

Universität Leipzig  
Wirtschaftswissenschaftliche Fakultät

# Konzeption und Realisierung eines wissensbasierten Protokollauswahlmoduls für die Hämatonkologie

Betreuer : Dr. Barbara Heller  
Universität Leipzig  
Institut für Medizinische Informatik, Statistik und Epidemiologie  
Liebigstr. 27

Bearbeiter : Michael Krüger  
Könneritzstr. 74  
04225 Leipzig  
Matr. Nr. 7556247

Stand : 23. November 1999

## Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die das Entstehen meiner Diplomarbeit in dieser Form ermöglicht haben.

Mein Dank gilt Herrn Prof. Dr. M. Löffler für die Bereitstellung der Rahmenbedingungen und Frau Dr. L. Mantovani, die die Anregung für diese Arbeit gab.

Für die intensive Betreuung sowie die zahlreichen kritischen und hilfreichen Hinweise während der Entstehungszeit der Arbeit bedanke ich mich besonders bei Frau Dr. B. Heller.

Desweiteren bedanke ich mich bei Volker Dötsch, Marc Junger, Kristin Lippoldt, Henrik Müller und Steffen Stundzig für die zahlreichen Diskussionen und Anregungen während der Entwicklungszeit des Systems.

Michael Krüger

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Motivation und Problemstellung . . . . .	1
1.2	Anforderungen an das zu erstellende System . . . . .	2
1.3	Zielsetzung . . . . .	4
<b>2</b>	<b>Grundlagen und Rahmenbedingungen</b>	<b>5</b>
2.1	Begriffe . . . . .	5
2.1.1	Grundbegriffe der Onkologie . . . . .	5
2.1.2	Grundbegriffe der Informatik . . . . .	6
2.2	Medizinische Grundlagen . . . . .	8
2.2.1	Diagnostik . . . . .	8
2.2.2	Therapieziel . . . . .	8
2.2.3	Therapiemöglichkeiten . . . . .	9
2.2.4	Onkologische Therapiestudien . . . . .	10
2.3	Informatische Einführung . . . . .	11
2.3.1	Künstliche Intelligenz . . . . .	11
2.3.2	Fallbasiertes Schließen . . . . .	12
2.4	Medizinisches Umfeld . . . . .	14
2.4.1	Vorarbeiten . . . . .	15
2.4.2	Einsatzszenario . . . . .	17
<b>3</b>	<b>Stand der aktuellen Forschung</b>	<b>20</b>
3.1	Anforderungen an ein Therapieauswahlssystem . . . . .	20
3.2	Bestehende Systeme . . . . .	21
3.2.1	Das Expertensystem ONCO-HELP . . . . .	21
3.2.2	Die PDQ Clinical Trails Database . . . . .	23
3.2.3	Das Anwendungssystem TheMPO . . . . .	26
3.3	Vergleichende Beurteilung . . . . .	27
<b>4</b>	<b>Anforderungen an das Studienauswahlmodul SAM</b>	<b>28</b>
4.1	Die Vorgehensweise des Arztes bei der Therapieauswahl . . . . .	28
4.2	Anforderungen an das zu entwickelnde Studienauswahlmodul SAM . . . . .	29
<b>5</b>	<b>Konzeption der wissensbasierten Therapieauswahl</b>	<b>30</b>
5.1	Abstraktion und Modellierung der Auswahlstrategien . . . . .	30
5.1.1	Das Modell der Vorauswahl . . . . .	30

5.1.2	Der detaillierte Fallvergleich . . . . .	34
5.2	Synonymverwaltung . . . . .	41
5.3	Wissensrepräsentation und Aufbau der Fallbasis . . . . .	42
5.4	Strategisches Auswahlwissen . . . . .	45
5.4.1	Modell der Ähnlichkeitsbestimmung . . . . .	45
5.4.2	Ähnlichkeit in der Vorauswahl . . . . .	46
5.4.3	Die lokale Ähnlichkeitsbestimmung . . . . .	47
5.4.4	Die globale Ähnlichkeitsbestimmung . . . . .	48
<b>6</b>	<b>Das Studienauswahlmodul SAM</b>	<b>50</b>
6.1	Architektur des wissensbasierten Studienauswahlmoduls SAM . . . . .	50
6.1.1	Funktionsumfang des Anwendungssystems . . . . .	50
6.1.2	Der Therapieauswahlprozeß . . . . .	51
6.2	Die Fallbasis . . . . .	52
6.2.1	Struktur der Fallbasis . . . . .	52
6.2.2	Vorauswahl beim Fallvergleich . . . . .	53
6.2.3	Detaillierter Fallvergleich . . . . .	56
6.3	Die Erklärungskomponente . . . . .	59
6.3.1	Allgemeine Anforderungen an eine Erklärungskomponente . . . . .	59
6.3.2	Inhalte der Erklärungskomponente von SAM . . . . .	60
6.4	Die Datenbankanbindung . . . . .	61
6.5	Die interaktive Benutzeroberfläche . . . . .	62
6.5.1	Das Hauptfenster . . . . .	62
6.5.2	Die Dialoge . . . . .	64
6.6	Integration in das Gesamtsystem . . . . .	67
<b>7</b>	<b>Diskussion des erstellten Studienauswahlmoduls SAM</b>	<b>68</b>
7.1	Akzeptanzkriterien des Anwenders . . . . .	68
7.1.1	Anwenderfreundlichkeit . . . . .	69
7.1.2	Stabilität und Zuverlässigkeit . . . . .	70
7.1.3	Zeit- und Ressourcenverhalten . . . . .	70
7.2	Wissensbasis . . . . .	71
7.2.1	Medizinisches Wissen . . . . .	71
7.2.2	Erweiterbarkeit der Fallbasis . . . . .	72
7.3	Erweiterbarkeit . . . . .	72
<b>8</b>	<b>Zusammenfassung und Ausblick</b>	<b>73</b>

**A Literatur**

**75**

**B Abkürzungsverzeichnis**

**78**

## Abbildungsverzeichnis

1	Aufbau von InKA . . . . .	17
2	Geplante Kommunikations- und Informationsstruktur im MedNet-Projekt . .	19
3	Eingabemaske für die PDQ Clinical Trails Database . . . . .	24
4	Die Oberfläche von SAM . . . . .	63
5	Der Paßwortdialog . . . . .	64
6	Dialog zum Laden von Patientendaten . . . . .	65
7	Dialog zur Angabe der Datenbankpfade . . . . .	66

