für blinde oder anderweitig behinderte Personen entsprechend den WCAG-Konformitätsrichtlinien des W3C [[↗ WCAG¹²](#)] ergonomisch zu gestalten. Diese Richtlinien legen die Eignung einer

¹² <http://www.w3.org/TR/WAI-WEBCONTENT/>

Webpräsenz für Behinderte fest. Die Richtlinien unterscheiden dabei drei Stufen: Stufe A, AA und AAA. Jeder Webentwickler muß lt. Richtlinien Stufe A, sollte Stufe AA und kann Stufe AAA implementieren. Jede Stufe inkludiert die Anforderungen ihres Vorgängers und fügt neue hinzu, AA inkludiert also A etc.

Zum Test der Eignung wurden mehrere Läufe mit einem Prüfprogramm [[↘ Bobby](#)¹³] ausgeführt, welches die Eignung von Webseiten für Behinderte mißt.

Da heutzutage jedoch noch keiner der verbreiteten Browser den HTML-4-Standard komplett und korrekt implementiert hat, ließen sich Kompromisse zwischen den Lösungen für „Sehende“ und „Nicht-Sehende“ nicht immer vermeiden. Aus diesem Grund konnte nicht für alle Seiten der Webpräsenz der höchste Standard (AAA) für Behinderte entsprechend den Richtlinien implementiert werden.

4.4.2.12 Clientanforderungen

Die Webseiten wurden in HTML Version 4.01 [[↘ HTML 4.01](#)¹⁴] und CSS Version 1 [[↘ CSS 1](#)¹⁵], partiell auch in CSS Version 2 [[↘ CSS 2](#)¹⁶] erstellt. Aktuelle W3C-konforme (*World Wide Web Consortium*) Browser verstehen diese Sprachen. Unter älteren Browsern bleiben die Seiten lesbar, nur einige Formatierungen gehen verloren. Getestet wurden *Internet Explorer* Version 4, 5 und 6, *Netscape Navigator* Version 4 und 6, *Opera* Version 5, 6 und 7 sowie *Konqueror*. Anzumerken ist, daß keiner dieser Browser HTML V. 4 und CSS komplett implementiert. Deshalb wurde auf den Einsatz von Formaten verzichtet, die von den aktuellen Browsern nur mangelhaft oder gar nicht unterstützt werden.

Ebf. Voraussetzung für die Benutzung des Meva-Dienstes ist ein Webbrowser, der formularbasierten Datei-Upload gemäß den *Request for Comments* [[↘ RFC 1867](#)¹⁷; [↘ RFC 2388](#)¹⁸] unterstützt. Dies ist die Regel bei allen HTTP/1.1-Clients [[↘ RFC 2616](#)¹⁹] und bei vielen HTTP/1.0-Clients [[↘ RFC 1945](#)²⁰], nicht jedoch bei alten HTTP/0.9-Clients [[↘ HTTP/0.9](#)²¹].

¹³ <http://bobby.watchfire.com/>

¹⁴ <http://www.w3.org/TR/html401/>

¹⁵ <http://www.w3.org/TR/1999/REC-CSS1-19990111>

¹⁶ <http://www.w3.org/TR/REC-CSS2/>

¹⁷ <http://www.ietf.org/rfc/rfc1867.txt>

¹⁸ <http://www.ietf.org/rfc/rfc2388.txt>

¹⁹ <http://www.ietf.org/rfc/rfc2616.txt>

²⁰ <http://www.ietf.org/rfc/rfc1945.txt>

²¹ <http://www.w3.org/Protocols/HTTP/AsImplemented.html>

4.4.2.13 Mehrsprachigkeit

Bei der Erstellung der Webseiten wurde Wert auf Mehrsprachigkeit gelegt, die Seiten sind in Deutsch und Englisch verfügbar. Dies trifft auch für Meva selbst zu. Die Sprachauswahl wird automatisch zwischen Client und Server ausgehandelt. Voraussetzung dafür ist, daß der Webserver *Content Negotiation* beherrscht (wie z.B. *Apache*) und der Client eine korrekte Sprache im Webbrowser eingestellt hat.

Letzteres ist häufig nicht der Fall, die Default-Installationen manch gängiger Browser, wie z.B. des Internet Explorers, tragen in Unkenntnis oder Ignoranz des entsprechenden RFC-Protokolls [[↘ RFC 2616](#)²²] häufig landesspezifische statt generischer Sprachen ein, z.B. *en-US* für US-Amerikaner oder *de-DE* für Deutsche, indem sie blind die im Betriebssystem eingestellte Sprache übernehmen. Nun verfügt aber kaum eine Webpräsenz über landessprachenspezifische Webseiten, allenfalls über sprachspezifische Seiten wie Englisch (*en*) oder Deutsch (*de*). Das ist auch verständlich: Keine Webpräsenz nimmt die Mühe auf sich, für Schweizer, Österreicher, Deutsche, Liechtensteiner etc. unterschiedliche deutsche Versionen vorzuhalten (die Unterschiede lägen in sprachlichen Feinheiten, Währungssymbolen und anderen Einheiten). Die Folge ist, daß wenn ein deutscher Webbrowser mit dem Eintrag *de-DE* eine solche Webpräsenz aufsucht, ein RFC-konformer Webserver nicht seine vorhandene deutsche Seite zurückliefert (wie man zu Recht annehmen könnte), sondern korrekt entsprechend dem RFC einen Client-Fehler nebst Auswahlmenü an verfügbaren Sprachen. Die meisten Webbenutzer sind damit überfordert und interpretieren diese Meldung als Serverfehler, indessen liegt der Fehler jedoch in ihren Browser-Einstellungen.

Lösungsansätze für dieses Dilemma werden in der Sektion weiter unten besprochen, die sich mit Problemen bei der Arbeit mit Meva beschäftigt.

Meva und die Webpräsenz unterstützen momentan Englisch und Deutsch als Auswahl Sprachen; andere Browsersprachen werden mit Englisch beantwortet.

²² <http://www.ietf.org/rfc/rfc2616.txt>

4.5 Entwicklungs-Architektur

4.5.1 Programmiersprache

Meva wurde in C programmiert. Die Wahl von C erfolgte aus folgenden Gründen:

- C ist eine der am häufigsten verwendeten Programmiersprachen (Pflege).
- Die erzeugten Programme sind durch die Maschinennähe hochperformant.
- Die Zergliederung des Programms ist sowohl auf Quellcode- als auch auf Modulebene möglich.
- Die Portabilität des Quellcodes ist hoch, sofern man sich an den [ANSI](#)- (American National Standards Institute) bzw. [ISO](#)-Standard (International Standards Organization) von C hält.

Als Alternativen wäre [PERL](#) (*Practical Extraction and Reporting Language*) denkbar gewesen oder andere serverbasierte Technologien wie [JSP](#) (*Java Server Pages*) bzw. [Servlets](#). Gerade letztere werden momentan sehr stark im Bereich des *E-Commerce* im Internet für Logins, Warenkörbe, Suchvorgänge und Benutzerbewertungen eingesetzt [[Eberhart et al. 2000](#), S. 135].

PERL ist nicht die bevorzugte Sprache, um darin Tausende Zeilen zu schreiben, und ist als interpretierte Sprache notwendigerweise langsamer. Dynamisch ladbare PERL-Module, wie es sie für den Apache-Webserver u.a. gibt, fangen dieses Manko etwas auf [[Tischer et al. 1997](#), S. 974]. Zudem leidet PERL an seiner kryptischen Notation.

Gegen JSP und Servlets spricht die Notwendigkeit, den Webserver mit dieser Technologie aufzurüsten. Als normale Serverapplikation nach dem [CGI](#)-Standard hingegen braucht man nur das Programm auf den Webserver aufzuspielen, um die gewünschte Dienstleistung zu implementieren.

Da die CGI-Schnittstelle nicht auf eine bestimmte Skriptsprache festgelegt ist, konnte weitgehend plattformunabhängig programmiert werden [[Brantner et al. 2001](#), S. 28]. Die Verwendung der CGI-Technologie hat jedoch auch Nachteile, da der Webserver für jede CGI-Anfrage einen eigenen Prozeß erzeugt, was sich negativ auf die Verfügbarkeit des Servers auswirken kann. Für den gleichzeitigen Zugriff sehr vieler Benutzer ist Meva somit nicht gedacht.

Alle Quelltexte von Meva wurden zum Zwecke einer hohen Portabilität im ANSI-Modus mit allen Warnungen kompiliert.

4.5.2 Compiler

Es wurden Compiler unter den Betriebssystemen Windows NT, Windows 2K und Linux verwendet, um die Portabilität sicherzustellen, im einzelnen:

- Borland C++ TM V. 5.5.1
- Bloodshed Dev-C++ V. 4.9.6.0
- GNU-C-Compiler egcs V. 2.91.66
- Logiciels/Informatique lcc-win32 V. 3.8
- Microsoft Visual Studio TM V. 6.0

4.5.3 Entwicklungskonzept

Die Entwicklung von Meva erfolgte nach dem Prinzip der strukturierten Programmierung. Der Code wurde in aufgabenbezogene Module zerlegt und die korrespondierenden Daten entsprechend modular gekapselt.

Die Nachteile des Strukturansatzes:

- Datenabhängigkeiten
- nicht-lokale Sprünge, z.B. zur Fehlerbehandlung
- Datenmodellierung

wurden wie folgt behandelt:

Datenabhängigkeiten wurden – wo es ging – vermieden, um [Seiteneffekte](#) auszuschließen. Die Bindung von Daten und Code wurde nach dem Prinzip der [lockeren Kopplung](#) vorgenommen. Eine nicht-lokale Fehlerbehandlung wurde durch Routinen der Sprache C teilweise zur Verfügung gestellt. Die Modellierung komplexer Daten (Tabellen etc.) wurde durch Verbunddatentypen erreicht, dem Analogon zu Objekten in der objektorientierten Programmierung.

4.5.4 Entwicklungsphasen

Aufgrund der steten Änderungen der Anforderungen der Benutzer an das Programm im Entwicklungszeitraum war ein linearer **Software Lifecycle** wie im *Wasserfallmodell* nicht realisierbar [↘ [Schiedermayer](#)²³]. Oft mußten grundlegende Neuformulierungen des Designs vorgenommen werden, was sich in einer verlängerten Entwicklungszeit niederschlug. Durch den zirkulären Charakter des Phasenablaufs bedingt läßt sich Mevas Entwicklung am besten durch ein *Spiralmodell* beschreiben. Im Spiralmodell folgen die Phasen

- Problem
- Analyse
- Entwurf
- Implementierung
- Programm
- Test
- Wartung

nicht linear aufeinander, sondern wiederholen sich mehrfach. Mit jedem Umlauf nähert man sich dem angestrebten Endprodukt näher an, bis letztendlich der Aufwand für die nächste Runde den zu erwartenden Fortschritt überwiegt.

4.5.5 Verwendete Algorithmen

In kurzer Zeit nach bestimmten Daten zu suchen, gehört zu den wichtigsten Aufgaben vieler Programme. In der Informatik gibt es viele, gut analysierte Algorithmen dafür. Großrechner sind bis zu 25% ihrer CPU-Zeit mit Suchen und Sortieren von Daten beschäftigt [[Hermann 1992](#), S. 279]. Für Meva beträgt dieser Anteil bereits weit über 50%! Grund genug, um sich im folgenden mit Implementation und Struktur dafür geeigneter Algorithmen zu beschäftigen.

4.5.5.1 Suchverfahren

Meva muß in diversen Daten suchen: in eingelesenen Benutzerdaten, in internen Tabellen als auch in externen Hilfsdateien wie z.B. den Indexdateien.

Bei der Wahl des Suchalgorithmus, der in den MeSH-Indexdateien verwendet wird, wurde besonderer Wert auf die Implementation eines hochperformanten Algorithmus gelegt, da externe Dateizugriffe in der Regel mit großen Zugriffszeiten zu Buche schlagen. Überschritte die Antwortzeit des CGIs die im

²³ <http://www.informatik.fh-muenchen.de/~schieder/programmieren-1-ws96-97/>

Server festgelegte Spanne, bräche der Server die Ausführung des CGI ab. Gleichfalls bräche auch ein Client nach einer Weile die Verbindung zum Server ab, käme keine Reaktion seitens des Servers.

Ein elementarer Suchalgorithmus wie die *Sequentielle Suche* innerhalb der 37.863 MeSH-Einträge (englischer MeSH, Stand 2002) ist mit einer mittleren Komplexität von $\Theta(n) = n/2$ offensichtlich nicht performant genug, da im Schnitt $n/2$ Elemente verglichen werden müssen, um ein bestimmtes Element zu finden. Daher empfiehlt sich die *Sequentielle Suche* nur in Prototypen oder für kleine Datenbestände und zeitunkritische Aufgaben. Für schnellere Suchfunktionen sollte die *Sequentielle Suche* durch ein anderes Verfahren ersetzt werden.

Höhere Stringsuchroutinen wie die Suche nach *Boyer-Moore* [Glogau 1994, S. 284; Hermann 1992, S. 293; Sedgewick 1997, S. 334] oder die Suche nach *Knuth-Morris-Pratt* [Glogau 1994, S. 282; Hermann 1992, S. 291; Sedgewick 1997, S. 329] erfüllen hier ebf. nicht ihren Zweck, da hauptsächlich nach einem Wort innerhalb einer Liste – und nicht in einem Text – gesucht werden muß.

Da die Initialen der MeSH-Deskriptoren ungleichmäßig verteilt sind, scheidet eine *Interpolationssuche* [Sedgewick 1997, S. 239; Hermann 1992, S. 287] ebf. aus, da deren Laufzeit bei ungleich verteilten Daten stark ansteigt. Die *Interpolationssuche* ist im Grunde eine weiterentwickelte *Binäre Suche* (s.u.), die Indexrechnungen zuläßt. Sie arbeitet statt mit einer einfachen binären Partitionierung der Daten mit einer zielgerichteten Suche, indem sie die vermutete Position des Suchtokens im Suchfeld nach seiner Stellung im Zeichenwertebereich kalkuliert:

$$i = \text{low} + (\text{high} - \text{low}) * (x - 'A') / ('Z' - 'A')$$

Versuche, die überprüften, ob die Verteilung der MeSH-Daten dennoch ausreichend ist, um eine Verbesserung mit einer Interpolation zu erzielen, ergaben lineare Laufzeiten und bestätigten somit die fehlende Eignung dieses Verfahrens. Zudem verbietet sich die Anwendung dieses Suchverfahrens, da zumindest in der Indextabelle mit den MeSH-Deskriptoren die Schlüsselwerte häufig mehrfach auftreten²⁴. Bei gleichverteilten und nicht multipel auftretenden Daten erreicht die *Interpolationssuche* hingegen eine Komplexität von $\Theta(n) = \text{ld}(\text{ld}(n))$.

Um die MeSH-Codes zu den MeSH-Strings und vice versa zu finden, wurde eine *Binäre Suche* [Sedgewick 1997, S. 236; Hermann 1992, S. 287; Ruhland 1991, S. 30] verwendet. Die *Binäre Suche* ist einfach zu implementieren und ist mit einer

²⁴ Da MeSH-Deskriptoren im MeSH-Baum sich oft an mehreren Positionen befinden, projiziert ein MeSH-Deskriptor häufig auf mehrere MeSH-Codes.

Komplexität von $\Theta(n) = \lg(n)$ hinreichend performant. Da bei jedem Schritt die Anzahl der zu prüfenden Elemente nach dem Prinzip *Divide et impera* halbiert wird, werden bei k Schritten insgesamt 2^k Elemente erfaßt. Das bedeutet, daß mit nur zehn Vergleichen 2^{10} Elemente durchsucht werden. Da die Initialen der Codes nicht gleichmäßig über den ASCII-Werteraum verteilt sind (Tab. 4), wird die Effizienz der *Binären Suche* allerdings etwas abgeschwächt:

Suchbereich	Vorliegende Verteilung	Gleichverteilung
'A' - 'N'	98,74%	53,8%
'O' - 'Y'	0%	42,3%
'Z'	1,26%	3,8%

Tab. 4: Verteilung der Initialen der englischen MeSH-Codes 2002

Dasselbe trifft für die Suche nach den Strings zu einem entsprechenden Code zu. Auch hier wird eine *Binäre Suche* verwendet, die aufgrund der ungleichmäßigen Verteilung der Stringinitialen nicht ideal zum Tragen kommt (Tab. 5):

Suchbereich	Vorliegende Verteilung	Gleichverteilung
Großbuchstaben	98,8%	41,94%
Kleinbuchstaben	0,58%	41,94%
Ziffern	0,57%	16,13%

Tab. 5: Verteilung der Initialen der englischen MeSH-Strings 2002

Abb. 5 veranschaulicht die inhomogene Häufigkeitsverteilung der Stringinitialen:

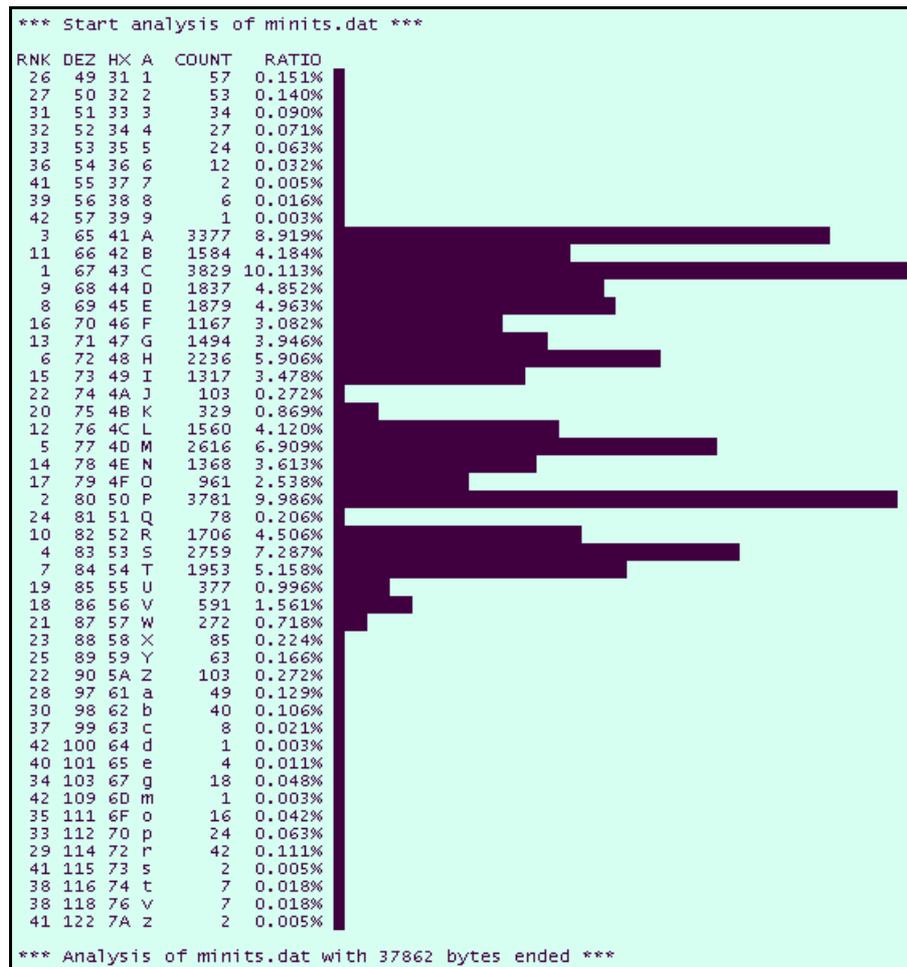


Abb. 5: Häufigkeiten der Initialen der englischen MeSH-Strings 2002

Wie man sieht, sind die Häufigkeiten für die Stringinitialen sehr ungleichmäßig verteilt: Die Buchstaben „C“ und „P“ sind in den MeSH-Strings mit jeweils ca. 10% vertreten, „J“ hingegen liegt unter einem Prozent.

Tab. 6 illustriert die Komplexitäten der Suchverfahren an einem konkreten Fall:

Suchverfahren	Erforderliche Zugriffe, um 1 von 37.862 Elementen zu finden
Sequentielle Suche	18.931
Binäre Suche mit vorliegenden Daten	14
Binäre Suche, gleichverteilte Daten	10

Tab. 6: Zugriffe von Suchverfahren im Vergleich

Für 55 zu suchende Codes mußte 798mal auf die Indexdatei zugegriffen werden, um die entsprechenden Codes von den gesamt 37.862 in der Indexdatei vorliegenden Codes zu finden. Das entspricht einem Schnitt von 14 Zugriffen pro Code. Eine *Binäre Suche* mit ausgewogenen Daten müßte hier im Schnitt nur

zehnmal zugreifen, eine *Sequentielle Suche* hingegen 18.931mal. Wie man sieht, ist die Suche nach den MeSH-Strings weniger performant als eine nach gleich verteilten Daten. Dennoch bleibt sie bei weitem unter der Laufzeit einer *Sequentiellen Suche*.

Eine geringe Optimierung ist mit einer modifizierten *Binären Suche* möglich, die vor dem Suchvorgang die Positionen jedes Anfangszeichenbereiches bestimmte und einmalig – in der ersten rekursiven Aufrufinstanz – als *Interpolationssuche* arbeitete. Da die bestehende Suchperformanz jedoch ausreichend ist und die Dateizugriffszeiten nicht maßgeblich auf die Gesamtausführungszeit durchschlagen, wurde auf eine weitere Optimierung verzichtet, um den Code nicht mit hochspezifischen Anpassungen zu überfrachten.

4.5.5.2 Sortierverfahren

Analog zum Suchen muß das Programm auch Sortieren: hier kommt es ebf. auf Schnelligkeit an, um vorzeitige Verbindungsabbrüche zu verhindern.

Meva muß diverse Daten sortieren, die es vom Benutzer einliest, angefangen von den Feldern bis hin zu den MeSH-Deskriptoren.

Als elementare Sortierverfahren mit quadratischer Komplexität scheiden *Bubble Sort*, *Insertion Sort*, *Shake Sort* oder *Selection Sort* [Ruhland 1991, S. 17; Hermann 1992, S. 296 ff.; Sedgewick 1997, S. 125 ff.] von vornherein aus. Zwar ist deren Sortierschema stabil (die relative Ordnung der Elemente wird im Sortierverfahren beibehalten), was die Weiterverarbeitung enorm erleichtern kann, aber die Laufzeiten sind für größere Elementanzahlen unakzeptabel. Da für Meva Elementanzahlen bis in den fünfstelligen Bereich auftreten können, ergibt sich als Obergrenze eine mittlere Zahl an Schritten (Vertauschen oder Vergleichen zweier Elemente) von $100.000 \text{ mal } 100.000 = 10 \text{ Mrd.}$ Ließe man einen Schritt auch nur 1ms dauern, betrüge die Ausführungszeit allein einer solchen Sortieroutine schon $10 \text{ Mrd. mal } 1\text{ms} \approx 116 \text{ Tage}$ – eine Spur zu lange für die meisten Benutzer. Etwas anschaulicher wird das Ganze, läßt man zehnmal höhere Elementanzahlen zur Verarbeitung zu: hier betrüge die Verarbeitungszeit dann schon ca. 32 Jahre.

Deutlich besser dagegen gestaltet sich der Einsatz höherer Verfahren wie z.B. des *Quick Sort* [Ruhland 1991, S. 19; Hermann 1992, S. 296 ff.; Sedgewick 1997, S. 145 ff.]. Seine Komplexität beträgt nur $\Theta(n) = n \lg(n)$, asymptotisch somit ein linearer Verlauf, so daß für ein $n = 10.000$ und 1ms Schrittzeit sich eine Ausführungszeit $t \approx 1.661\text{s} \approx 27\text{m}$ und für $n = 1.000.000$ sich ein $t \approx 19.932\text{s} \approx 5,54\text{h}$ ergäbe. Dies nur zur Veranschaulichung der Relationen: die heutigen Computer liegen deutlich unter einer Millisekunde pro Schritt, dementsprechend auch die Ausführungszeiten.

5 ERGEBNISSE

Im Kapitel Ergebnisse wird die Bedienung des Meva-Dienstes erläutert und beschrieben, welchen Benefit der Anwender aus Meva ziehen kann.

Folgende Punkte werden im einzelnen besprochen:

- Meva-Eingabe - wie der Benutzer Meva konsultiert
- Meva-Ausgabe - welche Informationen der Benutzer dem Meva-Resultat entnehmen kann
- Fallbeispiele, die den Nutzen von Meva zu verdeutlichen helfen sollen
- technische Grenzwerte von Meva
- flankierende Hilfsprogramme zu Meva
- Installation und Wartung von Meva durch den Administrator und
- mögliche Probleme und ihre Lösung in der Bedienung von Meva.

5.1 Meva-Eingabe

Das folgende Kapitel beschreibt die notwendigen Benutzereingaben in einer Meva-Konsultation, um interaktiv ein verdichtetes Analyseresultat zu erhalten.

Meva wird über mehrere Parameter im Konsultationsformular gesteuert, die dem Dienst mitteilen, welche bibliographische Felder der Benutzer analysieren und wie er das Analyseresultat dargestellt haben möchte (Abb. 6). Die Parameter sind in zwei Sektionen untergliedert, Eingabe und Ausgabe. Die zu setzenden Parameter gelten entweder einheitlich für beide zu analysierenden bibliographischen Felder (Feld 1 und Feld 2) oder separat nur für je eines, ausgedrückt durch die farbliche Unterlegung:

Parameter		Feld 1	Feld 2
Eingabe	Bibliogr. Feld	MeSH Terms	
	Suchfilter	<input type="checkbox"/> Groß-/Kleinschreibung ignorieren	<input type="checkbox"/> Groß-/Kleinschreibung ignorieren
		<input type="checkbox"/> Nur ganzes Wort	<input type="checkbox"/> Nur ganzes Wort
		<input type="checkbox"/> Nur Erstautor	
Min. Häufigkeit	1	1	
Ausgabe	Diagramme	<input checked="" type="checkbox"/> Top 10 skaliert mit auto	
	Details	<input checked="" type="checkbox"/> alphabetisch aufsteigend	<input type="checkbox"/> PMID immer anzeigen
	MeSH-Codes	<input type="checkbox"/>	
	MeSH-Baum	<input checked="" type="checkbox"/> Darstellungstiefe 12	
		A: Anatomy	als Teilbaum zur Häufigkeitsverrechnung bevorzugt
	Link-Restriktor		
	Datenformat	HTML	
	Kommentar		
Dateiname		Choose	
Konsultiere Meva		Löschen	

Abb. 6: Meva-Formular, über das der Benutzer Meva steuert

5.1.1 Eingabesektion von Mevas Konsultationsformular

Die Parameter der Eingabesektion bestimmen, welche bibliographischen Felder von Meva analysiert werden und welche Ausprägungen der Felder in das Meva-Resultat aufgenommen werden.

5.1.1.1 Feldname

Im Feldnamensfeld gibt der Benutzer die zu analysierenden Felder an. Wichtige Felder sind z.B. *Author*, *Affiliation* (Institution und Adresse des Erstautors), *Title*, *MeSH Term*, *Publication Type* oder *Journal Title Abbreviation*.

Meva stellt alle Feldeinträge des ersten gewählten Feldtyps von allen Artikeln mit zugehöriger PubMed-Artikelnummer (PMID) dar. Falls der Benutzer auch einen zweiten Feldnamen ausgewählt hat, werden entsprechend auch die im selben Artikel befindlichen Feldeinträge des 2. Feldtyps in Korrelation zum ersten angezeigt.

5.1.1.2 Suchfilter

Hiermit kann der Benutzer angeben, ob er Felder auf ein bestimmtes Muster einschränken will, das im Feldinhalt vorkommen muß. Der Benutzer kann z.B. einen Stern als Filter verwenden, um nur Hauptschlagwörter in den MeSH-Termen zu finden. Speziell bei MeSH-Termen wird dies generell empfohlen, um Datenmenge und Netzlast zu reduzieren.

5.1.1.2.1 Boolesche Suche

Zur Booleschen Suche können die logischen Operatoren *AND*, *OR* und *NOT* der Booleschen Algebra verwendet werden. Im Gegensatz zu PubMeds logischen Operatoren unterscheidet Meva nicht zwischen groß und klein geschriebenen Operatoren. Fehlt ein binärer logischer Operator zwischen zwei Operanden, werden diese mit *AND* verknüpft. Ein Ausdruck der Form *Colchicine blood* entspricht somit *Colchicine AND blood*.

Die Operatoren folgen keinen inhärenten Vorrangregeln, eine vorrangige Auswertung kann nur durch Klammerung der entsprechenden Teilausdrücke erzwungen werden. Anderenfalls erfolgt die Auswertung sequentiell von links nach rechts. Ein Ausdruck der Form *a AND b OR a AND c* entspricht damit implizit *((a AND b) OR a) AND c*. Die Suche nach *Colchicine AND blood OR pharmacokinetics* und die Suche nach *Colchicine AND (blood OR pharmacokinetics)* findet somit meist nicht dasselbe.

Die Booleschen Operatoren arbeiten im Gegensatz zu PubMeds logischen Operatoren auf Feld- und nicht auf Artikelebene, d.h. die verwendeten Operanden beziehen sich alle auf ein und dasselbe Feld. Sucht der Benutzer z.B. nach *Colchicine AND blood*, werden *Colchicine/blood*, *Colchicine/blood/pharmacokinetics* und auch *Colchicine/*analogs & derivatives/blood* gefunden, nicht aber Felder, die nur *Colchicine*, nur *Lysin/*analogs & derivatives/blood/pharmacokinetics* oder nur *Colchicine/pharmacokinetics* enthalten - auch wenn sie alle im selben Artikel vorkommen. Eine Verknüpfung auf Artikelebene ist besser in PubMed selbst zu realisieren, z.B. *Colchicine[mh] AND blood[mh]* (*mh* ist die Abkürzung für *Main Heading*).

5.1.1.2.2 Wörtliche Suche

Um Zeichenketten zu finden, die mehrere Wörter enthalten, kann der Benutzer den entsprechenden Teilausdruck in doppelte Anführungszeichen setzen. Die Suche nach "Cell Line" findet somit nur Zeilen, welche die komplette Zeichenkette einschließen, nicht aber Zeilen, die z.B. nur *Cell*, nur *Line* oder aber *Cell Proliferation Line* beinhalten.

5.1.1.2.3 Groß/Kleinschreibung

Hier kann der Benutzer einstellen, ob Groß- und Kleinschreibung beachtet oder ignoriert werden soll. Ist z.B. Ignorieren aktiviert, werden bei der Suche nach *age* sowohl *Age*, *age* als auch *AGE* gefunden.

5.1.1.2.4 Ganz- oder Teilwortsuche

Über diese Option kann der Benutzer festlegen, ob nur nach ganzen Wörtern oder auch nach Teilbegriffen gesucht werden soll. Als Wortgrenze gilt das Leerzeichen. Ist z.B. Ganzwortsuche aktiviert, werden bei der Suche nach *age* nur ganze Wörter wie z.B. *Age*, *age* als auch *AGE* gefunden (abhängig von der Groß/Kleinschreiboption), nicht aber Wörter wie *Aged* oder *Heritage*.

Bsp. 1: Der Benutzer will nach allen Zeichenketten suchen, die mit *Age* beginnen (*Age*, *Aged* oder *Aging*, aber nicht *Dosage* oder *Heritage*). Er lässt dazu das Groß/Kleinschreibfeld und das Ganzwortsuchefeld inaktiv (StandardEinstellung).

Bsp. 2: Der Benutzer sucht Wortzusammensetzungen mit *hydro* ungeachtet der Groß/Kleinschreibung, z.B. *Hydrofolate*, aber auch *Dehydrogenase*. Er aktiviert dazu das Groß/Kleinschreibfeld und lässt das Ganzwortsuchefeld inaktiv.

Bsp. 3: Der Benutzer will exakt nur *Age* finden. Er lässt dazu das Groß/Kleinschreibfeld inaktiv und aktiviert das Ganzwortsuchefeld.

Bsp. 4: Der Benutzer will *Age* als auch *age* finden, nicht aber *Aged*, *aged* etc. Er aktiviert dazu beide Filtermodusfelder.

5.1.1.2.5 Nur Erstautor

Mit dieser Option kann der Benutzer entscheiden, ob er sich bei der Suche nach Autoren auf Erstautoren beschränken will.

5.1.1.3 Minimale Häufigkeit

Der Benutzer kann die Anzeige auf Felder beschränken, die eine Mindesthäufigkeit aufweisen. Wenn er z.B. die Ziffer 3 angibt, werden nur Felder angezeigt, die mindestens dreimal im PubMed-Resultat vorkommen.

5.1.2 Ausgabesektion von Mevas Konsultationsformular

Die in der Ausgabesektion aufgeführten Parameter bestimmen, wie das Analyseresultat angezeigt wird.

5.1.2.1 Diagramme

Der Benutzer wählt mit dem Markierungsfeld, ob Histogramme der häufigsten Felder sowie Koinzidenztabelle angezeigt werden sollen.

5.1.2.2 Top-Wert

Mit dem Top-Wert kann er die Anzahl der Felder festlegen, die in den Diagrammen angezeigt werden sollen. Wünscht er z.B. eine Häufigkeitsdarstellung der drei häufigsten Felder, gibt er hier die Ziffer 3 an. Überschreitet der hier angegebene Wert die Anzahl der gefundenen Felder, wird er von Meva auf die Anzahl angepaßt. Maximal 50 Feldinhalte sind erlaubt.

5.1.2.3 Skalierung

Mit der Skalierung kann der Benutzer das Aussehen der Histogramme beeinflussen. Dieser Wert sollte in den meisten Fällen auf `auto` belassen werden. In den Fällen jedoch, in denen Histogramme mehrerer Resultatdateien verglichen werden sollen, kann es nützlich sein, die Skalierung auf einen absoluten Wert zu setzen, um eine direkte optische Vergleichbarkeit zu ermöglichen.

5.1.2.4 Details

Mit dem Markierungsfeld gibt der Benutzer an, ob er die Darstellung der Detailtabelle wünscht. Ohne diese Angabe werden nur die Histogramme, Korrelationsmatrizen und der MeSH-Baum angezeigt.

5.1.2.5 Sortierung

Er kann außerdem wählen, ob die gefundenen Felder in den Details alphabetisch oder nach Anzahl ihres Vorkommens sortiert werden sollen, und in welcher Reihenfolge: auf- oder absteigend.

5.1.2.6 PMID immer anzeigen

Mit der Option „PMID immer anzeigen“ kann der Benutzer erzwingen, daß die PMID-Spalte in den Details auch dann angezeigt wird, falls kein zweiter Feldtyp im Formular ausgewählt wurde. Falls die PMID aber nicht interessiert und man ein komprimiertes Listing, z.B. zum Ausdruck, benötigt, empfiehlt es sich, diese Option auf der Voreinstellung (inaktiv) zu lassen.

5.1.2.7 MeSH-Codes

Der Benutzer kann wählen, ob zu den MeSH-Termen in den einzelnen Sektionen auch die entsprechenden MeSH-Codes ausgegeben werden sollen.

5.1.2.8 MeSH-Baum

Hat der Benutzer im Feld 1 oder 2 MeSH-Terme ausgewählt, kann er einen MeSH-Baum ausgeben lassen, in dem alle gefundenen Terme mit ihren Häufigkeiten und entsprechend ihrer Codeposition dargestellt werden.

5.1.2.9 Darstellungstiefe

Der Benutzer kann zudem festlegen, bis zu welcher Tiefe der Baum dargestellt werden soll. So zeigt z.B. die Darstellungstiefe 1 nur die Hauptkategorien des Baumes an (s.a. Punkt [5.6.2.3.2.2](#)).

5.1.2.10 Bevorzugter Teilbaum

Bestimmte Deskriptoren finden sich nur einmal im MeSH-Baum. Ein Beispiel ist *3,4-Dichloro-N-methyl-N-(2-(1-pyrrolidinyl)-cyclohexyl)-benzeneacetamide, (trans)-Isomer* mit dem Code *D03.383.773.953* in der Kategorie *Chemicals and Drugs*.

Daneben gibt es aber auch Terme, die über mehrere Codes im Baum verfügen, also in mehreren Kategorien aufscheinen. Da PubMed nur die Zeichenkette eines Terms, nicht aber den entsprechenden Code liefert, und eine Zeichenkette unter mehreren Codes im Baum lokalisiert sein kann – *Visual Perception* findet sich z.B. unter *Psychiatry and Psychology* als *F02.463.593.932* als auch unter *Biological Sciences* als *G11.697.911.860* - stellt Meva alle zugehörigen Codes dar. Nun entsteht aber das Problem, die Häufigkeiten den Codes zuzuordnen. Wird *Visual Perception* z.B. neunmal gefunden, welcher Code soll die Neun tragen: *F02.463.593.932* oder *G11.697.911.860*? Meva versucht, dieses Problem mit mehreren Strategien aufzulösen:

5.1.2.10.1 Branch fit

Der Benutzer kann je nach Suchpräferenz einen MeSH-Teilbaum auswählen, in dem die Häufigkeiten bevorzugt erfaßt werden.

Beispiel: Der Benutzer wählt *F - Psychiatry and Psychology* als Teilbaum. Der Code *F02.463.593.932* bekommt somit alle Häufigkeiten von *Visual Perception* zugeschlagen.

Greift diese Strategie nicht – sei es, weil die ausgewählte Kategorie thematisch nicht paßt oder weil die Codes eines Deskriptors sich in derselben Kategorie

befinden – kommt eine Alternativstrategie zum Tragen:

5.1.2.10.2 Deepest fit

Hier wählt Meva den Code aus, der am tiefsten in der Baumhierarchie verankert ist, um die Spezifität des Resultates zu erhöhen. Mit wachsender Codelänge steigt die Spezifität eines Terms.

Beispiel: Der Term *17-Hydroxyprogesterone* findet sich z.B. unter *D06.472.798.773.478.400* als auch unter *D04.808.745.745.654.829.395.400* im MeSH-Baum. Meva schlägt hier alle Häufigkeiten von *17-Hydroxyprogesterone* dem Code *D06.472.798.773.478.400* zu.

Versagt auch dieser Ansatz, schaltet Meva auf die dritte Strategie um:

5.1.2.10.3 First fit

Hier wird der erste Code, der zu einer Zeichenkette gefunden wird, mit den Häufigkeiten des Strings belastet.

Beispiel: Die Häufigkeiten zu *2-Hydroxyphenethylamine*, welches *D02.033.100.291.410*, *D02.092.063.291.410*, *D02.092.211.215.581* und *D02.092.471.683.545* als Codes besitzt, würden allesamt *D02.033.100.291.410* zugeschlagen. Dies ist weniger eine Strategie als vielmehr die Ultima ratio.

5.1.2.11 Linkrestriktor

Der Benutzer kann hier einen Term angeben, der als zusätzlicher Suchausdruck in die Feldverknüpfungen des Meva-Resultats integriert wird. Da Meva ja nicht „weiß“, mit welchem Suchausdruck er in PubMed die Eingabedatei für Meva erstellt hat, kann er Meva auf diesem Wege die Information mitgeben. So kann z.B. sichergestellt werden, daß bei einer erneuten Suche in PubMed, die über eine Feldverknüpfung im Meva-Resultat angestoßen wird, die ursprüngliche Lösungsmenge nicht überschritten wird.

Ein Link in einem Meva-Resultat setzt sich somit aus dem Wert des angeklickten Feldes und dem des Linkrestriktors zusammen (Abb. 7).

Meva-Formular:

Linkrestriktor:

Meva-Resultat:

i	nc	n	MeSH Term	PMID	Author
5	12	5	Action Potentials/ drug effects/ *physiology	11960786	Spencer NJ, Hennig GW, Smith TK
				11856532	Taddese A, Bean BP

Verknüpfung: [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=search&term=SmithTK\[au\]Tetrodotoxin\[mh\]](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=search&term=SmithTK[au]Tetrodotoxin[mh])

Feldwert Linkrestriktor

Abb. 7: Aufbau eines PubMed-Links in Meva

Die Lösung, die PubMed beim initialen Suchen findet, wird im folgenden als *Primärmenge* oder *Primärlösung* bezeichnet. Die Lösung, die PubMed über das Aktivieren eines Links im Meva-Resultat findet, wird als *Sekundärmenge* oder *Sekundärlösung* bezeichnet.

Der Term muß der PubMed-Suchsyntax gehorchen, anderenfalls erhält der Benutzer falsche oder keine Resultate bei der Aktivierung einer Verknüpfung im Meva-Resultat. Um die Suche auf bestimmte Feldtypen zu beschränken, kann er hinter dem Suchausdruck in eckigen Klammern den Feldtyp angeben. Dies wird generell empfohlen, da es zu spezifischeren Ergebnissen führt. PubMed-Referenzen wie z.B. #1 sollten nicht als Linkrestriktor verwendet werden, da diese Referenzen nur eine kurze Gültigkeitsdauer haben.

Falls kein Linkrestriktor angegeben wurde, werden keine Verknüpfungen erzeugt (PMID und Korrelationsmatrizen ausgenommen), da die Lösungsmenge beim Aktivieren einer Verknüpfung ansonsten zu groß geriete: Eine globale Suche nach einem Autor namens `Smith` in PubMed ist natürlich wenig sinnvoll und belastet die PubMed-Datenbank unnötig.

Die folgenden drei Beispiele beschreiben die Möglichkeiten zur Beeinflussung der Sekundärlösung genauer. Die dabei verwendeten Abbildungen sind nicht Bestandteil des Meva-Resultates, sondern verdeutlichen die Lösungsmengen und ihre Häufigkeiten mit Stand Juni 2003. Die verwendeten Feldkürzel `MH`, `AU` und `DP` kennzeichnen MEDLINEs Feldtypen *MeSH Term*, *Author* und *Publication Date*:

5.1.2.11.1 Bsp. 1a - Verkleinerung der Primärlösung

Der Benutzer hat in PubMed initial nach `Tetrodotoxin[mh]` gesucht, um eine Eingabedatei für Meva zu erhalten. Im Meva-Formular hat er *Author* als Suchfeld ausgewählt. Er möchte nun, daß wenn er im Meva-Resultat auf einen der gefundenen Autoren klickt, PubMed nur Datensätze findet, in denen der Autor etwas über Tetrodotoxin geschrieben hat. In diesem Fall gibt er im Formular den ursprünglichen PubMed-Suchausdruck `Tetrodotoxin[mh]` als Linkrestriktor an.

Der linke Kreis in Abb. 8 illustriert die Primärmenge `Tetrodotoxin[mh]`, die schraffierte Schnittmenge hingegen die Sekundärlösung `Smith TK[au] AND Tetrodotoxin[mh]` mit 18 Artikeln, die PubMed beim Klick auf den Autor `Smith TK` im Meva-Resultat fände:

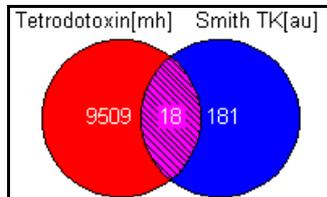


Abb. 8: Verkleinerung der Sekundärmenge in Mevas Verknüpfungen

Diese Konstellation stellt sicher, daß PubMed nur solche Datensätze findet, die einen Teil der ursprünglichen Lösungsmenge darstellen. Die Sekundärmenge ist somit eine Teilmenge der Primärmenge.

5.1.2.11.2 Bsp. 1b - Zusätzl. Verkleinerung der Primärlösung

Wie im Beispiel 1a hat der Benutzer in PubMed nach `Tetrodotoxin[mh]` gesucht und im Meva-Formular *Author* als Suchfeld ausgewählt. Er möchte nun mit der Aktivierung einer Verknüpfung alle Datensätze aus dem Jahre 2002 finden, in denen ein bestimmter Autor etwas über Tetrodotoxin geschrieben hat. In diesem Fall gibt er im Formular `Tetrodotoxin[mh] AND 2002[dp]` als Linkrestriktor an.

Der linke obere Kreis in Abb. 9 illustriert die Primärmenge `Tetrodotoxin[mh]`, die schraffierte Schnittmenge hingegen die Sekundärlösung `Smith TK[au] AND Tetrodotoxin[mh] AND 2002[dp]` mit 2 Artikeln, die PubMed beim Klick auf den Autor `Smith TK` im Meva-Resultat fände:

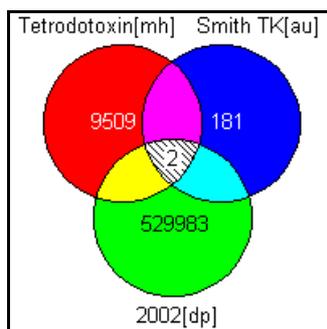


Abb. 9: Zusätzl. Verkleinerung der Sekundärmenge in Mevas Verknüpfungen

Diese Konstellation bietet sich an, wenn die Primärlösung sehr groß ausgefallen ist oder man den Suchradius weiter verfeinern möchte. Statt des Publikationsdatums kann natürlich auch ein anderer passender Restriktor gewählt werden.

5.1.2.11.3 Bsp. 2 - Teilweise Erweiterung der Primärlösung

Wie im Beispiel 1a hat der Benutzer in PubMed nach `Tetrodotoxin[mh]` gesucht

und im Meva-Formular *Author* als Suchfeld ausgewählt. Er möchte nun mit der Aktivierung einer Autor-Verknüpfung alle Datensätze des Autors aus dem Jahr 2002 finden (nicht nur solche zu Tetrodotoxin). In diesem Fall gibt er im Formular nur 2002[dp] als Linkrestriktor an.

Der linke obere Kreis in Abb. 10 illustriert die Primärmenge Tetrodotoxin[mh], die schraffierte Schnittmenge hingegen die Sekundärlösung Smith TK[au] AND 2002[dp] mit 15 Artikeln, die PubMed beim Klick auf den Autor Smith TK im Meva-Resultat fände:

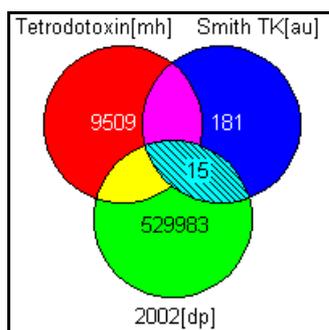


Abb. 10: Schnittmengenbildung in Mevas Verknüpfungen

Diese Konstellation bietet sich für Fragestellungen an wie „Über was hat der Autor noch geschrieben?“. Die Sekundärlösung ist hier nicht mehr eine Untermenge der Primärlösung, sondern eine Schnittmenge zweier Primärmengen. Der Linkrestriktor ist in diesem Falle nur notwendig, um Meva Verknüpfungen generieren zu lassen, da ansonsten die Sekundärmenge zu groß ausfiele. Statt des Publikationsdatums kann natürlich auch ein anderer passender Restriktor gewählt werden.

5.1.2.12 Datenformat

Der Benutzer kann angeben, ob er das Resultat im HTML- oder im datenbanktauglichen Textformat erhalten möchte.

Im HTML-Format besteht im Gegensatz zum Textformat die Möglichkeit zu Verknüpfungen und damit zur interaktiven Suche. Durch einen Bildschirmdruck in eine Grafikdatei (unter Windows z.B. mit *Alt-Druck*) kann so einfach Präsentationsmaterial gewonnen werden. Viele Browser erlauben auch einen Vollbildmodus ohne störenden Rahmen, meist über die Taste *F11*.

Im Textformat werden die Daten im *TSV-Format* (*Tabulator Separated Values*) übermittelt. In diesem Format besteht das Resultat aus einer durch Tabulatoren abgetrennten Liste der Feldinhalte aus Detail- oder MeSH-Sektion. Der Aufbau des Textformates ist somit ident zum *CSV-Format* (*Comma Separated Values*) [[Born](#)

1992, S. 53] mit der Ausnahme, daß Tabulatoren anstelle von Kommata treten und textuelle Feldinhalte nicht durch Hochkommata begrenzt werden.

Zusätzlich enthält die erste Zeile der gelieferten Datendatei die Feldnamen. Viele Datenbank- oder Chartprogramme (z.B. Oracle, MS-SQL-Server, MS Access) erlauben es, beim Einlesen von Texttabellen die erste Zeile als Liste von Feldnamen zu interpretieren.

Im Textformat können je nach Wahl des Benutzers entweder die Daten aus der Detailtabelle oder aus dem MeSH-Baum geliefert werden:

Bei der Detailtabelle im Textformat sind dieselben Ordnungskriterien und Benutzereinstellungen anwendbar wie bei der HTML-Ausgabe. Die Sortierung kann auf- oder absteigend geordnet sein und nach Häufigkeit oder alphabetisch erfolgen. Die Anzeige der PMID ist auch bei nur einem vom Benutzer ausgewählten bibliographischen Feld möglich.

Bei der Ausgabe des MeSH-Baums im Textformat gilt dasselbe, Ordnungs- und Ansichtskriterien können analog zum HTML-Format festgelegt werden. Die Tiefe des auszugebenden Baumes wie auch seine Anzeige mit oder ohne interpolierten Termen läßt sich über das Formular einstellen. Einfache und kumulierte Häufigkeiten werden gleichfalls ausgegeben.

5.1.2.13 Kommentar

Der Benutzer kann hier einen Kommentar eingeben, der von Meva ins Resultat eingebettet wird. Somit kann man später leicht die Suchergebnisse identifizieren, insbesondere wenn man mehrere Analysen durchführen will.

Es gibt zwei gute Gründe, dieses Feld nicht zu vernachlässigen:

Da der Kommentar von Meva in den Titel eingebettet wird und viele Browser den Titel verwenden, um beim Abspeichern einer HTML-Datei einen Dateinamen vorzuschlagen, kann man somit schnell die Datei unter einen aussagekräftigen Namen abspeichern, was zur späteren Identifikation hilfreich ist.

Außerdem erscheint der Kommentar beim Ausdrucken auf jeder Seite oben in der Kopfzeile. Insbesondere wenn man Ausdrücke über mehrere Seiten hat und diese nicht gleich abheftet, können dank des Kommentars die Seiten dann trotzdem noch richtig einander zugeordnet werden.

Wird kein Kommentar vergeben, wird der Linkrestriktor als Kommentar verwendet.

5.1.3 Weitere Parameter

5.1.3.1 Dateiname

Der Benutzer gibt hier den kompletten Pfad und Dateinamen zu der von der NLM heruntergeladenen Textdatei an, z.B. unter Windows `f:\in\ms20020617.txt` oder unter UNIX `/home/user/ms20020617.txt`. Anzumerken ist, daß die Datei im MEDLINE-Format von PubMed vorliegen muß.

5.1.3.2 Konsultiere Meva

Die Schaltfläche „Konsultiere Meva“ sendet die Formulardaten an Meva zur Auswertung.

5.1.3.3 Löschen

Die Schaltfläche „Löschen“ entfernt alle Formulardaten und setzt die Felder auf ihre Standardwerte zurück.