# 1 Einleitung

## 1.1 Motivation und Problemstellung

Die therapeutische Behandlung bösartiger Tumoren in der Erwachsenenonkologie ist durch eine Vielzahl existierender Behandlungsprotokolle charakterisiert. Einige dieser Behandlungsprotokolle werden nur in Studien angewendet, andere wiederum werden als Standardprotokolle betrachtet. Letztere werden auch als Goldstandard bezeichnet. Die Behandlungsprotokolle werden sowohl in der Fachliteratur als auch in Form von Studienprotokollen, welche von den studienleitenden Instituten herausgegeben werden, veröffentlicht. Momentan liegt ein Großteil dieser Behandlungsvorschriften nur in Papierform vor, so daß eine Informationsbeschaffung relativ aufwendig ist. Eine zielgerichtete und zudem komfortable Suche nach geeigneten Protokollen ist für den Arzt aufgrund der fehlenden rechentechnischen Unterstützung oft nicht möglich. Eine Unterstützung der Protokollauswahl scheint jedoch insbesondere bei seltenen Tumorerkrankungen notwendig, um einerseits schnell ein passendes Protokoll zur bestmöglichen Behandlung des Patienten zu finden. Andererseits ist es wichtig, den onkologischen Studien möglichst viele Patienten mit der entsprechenden Tumorerkrankung zuzuführen, um verlässliche Aussagen über die tatsächliche Qualität und Wirksamkeit der Therapie treffen zu können. Aussagen über die Wirksamkeit und die Qualität der Therapie spielen für den wissenschaftlichen und therapeutischen Fortschritt in der Onkologie eine große Rolle. Ein weiteres Problem ergibt sich aus der Vielzahl der existierenden Protokolle, welche z. T. geändert und aktualisiert werden, um neueste wissenschaftliche und therapeutische Erkenntnisse in den Behandlungsprozeß einfließen zu lassen und erfolgversprechende neue Strategien in der klinischen Anwendung zu untersuchen.

Neben den Studienprotokollen befinden sich auch die Bezeichnungen der Diagnosen sowie deren unterschiedliche Klassifikationen in einem allmählichen aber stetigen Wandel. Ein Beispiel für den Wandel der diagnostischen Klassifikation ist der Übergang von der Kiel-Klassifikation zur internationalen REAL-Klassifikation<sup>1</sup> in der Diagnostik der malignen Lymphome.

Aufgrund der existierenden Anzahl der Behandlungsprotokolle ist es für den Arzt oft schwierig, das für den Patienten am besten geeignete Behandlungsschema zu finden, da der Arzt in der Regel nicht alle Protokolle kennt. Diese Informationsdefizite auf der Seite des Arztes werden deutlich, wenn der relativ geringe Anteil der in den Therapiestudien erfaßten Lymphompatienten betrachtet wird. Nach [Heller & Löffler 1998] gehen Schätzungen hierbei von ca. 20 % bei den Non-Hodgkin-Lymphomen und von ca. 40 % bei den Hodgkin-Lymphomen aus. Diese Anteile unterliegen je nach histologischem Subtyp einigen Schwan-

---

<sup>1</sup>REAL-Klassifikation steht für Revised European-American Classification of Lymphoid Neoplasms.

kungen. Allerdings sind Unterschiede zwischen Nicht-Studienpatienten und Studienpatienten hinsichtlich der Qualität der Diagnostik, der Therapie und der Behandlungsergebnisse bisher ungeklärt<sup>2</sup>. Bei Nicht-Studienpatienten werden die Ergebnisse der Behandlung nicht dokumentiert, so daß die außerhalb von Studien gesammelten Erfahrungen dem wissenschaftlichen Erkenntnisprozeß nicht zur Verfügung stehen.

Aufgrund der im vorangegangenen Abschnitt geschilderten Problematik erscheint es sinnvoll, dem Arzt ein Softwarewerkzeug in die Hand zu geben, mit dessen Hilfe er schnell ein passendes Behandlungsprotokoll für den jeweiligen Patienten und darüber hinaus weitere studienrelevante Informationen finden kann. Wichtige Informationen in Bezug auf eine Studie sind z. B. die Adressen der Ansprechpartner, der Therapieplan und die erforderlichen Untersuchungen für eine Aufnahme in die Studie. Das Ziel eines solchen Softwarewerkzeuges besteht in der Reduzierung des Aufwandes für die Auswahl eines geeigneten Protokolls. Durch die Erleichterung der Protokollauswahl für die behandelnden Ärzte könnte die Zahl der Patienten, die innerhalb onkologischer Studien behandelt werden, erhöht werden. Dieser positive Effekt könnte indirekt durch einen verbesserten Informationsstand bei den behandelnden Ärzten verwirklicht werden.

Durch die rechentechnische Abbildung der Studien wird in Zukunft die Nutzung moderner Informations- und Kommunikationsstrukturen, wie z. B. das Internet, möglich. Im Vergleich zum Versenden der Studienprotokolle in Papierform könnte hierbei der Aufwand für die Verteilung der Informationen deutlich reduziert werden. Darüber hinaus ist durch eine rechnergestützte Aufbereitung bewerteter Listen mit geeigneten Protokollen mit einer Verringerung des Aufwandes für die Suche nach Protokollen zu rechnen.

## 1.2 Anforderungen an das zu erstellende System

Damit die Anforderungen an das zu erstellende System genau identifiziert werden können, ist zunächst die Betrachtung der Zielgruppe erforderlich. Die Zielgruppe besteht vorwiegend aus Ärzten und medizinischen Studienassistenten in onkologischen Fachpraxen und Kliniken. Dieser Personenkreis verfügt über ein hohes medizinisches Fachwissen. Im Umgang mit Computersystemen kann bei diesen Personen nur grundlegendes Bedienwissen vorausgesetzt werden. Hieraus ergeben sich hohe Anforderungen an eine komfortable Benutzerführung sowie an eine intuitiv bedienbare und leicht erlernbare Benutzeroberfläche. Mit Hilfe des zu entwerfenden Systems soll es möglich sein, dem Anwender aufgrund der Diagnose und weiteren, diagnosespezifischen Patientendaten ein oder mehrere potentiell geeignete Protokolle für die Behandlung vorzuschlagen. Für die Erstellung eines Therapievorschlages soll zusätzlich die Möglichkeit bestehen, nicht alle Daten zu erheben und aufgrund von unvollständigen