5.2 Meva-Ausgabe

Dieser Abschnitt beschreibt anhand eines beispielhaften Meva-Resultates, welche Informationen aus dem PubMed-Resultat Meva analysiert. Im vorliegenden Falle hat ein Benutzer in PubMed (Abb. 11) mit dem Suchterm `Optical Illusions[mh] 2000[dp]` nach Artikeln aus dem Jahr 2000 gesucht, die mit dem MeSH-Term `Optical Illusions` indiziert wurden:

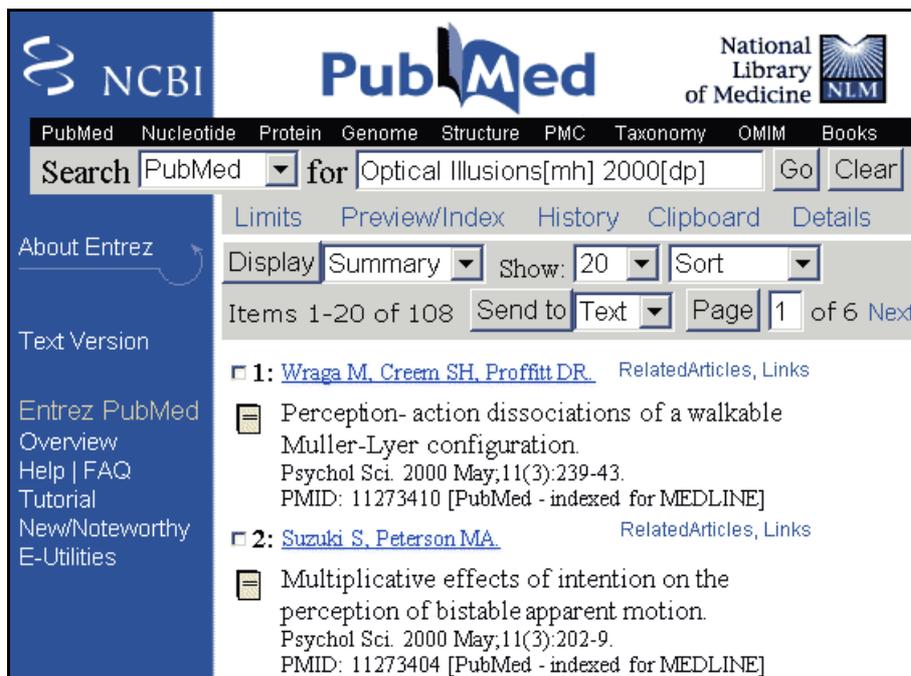


Abb. 11: Beispielein- und -ausgabe in PubMed

Nach dem Speichern des so gewonnenen Resultates im MEDLINE-Format (Display: MEDLINE, Send to: File) startete er eine Meva-Abfrage, die aus diesem Resultat MeSH-Terme (Feld 1) und Autoren (Feld 2) extrahieren sollte. Generische MeSH-Terme wie `Human`, `Animal` etc. sollten von der Analyse ausgeschlossen und nur Hauptschlagwörter gefunden werden, somit setzte er einen Stern (*) als Filter in das Feld 1. Es sollten nur MeSH-Terme berücksichtigt werden, die mindestens sechsmal auftraten, also setzte er die Minimale Häufigkeit für das Feld 1 im Formular auf den Wert Sechs. Von den acht häufigsten Feldern sollte ein Histogramm sowie eine Korrelationsmatrix ausgegeben werden, somit setzte er den Topwert im Formular auf den Wert Acht. Außerdem sollte ein Baum aus den MeSH-Einträgen gezeichnet werden, deren Häufigkeiten der Kategorie *F - Psychiatry and Psychology* zugerechnet werden sollten.

Abb. 12 illustriert die durch den Benutzer bereitgestellten Daten in Mevas Konsultationsformular:

Parameter	Feld 1	Feld 2
Bibliogr. Feld	MeSH Terms	Author
Eingabe	* <input type="checkbox"/> Groß-/Kleinschreibung ignorieren <input type="checkbox"/> Nur ganzes Wort <input type="checkbox"/> Nur Erstautor	 <input type="checkbox"/> Groß-/Kleinschreibung ignorieren <input type="checkbox"/> Nur ganzes Wort
	Min. Häufigkeit	6 1
	Diagramme	<input checked="" type="checkbox"/> Top <input type="text" value="8"/> skaliert mit <input type="text" value="auto"/>
Details	<input checked="" type="checkbox"/> alphabetisch <input type="text" value="aufsteigend"/> <input type="checkbox"/> PMID immer anzeigen	
MeSH-Codes	<input type="checkbox"/>	
Ausgabe	MeSH-Baum	<input checked="" type="checkbox"/> Darstellungstiefe <input type="text" value="10"/> F: Psychiatry and Psychology als Teilbaum zur Häufigkeitsverrechnung bevorzugt
	Link-Restriktor	Optical illusions[mh] 2000[dp]
Datenformat	HTML	
Kommentar	Aus Suche nach Optical illusions[mh] 2000[dp]	
Dateiname	Optical illusions[mh] 2000[dp].txt	<input type="button" value="Choose"/>
<input type="button" value="Konsultiere Meva"/> <input type="button" value="Löschen"/>		

Abb. 12: Beispielleingabe in Meva

Nach der Einstellung der erforderlichen Parameter in Mevas Konsultationsformular schickte der Benutzer die PubMed-Resultatdatei nebst den erforderlichen Parametern an Meva zur Konsultation. Die nachfolgenden neun Abschnitte beschreiben die gleichnamigen Sektionen des zurückgelieferten Meva-Resultates im Detail. Die Sektionen werden in der Reihenfolge der Ausgabe angezeigt.

5.2.1 Überschriftensektion

In Titel und Überschrift des Meva-Resultates finden sich die gesuchten Feldtypen und der vom Benutzer im Formular eingegebene Kommentar wieder (Abb. 13). Dies dient der schnellen Identifizierung der eigenen Resultate.

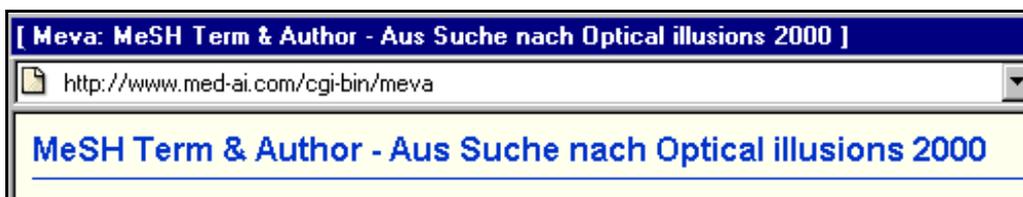


Abb. 13: Meva-Resultat - Überschriftensektion

5.2.2 Meldungssektion

Gemäß Abb. 14 gibt das System in dieser Sektion Warnungen, Empfehlungen zur Verbesserung der Suchstrategie oder Hinweise zu Verknüpfungen oder

vorgenommenen Datenkonversionen aus.

Hinweise und Warnungen	
1.	Mit einem Klick auf ein PMID-Feld erreichen Sie den Originalartikel in der PubMed [®] -Datenbank des NCBI.
2.	Ein Klick auf ein verlinktes Feld aktiviert eine weitere Suche in PubMed nach Feldwert und Linkrestriktor.

Abb. 14: Meva-Resultat - Meldungssektion

5.2.3 Parametersektion

In der Parametertabelle faßt Meva alle wichtigen Parameter der Konsultation zusammen, die der Benutzer im Formular festgelegt hat (Abb. 15).

Parameter			
Datum, Uhrzeit	06.06.2003, 21:29	Datei	n0108_Optical_illusions[mh]_2000[dp].meva
Maxima	In: 96/5120 KB; Out: 383/15000 Tags; TopVal: 48/1000	Top	8
Filter 1	* [groß ungleich klein, kein ganzes Wort]	Filter 2	-
Min. Häufigkeit 1	6	Min. Häufigkeit 2	1
Details	Alphabetisch, aufsteigend, PMID	Link Restrictor	Optical Illusions[mh] 2000[dp]
MeSH Codes	-	MeSH Tree	Bevorzugter Teilbaum F, Darstellungstiefe 10

Abb. 15: Meva-Resultat - Parametersektion

Im einzelnen sind dies die folgenden Parameter:

- Datum und Uhrzeit der Konsultation
- Mevas Grenzwerte
- Die Filter für Feld 1 und 2
- Den Filtermodus (Berücksichtigung der Groß- und Kleinschreibung und Ganzwortsuche)
- Die minimalen Häufigkeiten für Feld 1 und 2
- Ob Details gedruckt werden oder nicht
- Die Sortierung für die Details
- Die Sortierreihenfolge für die Details
- Ob Codes für MeSH-Terme gedruckt werden oder nicht
- Den Dateinamen der PubMed-Datei
- Den Top-Wert, bis zu dem die Histogramme und Korrelationsmatrizen Werte berücksichtigen
- Den Linkrestriktor
- Ob ein MeSH-Baum gedruckt wird oder nicht

- Die gewünschte Tiefe des MeSH-Baums
- Den gewünschten Teilbaum des MeSH-Baums, der zur Häufigkeitsverrechnung verwendet werden soll

Die Grenzwerte („Maxima“) verstehen sich wie folgt:

Maximum In: 202/5120 KB: Meva hat 202 KB von erlaubten 5.120 KB vom Benutzer erhalten. Die Notwendigkeit für diese Obergrenze liegt im Schutz des Servers und des Netzwerkes vor undifferenzierten großvolumigen Anfragen begründet.

Der Ressourcenengpaß besteht somit im Volumen des Netzwerkverkehrs.

Maximum Out: 383/15000 Tags: Meva druckt (ca.) 383 von 15.000 zulässigen Feldinhalten in der Detailsektion aus. Die Obergrenze begründet sich auf der Notwendigkeit des Schutzes des Benutzers vor sehr komplex strukturierten Antworten, die sehr schnell auftreten können, wenn der Benutzer nach multipel auftretenden Feldern in beiden Feldtypen sucht (quadratische Komplexität): Einschlägige Webbrowser sind bei Tabellen mit mehr als 15.000 Zellen überfordert.

Der Ressourcenengpaß besteht somit in der Darstellungsfähigkeit des Webbrowsers.

Maximum TopVal: 48/1000: Die maximal erlaubte Häufigkeit eines Feldes für Korrelationsmatrizen beträgt 1.000 und die aktuell größte gefundene im PubMed-Dokument beträgt 48. Die Notwendigkeit für diese Obergrenze liegt in der quadratischen Komplexität der zugrundeliegenden Auswertungsroutine begründet. Die Datenmaxima treffen nur für die HTML-Ausgabe zu, nicht für den Text-Modus. Aus der Tabelle ist ferner zu ersehen, daß generische MeSH-Terme wie *Human*, *Animal* etc. durch die Eingabe eines Sternes im 1. Filter ausgeschlossen wurden.

Der Ressourcenengpaß besteht somit in der Laufzeit von Meva.

5.2.4 Bilanzsektion

In der Bilanzsektion finden sich summarisch die Häufigkeiten der verarbeiteten Felder (Abb. 16), hier von MeSH-Termen und Autoren. Hier sollen beispielhaft nur die MeSH-Terme dargestellt werden:

Die 108 Artikel der PubMed-Datei enthielten 1.010 MeSH-Terme. 344 von den 1.010 ließen die Filter von Meva passieren. Von diesen erkannte Meva 116 als

distinkt. Davon siebte Meva alle mit einer Häufigkeit unter 6 (entsprechend dem Wert des Parameters *Minimale Häufigkeit 1* in der Parametersektion) aus, so daß nur 11 unterschiedliche Feldinhalte verblieben. Diese werden als *Datenbasis* bezeichnet und stellen das Ausgangsmaterial für alle folgenden Auswertungen!

Total			
Phase	Artikel	MeSH Term	Author
Eingelesen	108	1010	269
Die Filter passiert		344	269
Als distinkt erkannt		116	247
Min. Häufigkeiten erreicht und in Datenbasis aufgenommen		11	247

Abb. 16: Meva-Resultat - Bilanzsektion

5.2.5 Histogrammsektion

Meva verdichtet im Prozeß des *Data Mining* in der Histogrammsektion die bibliographischen Felder. Die häufigsten Felder, die sog. *Topwerte* (Top X), werden in einem Balkendiagramm dargestellt. Die Anzahl der zu analysierenden Felder wurde vorher vom Benutzer im Formular mit der Angabe des Top-Wertes festgelegt. Wurden zwei Felder vom Benutzer zur Analyse ausgewählt, werden auch zwei Histogramme aufgebaut.

5.2.5.1 Histogramm für Feldtyp 1

In Abb. 17 sieht man die acht häufigsten Schlagwörter der Datenbasis. (Acht war die Zahl, die der Benutzer angegeben hatte.) Wie man sieht, taucht **Optical Illusions* z.B. 48mal im PubMed-Resultat auf.

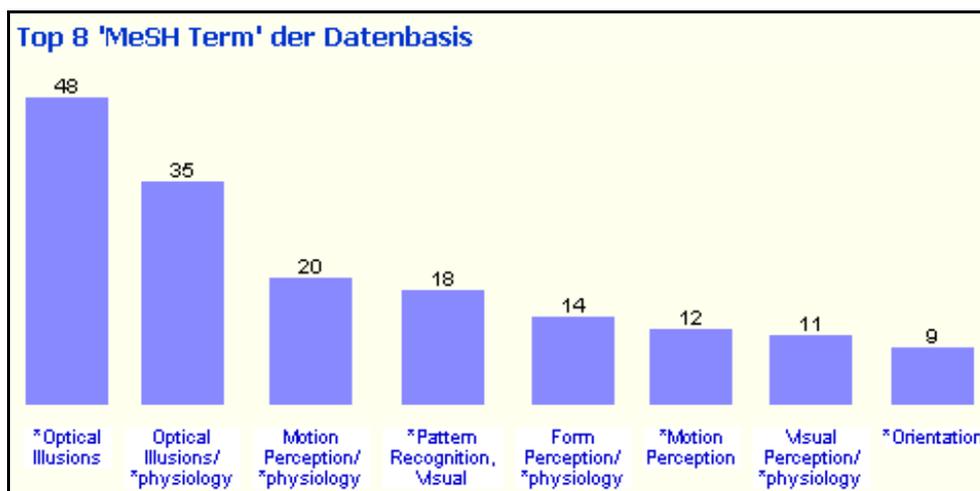


Abb. 17: Meva-Resultat - Histogrammsektion Feld 1

Interessant wird das Histogramm in diesem Zusammenhang, weil man mit seiner Hilfe wichtige Begriffe desselben Feldtyps findet, die man gezielt in PubMed weiterverfolgen kann. Auch bei unbekanntem Thema erkennt man sehr schnell,

welche Stichworte am wichtigsten sind. *Motion Perception* ist ein Begriff, der in diesem Beispiel mehrfach auftritt. Eine erste Hypothese ist somit die Schlußfolgerung, daß optische Täuschungen sehr viel mit der Art und Weise der Wahrnehmung von Bewegung zu tun haben. Deshalb ermöglicht Meva mit dem Anklicken eines Feldnamens im Histogramm gleich eine weitere Suche in PubMed nach Artikeln, die den entsprechenden Begriff und den Linkrestriktor enthalten (Bedingung dafür ist ein gesetzter Linkrestriktor vom Benutzer).

5.2.5.2 Histogramm für Feldtyp 2

In Abb. 18 sieht man entsprechend die acht häufigsten Felder der Datenbasis für den zweiten vom Benutzer angegebenen Feldtyp, die Autoren:

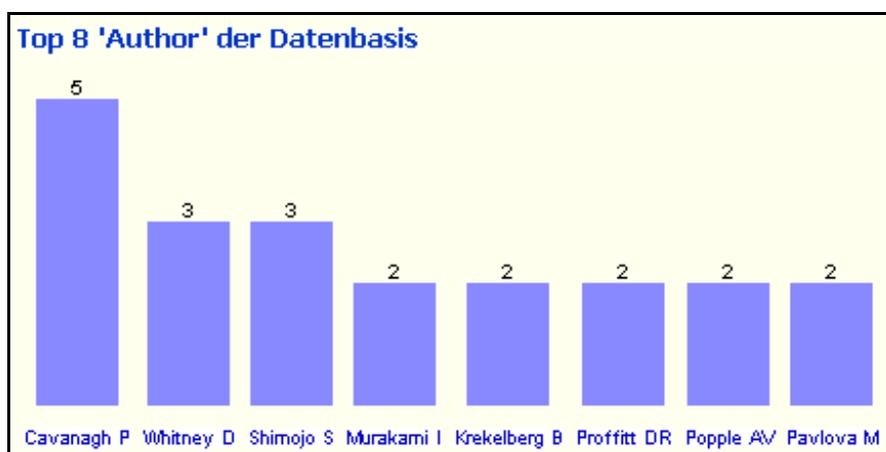


Abb. 18: Meva-Resultat - Histogrammsektion Feld 2

Auch hier ist interessant, wer sich am meisten mit diesem Thema beschäftigt hat. Man sieht sofort, daß Cavanagh P. sich am meisten mit dem gesuchten Thema auseinandergesetzt hat. Wurde ein Linkrestriktor vom Benutzer vergeben, stößt ein Klick auf seinen Namen im Histogramm wiederum eine Suche nach Artikeln von ihm in PubMed an.

5.2.6 Korrelationssektion

In der Korrelationssektion analysiert Meva im Prozeß des *Data Mining*, welche Felder innerhalb eines Artikels koinzidieren. Die Sektion weist entweder eine oder zwei Tabellen auf. Häufige Korrelationswerte sind dunkler gehalten, um eine schnelle optische Identifizierung der meistassoziierten Felder zu ermöglichen.

Durch das mehrfache Auftreten können Zeilen- oder Spaltensummen eines Feldes höher sein als seine Gesamthäufigkeit (abzulesen in der Detailsektion).

Dies soll durch das folgende Gedankenexperiment näher erläutert werden:

Eine PubMed-Datei bestehe aus fünf Artikeln. In jedem dieser Artikel tauche

sowohl ein Feld MH1, ein Feld AU1 als auch ein Feld AU2 auf. Die gemeinsame Häufigkeit für MH1 und AU1 beträgt somit fünf, die gemeinsame Häufigkeit für MH1 und AU2 ebf. fünf. Die Summe beider Korrelationswerte beträgt somit zehn – nichtsdestoweniger existiert das Feld MH1 nur fünfmal.

Die Summenwerte dieser Tabelle sagen also lediglich etwas über den Grad der Vergesellschaftung der häufigsten Felder aus, aber nichts über deren absolute Anzahl (die in der Bilanzsektion ausgewiesen ist).

5.2.6.1 Korrelation Feldtyp 1

Wenn das Feld 1 in einem Artikel mehrfach auftauchen kann wie z.B. MeSH-Term, Autor, Publikationstyp oder EC/RN-Nummer, wird die erste Tabelle aufgebaut (Abb. 19).

Koinzidenzen Top 'MeSH Term' - Top 'MeSH Term'									
(Mehrfachnennungen möglich)									
	*Optical Illusions	Optical Illusions/ *physiology	Motion Perception/ *physiology	*Pattern Recognition, Visual	Form Perception/ *physiology	*Motion Perception	Visual Perception/ *physiology	*Orientation	Total
*Optical Illusions			3	16	3	9	3	7	41
Optical Illusions/ *physiology			15		8		2		25
Motion Perception/ *physiology	3	15			2				20
*Pattern Recognition, Visual	16					2		5	23
Form Perception/ *physiology	3	8	2						13
*Motion Perception	9			2					11
Visual Perception/ *physiology	3	2							5
*Orientation	7			5					12
Total	41	25	20	23	13	11	5	12	150

Abb. 19: Meva-Resultat - Korrelationssektion Feld 1

Interessant wird die Korrelationsmatrix wie auch schon das Histogramm, da man durch sie schnell wichtige assoziierte Begriffe erkennen und gezielt weiterverfolgen kann. Dadurch wird es möglich, ein Thema weiter einzukreisen. Die Matrix ist symmetrisch aufgebaut, da nur Felder desselben Typs korrelieren. Wie man in Abb. 19 sieht, taucht **Optical illusions* z.B. sechzehnmal zusammen mit **Pattern Recognition, Visual* in einem Artikel auf. Durch einen Klick auf den Korrelationswert kann der Benutzer gezielt in PubMed nach dem gemeinsamen Auftreten beider Feldwerte und dem Linkrestriktor suchen, im vorliegenden Falle

also nach *Optical illusions[mh]* und **Pattern Recognition, Visual[mh]* sowie dem Linkrestriktor. Im Gegensatz zu PubMed erlaubt Meva somit eine bequeme schrittweise Verfeinerung des ursprünglichen Suchradius.

5.2.6.2 Korrelation Feldtyp 1 und 2

Hat der Benutzer einen zweiten Feldtyp angegeben, wird eine zweite Korrelationsmatrix (Abb. 20) aufgebaut, welche die Inzidenzen beider Feldtypen innerhalb eines Artikels darstellt: welche Felder des ersten Feldtyps mit welchen Feldern des zweiten Feldtyps innerhalb eines Artikels zusammenfallen. Diese Matrix ist im Gegensatz zur ersten asymmetrisch aufgebaut, da zwei unterschiedliche Feldtypen korreliert werden.

Koinzidenzen Top 'MeSH Term' - Top 'Author'									
(Mehrfachnennungen möglich)									
	Cavanagh P	Whitney D	Shimojo S	Murakami I	Krekelberg B	Proffitt DR	Popple AV	Pavlova M	Total
*Optical Illusions	1	1	1		1	1		1	6
Optical Illusions/ *physiology	3	1	1	1	1		2		9
Motion Perception/ *physiology	4	2	1	2	1			1	11
*Pattern Recognition, Visual						2			2
Form Perception/ *physiology									0
*Motion Perception	1	1	1		1			1	5
Visual Perception/ *physiology			1				1		2
*Orientation						2			2
Total	9	5	5	3	4	5	3	3	37

Abb. 20: Meva-Resultat - Korrelationssektion Feld 1 und 2

Auch hier können die Korrelationen wertvolle Aufschlüsse ergeben: Zu erkennen im Beispiel ist, daß P. Cavanagh in diesem Rahmen meistens über *Motion Perception* geschrieben hat. Analog zur vorgestellten Hypothese in der Histogrammsektion, daß optische Täuschungen sehr viel mit der Art und Weise der Wahrnehmung von Bewegung zu tun haben, hat sich P. Cavanagh auch diesem Nexus gewidmet.

5.2.7 Detailsektion

In der Detailsektion wird die komplette Datenbasis von Meva als Tabelle angezeigt: Die gefundenen Typ-1-Felder mit ihren Häufigkeiten, die dazugehörige PubMed-Artikelnummer und die im selben Artikel gefundenen Typ-2-Felder.

Die Detailtabelle erlaubt sowohl statistische Auswertungen der Häufigkeiten der

Felder als auch eine unmittelbare Sekundärsuche in PubMed durch Anklicken eines Feldes.

Die Tabelle ist nach dem ersten Feldtyp sortiert, aufsteigend oder absteigend, alphabetisch oder nach Häufigkeit. Abb. 21 zeigt einen kleinen Ausschnitt aus dieser alphabetisch sortierten Tabelle:

Details der Datenbasis					
i	nc	n	MeSH Term	PMID	Author
1	7	7	*Attention	11273404	Suzuki S, Peterson MA
				11273385	Crawford LE, Huttenlocher J, Engebretson PH
				11153859	Kincade S, Wilson AE
				11143444	Palmer SE, Nelson R
				11131742	Gunn DV, Warm JS, Dember WN, Temple JG
				11114232	Servos P, Carnahan H, Fedwick J
				10997045	Sokolov A, Pavlova M, Ehrenstein WH
2	19	12	*Motion Perception	11273404	Suzuki S, Peterson MA
				11219985	Sheth BR, Shimojo S
				11184992	Patel SS, Ogmen H, Bedell HE, Sampath V
				10997045	Sokolov A, Pavlova M, Ehrenstein WH
				10970215	Whitney D, Cavanagh P
				10970214	Krekelberg B, Lappe M
				10927113	Steinman RM, Pizlo Z, Pizlo FJ
				10910090	Sandor PS, Bachtold D, Henn V, Brugger P
				10820609	Ido K, Ohtani Y, Ejima Y
				10811162	DeLucia PR, Tresilian JR, Meyer LE
				10811159	Valdes-Sosa M, Cobo A, Pinilla T
				10720334	Eagleman DM, Sejnowski TJ
				Etc...	

Abb. 21: Meva-Resultat - Detailsektion

Der Spaltenname *i* bezeichnet den Tabellenindex, *n* die Anzahl der Ausprägungen eines Typ-1-Feldes und *nc* deren kumulierte Anzahl. Beträgt die minimale Häufigkeit 1, entspricht die kumulierte Häufigkeit der letzten Zeile der Anzahl der Typ-1-Felder, die den Filter passierten. Die Einträge der nach Häufigkeit sortierten Spalte erlauben Aussagen zu Perzentilen, z.B.: „Der Anteil an Feldern mit Häufigkeiten > p beträgt q% aller Felder.“

Durch einen Mausklick auf einen der PMID-Werte kommt der Benutzer direkt zum entsprechenden Artikel in PubMed. Hat er auch einen Linkrestriktor im Meva-Formular angegeben, sind auch die Feldeinträge wie z.B. MeSH-Terme oder Autoren verlinkt, ein Klick darauf löst eine PubMed-Suche nach dem Feldwert und dem Wert des Linkrestriktors aus. Klickt der Benutzer z.B. auf den zweiten MeSH-Term *Motion Perception* der obigen Detailliste, löst er damit eine Suche nach diesem Begriff und dem Linkrestriktor aus, in diesem Fall also nach

Motion Perception[mh] Optical illusions[mh] 2000[dp].

5.2.8 MeSH-Sektion

Die MeSH-Sektion wertet Schlagwörter aus. Da in den Schlagwörtern das in den Artikeln erfaßte Wissen hochkonzentriert und systematisch von den Bibliothekaren erfaßt ist, fallen Mevas Resultate problemspezifischer aus als im Vergleich zur Suche nach Titeln oder Abstracts. Die Menge des Wissens, das in den Schlagworten begründet ist, wird häufig unterschätzt. Durch die Darstellung der Schlagwörter können relevante Schlagwörter für ein Thema identifiziert werden und der Benutzer kann eine gezielte neue Suche in PubMed durchführen.

Die MeSH-Sektion untergliedert sich in die Bilanz zu den verarbeiteten MeSH-Termen, den Erfolg der angewandten Wichtungsstrategien und den MeSH-Baum, den Meva aus den Termen konstruiert.

Durch die Anzeige der MeSH-Terme in einer hierarchischen Struktur – dem MeSH-Baum – können leicht wichtige, assoziierte Terme gefunden werden, die in einer erneuten Suche in PubMed benutzt werden können. Als ein Beispiel nur sei die Anzeige des Terms *Visual Perception* im Baum genannt, der von Meva in Auswertung des Suchergebnisses für *Optical Illusions* mitaufgelistet wird.

Im Gegensatz zu PubMeds MeSH-Browser, der immer nur den Teilbaum für einen Term anzeigt, stellt Meva alle Terme der Suchmenge im Baum dar. Durch diese Darstellungsform bietet sich diese Anzeige auch für Präsentationen an, in denen die Hierarchie oder die Häufigkeit der verwendeten Schlagwörter verdeutlicht werden soll.

5.2.8.1 Bilanz

Die Bilanztafel der MeSH-Sektion (Abb. 22) zeigt, welche MeSH-Terme Meva verarbeitet, um einen Baum zu generieren.

MeSH Tree		
Phase	n	%
Unifizierte Strings	7	26%
Assoziierte Codes	12	46%
Interpolierte Codes	14	54%
Gesamte Codes	28	100%

Abb. 22: Meva-Resultat - MeSH-Sektion Bilanz

Zum Verständnis ist es wichtig zu wissen, daß Meva in seiner Verarbeitung Strings und dazugehörige Codes trennt. Dementsprechend sind drei Kategorien zu unterscheiden:

5.2.8.1.1 Unifizierte MeSH-Strings

Um die MeSH-Strings im MeSH-Baum zu finden und damit ihren Code zu erhalten, müssen die in der PubMed-Datei vorliegenden MeSH-Strings von führenden Sternchen (Vorzugsbezeichnungen) und anhängenden *Subheadings* (Qualifizierern) befreit werden, da der MeSH-Baum solche Ausprägungen nicht kennt. Die so behandelten Strings werden *unifizierte* Strings genannt. So ergeben z.B. **Optical illusions*, *Optical illusions* und *Optical illusions/physiology* trunkiert nur einen MeSH-Baum-Eintrag: *Optical illusions*. Generische Terme hingegen wie *Human* besitzen keinen Code und fallen weg.

Im Beispiel trunkierte Meva die elf gefundenen MeSH-Strings der Datenbasis (s. Bilanzsektion 5.2.4), so daß sich sieben unifizierte Strings im Baum fanden. Meva erkennt, ob ein unifizierter Term als Eintrag im MeSH-Baum existiert, indem es eine eigene MeSH-Datenbank abfragt.

Die Anzahl der unifizierten MeSH-Strings gibt somit Aufschluß darüber, wie viele MeSH-Terme der PubMed-Datei als Entitäten im MeSH-Baum existieren.

5.2.8.1.2 Assoziierte MeSH-Codes

Da ein unifizierter String mehrere assoziierte Codes im MeSH-Baum haben kann, ist deren Zahl hier auch angegeben. Meva findet die zugehörigen Codes zu den Strings, indem es eine eigene MeSH-String-Datenbank abfragt.

Orientation z.B. findet sich im MeSH-Baum sowohl unter F01.058.577 als auch unter F02.830.606 (vgl. auch die entsprechenden Ausführungen in der Beschreibung des Meva-Formulars unter Punkt 5.1.2.10 - [Bevorzugter Teilbaum](#)).

Offensichtlich ist die Zahl der assoziierten Codes zumeist größer als die der zugehörigen Strings, mindestens aber ebenso groß.

Diese Maßzahl gibt somit Aufschluß über die Redundanz der gefundenen Strings: wie oft sie mehrfach – unter verschiedenen Codes – im Baum existieren.

5.2.8.1.3 Interpolierte MeSH-Codes

Interpolierte Codes sind alle Codes, die zwischen einem gefundenen Term und der Baumwurzel liegen und von Meva nicht im PubMed-Suchresultat gefunden wurden. Meva findet die zugehörigen Strings zu den Codes, die es zu interpolieren gilt, indem es eine eigene MeSH-Code-Datenbank abfragt.

F01 - Behavior and Behavior Mechanisms ist z.B. ein Code, der entsprechend dem vorgestellten Beispiel nicht in der PubMed-Datei gefunden wurde und von Meva

dementsprechend interpoliert wurde.

Diese Maßzahl stellt somit die Anzahl der MeSH-Terme dar, die Meva benötigte, um den Baum ausgehend von seiner Wurzel bis hin zu den gefundenen Elementen zu bevölkern.

5.2.8.2 Wichtungsstrategien

Der Bilanztafel nachgeschaltet (Abb. 23) findet sich der Erfolg der angewandten Wichtungsstrategien (vgl. auch die entsprechenden Ausführungen in der Beschreibung des Meva-Formulars unter Punkt 5.1.2.10 - [Bevorzugter Teilbaum](#)):

Bewertungserfolg: Teilbaumstrategie (F): 100%, Tiefenstrategie: 0%.

Abb. 23: Meva-Resultat - MeSH-Sektion Strategien

Im vorliegenden Beispiel konstatiert Meva einen Erfolg von 100% für die Teilbaumstrategie (*Branch Fit*). Dies bedeutet, daß es Meva gelang, für jeden Term mindestens einen Code im angegebenen Teilbaum (hier *F – Psychiatry and Psychology*) zu finden und ihn mit der korrespondierenden Term-Häufigkeit zu beladen. *Visual Perception* findet sich z.B. doppelt (Abb. 24): 17mal unter F02.463.593.932 und einmal unter G11.697.911.860. Meva hat hier alle Häufigkeiten dem Teilbaum F zugeschlagen.

Da die Teilbaumstrategie vollständig erfolgreich war, kamen nachgelagerte Strategien (*Deepest Fit, First Fit*) gar nicht erst zur Anwendung.

5.2.8.3 MeSH-Baum

Der MeSH-Baum stellt alle MeSH-Terme, die im Baum gefunden wurden, semigraphisch in einer Baumstruktur dar (Abb. 24). Alle Elemente werden entsprechend ihren Codes hierarchisch im Baum mit ihren Häufigkeiten positioniert. Existiert ein String mehrfach als Code, werden alle zugehörigen Codes ausgegeben. Dies erhöht das Ausgabevolumen beträchtlich. Da Meva aber nicht weiß, welchen Code es bevorzugen soll, bringt es alle Codes zur Ausgabe. Lediglich die Häufigkeiten werden spezifisch dem Teilbaum zugerechnet, der vom Benutzer ausgewählt wurde (s. vorigen Absatz).

Die Häufigkeiten werden nach oben über jeden Term kumuliert, so daß die Häufigkeiten höherer Terme die niederer inkludieren, auch wenn jene vielleicht schon aus der benutzerdefinierten Darstellungstiefe herausfallen. Jeder Term wird in der Form

Code - String: Häufigkeit [kumulierte Häufigkeit]

dargestellt, interpolierte Terms sind farblich markiert.

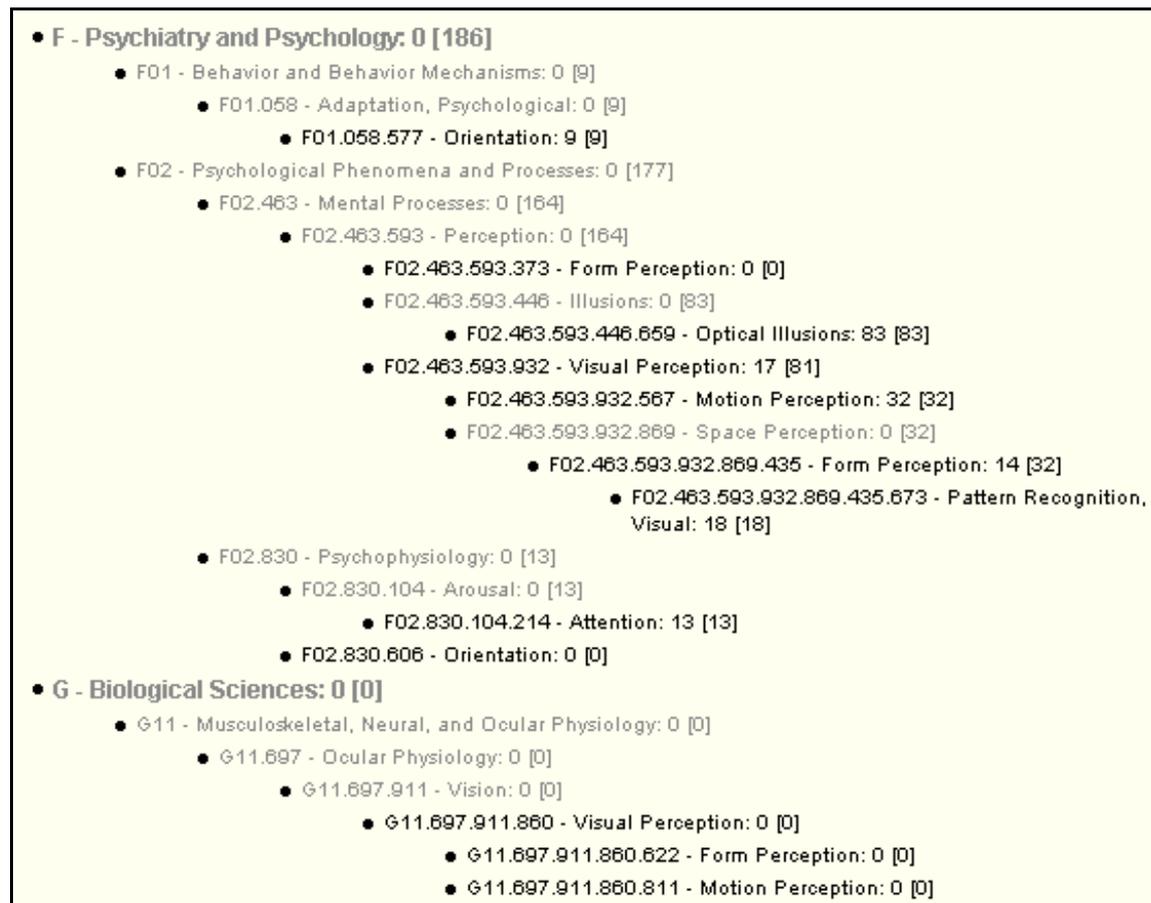


Abb. 24: Meva-Resultat - MeSH-Sektion Baum

Der Wert vor einer eckigen Klammer stellt die Häufigkeit dar, mit welcher der MeSH-Term gefunden wurde. Er bleibt null, wenn ein Term in mehreren Kategorien vorkommt und seine Häufigkeit in einer anderen Kategorie erfaßt wurde. *Visual Perception* findet sich z.B. doppelt: 17mal unter F02.463.593.932 und keinmal unter G11.697.911.860. Meva hat hier alle Häufigkeiten dem Teilbaum F zugeschlagen.

Die Zahl in der eckigen Klammer ist die kumulierte Häufigkeit aller Terme, die im entsprechenden Teilbaum vorhanden sind.

Durch das Auftreten eines Strings in mehreren Kategorien gewinnt man neue Anregungen, in welchen Wissensgebieten man noch nach dem Begriff suchen

kann. *Visual Perception* ist im Beispiel sowohl unter *Biologie – Physiologie des Auges* rubriziert als auch unter *Psychiatrie und Psychologie – Psychologische Phänomene und Prozesse*.

5.2.9 Fußnotensektion

In der Fußnotensektion findet sich ein optionales Logo oder ein Hinweis des jeweiligen Betreibers (Abb. 25). Am Schluß der Sektion wird die verwendete Meva-Version angezeigt, die benötigte Rechenzeit und die Anzahl der bisherigen Meva-Konsultationen.



Abb. 25: Meva-Resultat - Fußnotensektion

5.3 Fallbeispiele

Die nachfolgenden Fallbeispiele entstanden in der täglichen Arbeit mit Meva. Verdichtende Auswertungen, die ohne Meva vorher Wochen dauerten, konnten so innerhalb weniger Stunden erreicht werden. Durch das gezielte *Data Mining* ergab sich ein rascher Überblick zu den gesuchten Themenkreisen, wie die folgenden Beispiele näher ausführen.

Vorge stellt werden die Analysedaten zu zwei Autorenprofilen, zu einem Fachthema und die Gewinnung von Präsentationsmaterial für ein Tagungsposter.

5.3.1 Meva-Konsultation zum Profil von Filippi M.

Durch die Etablierung des *Sylvia Lawry Center for Multiple Sclerosis Research* am Institut für Medizinische Statistik und Epidemiologie der Technischen Universität München rückten die wissenschaftlichen Publikationen zur MS von Massimo Filippi in den Mittelpunkt des Interesses.

Zu Filippi M. fanden sich 316 Artikel in MEDLINE.

Durch Mevas Korrelationstabellen ergab sich, daß seine **Arbeitsstelle** die Abteilung für Neurologie am Ospedale San Raffaele, Universität Milano, Italien ist.

An den Artikeln waren 2.086 **Autoren** beteiligt (6,6 Autoren/Artikel), das sind 433 Personen. Mevas Histogramme zeigten, daß Filippi dabei 106mal als Erstautor erschien. Filippi veröffentlichte mit Comi G. (171), Rovaris M. (107), Rocca M.A. (61), Martinelli V. (52), Sormani M.P. (47), Miller D.H. (44) und Barkhof F. (39). Alle genannten Koautoren sind Mitarbeiter von Filippi in Mailand außer Miller und Barkhof, die auch in Neurologischen Instituten in London, Amsterdam und Genua arbeiteten.

Mevas Bilanztafel führte aus, daß die 316 Artikel 3.710 **MeSH-Terme** enthielten, darunter 1.088 Hauptterme, die sich als 511 Arten erwiesen. Die 12 häufigsten Hauptterme waren *Gehirnpathologie* (99), *MRI* (86), *MRI/Methoden* (76), *MRI/Instrumentation* (8), *Multiple Sklerose/Diagnose* (45), *MS/Diagnose/Pathologie* (10), *Rückenmarkspathologie* (18), *Kontrastmittel* (9) und *Interferon-beta/Therapieeinsatz* (7).

Lt. Mevas Korrelationstabellen traten folgende MeSH-Hauptterme **paarweise** auf: *MRI/Methoden/Instrumentation* sowohl mit *Multipler Sklerose/Diagnose/Pathologie/Instrumentation* als auch mit *MRI/Diagnose/Pathologie* (83), *Gehirnpathologie* sowohl mit *MRI* (46) als auch mit *MS* (57), *Gadolinium/Diagnostik* sowohl mit *MS* (6) als auch mit *MR* (6), sowie *Computerunterstützte Bildverarbeitung* mit *MRI* (6).

Mevas **MeSH-Baum** ergab folgende Schlagworthierarchien: Kategorie C des MeSH-Baums (*Krankheiten*, 346) enthielt die Subkategorie *Immunologische Krankheiten* (265), Kategorie E (*Diagnostik und Therapie*, 256) die Subkategorie *Diagnosen* (212), Kategorie D (*Chemikalien und Arzneimittel* 106) die Subkategorie *Chemische Reaktionen und Anwendungen* (28) und Kategorie A (*Anatomie*, 208) die Subkategorie *Nervensystem* (183).

Filippi veröffentlichte in 51 **Zeitschriften**: *Neurology* (51), *Journal of the Neurological Sciences* (39), *American Journal of Neuroradiology* (32), *Journal of*

Neurology, Neurosurgery, and Psychiatry (29), *Journal of Neurology* (27), *Magnetic Resonance Imaging* (15) und *Multiple Sclerosis* (14).

Dabei ließen sich 23 **Publikationstypen** unterscheiden: *Zeitschriftenartikel* (306), *Reviews* (48), *Review Tutorials* (44), *Klinische Studien* (16), *Multizenterstudien* (15) und *Kontrollierte Klinische Studien* (15).

Die Artikel wurden 1974 - 1994 mit einer **Publikationsrate** zwischen 1 - 5 Artikel/Jahr veröffentlicht, 1995 - 1997 zwischen 15 - 19, und 1998 - 2001 zwischen 39 - 54 (2002 und 2003 noch nicht vollständig). Das Maximum lag in 1999.

Fazit: Filippi Massimo, Mailand, Neuroradiologe, mit 316 Artikeln in MEDLINE vertreten, beschäftigt sich vorwiegend mit *Magnetic Resonance Interference* zur Berteilung des Verlaufes und Therapieerfolges bei Multipler Sklerose.

5.3.2 Meva-Konsultation zum Profil von Herfarth C.

Zu C. Herfarth fanden sich 585 Artikel in MEDLINE.

Durch Mevas Korrelationstabellen ergab sich, daß seine **Arbeitsstelle** die Chirurgische Universitätsklinik in Heidelberg ist.

An den Artikeln waren 2.930 **Autoren** beteiligt (5 Autoren/Artikel), das sind 708 Personen. Mevas Histogramme zeigten, daß Herfarth 101mal als Erstautor erschien. Herfarth veröffentlichte mit Klar E. (86), Schlag P. (69), Otto G. (61), Lehnert T. (55), Kraus T. (41), Mehrabi A. (39) und Gebhard M.M. (36). Die genannten Koautoren sind Mitarbeiter von Herfarth in der Chirurgischen Klinik Heidelberg.

Mevas Bilanztabelle führte aus, daß die 585 Artikel 7.122 **MeSH-Terme** enthielten, darunter 1.950 Hauptterme, die sich als 1.440 Arten erwiesen. Die 12 häufigsten Hauptterme waren *Lebertransplantation* (19), *Lebertransplantation/Physiologie* (14), *Pankreas/Blutersatz* (9), *Leber/Blutersatz* (8), *Leberdurchblutung* (6), *Kolitis ulzerative/Chirurgie* (19), *Morbus Crohn/Chirurgie* (13), *Wiederherstellende Proktokolektomie* (12), *Adenomatöse Polypose des Kolons/Chirurgie* (11), *Kolektomie* (7), *Kolektomie/Methode* (7), *Chirurgie/Ausbildung* (7).

Lt. Mevas Korrelationstabellen traten folgende MeSH-Hauptterme **paarweise** auf: *Kolitis ulzerative/Chirurgie* sowohl mit *Adenomatöse Polypose des Kolons/Chirurgie* (11) als auch mit *Morbus Crohn/Chirurgie* (4) als auch mit *Wiederherstellender Proktokolektomie* (4), sowie *Chirurgie/Ausbildung* mit *Computerhilfe* (5).

Mevas **MeSH-Baum** ergab folgende Schlagworthierarchien: Kategorie C des MeSH-Baums (*Krankheiten*, 830) enthielt die Subkategorie *Neubildung* (429) und *Erkrankungen des Digestiven Systems* (167), Kategorie E (*Diagnostik und Therapie*, 341) die Subkategorie *Operationen* (212), Kategorie D (*Chemikalien und Arzneimittel*, 332) die Subkategorie *Proteine* (67) und Kategorie A (*Anatomie*, 215) die Subkategorie *Verdauungsorgane* (109).

Herfarth veröffentlichte in 125 **Zeitschriften**: *Der Chirurg* (105), *Langenbecks Archives of Surgery* (64), *Tranplantation Proceedings* (47), *Annals of Surgery* (21), und *European Journal of Surgical Oncology* (18).

Dabei ließen sich 24 **Publikationstypen** unterscheiden (Mehrfachnennungen): 565 *Zeitschriftenartikel*, 51 *Reviews*, 36 *Klinischen Studien*, 15 *Randomisierte Kontrollierte Studien*, 14 *Editorials*, 8 *Kontrollierte Klinische Studien*.

Die Artikel wurden 1965 - 1976 mit einer **Publikationsrate** zwischen 1 - 9

Artikel/Jahr veröffentlicht, 1977 - 1981 zwischen 10 - 19, 1982 - 1987 zwischen 4 - 9, 1988 - 1989 zwischen 12 - 15 und 1990 - 2001 zwischen 28 - 67 (2002 und 2003 noch nicht vollständig). Das Maximum lag in 1998.

Fazit: Herfarth C., Heidelberg, mit 585 Artikel in MEDLINE vertreten, ist Chirurg im Bereich Allgemein Chirurgie und Transplantation. Er ist Spezialist für Lebertransplantationen allein oder verbunden mit gleichzeitiger Nieren- oder Pankreastransplantation und für Kolektomien bei chronischen entzündlichen Darmerkrankungen.

5.3.3 Meva-Konsultation zum Thema Lebertransplantation

Die Abfrage von *Liver Transplantation* in MEDLINE über PubMed ergab 18.456 Artikel.

An den Artikeln waren 105.103 **Autoren** beteiligt (5,7 Autoren/Artikel), das sind 32.386 Personen. Mevas Histogramme zeigten die Autoren mit den meisten Veröffentlichungen. Im Einzelnen waren dies zu diesem Thema Starzl T.E. (Abb. 26), Pittsburg, USA (523; 39mal Erstautor), Neuhaus P., Humboldt-Universität, Berlin (389; 29), Fung J.J., Pittsburg, USA (316; 17), Tanaka K., Kayoshima, Japan (267; 8), Todo S., Sapporo, Japan (250; 21), Williams R., Sydney, Australien (236; 11) und Pichlmayr R., Göttingen (230; 24). Herfarth C., Heidelberg (7; 3), sei des Bezuges halber erwähnt.

http://www.umc.pitt.edu/media/pcc010108/starzl.html

University of Pittsburgh

PITT

HOME | NEXT ARTICLE

Faisal Foundation Honors Starzl

Thomas E. Starzl, M.D., director and namesake of the UPMC Starzl Transplantation Institute, has been named cowinner of the King Faisal International Prize in Medicine for 2000. Starzl will be honored at a March ceremony in Riyadh.

He shares the award — and the accompanying \$200,000 prize — with two other transplantation pioneers: Sir Roy Yonke Calne, emeritus professor of surgery at the University of Cambridge, and Norman E. Shumway, professor and chairman of the Department of Cardiothoracic Surgery at Stanford University.

The King Faisal Foundation annually awards prizes in five categories "to show appreciation to dedicated men and women whose contributions make a positive difference. . ."

Abb. 26: Ausschnitt aus Webseite zu Starzl T.E., Juni 2003

Die 18.456 bei der Recherche zurückgegebenen Artikel enthielten 233.814 **MeSH-Terme**, darunter 62.954 Hauptterme, die sich in 22.996 Arten unterscheiden ließen. Die 15 häufigsten Hauptterme waren *Lebertransplantation* (6.082), *Lebertransplantation/ungünstige Ergebnisse* (1.175), *Lebertransplantation/Immunologie* (1.014), *Lebendspender* (465), *Lebertransplantation/Methode* (975), *Gewebe-spende* (888), *Lebertransplantation/Physiologie* (868), *Überlebenszeit des*

Transplantats (345), *Immunsuppressive Mittel/therapeutische Anwendung* (342), *Postoperative Komplikationen* (307), *Transplantationsabstoßung* (273), *Organpreservation* (267), *Nierentransplantation* (263), *Lebertransplantation/Mortalität* (231), *Tacrolimus/therapeutische Anwendung* (219).

Lt. Mevas Korrelationstabellen traten folgende MeSH-Hauptterme **paarweise** auf: *Lebertransplantation* mit jedem der aufgeführten 14 Terme, sowie *Lebertransplantation/Methoden* sowohl bei *Lebendspendern* (121) als auch bei *Hepatektomie/Methoden* (43) als auch bei *Gewebespendern* (30), sowie *Immundepressive Mittel/therapeutische Anwendung* sowohl bei *Tacrolimus/therapeutische Anwendung* (109) als auch bei *Lebertransplantation/Immunologie* (83), sowie *Organpreservation* bei *Lebertransplantation/Physiologie* (48), sowie *Hepatektomie/Methoden* bei *Lebendspendern* (38), sowie *Lebertransplantation/Immunologie* bei *Transplantatabstoßung* (38).

Mevas **MeSH-Baum** ergab folgende Schlagworthierarchien: Kategorie E des MeSH-Baums (*Diagnostik und Therapie*, 22.669) enthielt die Subkategorie *Operationen* (18.575), Kategorie C (*Krankheiten*, 16.516) die Subkategorie *Krankheiten des Digestiven Systems* (5.075, darunter 3.376 *Leberkrankheiten*, 867 *Neoplasmen* und 520 *DNA-Virusinfektionen der Verdauungsorgane*), Kategorie D (*Chemikalien und Medikamente*, 10.179) die Subkategorie *Immunologie und biologische Faktoren* (1.744) und Kategorie G (*Biologie*, 3.943) die Subkategorie *Immunität* (2.277).

Artikel über Lebertransplantation wurden in 1.438 **Zeitschriften** veröffentlicht: Die meistpublizierende Zeitschrift war *Transplantation Proceedings* (3.798), gefolgt von *Transplantation* (1.990), *Hepatology* (537), *Transplantation International* (513), *Liver Transplantation* (507), *Journal of Hepatology* (278), *Clinic of Transplantation* (231) und *Lancet* (203).

Dabei fanden sich 43 **Publikationstypen** (die nachfolgenden Angaben enthalten Mehrfachnennungen): Der häufigste Type war *Zeitschriftenartikel* (17.172), gefolgt von *Reviews* (2.660), *Review Tutorials* (1.085), *Klinische Studien* (1.085), *Letter* (780), *Kommentare* (576), *Randomisierte Kontrollierte Studien* (447), *Editorials* (355), *Multizenterstudien* (251), *Akademische Reviews* (193) und *Kontrollierte Klinische Studien* (160).