---

<sup>2</sup>Vgl. [Heller & Löffler 1998] S. 22

Patientendaten einen Vorschlag zu generieren. Das System sollte demnach mit fehlenden bzw. unvollständigen Daten umgehen können. Für die Einordnung der Daten in Pflichtdaten und Daten, welche für die Auswahl nicht unbedingt notwendig sind, muß der Aufwand zur Erhebung der jeweiligen Daten und der Einfluß, den die Patientendaten auf die Eignung in Bezug auf die Studie haben, betrachtet werden. Zwingend notwendige Patientendaten, wie z. B. die Diagnose, müssen eingegeben werden, damit die Suche nach einem geeigneten Behandlungsprotokoll gestartet werden kann. Eine wiederholte Suche nach Therapien mit vervollständigten Daten muß ebenfalls möglich sein. Mehrfacheingaben bei wiederholter Therapieauswahl sollen durch die Anbindung einer Datenbank für die Patientendaten vermieden werden. Weiterhin ist zu überdenken, ob eine Anbindung an ein bestehendes Krankenhausinformationssystem (KIS) sinnvoll ist und wie diese realisiert werden kann. Eine Anbindung an ein KIS wäre dann sinnvoll, wenn Mehrfacheingaben von Patientendaten in verschiedene Systeme vermieden werden könnten. Neben einer Zeitersparnis ist die Eliminierung zusätzlicher Fehlerquellen ein weiterer Vorteil für die Integration in ein KIS.

Aufgrund der Tatsache, daß viele potentielle Anwender über keinen permanenten Zugang zum Internet verfügen, ist es aus Gründen der Kostenreduktion notwendig, den Einsatz des Systems ohne permanenten Internetzugang zu ermöglichen. Der Verbindungsaufbau zur Studienzentrale, d. h. beispielsweise die Kommunikation über das Internet soll nur für den Abgleich der Studiendatenbank mit der lokalen Datenbank des Systems notwendig sein.

Neben den Anforderungen des Anwenders sind ebenfalls die Anforderungen aus Sicht des Softwareentwicklers an ein solches System zu beachten.

Hierzu ist die Oberfläche möglichst so zu gestalten, daß ein dynamischer Aufbau der Oberfläche zur Laufzeit in Abhängigkeit beispielsweise von der Diagnose realisiert werden kann. Weiterhin ist die Studiendatenbank so zu gestalten, daß neue Studien mit möglichst geringem Aufwand in die vorliegende Datenbankstruktur integrierbar sind. Diese ist so zu konzipieren, daß die auftretenden Varianzen zwischen den Studien ohne Probleme in dieser abgebildet werden können. Das Prozeßmodell für den Auswahlvorgang und den Auswahlalgorithmus ist so zu modellieren, daß fehlende Daten behandelt und neue Studien nach Aufnahme in die Datenbasis in den Auswahlvorgang mit einbezogen werden können.

Bei der Ergebnisrepräsentation ist darauf zu achten, daß das Ergebnis nur ein Vorschlag ist und das System nur entscheidungsunterstützend arbeiten soll. Die endgültige Entscheidung über die Behandlungsstrategie liegt beim behandelnden Arzt bzw. beim Studienarzt. Für die Repräsentation des Ergebnisses sollen die Studien entsprechend ihrer Eignung für den Patienten kategorisiert werden. Neben der übersichtlichen Repräsentation wird dem Anwender auch ein leistungsfähiger Mechanismus zum Nachvollziehen des Auswahlvorganges sowie zur Erklärung des Ergebnisses zur Verfügung gestellt. Dieser hat die Aufgabe, die ausschlag-

gebenden Merkmale für die entsprechende Kategorisierung zu identifizieren. Desweiteren soll der Anwender in die Lage versetzt werden, sich über die Studien zu informieren. Die oben angeführten Aufgaben können mit Hilfe einer Erklärungskomponente umgesetzt werden.

Zur Beschränkung des Umfanges der vorliegenden Diplomarbeit liegt der Fokus des Systems auf Studien aus dem Bereich der Hämatonkologie. Ein Hauptaugenmerk bei dem Entwurf und der Implementierung des Systems wird auf die Modularität der Wissensbasis in Hinblick auf die Erweiterbarkeit dieser um neue Studien gelegt.

### 1.3 Zielsetzung

Ziel dieser Arbeit ist die Konzeption und Realisierung eines wissensbasierten Systems für die Auswahl geeigneter Behandlungsprotokolle innerhalb der Hämatonkologie. Das Studienauswahlmodul SAM<sup>3</sup> hat die Aufgabe, den Arzt bzw. den Studienassistenten in einer Klinik oder in einer onkologischen Fachpraxis bei der Auswahl eines geeigneten Behandlungsprotokolls oder einer geeigneten Studie zu unterstützen. Ausgehend von der Vorgehensweise des Arztes soll hierzu ein Verfahren entwickelt werden, mit dessen Hilfe der Auswahlprozeß effizient abgebildet werden kann. Die vorliegende Arbeit beschreibt die Vorgehensweise bei der Entwicklung und Modellierung geeigneter Algorithmen und Strukturen.

Das System ist so zu entwerfen und zu realisieren, daß die im vorangegangenen Abschnitt umrissenen Anforderungen aus Sicht des Anwenders und des Entwicklers an ein solches System erfüllt werden. Diese Anforderungen werden ausführlich in den Abschnitten 3.1 und 4.2 beschrieben. Das System soll darüberhinaus dem Anwender die Möglichkeit geben, sich über Studien zu informieren und bei Bedarf sich Erklärungen zum Auswahlprozeß anzeigen zu lassen. Ein weiteres Ziel dieser Arbeit besteht darin, den Aufwand des Arztes bei der Suche nach einem geeigneten Studienprotokoll zu reduzieren und den Arzt über neue Studien zu informieren. Hierdurch erzielbare positive Nebeneffekte könnten einerseits in einem Kompetenzgewinn auf Seite des Arztes und andererseits in einer Erhöhung der Zahl der in Studien eingeschlossenen Patienten bestehen.

Die vorliegende Arbeit ist nicht isoliert, sondern im Kontext eines großen Verbundprojektes zum rechnerbasierten Studienmanagement im Bereich der Malignen Lymphome zu sehen. Während des Entwurfs und der Realisierung der vorliegenden Arbeit ist der Integrationsfähigkeit des Systems stets die notwendige Beachtung beizumessen. Ein mögliches Einsatzszenario für das zu erstellende System wird im Abschnitt 2.4.2 beschrieben. Das Studienauswahlmodul SAM soll als Softwarewerkzeug den Teilnehmern des Verbundprojektes „Maligne Lymphome“ nach einer Beta-Testphase zur Verfügung gestellt werden.

---

<sup>3</sup>SAM steht für StudienAuswahlModul.