Die **Publikationsrate** stieg in den Jahren 1988 bis 1992 von 552 auf 992 Artikel/Jahr und in den Jahren 1993 bis 2002 von 1.280 auf 1.544 Artikel/Jahr. Das Maximum lag mit 1.593 Artikeln/Jahr im Jahre 2001.

Fazit: Die Artikel zum Thema Lebertransplantation in MEDLINE fokussieren sich auf immunologische, methodische und physiologische Aspekte der Transplantation, auf Organpreservation und postoperative Komplikationen sowie auf Tacrolimus als Makrolid im therapeutischen Einsatz. Die Lebertransplantation wird eingesetzt bei Leberkrankheiten und bei Leberbeteiligungen im Rahmen von neoplastischen Prozessen oder DNA-Virusinfektionen.

5.3.4 Meva-Konsultation zur Gewinnung von Präsentationsmaterial

Durch die Verdichtung und optische Ausgabe der Analyseresultate kann Meva benutzt werden, um Präsentationsmaterial zu gewinnen.

Als Beispiel nur sei grob der Arbeitsgang zur Erstellung eines Posters 2002 zum Thema „Womit beschäftigt sich zur Zeit die Medizinische Informatik? Eine MEDLINE-Analyse“ aufgeführt [Thurmayr et al. 2002, S. 100]:

Bei der Suche nach *Medical Informatics* in PubMed konnte eine primäre Lösungsmenge gefunden werden. Durch die Aufbereitung der Daten mit Mevas Histogrammen, Korrelationstabellen und MeSH-Baum konnten wichtige weitere assoziierte Schlagwörter identifiziert werden, wie z.B. *Tomography*, *X-Ray Computed* etc.

Damit konnte die ursprüngliche Suchanfrage in PubMed verfeinert werden und eine sekundäre Lösungsmenge gewonnen werden.

Am Schluß dieses iterativen Verfeinerungsprozesses konnte durch die Ausgabe des MeSH-Baumes im Meva-Resultat eine **Hierarchie von assoziierten Schlagwörtern** mit ihren Häufigkeiten aufgestellt werden. Die „**Top 10**“ der **Schlagwörter** wurden gleichermaßen identifiziert. Durch die Korrelationstabellen konnte die **Stärke der Vergesellschaftung bestimmter Schlagwörter** auf dem Gebiet der Medizinischen Informatik quantifiziert und entsprechende **Autorenprofile** dargestellt werden. Die Histogramme erlaubten zudem die graphische Darstellung der **Evolution bestimmter Schlagwörter auf einer Zeitachse** (welche Schlagwörter wurden wann etabliert und benutzt).

Aus den Ergebnissen des *Data-Mining*-Prozesses ließ sich dann folgende Aussage ableiten:

„Die Med. Informatik beschäftigt sich zur Zeit vorwiegend mit Bildverarbeitung von MRI und dreidimensionalen Bildern, weiterhin mit Computersimulation biologischer, molekularer und neurologischer Modelle, außerdem mit Faktendatenbanken für Proteine und mit Algorithmen für Informationsdienste.“

Abb. 27 zeigt das entsprechende Poster, welches anlässlich der 47. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) 2002 in Berlin eingereicht wurde.

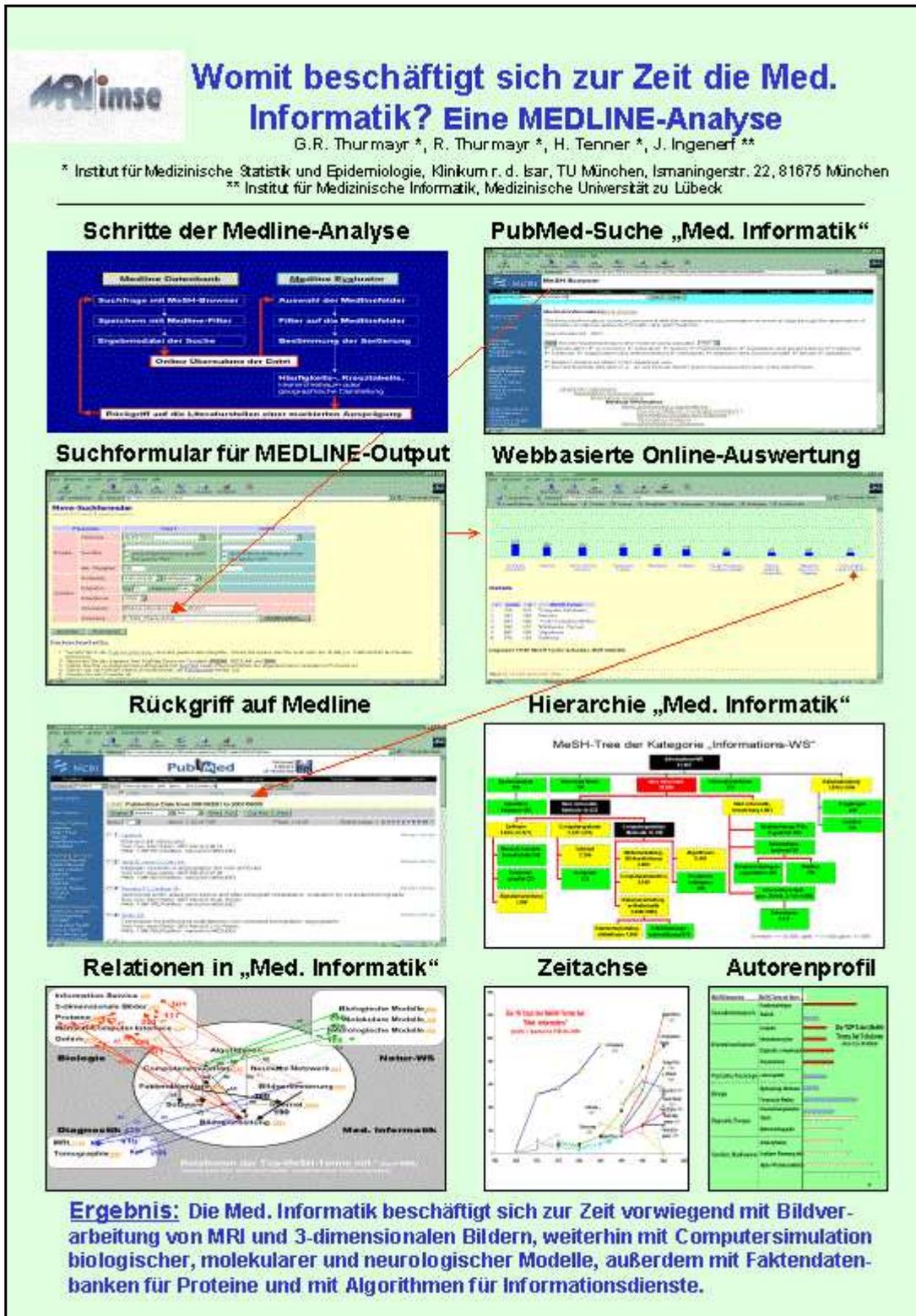


Abb. 27: Poster zur GMDS-Tagung 2002, Berlin

5.4 Technische Grenzwerte von Meva

Meva stellt durch die Verwendung hochperformanter Algorithmen ein sehr leistungsfähiges System dar. Jedoch kann es nicht immer seine Vorteile ausspielen, da es Rücksicht auf die beteiligten Systeme Webserver, Webclient und Netzwerk nehmen muß, um deren Überlastung vorzubeugen. Aus diesem Grunde sind maximale Belastungswerte für die ein- und ausgehenden Datenströme von Meva definiert, die im folgenden beschrieben werden.

5.4.1 Eingabemaxima

Meva definiert bestimmte Eingabegrößenmaxima, die einer Überlastung des Netzwerks und des Webserver vorbeugen sollen. Manche dieser Werte können vom Webadministrator mit eigenen Angaben übersteuert werden.

Sendet der Benutzer z.B. eine sehr große Datei an Meva zur Analyse, verweigert Meva den Dienst. In diesem Fall hat der Benutzer jedoch die Möglichkeit, alle nicht benötigten Daten (wie z.B. Abstracts) aus der PubMed-Resultatdatei zu entfernen, um die Dateigröße zu verringern.

Dazu wurde neben Meva ein zweites Programm namens *MePrep* (*Meva Preprocessor*) entwickelt (s. Kapitel [MePrep – Meva Preprocessor](#)). Die Kompression muß auf dem lokalen Benutzerrechner ausgeführt werden und erreicht in der Regel eine Reduktion auf durchschnittlich 1 - 20% der ursprünglichen Dateigröße. Ist Meva beispielsweise auf ein Dateimaximum von fünf MB eingestellt, können dennoch virtuelle 25 – 500 MB an Meva zur Verarbeitung geschickt werden. Dies ist die generell empfohlene Vorgehensweise bei Dateigrößen über fünf MB und bei wiederholten Anfragen mit denselben Daten, aber unterschiedlichen Parametern, da es die Netzlast reduziert und die Verarbeitungsgeschwindigkeit des Dienstes erhöht.

5.4.2 Ausgabemaxima

Unabhängig von der Größe der Eingabedatei definiert Meva auch Ausgabegrößenmaxima, da in vielen Fällen die zurückgelieferte HTML-Datei so groß geriete, daß Webbrowser sie nicht mehr vernünftig darstellen würden. Dies kann passieren, wenn entweder Mevas Eingabedatei vorher mit MePrep hoch komprimiert wurde oder aber eine Datei mit hochkombinanten Feldern, z.B. mit MeSH-Termen und Autoren, zwar Mevas Eingabemaximum unterschritt, jedoch die Anzahl der darzustellenden Felder im Meva-Resultat kombinatorisch explodieren würde. Im Text-Modus trifft diese Beschränkung jedoch nicht zu.

5.5 MePrep - Meva Preprocessor

5.5.1 Einführung

Meva verweigert ab einer bestimmten, vom Administrator definierbaren Eingabedateigröße die Verarbeitung, um das Netzwerk nicht zu überlasten. In diesem Fall kann der Benutzer jedoch die Daten mit *MePrep*, dem *Meva Preprocessor*, lokal auf seinem Rechner vor dem Absenden an Meva komprimieren.

Die Kompression erreicht in der Regel eine Reduktion auf durchschnittlich 1 - 20% der ursprünglichen Dateigröße, wenn nur eins bis zwei Felder extrahiert werden, und auf durchschnittlich 30 - 40%, wenn alle Felder außer Abstracts extrahiert werden. Die PubMed-ID wird immer extrahiert. Ist Meva beispielsweise auf ein Dateimaximum von zehn MB eingestellt, kann der Benutzer so dennoch virtuelle 50 – 1000 MB an Meva zur Verarbeitung schicken. Dies wird generell bei sehr großen Dateigrößen oder bei wiederholten Anfragen mit ein und derselben empfohlen, da es die Netzlast reduziert und die Verarbeitungsgeschwindigkeit deutlich erhöht.

Anzumerken ist, daß Meva unabhängig von der Größe der Eingabedatei nur eine Maximalzahl an Feldern ausgibt, da ansonsten die zurückgelieferte HTML-Datei so groß geriete, daß die Webbrowser sie nicht mehr darstellen könnten. Im Text-Modus trifft diese Beschränkung jedoch nicht zu.

5.5.2 Installation

Der Benutzer lädt sich die ausführbare Datei *meprep.exe* (sieben KB) auf seinen Rechner. Eine besondere Installation ist nicht nötig. Er muß lediglich das Programm in ein Verzeichnis seiner Wahl legen, z.B. in das Verzeichnis, in dem er auch die PubMed-Dateien abspeichert.

Falls die von PubMed heruntergeladenen Dateien immer die Dateierweiterung „.fcgi“ besitzen (Vorgabewert von PubMed), kann man sich die spätere Handhabung erleichtern, wenn man im Betriebssystem MePrep als ausführbares Programm für fcgi-Dateien festlegt. Dies führt dazu, daß der Benutzer später nur auf eine fcgi-Datei doppelt klicken muß, um ihre Verarbeitung durch MePrep zu starten. (Als Beispiel nur kurz für Windows-Betriebssysteme: Auf eine fcgi-Datei doppelt klicken, im Dialogfenster „Öffnen mit“ den Punkt „Andere...“ anwählen und den Pfad zu MePrep wählen.)

5.5.3 Bedienung

MePrep muß wissen, aus welcher Datei es welche Felder extrahieren soll, und in welche Datei das Resultat geschrieben werden soll. Dies wird dem Programm über vier Parameter mitgeteilt:

1. Name der Eingabedatei, also der von PubMed heruntergeladenen Datei, z.B. `query.fcgi`.
2. Kürzel des 1. Suchfeldes, z.B. `AD` für *Affiliation*. Läßt man den Parameter frei, nimmt MePrep standardmäßig an, daß man lediglich die *Abstracts* entfernen will, d.h. alle anderen Felder werden kopiert. (Abstracts können in Meva ohnehin nicht abgefragt werden, belegen aber viel Platz.) Trotz der Reduktion der Dateigröße kann man mit dieser Datei Meva also immer noch zu allen Feldern konsultieren.
3. Kürzel des 2. Suchfeldes, z.B. `SO` für *Source*. Der Defaultwert ist hier `NONE`, d.h. es wird kein zweites Feld kopiert.
4. Name der Ausgabedatei, also der Datei, die dann als Eingabe für Meva dient, z.B. `ad-so.txt`. Standardmäßig wird der Name der Eingabedatei verwendet, vermehrt um die Feldkürzel (die Feldkürzel finden sich im [Anhang B: MEDLINE-Felder](#) aufgelistet).

5.5.3.1 Interaktive Bedienung

Der Benutzer zieht die Datei, die MePrep komprimieren soll, per Drag and Drop auf das Programm (oder klickt doppelt auf die Datei, sofern er es in der Installation so eingerichtet hat - s.o.). MePrep wird damit gestartet und fragt die ausstehenden Parameter ab: Der Benutzer drückt entweder die ENTER-Taste, um die in eckigen Klammern von MePrep angezeigten voreingestellten Werte zu übernehmen, oder er gibt eigene Werte an.

In PubMed wurde nach Veröffentlichungen zu *Multiple Sclerosis* im Jahr 2002 gesucht und das Ergebnis unter dem Dateinamen `MS_2000.fcgi` abgespeichert. Diese Datei soll nun als Eingabedatenstrom für MePrep dienen, aus dem eine komprimierte Ausgabedatei für Meva erstellt werden soll. (Hinweis: `File to read [query.fcgi]`: wird nicht abgefragt, wenn die Datei im File Explorer doppelt geklickt oder auf MePrep gezogen wird.)

Nachfolgend drei unterschiedliche Szenarien mit ihren entsprechenden Kompressionsraten:

5.5.3.1.1 Beispiel 1 - Extraktion aller Feldtypen außer Abstracts

Im ersten Beispiel werden alle Felder außer Abstracts extrahiert. Dies ist sinnvoll, wenn man mit ein und derselben (komprimierten) Datei mehrere Anfragen an Meva nach unterschiedlichen Feldern richten möchte. Die Kompression ist allerdings nicht so gut, als wenn spezifische Felder extrahiert würden (hier im Beispiel 44%):

```
* Meva Preprocessor V1.0. Type meprep -? for help.

File to read [query.fcgi]: MS_2000.fcgi
Field 1 [Delete only AB's]:
File to write [MS_2000.meva]:

Deleted only AB's.
Compressed 10238kB of MS_2000.fcgi onto 4576kB (44%) in MS_2000.meva.

Press a key to continue..
```

5.5.3.1.2 Beispiel 2 - Extraktion zweier Feldtypen

Im zweiten Beispiel werden nur MeSH-Terme (MH) und Autoren (AU) extrahiert, der Rest wird verworfen. Die Kompression ist mit 15% besser als im Beispiel 1. Wie man sieht, schlägt MePrep einen Zieldateinamen vor, der zur schnelleren Identifikation die Namen der extrahierten Feldtypen angehängt bekommt:

```
* Meva Preprocessor V1.0. Type meprep -? for help.

File to read [query.fcgi]: MS_2000.fcgi
Field 1 [Delete only AB's]: MH
Field 2 [NONE]: AU
File to write [MS_2000-MH-AU.meva]:

Counted 3908 records.
Extracted 41488 'MH's.
Extracted 17267 'AU's.
Compressed 10238kB of MS_2000.fcgi onto 1586kB (15%) in MS_2000-MH-
AU.meva.

Press a key to continue..
```

5.5.3.1.3 Beispiel 3 - Extraktion eines Feldtyps

Im letzten Beispiel werden nur die *Länder* (CY) extrahiert. Die Kompression ist jetzt mit 100:1 hervorragend, allerdings können mit dieser Datei nur noch Anfragen nach Ländern an Meva gerichtet werden. Sollen andere Feldtypen abgefragt werden, muß MePrep erneut mit der Originaldatei gestartet werden:

```
* Meva Preprocessor V1.0. Type meprep -? for help.

File to read [query.fcgi]: MS_2000.fcgi
Field 1 [Delete only AB's]: CY
Field 2 [NONE]:
File to write [MS_2000-CY.meva]:

Counted 3908 records.
Extracted 3853 'CY's.
Compressed 10238kB of MS_2000.fcgi onto 127kB (1%) in MS_2000-CY.meva.

Press a key to continue..
```

5.5.3.2 Automatische Bedienung

Alternativ kann der Benutzer die Parameter auch direkt auf der Kommandozeile (MS-DOS-Eingabeaufforderung) übergeben; fehlende Parameter fragt MePrep nach. Die Eingabe von `meprep -?` auf der Kommandozeile druckt einen kurzen Hilfetext, der die erforderlichen Parameter noch einmal darstellt:

```
Meva Preprocessor V1.0.

Usage: meprep [-?] [infile pubmed_field1 pubmed_field2 outfile]

Example to extract MeSH Terms and Authors from query.fcgi into mh-au.txt:

    meprep query.fcgi MH AU mh-au.txt

For missing parameters meprep will ask interactively.
```

5.6 Administration von Meva

Das folgende Kapitel führt in die Administration von Meva ein. Insbesondere wird ein Augenmerk auf die Installation und Wartung von Meva im Zusammenhang mit dem erforderlichen Webserver gelegt.

5.6.1 Installation

5.6.1.1 CGI

Meva muß mit seinen assoziierten Pflichtdateien (Konfigurationsdatei und Indexdateien) in ein Verzeichnis mit Web- und Systemausführungsrechten kopiert werden (in der Regel *cgi* oder *cgi-bin*). Evt. müssen Konfigurationsdatei und HTML-Formulare angepaßt werden bzw. Indexdateien erstellt werden, sofern sie noch nicht vorhanden sind. Dies wird nachfolgend spezifiziert.

5.6.1.2 Counter-Datei

Meva sucht beim Start nach einer Counter-Datei namens *meva.cnt*. Existiert diese Datei noch nicht im aktuellen Pfad, legt Meva eine dort an. Kann sie nicht angelegt werden, gibt Meva einen Fehler aus.

5.6.1.3 Logdatei

Ist Logging aktiviert, schreibt Meva die Benutzerzugriffe in *meva.log* mit. Das Format lehnt sich an [NCSAs Combined Format](#) an, zusätzlich werden Benutzerdatengröße, Browsersprache und verlangte Feldtypen ausgegeben. Über `?cmd=1` kann das Log formatiert ausgegeben werden. Ist ein Paßwort xyz gesetzt, kann via `?cmd=1&pwd=xyz` das Log ausgegeben werden.

5.6.1.4 Konfigurationsdatei

Meva sucht beim Start nach einer Datei namens *meva.ini*. Existiert diese Datei nicht im aktuellen Pfad, ignoriert Meva dies und nimmt Standardeinstellungen; es wird keine Warnung ausgegeben (Tab. 7).

Eintrag	Pflicht?	Default	Bedeutung
trailer	Nein	Leer	Trailer, der von Meva an den Schluß des Ergebnisses gehangen wird. Kann z.B. für ein Betreiberlogo genutzt werden. Achtung: alle URIs müssen hier absolut formuliert werden! Bsp.: trailer=<p></p>
maxin	Nein	Mevas internes Maximum	Maximale KB, die Meva vom Benutzer einliest. Dieser Wert kann allerdings nicht Mevas internes Maximum (ablesbar unter der Rubrik „Etc“ im Diagnostikmodus) überschreiten. Insbesondere wenn Netzwerkverkehr eine kritische Größe darstellt, kann hier Meva gedrosselt werden. Bsp. für 2000 KB: maxin=2000
log	Nein	0	Falls 1, loggt Meva die Benutzerzugriffe in meva.log mit. Bsp.: log=1
pwd	Nein	Leer	Gibt ein erforderliches Paßwort für Diagnostik- und Logauswertungsmodus an. Bsp.: pwd=xyz

Tab. 7: Konfigurationsparameter von Meva

5.6.1.5 Indexdateien

Meva sucht beim Start nach den Dateien namens *meva-ms.ndx* und *meva-mc.ndx*, sofern der Benutzer MeSH-Codes oder den MeSH-Baum ausgegeben haben will. Eine erfolglose Suche wird mit einer Fehlermeldung quittiert. Sind die Optionen zur MeSH-Ausgabe nicht angewählt worden, werden hingegen fehlende Index-Dateien ignoriert.

Zur Installation und jährlichen Aktualisierung dieser Dateien wird Mevas Indexgenerator benötigt.

5.6.1.6 HTML-Dateien

Das HTML-Dateiset kann in ein beliebiges Webverzeichnis gestellt werden, typischerweise mit Nur-Lese-Rechten. Existieren die HTML-Dateien als mehrsprachige Varianten, muß beim Webserver das Servermerkmal *Content Negotiation* aktiviert werden. Die URL-Referenzen (*Uniform Resource Locator*) des

Action-Parameters in den HTML-Formularen müssen auf den Pfad des CGI entsprechend angepaßt werden.

5.6.1.7 Kompatibilität

Meva wurde mit Apache unter Windows NT, Windows 2000 und Linux getestet, und mit OmniHTTPD unter Windows NT und Windows 2000.

Ein Webserver sollte in der Regel das aktuelle Verzeichnis beim Starten eines CGI auf den CGI-Pfad legen. Experimentelle Webserver bzw. nur rudimentär implementierte Webserver können dies versäumen, Meva fände dann seine assoziierten Dateien nicht.

Einige ältere HTTP/1.0-Webserver bzw. nur rudimentär implementierte Webserver verstehen den MIME-Typ *multipart/form-data* entsprechend RFC 1867 nicht, den Meva zum Hochladen der PubMed-Datei benötigt. Meva wird dann die Ausführung mit einer entsprechenden Fehlermeldung quittieren. Dies trifft auch für ältere HTTP/1.0-Webclients bzw. nur rudimentär implementierte Webclients zu.

5.6.2 Wartung

5.6.2.1 Diagnostikmodus

Die wichtigsten Parameter, die Meva steuern, kann man durch den Aufruf von Meva in der URL-Leiste des Browsers mit dem Parameter `?cmd=d` abfragen. Ist ein Paßwort *xyz* gesetzt, kann via `?cmd=d&pwd=xyz` die Diagnose ausgegeben werden. So lassen sich oft schnell die Ursachen von Problemen finden. Ausgegeben werden die Versionsnummern, die CGI-Umgebung, der Inhalt der Konfigurations- und der Counter-Datei u.a. Daten.

Aus Sicherheitsgründen sollte dieser Aufruf dem Webadministrator vorbehalten bleiben, die Benutzer sollten diesen Zugang nicht kennen oder nutzen können.

5.6.2.2 Systemmeldungen

5.6.2.2.1 Mevas Fehlermeldungen

Fehler führen zum Abbruch des Programms und werden in client- (Tab. 8) und serverseitige Fehler (Tab. 9) unterschieden. Alle Fehler werden dem Benutzer gemeldet, Serverfehler werden zusätzlich in das Serverlog geschrieben (Apache).

Code	Meldung	Bedeutung
204	No Content	Leere Anfrage des Client
400	CONTENT_TYPE missed	Fehlerhafter Client
400	QUERY_STRING missed	Fehlerhafter Client
400	CONTENT_LENGTH does not match length of data	Fehlerhafter Client
401	Unauthorized	Falsches Paßwort.
405	Method not allowed	Fehlerhafter Client
411	Length required	Fehlerhafter Client
413	Request too large	Clientdaten zu groß
415	CONTENT_TYPE unknown	Fehlerhafter Client
415	Boundary missed	Fehlerhafter Client
415	Preamble line (1) missed	Fehlerhafter Client
415	Invalid preamble	Fehlerhafter Client
415	Content-Disposition line (2) missed	Fehlerhafter Client
415	Content-Disposition field name missed	Fehlerhafter Client
415	Line (3) after Content-Disposition missed	Fehlerhafter Client
415	Blank line (4) missed	Fehlerhafter Client
415	Field line missed	Fehlerhafter Client
415	Closing pre- or postamble missed	Fehlerhafter Client
415	HTML form V ... required	Falsche Formular-Version
415	Form variable ... missing	Ungültiges Meva-Formular
415	... value unknown	Ungültiges Meva-Formular
415	I got no PubMed file	Leere/fehlende PubMed-Datei
415	Wrong syntax in search filter	Fehlerhafte Suchfiltersyntax

Tab. 8: Mevas Clientfehlermeldungen

Häufig auftretende Clientfehler können „Request too large“ (Zu große PubMed-Datei) oder „HTML form V ... required“ (falsches Konsultationsformular) sein.

Mevas Serverfehler sind nur durch den Administrator zu beheben (Tab. 9).