## 2 Grundlagen und Rahmenbedingungen

### 2.1 Begriffe

#### 2.1.1 Grundbegriffe der Onkologie

Die folgenden Begriffe wurden in Anlehnung an [Ostendorf und Seeber 1998] sowie [Classen, Diehl und Kochsiek 1998] definiert.

**Onkologie** ist die Lehre der Geschwulsterkrankungen, insbesondere der bösartigen Formen.

Sie befaßt sich mit den Ursachen, den Eigenschaften und der Verbreitung der Geschwülste, sowie deren Vorbeugung, Früherkennung, Behandlung und Nachsorge.

**Maligne Tumore** sind bösartige Tumore, welche unbehandelt zum Tod des Patienten führen und sich durch ein überschießendes Wachstum körpereigener Zellen auszeichnen. Das Wachstum vollzieht sich autonom<sup>4</sup> und progressiv<sup>5</sup>. Maligne Tumore unterscheiden sich im Gegensatz zu den benignen Tumoren durch ein schnelles, invasives<sup>6</sup> und destruierendes<sup>7</sup> Wachstum. Häufig bilden sie auch Metastasen<sup>8</sup>.

**Maligne Lymphome** sind Neoplasien des lymphatischen Systems und werden nach histologischen Kriterien in die Gruppe der Hodgkin-Lymphome und der Non-Hodgkin-Lymphome unterteilt. In erster Linie sind die lymphatischen Organe<sup>9</sup> betroffen. Andere Organe können jedoch durch unmittelbare Tumordinfiltration des benachbarten Gewebes von einem erkrankten Lymphknoten aus befallen werden. In fortgeschritteneren Stadien ist eine Infiltration auch über hämatogene Wege möglich.

**Hodgkin-Lymphome** sind maligne Erkrankung des lymphatischen Systems. Kennzeichnend für das Hodgkin-Lymphom ist der histologische Nachweis von wenigen ein- und mehrkernigen malignen Riesenzellen, sogenannter Hodgkin- und Sternberg-Reed-Zellen, welche von einer Vielzahl reaktiver, vorwiegend lymphastischer Zellen umgeben sind. Die Ausbreitung erfolgt erst lymphastisch, im Spätstadium auch hämatogen.

**Non-Hodgkin-Lymphome** sind maligne Lymphome, die sich histologisch von den Hodgkin-Lymphomen durch den Nachweis von Hodgkin- und Sternberg-Reed-Zellen abgren-

---

<sup>4</sup>nach [Pschyrembel 1997]: unabhängig von den übergeordneten Regulationssystemen des Organismus

<sup>5</sup>nach [Pschyrembel 1997]: weiteres Wachstum des Tumors auch nach Beendigung des auslösenden Reizes

<sup>6</sup>nach [Pschyrembel 1997]: in Organe, Gefäße und Gewebe einbrechend

<sup>7</sup>nach [Pschyrembel 1997]: das ursprüngliche Gewebe zerstörend

<sup>8</sup>nach [Pschyrembel 1997]: im engeren Sinne Geschwulstmetastasen, insbesondere bei malignen Tumoren durch Verschleppung von Tumorzellen (sogen. Tochtergeschwulst) entstehend

<sup>9</sup>nach [Pschyrembel 1997]: werden unter dem Begriff des Lymphatischen Organs die Einheit von Lymphknoten und Milz verstanden

zen lassen. Non-Hodgkin-Lymphome bilden eine heterogene Gruppe lymphatischer Neoplasien, welche nach histologischen Kriterien eingeteilt werden.

**Behandlungsstrategie** beinhaltet die Aussage darüber, mit welcher der Behandlungsalternativen aus Operation, Bestrahlung und Internistischer Therapie bzw. einer Kombination daraus der Patient behandelt wird.

**Onkologisches Therapieschema** bzw. Therapieprotokoll, im folgenden auch kurz als Protokoll bezeichnet, umfaßt eine Kette von Behandlungsschritten aus Operation, Radio- und/oder Chemotherapie.

**Onkologische Therapiestudien**, im folgenden nur noch als Studien bezeichnet, bestehen aus zwei oder mehr onkologischen Therapieschemata, die vergleichend auf ihre Wirksamkeit untersucht werden. Ein Therapieschema wird hier als Studienarm bezeichnet. In der Regel handelt es sich bei onkologischen Therapiestudien um kontrollierte, prolektive<sup>10</sup>, klinische Studien. Die Zuordnung erfolgt meist randomisiert. Die Randomisation soll eine zufällige Aufteilung der Patientengruppen zur Sicherung der Strukturgleichheit gewährleisten. Die Strukturgleichheit der Gruppen ist für die Auswertbarkeit der Studien notwendig.

**Studienprotokoll** ist die Behandlungsvorschrift für die an der Studie teilnehmenden Patienten.

**Ein- und Ausschlußkriterien** sind Kriterien, welche die Patientenzielgruppe der Studie beschreiben. Solche Kriterien können sich auf direkte Patientenmerkmale, z. B. Alter, Diagnose, Geschlecht usw., und/oder auf Vor- und Begleiterkrankungen bzw. Vorbehandlungen beziehen.

### 2.1.2 Grundbegriffe der Informatik

In diesem Abschnitt sollen einige wichtige Begriffe aus dem Bereich der Informatik und insbesondere aus dem Teilbereich der Künstlichen Intelligenz kurz beschrieben werden:

**Künstliche Intelligenz** ist ein Teilgebiet der Informatik und verfolgt nach [Mainzer 1999] die Idee, daß das menschliche Problemlösen mit Hilfe von programmgesteuerten Computern simuliert werden kann.

---

<sup>10</sup> Aufteilung der Patientenpopulation in eine Interventionsgruppe, welche die zu testende Therapie erhält, und einer Kontrollgruppe, welche die Standardtherapie erhält. Nach Behandlungsende wird der Erfolg der Therapien miteinander verglichen.

**Wissen** ist ein extrahiertes Abbild der realen Welt. Innerhalb von Softwaresystemen wird Wissen in ein Teilmodell der realen Welt abgebildet. Dieses Teilmodell wird häufig auch als Weltausschnitt bezeichnet<sup>11</sup>.

**Wissensrepräsentation** ist die symbolische Darstellung von Wissen innerhalb eines rechnergestützten Systems für die automatische Wissensverarbeitung und Manipulation auf syntaktischer Ebene<sup>12</sup>.

**Eine Wissensbasis** ist eine Datenstruktur, welche als Beschreibung von Zuständen der Welt bzw. eines Weltausschnittes interpretiert werden kann. Die Wahl des geeigneten Repräsentationsschemas ist stark vom Anwendungsgebiet und den zu lösenden Problemen abhängig<sup>13</sup>.

**Wissensbasierte Systeme** sind rechnergestützte Anwendungssysteme zur Lösung von Problemen, bei den programmtechnisch Problemlösungsmethoden und Wissen getrennt sind<sup>14</sup>.

**Fall** ist die Beschreibung einer Problemsituation zusammen mit den Erfahrungen, die bei der Problemlösung gewonnen werden konnten. Häufig wird ein Fall nur als ein Paar aus Problembeschreibung und Problemlösung verstanden<sup>15</sup>.

**Fallbasis** ist eine Sammlung von Fällen, die auf eine geeignete Art und Weise organisiert ist<sup>16</sup>.

**Fallbasiertes Schließen** ist ein Ansatz zur Problemlösung, bei dem Lösungen ähnlicher, schon einmal gelöster Problemfälle auf das aktuelle Problem übertragen werden. Der Grundgedanke dieses Ansatzes besteht in der Annahme, daß ähnliche Probleme (Fälle) auch ähnliche Lösungen besitzen. Ziel dieses Ansatzes ist es, Wissen erneut zu verwenden<sup>17</sup>.

**Ähnlichkeitsmaß** ist eine (mathematische) Funktion zur Berechnung der Ähnlichkeit zweier Fälle. Die Ähnlichkeit der Fälle kann dabei implizit, durch die Angabe von Funktionswerten, oder explizit, durch die Angabe einer Funktion, erfolgen<sup>18</sup>.

---

<sup>11</sup>Vgl. [Petersen 1996], S. 18

<sup>12</sup>Vgl. [Petersen 1996], S. 20

<sup>13</sup>Vgl. [Helbig 1996], S. 249

<sup>14</sup>Vgl. [Puppe 1990], S. 238

<sup>15</sup>Vgl. [Stolpmann & Weiß 1999], S. 67

<sup>16</sup>Vgl. [Weiß 1996], S. 26

<sup>17</sup>Vgl. [Weiß 1996], S. 28

<sup>18</sup>Vgl. [Petersen 1996], S. 31

**Auswahlstrategie** beschreibt den Prozeß der Suche nach ähnlichen Fällen und umfaßt die Interpretation der Fallbeschreibung, die Verfolgung der Indexstruktur der Datenbank, den Ähnlichkeitsvergleich sowie das Ranking<sup>19</sup>.

## 2.2 Medizinische Grundlagen

### 2.2.1 Diagnostik

Die Stellung einer zuverlässigen Tumordiagnose ist die Voraussetzung für die Durchführung einer Therapie. Dabei gehören zur Diagnosestellung die Anamnese, die klinische Untersuchung, die Laboruntersuchung, die Anwendung diagnostischer Verfahren, wie z. B. der Computertomographie, Sonographie u.a., sowie die histologische Absicherung der Diagnose. Dadurch sollen der Malignitätsgrad, die Ausbreitung und die Metastasierung festgestellt, sowie das Ursprungsgewebe identifiziert werden. Auf der Grundlage der Ergebnisse dieser Daten erfolgt die Klassifikation der Krankheit und die Stadieneinteilung. In der Hämatonkologie, speziell für die Einteilung von Hodgkin-Lymphomen bzw. Non-Hodgkin-Lymphomen, werden die Klassifikationen nach der Rye-Konferenz<sup>20</sup> bzw. die Kiel- und die REAL-Klassifikation<sup>21</sup> verwendet. Allerdings existieren noch eine ganze Reihe weiterer Klassifikationen, so daß eine allgemeingültige Klassifizierung nicht möglich ist. Die Einteilung der Stadien wird in der Hämatonkologie meist mit Hilfe der Klassifikation nach Ann Arbor vorgenommen, wobei es auch hier wieder zu Abweichungen kommen kann.

Für die Entscheidung über eine geeignete Therapie reicht meist eine grobe Klassifikation der Diagnose aus<sup>22</sup>. Die Erstellung einer Differentialdiagnose ist dennoch notwendig, da einzelne Subtypen unterschiedliche Prognosen besitzen. Weiterhin ist die Differentialdiagnose wichtig für die Dokumentation, insbesondere für die Studiendokumentation.

### 2.2.2 Therapieziel

Die Wahl einer speziellen Therapie sowie deren Art und Intensität ist vom Behandlungsziel unter Berücksichtigung des allgemeinen Zustandes des Patienten abhängig. Eine kurative Therapie hat die Heilung des Patienten und damit die definitive Beseitigung der Krankheit zum Ziel. Potentiell auftretende Nebenwirkungen und bleibende Schäden, welche durch die Therapie verursacht werden können, werden hier bewußt in Kauf genommen. Für den Fall, daß die Krankheit bereits so weit fortgeschritten ist, daß eine Heilung des Patienten nicht mehr zu erreichen ist, wird eine palliative Therapie angewendet. Ziel einer palliativen Therapie

---

<sup>19</sup>Vgl. [Petersen 1996], S. 32

<sup>20</sup>Vgl. [Diehl & Pfreundschuh 1997], S. 536

<sup>21</sup>Vgl. [Engelhard et al. 1997], S. 547 und S. 552 f.

<sup>22</sup>Vgl. [Diehl & Pfreundschuh 1997], S. 540 und [Engelhard et al. 1997], S. 550 und S. 554

ist es, die Lebensqualität zu verbessern, die Beschwerden zu lindern, den Progreß der Krankheit aufzuhalten und die Überlebenszeit des Patienten zu verlängern. Neben den palliativen und kurativen Therapien werden auch die adjunktiven Therapien angewendet. Adjunktive Therapien werden oft im Anschluß an potentiell kurative Therapien durchgeführt und dienen meist der Rezidivprophylaxe<sup>23</sup>.

### 2.2.3 Therapiemöglichkeiten

Für die Behandlung von Tumorerkrankungen stehen prinzipiell drei verschiedene Behandlungsmethoden zur Verfügung: die operative Behandlung, die Strahlentherapie sowie die internistische Therapie, welche sich wiederum in Chemotherapie, Immuntherapie und Hormontherapie unterteilen läßt<sup>24</sup>. Ausschlaggebend für die Wahl der therapeutischen Methoden sind neben der Art des Tumors auch das Stadium, der Differenzierungsgrad, die Lokalisation des Tumors sowie Alter und Gesamtzustand des Patienten. Hierbei ist allerdings zu beachten, daß in der Hämatookologie die operative Behandlung aufgrund der Pathologie der Tumore meist nicht in Frage kommt. Eine Ausnahme bilden hier die primären gastrointestinalen Non-Hodgkin-Lymphome. Meist wird ein Ansatz gewählt, der zwei oder mehr Behandlungsmethoden miteinander kombiniert damit eine möglichst große Wirkung bei Behandlung von Tumoren erzielt werden kann.