Code	Meldung	Bedeutung
500	Cannot open meva.cnt	Counter-Datei nicht lesbar oder anlegbar
500	Bad ini syntax	Konfigurationsdateisyntax fehlerhaft
500	Cannot allocate I/O buffer	Interner Server-Fehler
500	Cannot switch into binary mode	Interner Server-Fehler (nur Windows)
500	Cannot open file ...ndx	Mevas Indexdatei kann nicht geöffnet werden.
500	Bad file version: ...ndx	Mevas Indexdatei entstammt einer älteren Version oder ist korrupt
500	Corrupted file: ...ndx	Mevas Indexdatei ist korrupt
500	Different index files	Mevas Indexdateien entstammen unterschiedlichen Jahren
501	This server declares the POST data to be URL encoded, but data start with '--' like multipart (RFC1867). In case you did not use '--' as a variable name, this server is either pretty old or buggy.	Interner Serverfehler oder älterer HTTP/1.0-Webserver

Tab. 9: Mevas Serverfehlermeldungen

Ein häufig auftretender Serverfehler kann "Internal Server error" (unspezifizierter Fehler des Webservers) sein.

5.6.2.2.2 Mevas Warnungen

Mevas Warnungen werden dem Benutzer ausgegeben. Sie führen nicht zum Abbruch des Programms, indizieren aber eine grenzwertige Konstellation (Tab. 10).

Meldung
Sie haben viele Felder erhalten. Versuchen Sie, Filter oder Minimale Häufigkeiten zu verschärfen.
Sie haben zu viele Felder erhalten. Die Details wurden deaktiviert. Versuchen Sie, Filter oder Minimale Häufigkeiten zu verschärfen.
Sie haben zu hohe Top-Werte erhalten. Die Korrelationsmatrizen wurden deaktiviert. Versuchen Sie, die Filter zu verschärfen.
Die Feldverknüpfungen wurden infolge fehlenden Linkrestriktors deaktiviert.
Die MeSH-Codes sind aus dem Jahr ... Veranlassen Sie den Administrator, neue Codes einzuspielen.

Tab. 10: Warnungen von Meva

Wie neue MeSH-Codes entsprechend der letzten Fehlermeldung eingespielt werden, wird unter Punkt [5.6.2.3.5](#) näher beschrieben.

5.6.2.2.3 Hinweise

Alle Hinweise von Meva werden dem Benutzer ausgegeben (Tab. 11). Sie dienen der schnellen Orientierung des Benutzers.

Meldung
Mit einem Klick auf ein PMID-Feld erreichen Sie den Originalartikel in der PubMed®-Datenbank des NCBI.
Mit einem Klick auf ein verlinktes Feld aktivieren Sie eine weitere Suche in PubMed nach dem entsprechenden Begriff und dem Linkrestriktor.
Die Publikationsdaten wurden auf das Jahr reduziert.
Die Sprachangaben wurden auf Kleinschreibung gesetzt.
Die Landesangaben wurden auf Großschreibung gesetzt.

Tab. 11: Hinweise von Meva

5.6.2.3 MeSH2NDX - Meva Index Generator

5.6.2.3.1 Beschreibung

Zur Anzeige der MeSH-Codes benötigt Meva in seinem Verzeichnis entsprechende Indexdateien (*meva-ms.ndx* und *meva-mc.ndx*), die es mit den entsprechenden Codes der gefundenen MeSH-Deskriptoren versorgt. Diese

Indexdateien müssen von dem Programm *mesh2ndx* aus einer MeSH-Datei erstellt werden. Die aktuelle MeSH-Datei („*MeSH Trees*“) können vom Administrator von der NLM heruntergeladen werden. Das jährliche Update der Indexdateien ist in technischer Hinsicht erforderlich, wenn Meva stets alle Codes zur Verfügung haben soll, und in legaler Hinsicht infolge des Lizenzvertrages, der beim Herunterladen der MeSH-Codes mit der NLM eingegangen wird und der hier auszugsweise wiedergegeben wird:

„At least annually, the data in the downloaded product/service containing searchable MEDLINE-derived data must be corrected to incorporate the corrections NLM has made to these records during the year and in year-end maintenance. NLM data in the product/service must be carefully checked with that maintained on NLM’s version of the database at the end of each calendar year and corrections made. This does not apply to products/services containing MEDLINE-derived data which itself is not searchable (e.g., MEDLINE-derived references and abstracts in the bibliography of an electronic textbook). The MEDLINE-derived data should be updated when a new edition of the product is released. In this case, users should be informed that the MEDLINE-derived data may change.“

Zusätzlich wird eine Datei mit den Hauptkategorien des MeSH-Baums benötigt.

5.6.2.3.2 Aufbau der Code-Dateien

5.6.2.3.2.1 MeSH-Code-Datei

2002 bestand die MeSH-Code-Datei aus 37.862 Einträgen, die wie folgt aufgebaut waren:

```
1,2-Dimethylhydrazine;D02.442.600.400.200
1,2-Dipalmitoylphosphatidylcholine;D10.516.570.755.375.760.400.800.224
...
zeta Carotene;D02.455.849.131.912
zeta Carotene;D11.557.261.925
```

Die MeSH-Code-Datei muß aus Zeilen bestehen, die aus einem MeSH-String bestehen, gefolgt vom Semikolon und dem entsprechenden MeSH-Code. Ein MeSH-Code beginnt mit dem Buchstaben der Hauptkategorie, gefolgt von weiteren Hierarchiestufen, die jeweils durch drei oder zwei Ziffern dargestellt werden, getrennt durch einen Punkt.

5.6.2.3.2.2 MeSH-Kategorien-Datei

Für 2002 waren die 15 Kategorien des MeSH-Baums wie folgt:

```
Anatomy;A
Organisms;B
Diseases;C
Chemicals and Drugs;D
Analytical, Diagnostic and Therapeutic Techniques and Equipment;E
Psychiatry and Psychology;F
Biological Sciences;G
Physical Sciences;H
Anthropology, Education, Sociology and Social Phenomena;I
Technology and Food and Beverages;J
Humanities;K
Information Science;L
Persons;M
Health Care;N
Geographic Locations;Z
```

Jede Zeile der MeSH-Kategorien-Datei muß aus dem Kategoriennamen bestehen, gefolgt vom Semikolon und dem Kategorienbuchstaben.

5.6.2.3.3 Installation des Indexgenerators

Eine Installation des Indexgenerators ist nicht notwendig. Das Programm kann von jedem beliebigen Ort ausgeführt werden.

5.6.2.3.4 Programmstart des Indexgenerators

Der Administrator startet eine Konsole. Er führt folgenden Befehl in der Konsole aus:

```
mesh2ndx <mesh_code_file> <mesh_main_code_file> <mesh_release_year>
```

Beispiel: Der Administrator hat 2002 die aktuelle MeSH-Code-Datei der NLM heruntergeladen und mesh2002.txt genannt. Um nun Indexdateien für Meva daraus zu erstellen, gibt er folgendes ein:

```
mesh2ndx mesh2002.txt main.txt 2002
```

Anmerkung: Der Aufruf des Programms mit

```
mesh2ndx /?
```

gibt außerdem eine weitere Hilfestellung aus.

5.6.2.3.5 Sicherstellung der Aktualität der MeSH-Codes

Das Ausgabejahr der MeSH-Codes, mit denen die Indexdateien erstellt wurden, kann festgestellt werden, indem in einer DOS-Konsole der Befehl

```
type <index_file>
```

oder in einer UNIX-Konsole der Befehl

```
head -n 1 <index_file>
```

eingeben wird, der folgende o.ä. Ausgabe zur Folge hat:

```
Meva NDX file V.1, 2002
```

Ersichtlich sind Indexversion und Jahrgang.

Meva erzeugt eine Benutzerwarnung, wenn Indexdateien verwendet werden, die nicht dem aktuellen Jahr entsprechen. Werden Indexdateien zwischenzeitlich korrumpiert oder entstammen nicht demselben Jahrgang, bricht Meva die Ausführung mit der Fehlernummer 500 ab.

5.7 Probleme und deren Lösungen in der Arbeit mit Meva

Durch fehlerhafte Einstellungen des Browsers oder des Webservers können Probleme in der Arbeit mit Meva entstehen. Einige der häufigeren werden mit ihren Lösungen nachfolgend beschrieben.

5.7.1 Falscher Rückgabedatentyp

5.7.1.1 Auftreten

Falls der Benutzer das Formular mit der Option Text abschickt, kann es passieren, daß der Browser kein Resultat anzeigt, sondern fragt, ob die Datei geöffnet oder gespeichert werden soll. Dies wird verursacht, wenn der Browser in der Standardeinstellung annimmt, daß durch CGIs generierte Daten vom Typ *application/octet-stream* sind, statt den MIME-Typ der Kopfzeilen auszuwerten.

5.7.1.2 Lösung

Um dies auszuschalten, muß die Option

Bestimme Aktion durch MIME-Typ

im Browser eingestellt werden.

5.7.2 Zu großes Resultat

5.7.2.1 Auftreten

Der Benutzer kann bei der Konsultation die Fehlermeldung *413 Request entity too large* erhalten. In diesem Fall hat er die erlaubte Eingabegröße überschritten. Meva erlaubt keine beliebig großen Anfragen, um das Netzwerk nicht zu überlasten. Es ist meist auch wenig sinnvoll, solche rohen Datenmengen zu analysieren.

5.7.2.2 Lösung

Lösungsansätze sind:

- PubMed-Abfrage mit differenzierteren Suchmustern wiederholen, um weniger an Daten zu erhalten
- MePrep zur Kompression verwenden.

5.7.3 MS-Access-Fehlermeldung

5.7.3.1 Auftreten

Importiert der Benutzer ein im Textformat zurückgeliefertes Resultat in MS-Access, kann er die Fehlermeldung bekommen: *Das Microsoft-Jet-Datenbankmodul konnte das Objekt xxx nicht finden oder der Tabellename ist ungültig.*

5.7.3.2 Lösung

Dies ist ein Fehler im MS-Access 2000. Der Benutzer muß in diesem Fall die Datei im Browser mit maximal einem Punkt im Dateinamen abspeichern, z.B. `result.txt`.

5.7.4 Fehlende Grafiken

5.7.4.1 Auftreten

Bei manchen Benutzern sind die Balken oder die Hintergründe von Tabellen im Ausdruck des Meva-Resultates verschwunden.

5.7.4.2 Lösung

Bei fehlenden Grafiken muß im Webbrowser Druck von Hintergründen aktiviert werden. Einstellbar bei Opera unter

Datei | Druckoptionen,

bei Internet Explorer unter

Extras | Internetoptionen | Erweitert | Drucken | Hintergrundfarben und -bilder drucken

und bei Netscape unter

Datei | Seite einrichten.

Netscape V.4 hat dennoch massive Schwierigkeiten damit, hier empfiehlt sich ohnehin der Umstieg auf einen modernen und halbwegs fehlerfreien Browser, z.B. Opera V.6 oder Internet Explorer V.6.

5.7.5 Zerrissener Text im Ausdruck

5.7.5.1 Auftreten

Manche Benutzer erhalten im Ausdruck unregelmäßige Schriftgrößen. Der Text schaut zerissen aus, teils groß, teils klein.

5.7.5.2 Lösung

Ein unregelmäßig ausschauender Text kommt u.a. beim Browser Netscape 4 vor, der enorme Schwierigkeiten hat, CSS korrekt zu verarbeiten und Bugs im Druckertreiber aufweist. Wie schon vorher erwähnt, hilft hier nur der Umstieg auf einen halbwegs aktuellen Browser, der HTML 4 und mindestens CSS 1 beherrscht.

5.7.6 Zerrupftes Tabellenformat im Textresultat

5.7.6.1 Auftreten

Wählt der Benutzer als Datenformat Text aus, können die Daten, die Meva zurückschickt, zerrupft ausschauen statt tabellarisch.

5.7.6.2 Lösung

Eine Lösung für ein zerrupft wirkendes Datenformat ist nicht erforderlich. Die Felder der zurückgeschickten Textdatei sind mit Tabulatoren getrennt, die Browser arbeiten aber nur mit festen und nicht mit benutzerdefinierbaren Tabulatorpositionen. Dies schränkt aber die Verwertbarkeit der Daten nicht ein: jedes Datenbanksystem liest diese Daten korrekt ein. Der Benutzer kann aber auch ohne Datenbanksystem eine optisch korrekte Darstellung der Daten erreichen, indem er diese in ein beliebiges Textverarbeitungsprogramm lädt, das die Adjustierung der Tabulatorpositionen erlaubt.

5.7.7 Falsche Sprache der Webpräsenz

5.7.7.1 Auftreten

Formular, Resultat und die anderen HTML-Dateien können für einen deutschen Benutzer in Englisch erscheinen und umgekehrt.

5.7.7.2 Lösung

Die möglichen Gründe für eine falsche Sprachauswahl und ihre Lösung werden nachfolgend beschrieben (aufgelistet nach der Häufigkeit des Auftretens):

5.7.7.2.1 Falsch konfigurierter Browser

Webserver und Meva bestimmen die Sprache nach der im Browser des Benutzers eingestellten Landessprache. Diese ist oft falsch eingestellt. Bei den meisten Browsern kann dies jedoch auch nachträglich durch den Benutzer in einer Liste geändert werden:

Bei Opera kann dies eingestellt werden unter

Datei | Einstellungen | Sprachen,

bei Netscape unter

Bearbeiten | Einstellungen | Navigator | Sprachen

und bei Internet Explorer unter

Tools

oder unter

Extras | Internet-Optionen | Allgemein | Sprachen.

Ein Deutsch Sprechender sollte mindestens ein „de“ in dieser Liste haben. Landeserweiterungen wie z.B. in „de-CH“ sollten auf keinen Fall hinzugefügt werden! Fordert nämlich der Browser z.B. eine Datei in „de-CH“ an (Schweizer Deutsch) und der Server hat diese Datei nur in „de“, wird er dem Benutzer nicht die deutsche Variante schicken, sondern einen Fehler 413 zurückgeben. Dies klingt unlogisch, ist aber die vorgeschriebene Vorgehensweise, wie sie im entsprechenden RFC [↗ [RFC2616](http://www.ietf.org/rfc/rfc2616.txt)²⁵] definiert ist. Einige Webbrowser wie IE und andere stellen irrtümlicherweise bei der Installation eine Landeserweiterung ein (abhängig von den Spracheinstellungen des zugrundeliegenden Betriebssystems des Benutzers). In diesem Fall sollte dieser Eintrag gelöscht werden!

Außerdem ist eine gute Idee, als Deutschsprachiger neben Deutsch („de“) Englisch („en“) als letzte Sprache einzutragen, anderenfalls werden dem Benutzer manche Webserver ohne Fallback-Sprache den Fehler 413 *Not acceptable* zurückgeben, falls das vom Benutzer angeforderte Dokument nicht in Deutsch vorhanden ist.

Meva hat hingegen eine Fallback-Sprache (englisch) installiert: Scheitert die automatische Verhandlung der Sprachauswahl zwischen dem Browser eines deutschen Benutzers und dem Webserver, weil z.B. in seinem Browser „de-DE“ als einzige Sprache eingetragen ist, der Webserver jedoch nur „de“ als Sprache anbietet, greift der Server auf die Fallback-Sprache zurück und liefert dem Benutzer die angeforderte HTML-Seite in englisch. Dann kann der Benutzer zumindest händisch die entsprechende deutsche Seite im Menü von Mevas Webseiten anfordern.

5.7.7.2.2 Falsch konfigurierter Proxy

Eine weitere Ursache für eine fehlerhafte Sprachpräsentation ist ein falsch konfigurierter Proxy. Dies kommt immer öfter vor, da viele ISPs (Internet Service Provider) Caching als Mittel zur Reduktion des Webverkehrs einsetzen. In diesem Fall kann der Benutzer den Proxy im Browser ausschalten.

²⁵ <http://www.ietf.org/rfc/rfc2616.txt>

Ist hingegen ein transparenter Proxy des ISP die Fehlerursache, verfängt dies nicht. In jenem Fall kann der Benutzer jedoch immer noch händisch seine Sprache im Menü von Mevas Webseiten auswählen.

5.7.7.2.3 Falsch konfigurierter Server

Eine fehlerhafte Sprachpräsentation kann auch durch den Webserver bedingt sein, wenn das Servermerkmal *Content Negotiation* durch den Webadministrator falsch installiert wurde. In diesem Fall kann der Benutzer entweder händisch seine Sprache aus dem Menü von Mevas Webseiten auswählen oder die am meisten passende Repräsentation aus der Auswahlliste selektieren, die der Server mit einem Fehler 413 mitsendet.

5.7.8 Skriptmeldungen

5.7.8.1 Auftreten

Manche Browser, sofern mit strengeren Sicherheitsrichtlinien ausgestattet, stellen den Benutzer beim Laden des Formulars die Frage: „Skripte sind normalerweise sicher. Soll die Ausführung von Skripten zugelassen werden?“

5.7.8.2 Lösung

Die Frage des Skripts kann mit „Ja“ beantwortet werden. Das im Formular enthaltene Skript überprüft lediglich, ob die Pflichtfelder (z.B. der Dateiname) ordnungsgemäß angegeben wurden. Das Formular würde auch ohne Skript funktionieren. Meva finge diese Benutzerfehler auch ab, allerdings auf Kosten des Netzwerkverkehrs. Das Skript gibt darüber hinaus bestimmte Warnungen bzw. Hinweise aus und arbeitet rein lokal, d.h. das Netzwerk wird nicht unnötig belastet.

Steht der Meva-Server z.B. im lokalen Intranet, kann der Benutzer das stete Nachfragen des Browsers auch verhindern, indem er den Meva-Server als vertrauenswürdigen oder lokalen Server im Browser einträgt und ggf. die Sicherheitsstufe anpaßt. Für den Internet Explorer z.B. unter

Extras|Internetoptionen|Sicherheit|Sites (Intranet)

bzw. unter

Extras|Internetoptionen|Sicherheit|Stufe anpassen (Intranet)|Scripting|Active Scripting

5.7.9 Lange Sendezeit

5.7.9.1 Auftreten

Manche Browser benötigen sehr lange, ehe sie die Daten an Meva abgesandt haben.

5.7.9.2 Lösung

Für die lange Sendezeit gibt es drei mögliche Ursachen: Entweder will der Benutzer eine sehr große Datei an Meva schicken, oder er hat eine langsame Verbindung oder der Server ist überlastet. Im ersten Fall gibt es Abhilfe: Er kann die Daten mit MePrep vor dem Absenden um den Faktor 5 - 100 komprimieren. Dies beschleunigt die Anfrage.

5.7.10 Lange Reaktionszeit

5.7.10.1 Auftreten

Manche Browser benötigen sehr lange, ehe sie das Meva-Resultat darstellen.

5.7.10.2 Lösung

Speziell wenn der Benutzer Felder kombiniert, die mehrfach in einem Datensatz auftreten (wie z.B. MeSH-Terme und Autoren), explodiert die Zahl der zurückgelieferten Tabelleneinträge kombinatorisch: das Volumen des zu sendenden Resultates wächst damit sehr schnell. Viele Browser können Daten, die größer als ein Megabyte sind, nur sehr langsam darstellen; sogar Browserabstürze können auftreten. Mögliche Lösungsansätze sind:

- Verwendung von *Filtern* im Meva-Formular
- Verschärfung der *Minimalen Häufigkeiten* im Meva-Formular
- Auslassung des *Linkrestriktors* im Meva-Formular, dies inaktiviert die Links, reduziert somit die zurückgelieferte Datenmenge und die

Antwortzeit des Browsers

- Auslassung der MeSH-Codes im Meva-Formular, insbesondere wenn auch kein MeSH-Baum vom Benutzer angefordert wurde
- Reduktion der Darstellungstiefe oder Auslassung des MeSH-Baums im Meva-Formular
- Auslassung des zweiten Suchfeldes im Meva-Formular
- Auslassung der *PMID*-Anzeige im Meva-Formular, falls nur ein Suchfeld ausgewählt wurde; dies verringert deutlich die Größe des Resultates
- Auslassung der *Details* im Meva-Formular
- Umschaltung auf Textformat im Meva-Formular

Dessen ungeachtet kann eine verzögerte Browserreaktion auch an der niedrigen Performanz des Benutzersystems liegen. Jedoch greifen auch in diesem Fall die o.g. Punkte zur Beschleunigung.

Unabhängig von den aufgeführten Maßnahmen kann auch das Volumen der initialen PubMed-Datei reduziert werden, um eine reduzierte Eingangsdatei für Meva zu erhalten und damit eine reduzierte Meva-Lösung. Ist man an Artikeln zu Cholesterin im Zusammenhang mit hohem Blutdruck interessiert, sollte man in PubMed statt einer undifferenzierten Suche nach `Cholesterol` eine Suche nach `Hypertension AND Cholesterol` durchführen.

6 DISKUSSION

In der Diskussion wird die aufgezeigte Lösung zur Datenverdichtung und Präsentation von Suchergebnissen bei der medizinischen Literatursuche in MEDLINE – der Einsatz von Meva als *Data-Mining*-Werkzeug – besprochen.

Vor- und Nachteile der Lösung werden einzeln analysiert, des weiteren wird die Eignung von Meva auch für andere Datenbanken als PubMed kritisch untersucht.

6.1 Vorteile von Meva

Das nachfolgende Kapitel unterstreicht noch einmal die Vorteile, die Meva dem Benutzer durch die Automatisierungs- und Verdichtungsaspekte bietet.

6.1.1 Allgemeine Vorteile

Durch die zentrale Vorhaltung und Pflege des Dienstes durch den Webadministrator kann Meva ökonomisch und einheitlich für alle Benutzer konfiguriert werden. Dies betrifft z.B. die Einspielung jährlich neuer MeSH-Terme oder die Festlegung von Eingabemaxima für Benutzer.

Dasselbe gilt für Änderungen am Code von Meva: muß der Code von Meva angepaßt werden, benötigt man keine Änderung auf jedem Benutzerrechner, sondern nur eine Änderung des Dienstes zentral auf dem Webserver.

Durch das Aufsetzen von Meva auf dem kostenlos verfügbaren PubMed-Dienst ist die freie und kostenlose Verfügbarkeit von Meva im Netz gewährleistet: andere MEDLINE-Distributionen als PubMed werden in der Regel kostenpflichtig vertrieben.

Des weiteren ist keine Installation von Meva beim Benutzer notwendig: der Aufruf des Dienstes über einen beliebigen Browser ist alles, was der Benutzer benötigt, um den Dienst in Anspruch zu nehmen.

Der Benutzer kann durch eine Vielzahl an Konfigurationsparametern den Prozeß des *Data Mining* breit steuern und besitzt somit wesentlichen Einfluß auf Form und Inhalt des Analyseresultates.

Meva bietet sich an, um einen schnellen Überblick zu einem Thema zu erhalten. Eine analoge Möglichkeit gibt es in PubMed nicht.

Die Verknüpfung des Meva-Resultates mit PubMed ermöglicht eine bequeme iterative Verfeinerung und Erhöhung der Relevanz der ursprünglichen Suchmenge.

Der Benutzer kann das Analyseresultat ausdrucken, off-line für Präsentationen weiterverwenden oder aber im datenbanktauglichen Textformat abspeichern, um es maschinell weiter zu verarbeiten.

6.1.2 Histogramme

Mevas Histogramme bieten im Gegensatz zu PubMed die Möglichkeit, die Suche nach Informationen auf das Wichtigste zu beschränken (Top-Werte) und die

Daten sortiert nach Häufigkeit und graphisch als Balkendiagramm ausgeben zu lassen. So können sich Benutzer schnell einen Überblick über die wichtigsten Autoren oder Schlagwörter eines interessierenden Gebietes verschaffen. Eine wichtige Anwendung ist z.B. die Frage, welcher Autor die meiste klinische Erfahrung auf einem bestimmten Gebiet hat.

Auch lassen sich durch die Kombination eines bibliographischen Feldes mit dem Publikationsdatum zeitliche Verläufe, z.B. die Evolution und Verwendung von Schlagwörtern oder Autoren darstellen.

6.1.3 Korrelationsmatrizen

Durch Mevas Korrelationsmatrizen kann ein Benutzer quantitative Bezüge zwischen Feldern herstellen – ein Leistungsmerkmal, über welches PubMed leider nicht verfügt. Interessierende Bezüge, um nur einige aufzuführen, können z.B. Land – Schlagwort, Institution – Schlagwort, Institution – Autor, Schlagwort – Publikationsdatum (Zeitachse), Schlagwort – Sprache, Schlagwort – Publikationstyp o.a. Schlagwort – Autor sein.

So lassen sich über die Korrelationsmatrizen auch Fragestellungen wie „Welcher Therapieansatz zeitigte die besten therapeutischen Ergebnisse?“ schnell beantworten.

Auch medizinische Laien können darüber wertvolle Informationen gewinnen, z.B. Fragestellungen „Wo sitzt der Spezialist für ein bestimmtes Krankheitsbild und wie heißt er?“

Aus den Korrelationen zwischen Autor und Schlagwörtern lassen sich des weiteren Autorenprofile ableiten, die zur Einordnung und Positionierung eines Autors in der wissenschaftlichen Landschaft unerlässlich sind.

Autorenprofile sind ebf. interessant, um die medizinischen Forschungsschwerpunkte eines Bewerbers für eine medizinische Forschungsstelle zu eruieren bzw. umgekehrt auch, um als Bewerber die Forschungsschwerpunkte der weiterbildungsermächtigten Lehrkraft zuvor kennenzulernen.

6.1.4 Filter

Durch die Möglichkeit in Meva zur Filterung von Feldinhalten nach bestimmten Ausdrücken kann der Benutzer eine feinere Abstufung der Analyse vornehmen.

Durch die Beschränkung der Analyse auf Erstautoren erlaubt es Meva zudem, die federführenden Autoren der interessierenden Artikel zu extrahieren.