Die Operation ist die älteste Form der Behandlung von Tumoren und kann sowohl palliativ als auch kurativ eingesetzt werden. Oft wird die Operation auch zur Reduktion der Tumormasse angewendet damit die Effektivität einer anschließenden Chemo- und/oder Strahlentherapie erhöht werden kann. Bei Operationen besteht jedoch die Gefahr der intraoperativen Tumorverschleppung.

Bei inoperablen Tumoren besteht die Möglichkeit der Behandlung mit ionisierender Strahlung. Diese Methode der Behandlung ist möglich, da die malignen Tumore aufgrund ihres geringen Differenzierungsgrades sowie ihrer, durch das schnellere Wachstum bedingten, höheren Zellteilungsrate strahlenempfindlicher sind als normales Gewebe. Die verwendete Strahlung, die Art der Bestrahlung, sowie deren Stärke sind von der Lokalisation des Tumors und der Tumorart und der Größe des Tumors abhängig. Die Effektivität der Strahlentherapie sinkt mit zunehmender Tumormasse, da auch die Zellteilungsrate mit zunehmender Tumormasse abnimmt. Die Strahlentherapie kann ebenso wie die Operation palliativ und kurativ eingesetzt werden.

Die dritte Behandlungsform ist die internistische Therapie. Die internistische Therapie wirkt im Gegensatz zu den Therapiemöglichkeiten der Operation und der Strahlentherapie

---

<sup>23</sup>Siehe hierzu auch [Classen, Diehl & Kochsiek 1998], S. 120 f.

<sup>24</sup>Vgl. [Classen, Diehl & Kochsiek 1998], S. 121 ff.

nicht nur lokal begrenzt, sondern systemisch. Bei diesem Therapieansatz kommen Zytostatika zum Einsatz. Die Behandlung mit Zytostatika soll bei maximaler Schonung der Normalzellen möglichst selektiv die Krebszellen abtöten. Auf die Behandlung mit Zytostatika sprechen insbesondere Zellen in der Zellteilungsphase an. Hieraus resultiert eine hohe Wirksamkeit der Behandlung auf Tumore mit einer hohen Zellteilungsrate. Bei kleinen Tumoren ist die Zellteilungsrate besonders hoch und damit auch die Wirksamkeit der Zytostatika. Mit zunehmender TumorgroÙe sinkt die Wachstumsrate und damit auch die Zellteilungsrate. Die sinkende Wachstums- und Zellteilungsrate hat eine verminderte Wirksamkeit der Zytostatika zur Folge. Meist findet eine Kombination verschiedener Medikamente bei der internistischen Therapie Anwendung, um einen additiven Effekt bei relativ geringer Summation der Nebenwirkungen zu erzielen. Eine solche kombinierte Therapie hat folgende Vorteile:

- Maximale Abtötung der Tumorzellen, da es möglich ist, für jede einzelne Substanz den möglichen Toxizitätsbereich voll auszuschöpfen
- Abtötung von Zellen, welche bereits gegen eine Substanz resistent sind
- Verlangsamung der Entwicklung resistenter Zellmutationen

Der Erfolg einer internistischen Therapie ist stark von der richtigen Dosierung und Auswahl der Substanzen, sowie dem Zeitplan für deren Verabreichung abhängig. Für die meisten Tumorbehandlungen, insbesondere mit einer kurativen Zielstellung, gilt es, die maximal vertretbare Dosis zu verabreichen. Eine Verringerung der Dosis bewirkt in diesem Zusammenhang meist eine Verringerung der Heilungschancen. Die Reduktion der Dosis wird oft nur bei starken Nebenwirkungen oder einer Verschlechterung des Gesamtzustandes des Patienten vorgenommen. Die internistische Therapie kann sowohl palliativ als auch kurativ eingesetzt werden.

Nähere Ausführungen zu den einzelnen Therapiemöglichkeiten sind in [Classen, Diehl & Kochsiek 1998] zu finden.

#### **2.2.4 Onkologische Therapiestudien**

Onkologische Therapiestudien sind nicht mit den speziellen Therapiemöglichkeiten, wie im Abschnitt 2.2.3 beschrieben, gleichzusetzen, sondern bilden den organisatorischen Rahmen für die Entwicklung und die Überprüfung der Effizienz neuer Behandlungsstrategien bzw. Therapieschemata. Diese Studien dienen in erster Linie dem therapeutischen und wissenschaftlichen Fortschritt in der Onkologie.

Meist werden innerhalb dieser Studien zwei oder mehr alternative Behandlungsschemata miteinander verglichen. Die Behandlungsschemata werden im Studienkontext als Studienar-

me bezeichnet. Anhand der Studie sollen gemäß den Studienzielkriterien Aussagen über die Qualität der einzelnen Behandlungsschemata gemacht werden. Studienzielkriterien können unter anderem Zielgrößen sein, wie Ansprechen der Therapie, rezidivfreies Überleben, Gesamtüberleben, Lokalisation der Rezidive und Langzeittoxizität. Nach Abschluß der Studie wird diese entsprechend den Studienzielkriterien ausgewertet. Allerdings werden bereits während der Laufzeit der Studie Zwischenauswertungen durchgeführt, um beim Auftreten von signifikant besseren oder schlechteren Ergebnissen reagieren zu können. Reaktionen werden bei einer Häufung therapiebedingter Todesfälle bzw. starker therapiebedingter toxischer Nebenwirkungen, sowie signifikanter Überlegenheit oder Unterlegenheit eines Therapiearmes notwendig.

Über den Einschluß eines Patienten in die Studie entscheiden bestimmte Ein- und Ausschlußkriterien. Diese Kriterien sollen die Patientenzielgruppe für eine Studie beschreiben und können sich auf direkte Patientenmerkmale, z. B. Alter, Diagnose, Geschlecht usw., und/oder auf Vor- und Begleiterkrankungen bzw. Vorbehandlungen beziehen. Der Ausschluß eines Patienten sagt allerdings nicht zwangsläufig aus, daß er nicht nach diesem Studienprotokoll behandelt werden kann, sondern nur, daß seine Daten nicht mit in die Auswertung der Studie einbezogen werden, um die Ergebnisse der Auswertung nicht zu verfälschen.

## 2.3 Informatische Einführung

### 2.3.1 Künstliche Intelligenz

Im Abschnitt 2.1.2 wurde bereits der Begriff der Künstlichen Intelligenz (KI) kurz beschrieben. An dieser Stelle soll näher auf den Begriff der Künstlichen Intelligenz eingegangen werden.