6.1.5 Details

Mevas Detailtabelle erlaubt es, vergesellschaftete bibliographische Felder anzuzeigen und somit Zusammenhänge herzustellen. Im Gegensatz zu den Korrelationsmatrizen wird hier unverdichtet die komplette Datenbasis des PubMed-Resultates ausgegeben.

Dadurch sind auch Statistikauswertungen möglich wie z.B. die Ermittlung der Perzentilen von Häufigkeiten bibliographischer Felder.

6.1.6 MeSH

Durch die spezielle Behandlung, die Meva den Schlagwörtern in der Analyse zukommen läßt, kann die Relevanz einer erneuten Suche in PubMed deutlich gesteigert werden, indem wichtige Schlagwörter für ein Thema im MeSH-Baum identifiziert werden. Durch die klinischen Filter und Systematischen Reviews können in PubMed zwar für quantitative Studien allgemeine Filter gesetzt werden; die Suchbegriffe, die das Thema jedoch abstecken, muß der Benutzer immer noch selbst definieren.

Noch stärker wird die Problematik, die Spreu vom Weizen zu trennen, deutlich, sucht man nach hochwertigen qualitativen Studien, die vorwiegend deskriptiver Natur sind und z.B. die Arzt-Patienten-Beziehung oder Coping-Mechanismen des Patienten thematisieren. Qualitative Forschung spielt eine wichtige Rolle in der evidenzbasierten Gesundheitsfürsorge, indem sie die menschlichen Dimensionen und Erfahrungen des betroffenen Personenkreises widerspiegelt. Solche Studien standen bisher nicht im Zentrum der EBM-Bewegung und sind daher in Bezug auf Klassifikation und Suchmöglichkeiten noch nicht so stark formalisiert wie ein Randomized Control Trial oder Systematic Review für klinisch „harte“ Studien [Evans 2002, S. 290 - 293]. Die Suche nach solcher Literatur gestaltet sich somit ungleich schwieriger. Durch die Möglichkeit in Meva, Felder in Bezug zueinander zu setzen, können jedoch auch für „weiche“ Studien wichtige Suchbegriffe via Meva extrahiert werden, z.B. die Extraktion weiterführender Schlagwörter oder die Bestimmung der wichtigsten Institutionen.

Durch die Anzeige eines kombinierten MeSH-Baumes für alle Schlagwörter eines PubMed-Resultates ist ein großer Vorteil gewonnen: PubMeds MeSH-Browser zeigt immer nur den Baum für ein Schlagwort an. Dadurch bietet sich die Ausgabe des MeSH-Baumes auch für eine statistisch-hierarchische Auswertung von Schlagwörtern zu einem Thema an bzw. zur Präsentation.

6.2 Nachteile von Meva

Die Nachteile von Meva liegen in der Einarbeitungszeit zum Programm begründet. Ohne eine grundlegende Kenntnis der Parameter, die Meva steuern, können die Möglichkeiten des Dienstes nicht ausgereizt werden. In dieser Hinsicht teilt es diesen Nachteil mit jeder Benutzeroberfläche, über die komplexe Vorgänge gesteuert werden können, wie u.a. PubMeds Oberfläche oder die Masken anderer Suchprogramme.

Allerdings ist Mevas Konsultationsformular so aufgebaut, daß auch ohne umfangreiche Parametrisierung der Analyse durch den Benutzer in den meisten Fällen sinnvolle Ergebnisse geliefert werden: die Standardwerte der Formularparameter sind darauf ausgelegt, auch Anfängern schnell zu einem „ersten Erfolg“ zu verhelfen.

6.3 Mevas Anwendbarkeit auf andere biomedizinische Datenbanken

Die Datenbanken *EMBASE*, *Current Contents*, *BIOSIS* und *Science Citation Index* verfügen entweder über kein oder ein zu MEDLINEs MeSH differentes Vokabular. Aus diesem Grunde sind diese Datenbanken für eine Schlagwortauswertung nur sehr eingeschränkt oder gar nicht mit Meva benutzbar.

Prinzipiell gilt, daß jede Datenbank mit Meva benutzt werden kann, solange sie dem Benutzer ein Exportformat im MEDLINE-Format anbietet, da Meva nur dieses Format als Eingabeformat akzeptiert.

Ist das Exportformat einer Datenbank nicht ident zum MEDLINE-Format, jedoch an dieses angelehnt (z.B. DIMDI's Ausgabeformat für MEDLINE-Suchresultate, welches auch deutsche MeSH-Terme und andere Feldnamen beinhaltet), kann eine Konversion des Formats in das MEDLINE-Format über ein Filterprogramm unternommen werden, um eine Kompatibilität zu Meva zu erhalten. Das Filterprogramm kann ein einfacher DOS- [Tischer 1994, S. 816] oder UNIX-Filter [Brown 1994, S. 445] sein, der die Daten vom Standardeingabekanal einliest, transformiert und auf den Standardausgabekanal schreibt.

So ist z.B. das Text-Exportformat der *Cochrane Library* dem MEDLINE-Format sehr ähnlich. Ein *Data Mining* durch Meva in den exportierten Artikeln aus der *Cochrane Library* ist somit möglich, wenn durch einen Konverter die Feldbezeichnungen dieses Exportformats an das MEDLINE-Format angepaßt werden.

Falls die Datenbank über ein ähnliches Exportformat wie das MEDLINE-Format verfügt, jedoch über einen differenten Thesaurus, ist ein nachgeschaltetes *Data Mining* mit Meva möglich, sofern ein entsprechender Konverter die Datentransformation in das MEDLINE-Format übernimmt und der Benutzer auf die Auswertung von Schlagwörtern verzichtet.

7 ZUSAMMENFASSUNG

Im Mittelpunkt der vorliegenden Arbeit standen die Probleme, mit denen ein Anwender konfrontiert ist, der via PubMeds Schnittstelle zu MEDLINE wichtige Erkenntnisse und Lösungsansätze zu medizinischen Fragen gewinnen will.

Aus diesem Grunde wurde eine webbasierte und universelle *Data-Mining*-Lösung namens Meva entwickelt, die es erlaubt, Methoden zur Verdichtung und optischen Aufbereitung von Daten einzusetzen, die PubMed bei einer Recherche liefert. Insbesondere die folgenden Aspekte prädestinieren Meva als Lösung:

Allgemeine Vorteile von Meva sind die freie Verfügbarkeit, die kostenlose Nutzung, die zentrale Vorhaltung und Pflege des Dienstes, die fehlende Notwendigkeit einer Installation beim Benutzer, die breite Steuerbarkeit der Analyse durch den Benutzer sowie die Möglichkeit zur bequemen iterativen Verfeinerung der ursprünglichen PubMed-Lösungsmenge, um damit die Relevanz und die Nachweisquote des Suchresultats zu erhöhen.

Spezifische Vorteile, die Meva als Folge der Verdichtung von PubMed-Daten aufweist, sind u.a. die Anzeige von Histogrammen und Korrelationsmatrizen, die Möglichkeit der Filterung nach bestimmten Ausdrücken oder nach Erstautoren, die Anzeige vergesellschafteter bibliographischer Felder sowie die Darstellung eines MeSH-Baumes für alle gefundenen Felder. So lassen sich z.B. schnell Autorenprofile, Spezialisten auf einem Gebiet oder Therapieansätze erkennen.

Somit ist Meva eine geeignete *Data-Mining*-Lösung im Sinne der Problemstellung.

Anzumerken ist, daß die Möglichkeiten zur Verdichtung durch Meva bei weitem noch nicht ausgeschöpft sind. Wie bei jedem Programm lassen sich immer weitere Optimierungen finden. So wären Verbesserungen denkbar wie z.B. der Einsatz eines artikelbasierten Suchfilters statt eines zeilenweisen, die Anzeige eines vierten Feldes in der Auflistung, die optische Anzeige der räumlichen Verteilung von Feldern (z.B. von Autoren oder publizierenden Institutionen) auf einer Weltkarte, die Sortierung von Zitaten nach Impact-Punkten etc.

Doch ist mit der aktuellen Version von Meva ein deutlicher Fortschritt im Bereich des *Data Mining* innerhalb von MEDLINE gelungen und der Benefit, den der Anwender mit dem Dienst gewinnt, groß genug, um ihn den in der Problemstellung vorgestellten Methoden zur Datenextraktion vorzuziehen.

8 STICHWORTVERZEICHNIS

Abstract 18, 19, 26, 37, 38, 41, 83, 98, 99, 100, 101, 141, 147, 149
Ahornzuckerkrankheit 11
Algorithmen 57, 96, 98, 132, 133
American National Standards Institute *siehe* ANSI
ANSI 46, 55, 139, 140
Aries/Nova Idea 18, 141
Autorenprofile 88
Autorenverzeichnis 25
Bayerisches Gesundheitsnetz 9
Behinderte Personen 52
Big Endian 46
BIOSIS 15, 36, 38, 124, 149
Boolesche Suche 25, 64
Branch Fit 67
Brief 19
Bubble Sort 61
C 55
Caching 46
Cascading Style Sheets 44, 47, 135, 139
CD-Plus 18, 141
CGI 46, 47, 55, 58, 105
Clinical Queries 28, 29
Cochrane Controlled Trials Register 37
Cochrane Library 15, 29, 37, 124
Common Gateway Interface *siehe* CGI
Compiler 56
Computergestütztes Lernen 16
Content Negotiation 54, 116
Controlled Terms *siehe* Deskriptoren
Controlled Vocabulary *siehe* MeSH
CSS 53
Cumulated Index Medicus 18
Current Contents 15, 36, 38, 124
Data Mining 10, 16, 78, 79, 88, 120, 124, 125
DataStar 18, 37, 38, 39, 141
Deepest Fit 68
Deskriptoren 31
Diabetes 11
Dienst 17, 41, 43, 63, 120
DIMDI 20, 37, 38, 39, 132, 141, 149
Divide et impera 59

DRG 9
EBM 29, 30, 122
Editorial 19
EMBASE 36, 37, 38, 124, 132, 134, 149
Embase Alert 36
EMTree *siehe* EMBASE
Entry Terms *siehe* MeSH-Querverweise
Erstautor 65
Evidence based Medicine *siehe* EBM
Excerpta medica 15
Excerpta Medica *siehe* EMBASE
Felder, typisierte 26
First Fit 68
Fleschs Formel 51
Free grips WebSearch 20
Freitextsuche 23, 26, 31, 32
Genauigkeitsmaß *siehe* Relevanz
Granulated heap allocation 47
Histogramme 66, 78
HTML 53
HTML-Validatoren 49
HTTP 22, 43, 46, 47, 53, 105, 140
Hypertext Transfer Protocol *siehe* HTTP
Index Medicus 18
Index to Dental Literature 18
Insertion Sort 61
International Nursing Index 18
International Standards Organization *siehe* ISO
ISO 55, 139, 140
Java Server Pages *siehe* JSP
JavaScript 20, 24, 52
Journalverzeichnis 11, 19, 24, 25, 63, 89, 90, 91, 94, 146, 147, 148
JSP 55
Just the right time learning 8
Klinische Fragestellungen 27
Klinische Richtlinien 29
Knowledge Finder 15, 18, 141
Koinzidenztabelle 66, 79
Kompression 99
Konsensuskonferenzen 29
Konsultationsformular 63
Korrelationstabellen *siehe* Koinzidenztabelle
Künstliche Intelligenz 16
Lasttrennung 43

- Laufzeitoptimierung 46
- Lesbarkeit 49
- Limits, PubMed-Suche 27
- Linkrestriktor 68, 79
- Little Endian 46
- Main heading 31
- Maple Syrup Urine Disease *Siehe* Ahornzuckerkrankheit
- Massendaten 10, 16, 139
- Medical Literature, Analysis, and Retrieval System Online *siehe* MEDLINE
- Medical Subject Headings *siehe* MeSH
- MEDLINE 15, 18, 36
- MEDLINE Direct 20
- MEDLINE Evaluator *siehe* Meva
- MEDLINE-Format 41, 124
- MEDLINEplus 19
- MEDLINE-Postprozessor *siehe* Meva
- Mehrsprachigkeit 54, 114
- MePrep 99
- MeSH 31
- MeSH Annotated Alphabetic List 34
- MeSH Strings, unifizierte 84
- MeSH Supplementary Concepts Records 35
- MeSH Tree Structures 34, 141
- MeSH-Baum 67, 85
- MeSH-Browser 31
- MeSH-Codes, assoziierte 84
- MeSH-Codes, interpolierte 84
- MeSH-Querverweise 32
- Metaanalysen 29
- Meva 41
- Meva Preprocessor *siehe* MePrep
- Meva-Administration 103
- Meva-Datenformat 71
- Meva-Fragen 112
- Meva-HTML-Set 104
- Meva-Indexdateien 104
- Meva-Indexgenerator 108
- Meva-Kommentar 72
- Meva-Konfiguration 104
- Meva-Logging 103
- Meva-Wichtungsstrategien 85
- MIME 46, 105, 112, 142
- Mooreches Gesetz 8
- Multimedia Internet Mail Extensions *siehe* MIME

Multitasking 46
Multithreading 47
Nachweisquote 17
National Center for Biotechnology Information *siehe* NCBI
National Institute of Health *siehe* NIH
National Library of Medicine *siehe* NLM
NCBI 22
Neuronale Netzwerke 16, 139
NIH 22, 142
NLM 19, 22, 31, 36, 109, 142
NPH *Siehe* Non-parsed Header
OLAP 16
OLDMEDLINE 19
On-line Analytical Processing *siehe* OLAP
OVID 10, 15, 18, 141
Page Ranking 44
Paging 46
Paginierung 19
Parsed Header CGI 46
Patient-oriented evidence that matters *siehe* POEM
PERL 55
Permuted MeSH 34
POEM 29
Portabilität 46
Precision *siehe* Relevanz
PreMEDLINE 22, 23
Primärmenge 69
Proxy 115
PubMed 22, 36
PubMedCentral 23, 143
Quick Sort 61
Randomized Control Trial 29, 122
Recall *siehe* Nachweisquote
Regelinduktion 16, 139
Related Articles, PubMed *siehe* Verknüpfte Artikel
Relevanz 17
Review 19, 90, 94, 147
RFC 53, 54, 105, 115, 143
Schlagwortverzeichnis 25
Science Citation Index 15, 36, 38, 124
Sekundärmenge 69
Selection Sort 61
Sensitivität 28, 29
Service *siehe* Dienst

Servlets 55
Shake Sort 61
SilverPlatter 18, 141
Software Lifecycle 57, 144
Sortierverfahren 61
Spezifität 28, 68
Spiralmodell 57
SPIRS 18, 141
STN 18, 37, 132, 141
Strukturierte Programmierung 56
Subheading 31
Subject Headings *siehe* Deskriptoren
Suche n. Boyer-Moore 58
Suche n. Knuth-Morris-Pratt 58
Suche, binäre 58, 59, 60
Suche, interpolierte 58, 61
Suche, sequentielle 58, 60, 61
Suche, typisierte 26
Suche, wörtliche 26, 65
Suchfilter 64
Suchmaschinen 8, 43, 44, 45, 48, 50
Suchmaße 16
Suchverfahren 57
Supplementary Concept Records 32
Swapping 46
Systematic Reviews 29, 122
Telematik 9
Term Mapping 25
Tetrodotoxin 69
Trunkierung 26
Typisierte Feldsuche 24
Uniform Resource Identifier *siehe* URI
URI 22, 46, 145
Verknüpfte Artikel, PubMed 27
Video Conferencing 9
Vollständigkeit *siehe* Nachweisquote
Wasserfallmodell 57
Web of Science 36, 38, 39
Webkataloge 43
WinSPIRS 10, 15
World Wide Web Consortium 20, 53, 145

9 LITERATURVERZEICHNIS

- [1] Bachmann, L.M., Coray, R., Estermann, P., Ter Riet, G.
[Identifying diagnostic studies in MEDLINE: reducing the number needed to read.](#)
J. Am. Med. Inform. Assoc. 9 (2002) 653 - 658
- [2] Blieberger, J., Schildt, G.H., Schmid, U., Stöckler, S.
Informatik
Springer-Verlag, Wien – New York, 1992, 2. Neubearb. Aufl.
- [3] Born, G.
Referenzhandbuch Dateiformate – Grafik, Text, Datenbanken, Tabellenkalkulation
Addison-Wesley-Verlag, Bonn – München – Paris, 1992, 2. überarb. Aufl.
- [4] Brantner, M., Schmidt, S., Wabnitz, B., Wabnitz, D.
Java Server Pages und Servlets
Data-Becker-Verlag, Düsseldorf, 2001, 1. Aufl.
- [5] Brown, C.
Programmieren verteilter UNIX-Anwendungen
Prentice-Hall-Verlag, München, 1994, 1. Aufl.
- [6] Coletti, M.H., Bleich, H.L.
[Medical Subject Headings Used to Search the Biomedical Literature.](#)
J. Am. Med. Inform. Assoc. 8 (2001) 317 – 323
- [7] Eberhart, A., Fischer, S.
Java-Bausteine für E-Commerce-Anwendungen. Verteilte Anwendungen mit Servlets, CORBA und XML
Hanser-Verlag, München – Wien, 2000, 1. Aufl.
- [8] Evans, R.N.
[Database searches for qualitative research.](#)
J. Med. Libr. Assoc. 90 (2002) 290 - 293
- [9] Glogau, R.
Muster erkannt - Gefahr gebannt.
DOS International 10 (1994) 280 - 284
- [10] Grandage, K., Slawson, D., Barnett, B., Shaughnessy, A.F.
[When less is more: a practical approach to searching for evidence-based answers.](#)
J. Med. Libr. Assoc. 90 (2002) 298 - 304

- [11] Greenhalgh, T.
How to read a paper – the MEDLINE database.
B.M.J. 315 (1997) 180 - 183
- [12] Gretz, M., Schmitt, R.D., Thomas, M.
EMBASE remains EMBASE - on every host? A comparative analysis of EMBASE on the hosts DATA-STAR, DIALOG, DIMDI, and STN.
Comput. Biomed. Res. 29 (1996) 494 - 506
- [13] Haynes, R.B., Wilczynski, N., McKibbin, K.A., Walker, C.J., Sinclair, J.C.
Developing optimal search strategies for detecting clinically sound studies in MEDLINE.
J. Am. Med. Inform. Assoc. 1 (1994) 447 - 458
- [14] Hermann, D.
Algorithmen-Arbeitsbuch
Addison-Wesley-Verlag, Bonn - München - Paris, 1992, 1. Aufl.
- [15] Kientzle, T.
Internet-Dateiformate – Windows, DOS, UNIX & Mac
International Thomson Publishing, Bonn, 1996, 1. Aufl.
- [16] Klare, G.
Assessing Readability.
Reading Research Quarterly 10 (1974) 62 - 102
- [17] Luhmann, N.
Rechtssoziologie
Westdeutscher Verlag, Opladen, 1983, 2. erw. Aufl.
- [18] Lusti, M.
Data Warehousing und Data Mining: Eine Einführung in entscheidungsunterstützende Systeme
Springer-Verlag, Berlin – Heidelberg – New York, 2002, 2. überarb. und erw. Aufl.
- [19] Monadjemi, P.
PC-Programmierung in Maschinensprache: eine Einführung mit vielen Beispielen
Verlag Markt & Technik, Haar bei München, 1991, 2. erw. und überarb. Aufl.
- [20] Obst, O.
Wo sind die aktuellsten medizinischen Informationen zu finden?
Medizin-Bibliothek-Information 1 (2001) 34

-
- [21] Olsen, O., Middleton, P., Ezzo, J., Gotzsche, P.C., Hadhazy, V., Herxheimer, A., Kleijnen, J., McIntosh, H.
[Quality of Cochrane reviews: assessment of sample from 1998.](#)
B.M.J. 323 (2001) 829 - 832
- [22] Ruhland, J.
TOOL Informatik: Datenstrukturen und elementare Algorithmen
Vogel-Verlag, Würzburg, 1991, 1. Aufl.
- [23] Schöning, U.
Logik für Informatiker
Wissenschaftsverlag, Mannheim – Leipzig – Wien - Zürich, 1992, 3. überarb. Aufl.
- [24] Sedgewick, R.
Algorithmen
Addison-Wesley-Verlag, Bonn – München – Reading, Mass., 1997, 3. unveränd. Nachdruck
- [25] Shaughnessy, A.F., Slawson, D.C., Bennett, J.H.
[Becoming an information master: a guidebook to the medical information jungle.](#)
J. Fam. Pract. 39 (1994) 489 - 499
- [26] Shea, S.E.
[Diversion. PubMed and me: Why my coughs are more productive than my searches.](#)
C.M.A.J. 12 (2002) 167
- [27] Shojania, K.G., Bero, L.A.
[Taking advantage of the explosion of systematic reviews: an efficient MEDLINE search strategy.](#)
Eff. Clin. Pract. 4 (2001) 157 - 162
- [28] Suarez-Almazor, M.E., Belseck, E., Homik, J., Dorgan, M., Ramos-Remus, C.
[Identifying clinical trials in the medical literature with electronic databases: MEDLINE alone is not enough.](#)
Control. Clin. Trials 21 (2000) 476 - 487
- [29] Thurmayr, G.R., Thurmayr, R., Tenner, H., Ingenerf, J.
Womit beschäftigt sich zur Zeit die Med. Informatik? Eine MEDLINE-Analyse.
Informatik, Biometrie und Epidemiologie in Medizin und Biologie 33 (2002) 100

- [30] Tischer, M.
PC Intern – Systemprogrammierung
Data-Becker-Verlag, Düsseldorf, 1994, 1. Aufl.
- [31] Tischer, M., Jennrich, B.
Internet intern - Technik und Programmierung
Data-Becker-Verlag, Düsseldorf, 1997, 1. Aufl.
- [32] Topfer, L.A., Parada, A., Menon, D., Noorani, H., Perras, C., Serra-Prat, M.
Comparison of literature searches on quality and costs for health technology assessment using the MEDLINE and EMBASE databases.
Int. J. Technol. Assess. Health Care 15 (1999) 297 - 303
- [33] White, P.J.
Evidence-based medicine for consumers: a role for the Cochrane Collaboration.
J. Med. Libr. Assoc. 90 (2002) 218 - 222
- [34] Wiedmann, K., Buckler, F., Buxel, H.
Data Mining – ein einführender Überblick.
In: „Neuronale Netze im Marketing-Management. Praxisorientierte Einführung in modernes Data-Mining“, Wiedmann, K., Buckler, F. (Hrsg.), Gabler-Verlag, Wiesbaden, 2001, 1. Aufl., 15 - 33
- [35] Wiessens, N., Müller, B.
Sortieralgorithmen in Pascal.
DOS International 12 (1994) 254 - 258

10 INTERNET-QUELLENVERZEICHNIS

Hinweis: Die Quellen datieren per 05.06.2003, sofern nicht anders angegeben.