In der Literatur sind viele unterschiedliche Definitionen zum Gebiet der Künstlichen Intelligenz zu finden. Eine der Definitionen von Partrick Winston, Leiter des KI-Labors des Massachusetts Institute of Technology besagt, daß die Künstliche Intelligenz die Untersuchung von Ideen ist, die es Computern ermöglichen, intelligent zu sein. Eine Definition im gewöhnlichem Sinn erscheint aber unmöglich, denn die Intelligenz kann als eine Mischung vieler informationstechnischer und -verarbeitender Talente aufgefaßt werden.

Dennoch können die Ziele der Künstlichen Intelligenz in Anlehnung an Winston [Winston 1989] wie folgt definiert werden:

- Ein Hauptziel der Künstlichen Intelligenz besteht darin, Computer nützlicher zu machen.
- Ein weiteres zentrales Anliegen besteht darin, die Prinzipien zu verstehen, die Intelligenz ermöglichen.

Die Künstliche Intelligenz beschäftigt sich mit der maschinellen Repräsentation und Verarbeitung menschlicher Wissensstrukturen als Grundlage für komplexe Informationsverarbeitungsprozesse. Mit Hilfe von Verfahren der Künstlichen Intelligenz können zum einen relevante Aspekte intelligenten Verhaltens wie beispielsweise Schlußfolgerungsprozesse bezogen auf anspruchsvolle informationstechnische Anwendungen maschinell verfügbar gemacht werden. Zum anderen können nach [Winston 1989] und vielen anderen Autoren aus dem Bereich der Künstlichen Intelligenz Methoden der Künstlichen Intelligenz zur Herleitung neuer wissenschaftlicher Erklärungsmodelle für kognitive Leistungen herangezogen werden.

Bei der Realisierung von Systemen mit Problemlösungsmethoden aus dem Bereich der KI werden zwei grundlegende Strategien unterschieden. Die erste Methode versucht das menschliche Denken und Vorgehen beim Lösen von Problemen zu erforschen und anschließend nachzubilden. Diese Herangehensweise wird oft auch als Simulationsmethode bezeichnet. Zu den Simulationsmethoden werden beispielsweise die künstlichen neuronalen Netze und z. T. auch das fallbasierte Schließen gezählt. Die andere Methode versucht sich an den rein phänomenologischen Merkmalen der Intelligenz zu orientieren. Die auftretenden Phänomene werden dabei nicht primär auf die Ursache ihres Auftretens bzw. ihres Zustandekommens hin untersucht. Das Ziel dieser Methode wird erreicht, wenn das entwickelte System sich intelligent verhält, ohne dabei die Denkprozesse des Menschen zu simulieren. Die Vorgehensweise wird als phänomenologische Methode bezeichnet und bildet beispielsweise die Grundlage für Deduktionssysteme.

### 2.3.2 Fallbasiertes Schließen

Das fallbasierte Schließen (Case-Based Reasoning, kurz CBR) ist ein Problemlösungsansatz, bei dem Wissen zur Lösung eines aktuellen Problems aus einer früheren konkreten Situation herangezogen wird. Dieses Wissen besteht im Gegensatz zu anderen Problemlösungsverfahren aus dem Gebiet der Künstlichen Intelligenz weniger aus allgemeinen Prinzipien und Regeln, sondern vielmehr aus spezifischem Wissen in Form von situations- und kontextbezogenen Erfahrungen<sup>25</sup>. Die folgenden Beispiele sollen den Grundgedanken des fallbasierten Schließens verdeutlichen<sup>26</sup>:

- Der Arzt untersucht einen Patienten, der eine ungewöhnliche Kombination von Symptomen aufweist. Er erinnert sich an einen Patienten mit vergleichbaren Symptomen, an die dort gestellte Diagnose, die durchgeführte Therapie und den Verlauf der Behandlung. Mit Hilfe dieses Wissens wird in der jetzigen Situation eine vergleichbare Diagnose gestellt.

---

<sup>25</sup>Vgl. [Kolonder 1993], S. 14

<sup>26</sup>Vgl. [Schulz 1998], S. 97

- Ein Rechtsanwalt vertritt einen Mandanten vor Gericht. Er erinnert sich an einen ähnlichen Fall, sowie die damals verwendete Strategie der Verteidigung. Die aktuelle Strategie der Verteidigung gründet sich auf die Strategie und die Argumentation im Präzedenzfall.
- Ein Servicetechniker erinnert sich aufgrund der Fehlerbeschreibung an einen früheren Defekt eines Gerätes. Bei dem aktuellen Problem geht er analog zu der damals verwendeten Strategie vor.

Die Vorgehensweise des fallbasierten Schließens wird, wie diese Beispiele verdeutlichen, vom Menschen im alltäglichen Leben verwendet. Das fallbasierte Schließen wird daher als natürlicher Bestandteil des menschlichen Denkens und Problemlösens angesehen und bildet aus Sicht der Kognitionswissenschaften ein Modell für einen Teil des menschlichen Problemlösungsverhaltens. Daher ist es naheliegend, diese Strategie auch auf rechnergestützte Problemlösungsprozesse zu übertragen. Die Grundidee des fallbasierten Schließens liegt in der Annahme, daß ähnlich gelagerte Probleme auch ähnliche Lösungen besitzen<sup>27</sup>.

Der Prozeß des fallbasierten Schließens läßt sich in vier Phasen unterteilen:

- Retrieve: Bestimmen des ähnlichsten Falles aus der Fallbasis
- Reuse: Adaption der gefundenen Lösung und deren Anwendung
- Revise: Verifizieren der gefundenen Problemlösung
- Retain: Speicherung des neu gewonnenen Falles für eine spätere Wiederverwendung

In der ersten Phase wird die Fallbasis durchsucht. Dieser Vorgang entspricht dem menschlichem Erinnern an einen früheren Fall. Ziel dieses Vorganges ist es, einen gespeicherten Fall, welcher dem aktuellen Problem am ähnlichsten ist, zu finden. Die Ähnlichkeit wird mit Hilfe eines Ähnlichkeitsmaßes bestimmt. Dieses Ähnlichkeitsmaß ist eine spezielle Funktion, welche im Prinzip auf alle Fälle angewandt wird. Die Fälle werden anschließend entsprechend der berechneten Ähnlichkeitswerte sortiert. Das verwendete Ähnlichkeitsmaß sowie die Struktur und Organisation der Fallbasis ist hierbei von der gegebenen Problemstellung abhängig.