- [API-Studie] LA-MED
API-Studie. Reichweitenstudie für medizinische Fachtitel.
URL: <http://www.dgn-service.de/Neues/api.html>
- [Bobby] **Bobby** - Testprogramm zur Überprüfung einer Webpräsenz auf Behindertentauglichkeit.
URL: <http://bobby.watchfire.com/>
- [CSS 1] Lie, H.W., Bos, B.
Cascading Style Sheets V. 1, 1996
URL: <http://www.w3.org/TR/1999/REC-CSS1-19990111>
- [CSS 2] Bos, B., Lie, H.W.
Cascading Style Sheets V. 2, 1998
URL: <http://www.w3.org/TR/REC-CSS2/>
- [CSS-Validator] W3C-Konsortium
CSS-Validator
URL: <http://jigsaw.w3.org/css-validator/>
- [CGI 1.1] **Spezifikation des Common Gateway Interface V 1.1**
URL: <http://www.w3.org/CGI/>
- [HTML 4.01] W3C-Konsortium
HTML V. 4.01 Spezifikation, 1999
URL: <http://www.w3.org/TR/html401/>
- [HTML-Validator] W3C-Konsortium
HTML-Validator
URL: <http://validator.w3.org/>
- [HTTP/0.9] CERN
Original HTTP – HTTP/0.9, 1991
URL:
<http://www.w3.org/Protocols/HTTP/AsImplemented.html>
- [Nielsen] Nielsen, H.F., Gettys, J., Baird-Smith, A., Prud'hommeaux, E., Lie, H.W., Lilley, C.
Network Performance Effects of HTTP/1.1, CSS1, and PNG, W3C Note 1997
URL:
<http://www.w3.org/Protocols/HTTP/Performance/Pipeline>

- [RFC 1867] Nebel, E., Masinter, L.
Form-based File Upload in HTML, 1995
URL: <http://www.ietf.org/rfc/rfc1867.txt>
- [RFC 1945] Berners-Lee, T., Fielding, R. et al.
Hypertext Transfer Protocol – HTTP/1.0, 1996
URL: <http://www.ietf.org/rfc/rfc1945.txt>
- [RFC 2388] Masinter, L.
Returning Values from Forms: multipart/form-data, 1998
URL: <http://www.ietf.org/rfc/rfc2388.txt>
- [RFC 2616] Fielding, R. et al.
Hypertext Transfer Protocol - HTTP/1.1, 1999
URL: <http://www.ietf.org/rfc/rfc2616.txt>
- [Schallhorn] Schallhorn, K.
Eintrag, Optimierung, Ranking & Suchbegriffe, 2003
URL: <http://www.kso.co.uk/>
- [Schiedermeier] Schiedermeier, R.
**Elektronische Materialsammlung zur Vorlesung
„Programmieren I“**, FB 07 Informatik/Mathematik der FH
München, 1997
URL: [http://www.informatik.fh-
muenchen.de/~schieder/programmieren-1-ws96-97/](http://www.informatik.fh-muenchen.de/~schieder/programmieren-1-ws96-97/)
- [WCAG] W3C-Konsortium
Web Content Accessibility Guidelines, W3C-
Empfehlung 1999
URL: <http://www.w3.org/TR/WAI-WEBCONTENT/>

11 ABBILDUNGSVERZEICHNIS

Abb. 1: Ausschnitt aus PubMeds Suchformular	22
Abb. 2: Ausschnitt aus einem PubMed-Resultat.....	24
Abb. 3: Ausschnitt aus einem MeSH-Resultat.....	32
Abb. 4: Operative Architektur von Meva.....	42
Abb. 5: Häufigkeiten der Initialen der englischen MeSH-Strings 2002.....	60
Abb. 6: Meva-Formular, über das der Benutzer Meva steuert	63
Abb. 7: Aufbau eines PubMed-Links in Meva	69
Abb. 8: Verkleinerung der Sekundärmenge in Mevas Verknüpfungen.....	70
Abb. 9: Zusätzl. Verkleinerung der Sekundärmenge in Mevas Verknüpfungen ...	70
Abb. 10: Schnittmengenbildung in Mevas Verknüpfungen.....	71
Abb. 11: Beispielin- und -ausgabe in PubMed	74
Abb. 12: Beispielingabe in Meva.....	75
Abb. 13: Meva-Resultat - Überschriftensektion	75
Abb. 14: Meva-Resultat - Meldungssektion.....	76
Abb. 15: Meva-Resultat - Parametersektion	76
Abb. 16: Meva-Resultat - Bilanzsektion.....	78
Abb. 17: Meva-Resultat - Histogrammsektion Feld 1	78
Abb. 18: Meva-Resultat - Histogrammsektion Feld 2	79
Abb. 19: Meva-Resultat - Korrelationssektion Feld 1	80
Abb. 20: Meva-Resultat - Korrelationssektion Feld 1 und 2.....	81
Abb. 21: Meva-Resultat - Detailsektion	82
Abb. 22: Meva-Resultat - MeSH-Sektion Bilanz.....	83
Abb. 23: Meva-Resultat - MeSH-Sektion Strategien	85
Abb. 24: Meva-Resultat - MeSH-Sektion Baum	86
Abb. 25: Meva-Resultat - Fußnotensektion.....	87
Abb. 26: Ausschnitt aus Webseite zu Starzl T.E., Juni 2003.....	93
Abb. 27: Poster zur GMDS-Tagung 2002, Berlin.....	97

12 TABELLENVERZEICHNIS

Tab. 1: Artikelmenen beim Suchen.....	17
Tab. 2: Sensitivität kombinierter Suchbegriffe für klinische Studien 1991 nach Haynes et al.	28
Tab. 3: Lesbarkeitsgrad nach Flesch	50
Tab. 4: Verteilung der Initialen der englischen MeSH-Codes 2002	59
Tab. 5: Verteilung der Initialen der englischen MeSH-Strings 2002	59
Tab. 6: Zugriffe von Suchverfahren im Vergleich	60
Tab. 7: Konfigurationsparameter von Meva	104
Tab. 8: Mevas Clientfehlermeldungen	106
Tab. 9: Mevas Serverfehlermeldungen.....	107
Tab. 10: Warnungen von Meva	108
Tab. 11: Hinweise von Meva.....	108

13 ANHANG A: GLOSSAR

Abstract	Kurze inhaltliche Zusammenfassung eines Artikels, die es erlaubt, einen schnellen Überblick zum Artikel zu erhalten..
Algorithmus	Abgeleitet vom Namen des arabischen Mathematikers Muhammed Ibn Musa al-Kwarizmi (ca. 813-864). Ein Algorithmus ist in der Informatik eine Anleitung zur Lösung eines berechenbaren Problems.
ANSI	American National Standards Institute . Amerikanisches Standardisierungsgremium. Sitz: Washington. Beispiele für ANSI-Normen sind der ANSI-Zeichensatz (ISO 8859-1), ANSI ASCII (ISO 6), ANSI-C, ANSI-Pascal etc.
Assembler	Programm, um Assemblersprache in Maschinensprache zu übersetzen.
CGI	Common Gateway Interface . Ein Protokoll, welches festlegt, wie ein serverseitiges Programm mit einem Webserver Informationen austauscht. Aktuelle Version ist V. 1.1.
Client	Arbeitsplatzrechner, an dem Benutzer lokal arbeiten im Gegensatz zu einem Server, der seine Dienste über ein Netzwerk zur Verfügung stellt.
Client-Server-Modell	Ein Modell, nach dem ein Teil der Funktionalität, die einem Client zur Verfügung steht, nicht vom Client, sondern vom Server erbracht wird. Die Benutzung des World Wide Web ist ein Beispiel für ein solches Modell.
CSS	Cascading Style Sheets. Technologie, um Webseiten auszuzeichnen. Basiert auf einer logischen Formatierung in HTML und ersetzt damit die frühere physische HTML-Formatierung.
CSV	<i>Comma Separated Values</i> . Datenbanktaugliches Textdateiformat, in dem die Feldinhalte durch Kommata abtrennt sind.
Data Mining	Nichttriviale und automatische Suche nach Wissen in Massendaten. <i>Data Mining</i> i.e.S. ermöglichen vor allem die Statistik, die Visualisierung, die Regelinduktion und Neuronale Netzwerke. <i>Data Mining</i> i.w.S. umfaßt auch <i>OLAP</i> und konventionelle Abfrage- und Tabellenkalkulations-Funktionen.

Deskriptor	Schlag- bzw. Suchwort. Ein Deskriptor beschreibt inhaltlich einen Datensatz einer Datenbank. Er dient dem Wiederauffinden der gewünschten Information. Sucht man beispielsweise in einer Literaturdatenbank wie MEDLINE nach einer bestimmten Publikation, so kann man als Deskriptor beispielsweise den Autorennamen oder ein Schlagwort aus dem Thesaurus eingeben. S.a. PubMeds bibliographische Felder .
DIMDI	Deutsches Institut für Medizinische Dokumentation und Information, gehört zum Gesundheitsministerium. Staatliches Institut in Köln, welches als wichtigster Host auf dem Bereich der Medizin in Deutschland etwa 100 biomedizinische Datenbanken (z.B. MEDLINE u.a.) meist gebührenpflichtig on-line anbietet.
HTML	Hyper Text Markup Language. Auszeichnungssprache zur Formatierung von Webseiten. Der aktuelle Standard ist HTML Version 4.01.
HTTP	Hyper Text Transfer Protocol. Ein Übertragungsprotokoll des World Wide Web.
Index Medicus ®	Durch das NLM seit über 100 Jahren gepflegtes, gedrucktes bibliographisches Verzeichnis aller Artikel der meisten biomedizinischen Zeitschriften und Journale dieser Welt.
ISO	International Standards Organization . Internationales Standardisierungsgremium, Pendant zum ANSI, 1947 gegründet. Beispiele: OSI-Referenzmodell für offene Kommunikation, ISO-C, ISO-Pascal etc.
ISP	Internet Service Provider. Ein ISP ermöglicht einem Benutzer den Zugang zum Internet.
Java Server Pages	Technologie, welche die Fähigkeiten des Webservers erweitert, indem die HTML-Seiten vor dem Verschicken vom Webserver geparkt werden, was Aufrufe von in Java geschriebenen Applikationen im Serverkontext zur Folge hat. Somit ähnelt diese Technologie Microsofts ASP.
Kopplung, lockere	Begriff aus der Programmierung, der die meisten Variablen lokal vorhält, jedoch Seiteneffekte für wenige ausgezeichnete globale Variablen zulässt.
Main heading	Ein MeSH-Term ohne Qualifizierer, z.B. Multiple sclerosis.

- MEDLINE ® Bedeutend(st)e englische Literaturdatenbank für die gesamte Medizin. MEDLINE ist die seit 1971 verfügbare, durch das [NLM](#) gepflegte elektronische Version des [Index Medicus](#) und einiger anderer gedruckten Bibliographien. Die Datenbank enthält die bibliographischen Angaben sowie die [Abstracts](#) der wichtigsten internationalen medizinischen Publikationen (ca. 4.600) seit 1966 und wird wöchentlich auf den neuesten Stand gebracht. Das derzeitige Volumen beträgt knapp 12 Mio. Artikel aus den USA und 70 anderen Staaten. Suchbar sind bibliographische Angaben, [Deskriptoren](#) (Schlagwörter) und [Abstracts](#) (76%). MEDLINE enthält Nachweise der internationalen Literatur aus allen Bereichen der Biomedizin, einschließlich der Zahn- und Veterinärmedizin, Psychologie und des öffentlichen Gesundheitswesens sowie einiger Randgebiete. Die Datenbank wird an der [National Library of Medicine](#) zusammengestellt und von verschiedenen Anbietern als Datenbank auf CD-ROM vertrieben oder von Datenbank-Hosts als On-line-Datenbank angeboten. Anbieter, die MEDLINE auf CD-ROM mit jeweils unterschiedlichen Systemen anbieten, sind z.B. SilverPlatter (System: SPIRS), Aries/Nova Idea (System: Knowledge Finder) und CD-Plus (System: OVID). Hosts, die MEDLINE [on-line](#) aufliegen haben, sind z.B. [DIMDI](#), STN und DataStar.
- Der Anwender kann den [MeSH](#) benutzen, um bestimmte Artikel zu einem Thema zu finden. In diesem MEDLINE-spezifischen [Thesaurus](#) sind alle Schlagwörter hierarchisch aufgelistet, mit denen die Artikel indiziert wurden. DIMDI übersetzt MEDLINE jedes Jahr ins Deutsche. Die deutsche Datenbank ist über *Free grips WebSearch*® [on-line](#) entgelt- und lizenzfrei abrufbar.
- MeSH ® Medical Subject Headings. Übliche Kurzbezeichnung für den [Thesaurus](#) der Datenbank [MEDLINE](#). Er ist gedruckt und als Datenträgerversion erhältlich (jeweils in verschiedenen Sprachen), aber auch [on-line](#) abrufbar.
- Der MeSH ist polyhierarchisch strukturiert und besteht aus vier Teilen:
1. Hierarchischer Teil (systematisch, *MeSH Tree Structures*).
 2. Alphabetischer Teil (*MeSH Annotated Alphabetic List*).
 3. *Permuted MeSH*. Dies ist das permutierte Verzeichnis des MeSH. Es entlastet das alphabetische Verzeichnis bei den Mehrwortbezeichnungen durch die Auflistung der zu einem Teilbegriff vorhandenen Deskriptoren.
 4. *Supplementary Chemical Records* (von DIMDI nicht

	übersetzt). In den <i>Supplementary Chemical Records</i> sind die Bezeichnungen chemischer Substanzen aufgelistet.
Meva	MEDLINE-Evaluator . Ein Postprozessor zur Verdichtung und optischen Aufbereitung von Daten, die bei einer MEDLINE-Recherche geliefert werden.
MH	S. Main Heading.
MIME	<i>Multimedia Internet Mail Extensions</i> . Standard, der die Verwendung von Multimedia-Inhalten beim Mailversand regelt.
Mooresches Gesetz	Gordon Moore war 1965 Forschungsdirektor bei Fairchild Industries und später Mitgründer von INTEL. 1965 fand er die nach ihm benannte Gesetzmäßigkeit, daß sich die Anzahl der Transistoren je Quadratzoll pro Prozessor verdoppelt in 18 Monaten verdoppelt. Moore vermutete nun, daß diese Entwicklung sich auch in der Zukunft fortsetzen würde. Dies war korrekt: Seine Vorhersage hat bis heute nichts von ihrer Gültigkeit verloren.
Multithreading	Parallele Ausführung mehrerer Programmteile auf einem Rechner, sog. <i>Threads</i> .
NCBI	National Center for Biotechnology Information , Bethesda, Maryland, USA, Teil der NLM . 1988 als Forschungszentrum für Molekularbiologie gegründet. Das NCBI pflegt u.a. auch die PubMed -Datenbank.
NCSA	National Center for Supercomputing Applications. 1986 als Forschungszentrum für Supercomputer von der National Science Foundation und einer Abteilung an der University of Illinois, Urbana-Champaign gegründet.
NIH	National Institute of Health , Bethesda, Maryland, USA. Größtes staatliches medizinisches Forschungsinstitut des amerikanischen Gesundheitsministeriums. Zu ihm gehören 27 separate Institute und Zentren, u.a. die NLM .
NLM	National Library of Medicine , Bethesda, Maryland, USA, Teil des NIH .
NPH	Non Parsed Header. Aufrufform von CGIs, bei denen das CGI den vollen Satz an Kopfzeilen rückliefert.
OLAP	On-line Analytical Processing. Abfragemethode, die Endbenutzern einen multidimensionalen Zugriff und eine

	benutzerfreundliche interaktive Analyse von Datamarts erlaubt.
PERL	Practical Extraction and Reporting Language. Interpretierte Sprache, die von Larry Wall speziell für die Verarbeitung von Zeichenketten entwickelt wurde.
Programmierung, strukturierte	Die strukturierte Programmierung beschreibt ein seit Ende der 60er Jahre in der Informatik bestehendes Konzept, nach dem der Programmablauf durch wenige, aber einheitliche Kontrollstrukturen definiert wird, graphisch ausdrückbar in Struktogrammen. Diese Vorgehensweise löste damit die zuvor weit verbreitete Form der „Spaghettiprogrammierung“ ab, in welcher der lineare Programmfluß nur durch Sprünge zu Marken unterbrochen wurde.
Proxy	Zwischenspeicher, der die Zugriffszeiten auf ein Primärmedium senken soll.
PubMed ®	<p>Öffentlich und kostenlos abfragbare Datenbank, welche die Daten von 1. MEDLINE enthält sowie einige weitere Datenbestände:</p> <ol style="list-style-type: none"> 2. Nichtmedizinische Artikel aus MEDLINE-Journalen (z.B. zu Astrophysik oder Plattentektonik) 3. Bibliographische Daten zu Artikeln aus MEDLINE-Journalen, die vor der ersten Erfassung dieser Journale in MEDLINE liegen 4. Volltextversionen einiger Journale, die den Volltext elektronisch PubMedCentral™ zur Verfügung stellen und durch die NLM qualitativ geprüft werden. <p>Die Bibliothekare der NLM verschlagworten diese Daten mittels genormter Schlagworte (Deskriptoren = <i>Subject headings</i> = <i>Controlled Terms</i> = CT). Über ein MeSH-Schlagwort/Nummer kann man somit die bibliographischen Daten aller zugehörigen Artikel finden. Diese Deskriptoren sind noch einmal zusammengefaßt im MeSH (<i>Controlled Vocabulary</i>).</p>
Qualifier	S. Subheading.
RFC	Request For Comment. Webstandard.
Seiteneffekt	Allgemeiner Begriff für jede Art von bleibender Veränderung, die nach Verarbeitung eines lokalen Blockes auf der Ebene des Aufrufers (oder höher) in einem Programm bestehen bleibt.
Server	Rechner, der einen Dienst (Service) für Clients anbietet. In der

	Regel arbeiten die Benutzer nicht lokal am Server, sondern benutzen die Funktionalität des Dienstes über ein Netzwerk.
Service	Eine bestimmte Funktionalität, die von einem Server für Clients angeboten wird. Ein Service ist z.B. ein Druckservice, über den Benutzer auf einen zentral installierten Netzwerkdrucker zugreifen können.
Servlets	Servlets sind Java-Programme analog zu Java-Applets, werden aber im Webserver-Kontext statt in der lokalen Umgebung ausgeführt.
SH	S. Subheading.
Software Lifecycle	Gesamtheit der Lebensphasen eines Projektes, bestehend aus Problem, Analyse, Entwurf, Implementierung, Programm, Test und Wartung.
Subheading	Ein an MeSH-Termen angehängter Qualifizierer, der eine feinere Abstufung in der Kategorisierung eines Artikels erlaubt, z.B. Multiple sclerosis/etiology.
Thesaurus	<p><über lat. Thesaurus aus gr. Thesaurós „Schatz(kammer)“, eigtl. „Ort zum Sammeln und Aufbewahren“, zu tithénai, vgl. These></p> <ol style="list-style-type: none"> 1. Titel wissenschaftlicher Sammelwerke, besonders großer Wörterbücher der alten Sprachen. 2. alphabetisch und systematisch geordnete Sammlung von Wörtern eines bestimmten (Fach-)Bereichs. So ist z.B. dieses Glossar auch ein Thesaurus. 3. (Antike:) kleineres Gebäude in einem Heiligtum zur Aufbewahrung von kostbaren Weihegaben. 4. (Informatik:) Synonymwörterbuch in vielen Textverarbeitungen. Auf Anfrage schlägt der Thesaurus sinnverwandte Wörter vor, die der Anwender bei Bedarf übernehmen kann. 5: (Informatik:) Zusammenstellung der Schlagwörter oder Deskriptoren einer Datenbank. Beispiel: Der MeSH ist der Thesaurus der Datenbank MEDLINE. Der Anwender kann Begriffe aus dem Thesaurus auswählen, und wird dann durch die Retrievalsoftware auf die zu diesem Begriff passenden und in der Datenbank vorhandenen Datensätze geführt.
Thin Client	Ein auf dem Client-Server-Modell aufsetzendes Konzept, bei dem statt ressourcenfressender Client-Software nur ein kleiner Client mit Minimalfunktionalität installiert wird. Der Großteil

des Codes liegt hingegen auf dem Server.

- TSV *Tabulator Separated Values*. Datenbanktaugliches Textdateiformat, in dem die Feldinhalte durch Tabulatoren abtrennt sind.
- URI *Uniform Resource Identifier*. 1990 von der *WWW Global Information Initiative* begonnenes und 1994 von T. Berners-Lee weiterentwickeltes generisches **Konzept** zur universalen Adressierung von Objekten (RFC 1630, 1994). Eine Webadresse wie <http://www.med-ai.com/meva/> ist z.B. ein URI.
- URL *Uniform Resource Locator*. 1994 von der *URI Working Group* der *Internet Engineering Task Force* standardisiertes **Konzept** zur Lokalisierung von Ressourcen (RFC 1738, 1994). RFC 1630 führt weiter aus, daß ein URL Objekte mit Hilfe von existenten Protokollen universell adressiert. Hierbei wird auf die Zugriffsform abgestellt, während ein URI als abstrakteres Konzept (RFC 2396, 1998) auch Ressourcen bezeichnen kann, die nicht über ein Netzwerk adressiert werden können, wie z.B. Menschen, Firmen oder gebundene Bücher in einer Bibliothek. Für das Internet besteht jedoch somit momentan begriffliche Deckungsgleichheit zwischen einem URI und einem URL (dies in Abgrenzung zu Aussagen wie z.B. in [Tischer 1997](#), S. 919, ein URI sei immer Teil eines URL - und zwar ein URI ohne Angabe von Protokoll, Server und Port). Eine Webadresse wie z.B. <http://www.med-ai.com/meva/> ist z.B. ein URL.
- W3C **World Wide Web Consortium**. Gremium, das Aufbau und Struktur des WWW regelt.

14 ANHANG B: MEDLINE-FELDER

Die folgende, von PubMed stammende Tabelle dokumentiert die bibliographischen MEDLINE-Felder, die *Meva* verwendet:

Tag	Name	Beschreibung
AB	Abstract	Abstract.
AD	Affiliation	Institutional affiliation and address of the first author, and grant numbers.
AID	Article Identifier	Article ID values may include the pii (controlled publisher identifier) or doi (Digital Object Identifier).
AU	Author	Authors.
CI	Copyright Information	Copyright statement.
CIN	Comment In	Reference containing a comment about the article.
CN	Collective Name	Corporate author or group names with authorship responsibility.
CON	Comment On	Reference upon which the article comments.
CY	Country	The place of publication of the journal.
DA	Date Created	Used for internal processing at NLM.
DCOM	Date Completed	Used for internal processing at NLM.
DEP	Date of Electronic Publication	Electronic publication date.
DP	Publication Date	The date the article was published.
EDAT	Entrez Date	The date the citation was added to PubMed.
EIN	Erratum In	Reference containing a published erratum to the article.
GS	Gene Symbol	Abbreviated gene names (used 1991 through 1996).
ID	Identification Number	Research grant numbers, contract numbers, or both that designate financial support by any agency of the US PHS (Public Health Service).
IP	Issue	The number of the issue, part, or supplement of the journal in which the article was published.
IS	ISSN	International Standard Serial Number of the journal.
JC	Journal Title Code	MEDLINE unique three-character code for the journal.
JID	NLM Unique ID	Unique journal ID in NLMs catalog of books, journals, and audiovisuals.
LA	Language	The language in which the article was published.

Tag	Name	Beschreibung
LR	Last Revision Date	The date a change was made to the record during a maintenance procedure.
MH	MeSH Terms	NLMs controlled vocabulary.
MHDA	MeSH Date	The date MeSH terms were added to the citation. The MeSH date is the same as the Entrez date until MeSH are added.
PG	Pagination	The full pagination of the article.
PHST	Publication History Status Date	History status date.
PMID	PubMed Unique Identifier	Unique number assigned to each PubMed citation.
PS	Personal Name as Subject	Individual is the subject of the article.
PST	Publication Status	Publication status.
PT	Publication Type	The type of material the article represents.
RF	Number of References	Number of bibliographic references for Review articles.
RIN	Retraction In	Retraction of the article.
RN	EC/RN Number	Number assigned by the Enzyme Commission to designate a particular enzyme or by the Chemical Abstracts Service for Registry Numbers.
ROF	Retraction Of	Article being retracted.
RPF	Republished From	Original article.
RPI	Republished In	Corrected and republished article.
SB	Journal Subset	Code for a specific set of journals.
SI	Secondary Source Identifier	Identifies a secondary source that supplies information, e.g., other data sources, databanks and accession numbers of molecular sequences discussed in articles.
SO	Source	Composite field containing bibliographic information.
TA	Journal Title Abbreviation	Standard journal title abbreviation.
TI	Title	The title of the article.
TT	Transliterated / Vernacular Title	Non-Roman alphabet language titles are transliterated.

Tag	Name	Beschreibung
UI	MEDLINE Unique Identifier	Unique number assigned to each MEDLINE citation.
UIN	Update In	Update to the article.
UOF	Update Of	The article being updated.
URLF	URL Full-Text	Link to the full-text of article at providers website. Links are incomplete.
URLS	URL Summary	Link to the article summary at providers website. Links are incomplete.
VI	Volume	Journal volume.

15 ANHANG C: DIMDIS DATENBANKZUGÄNGE

Nachfolgend sind die Datenbanken aufgeführt, zu denen DIMDI einen (zumeist kostenpflichtigen) Zugang gewährt:

- *ABDA-AKTUELLE INFO*
- *ABDA-ARZNEISTOFFE*
- *ABDA-FERTIGARZNEIMITTEL*
- *ABDA-INTERAKTIONEN*
- *ABDA-WIRKSTOFFDOSSIERS*
- *AMED*
- *AMIS - Bezeichnungsverordnung*
- *AMIS - Öffentlicher Teil*
- *AnimAlt-ZEBET*
- *BGI-PRESSEDIENSTE*
- *Bibliotheksverbund Bayern*
- *BIOSIS PREVIEWS*
- *BIOTECHNOBASE*
- *BMGS-PRESSEMITTEILUNGEN*
- *Bundesanzeiger (amtl. Teil Ressort BMG)*
- *CAB ABSTRACTS*
- *CAB HEALTH*
- *CANCERLIT*
- *CATFILEplus*
- *CCRIS*
- *CHEMID*
- *CHEMLINE*
- *CIVS*
- *Cochrane - CDSR*
- *Cochrane - CENTRAL*
- *Cochrane - DARE*
- *DERWENT BIOTECH*
- *DERWENT DRUG BAKCFILE*
- *DERWENT DRUG FILE*
- *DIQ*
- *DITR*
- *Elsevier BIOBASE*
- *EMBASE*
- *EMBASE ALERT*
- *ETHMED*
- *EUROETHICS*
- *FSTA*
- *GEROLIT*
- *HEALTH DEVICES ALERTS (HDA)*
- *HEALTH DEVICES SOURCEBASE (HDS)*
- *HECLINET*
- *HSDB*
- *INT. HEALTH TECHNOLOGY ASSESSMENT (IHTA)*
- *INTOX*
- *IPA International Pharmaceutical Abstracts*
- *ISTP/ISSHP*
- *ISTPB*
- *Kluwer-Verlagsdatenbank*
- *Kontaktallergene*
- *Kunststoff-Empfehlungen*
- *LEBENSMITTEL-MONITORING (LM50)*
- *MEDIKAT*
- *MEDITEC*
- *Medizinprodukte - Öffentlicher Teil*
- *MEDLINE*
- *MEDLINE Alert*
- *OldMEDLINE*
- *PSYCINFO*
- *PSYNDEX*
- *RTECS*
- *SCISEARCH*
- *SERLINE*
- *SOCIAL SCISEARCH*
- *SOMED*
- *Springer-Verlagsdatenbank*
- *Springer-Verlagsdatenbank PrePrint*
- *Thieme-Verlagsdatenbank*
- *TOXBIO*
- *TOXCAS*
- *TOXLINE siehe XTOXLINE*
- *XTOXLINE*
- *ZEBET*

16 DANKSAGUNG

Am Ende dieser Arbeit möchte ich mich bei allen bedanken, die zum Gelingen dieser Arbeit beigetragen haben.

Besonders herzlichen Dank möchte ich meinen Doktoreltern, Hr. Univ.-Prof. Dr. Thurmayr und Fr. Prof. Dr. Thurmayr, für die hochschulseitige Betreuung dieser Arbeit aussprechen. Sie gaben mir wertvolle Hinweise und Anregungen und gewährten mir zugleich den nötigen Freiraum zur Verwirklichung meiner Ideen. Der rege Gedankenaustausch mit ihnen hat wesentlich zum Gelingen dieser Arbeit beigetragen.

Holger Tenner, Juni 2003