Während der Phase des Reuse werden die gefundenen Lösungen auf das aktuelle Problem übertragen. Diese Phase entspricht einem Wiederverwenden mit Anpassen. Die Anpassung der Lösung kann in verschiedenen Stufen und Formen erfolgen. Im trivialen Fall, d. h. eine bereits vorhandene Lösung kann direkt auf einen neuen Fall übertragen werden, ist keine Anpassung der Lösung notwendig. In diesem Falle spricht man auch von einer Null-Adaption.

---

<sup>27</sup>Vgl. [Schulz 1998], S. 97 f.

Ist hingegen eine Adaption notwendig, kann sie entweder interaktiv mit dem Nutzer erfolgen oder aufgrund der im Wissensmodell gespeicherten Strategien und Verfahren vorgenommen werden.

In der dritten Phase wird eine Lösungsverifikation vorgenommen. Die Lösung soll bestätigt werden. Diese Phase dient meist der Qualitätskontrolle der gefundenen Lösung. Desweiteren kann der Vorgang der Adaption überwacht werden.

Die abschließende Phase des Retain kann dem menschlichen Lernen gleichgesetzt werden. An dieser Stelle wird eine erfolgreich angepaßte und verifizierte Problemlösung in die Fallbasis aufgenommen. Mit Hilfe dieser Phase ist es möglich, die Problemlösungskompetenz eines fallbasierten Systems zu erhöhen. Allerdings kann sich durch eine Aufnahme neuer Fälle auch die Notwendigkeit einer Umstrukturierung der Fallbasis ergeben. Fallbasierte Systeme unterscheiden sich von anderen Methoden der künstlichen Intelligenz durch wichtige Kriterien, welche den praktischen Nutzen wissensbasierter Systeme entscheidend beeinflussen<sup>28</sup>:

- Vermeidung eines hohen Wissensakquisitionsaufwandes
- einfache Wartbarkeit der Wissensbasis
- höhere Effizienz beim Problemlösungsvorgehen
- bessere Ausnutzung bestehender Datenbestände, insbesondere durch unscharfe Anfragen und Verwendung von mehr Wissen
- gesteigerte Benutzerakzeptanz durch bessere Nachvollziehbarkeit der Ergebnisfindung

Mit Hilfe von fallbasierten Strategien und Problemlösungsmethoden ist es oft möglich, mit relativ geringem Aufwand ein System zu entwickeln, das bei Routinetätigkeiten den Anwender bei der Entscheidungsfindung unterstützen kann. Ein fallbasiertes System liefert zu einem aktuellen Problem alle ähnlichen Fälle der Fallbasis. Dies ist zwar keine Garantie, daß sich die gesuchte Lösung unter diesen Fällen befindet. Der Lösungsraum kann jedoch oft so eingeschränkt werden, daß die endgültigen Lösungsfindung nur noch geringen Aufwand für den Anwender bedeutet.

Das fallbasierte Schließen hat sich in der letzten Zeit zu einer wichtigen Problemlösungsmethode bei der Entwicklung leistungsfähiger wissensbasierter Softwarewerkzeuge entwickelt.

## 2.4 Medizinisches Umfeld

Im Studienkontext arbeiten zur Zeit eine Vielzahl verschiedener Einrichtungen. So sind neben den Kliniken und onkologischen Fachpraxen eine Reihe von Referenzzentren, Studienzentra-

---

<sup>28</sup>Vgl. [Stolpmann & Weiß 1999], S. 72

len bzw. Studiensekretariaten und biometrischen Zentren in den Studienkontext involviert. Aufgrund dieser Vielzahl von Einrichtungen, die multizentrisch und interdisziplinär arbeiten, ergibt sich ein hoher Kommunikationsbedarf. Ein Großteil der Kommunikation zwischen den verschiedenen Instituten, Einrichtungen und Kliniken erfolgt in Papierform. Die Abläufe innerhalb der Studiendurchführung werden größtenteils noch nicht informationstechnisch unterstützt. Dies gilt auch für die Bereitstellung vieler Informationen, welche oft nur in Papierform vorliegen. So ist z. B. eine rechnergestützte Aktualisierung der Studienprotokolle und Leitlinien nicht möglich. Die Ursache hierfür ist häufig darin zu sehen, daß die Studienprotokolle nur selten in elektronischer Form vorliegen und somit eine ständige Aktualisierung der Studienprotokolle sehr aufwendig und teuer ist.

Aufgrund dieses Mangels besteht ein hoher Bedarf, die Datenkreisläufe und die Informationsbereitstellung zu optimieren und mit Hilfe von modernen informationstechnischen Methoden effizient zu gestalten und zu unterstützen. Weitere Bestrebungen gehen dahin, die Abläufe innerhalb der Studiendurchführung zu optimieren und zu standardisieren. Darüberhinaus besteht die Notwendigkeit, die Therapieoptimierungsstudien im Bereich der Hämatonkologie auf einen einheitlichen Qualitätsstandard zu bringen und die Abläufe informationstechnisch zu unterstützen.

#### 2.4.1 Vorarbeiten

Zur informationstechnischen Unterstützung der Arbeit im Studienkontext wurden bereits zwei Softwaremodule am Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE) der Universität Leipzig entwickelt. Dabei handelt es sich um das Therapieplanungsmodul ThEO<sup>29</sup> und das Softwaremodul InKA<sup>30</sup>.

#### ThEO

ThEO ist ein wissensbasiertes Werkzeug zur Dosismodifikation und zur Erstellung von individuellen Therapieplänen und wurde im Rahmen einer Diplomarbeit an der Universität Leipzig<sup>31</sup> entwickelt. Zur Zeit befindet sich ThEO in einer Beta-Testphase an unterschiedlichen Kliniken und Praxen in Deutschland.

Mit Hilfe von ThEO sollen auf Grundlage der gestellten Diagnose Therapievorschlüsse abgerufen und patientenspezifische Therapiepläne entwickelt werden können. Die Kommunikation mit dem Anwender erfolgt über eine interaktive Benutzerschnittstelle, welche sich durch eine intuitive Bedienbarkeit auszeichnet.

---

<sup>29</sup>Therapieplanung in der Erwachsenen-Onkologie, vgl. [Berger 1996]

<sup>30</sup>InternetKommunikationApplikation, vgl. [Forster & Krüger 1998]

<sup>31</sup>Vgl. [Berger 1996]

ThEO besteht aus zwei voneinander getrennten Programmodulen:

- Anwendungsmodul
- Definitionsmodul

Das Anwendungsmodul unterstützt den Anwender bei der Auswahl geeigneter Studien und bei der Erstellung des individuellen Therapieplans. Desweiteren soll es die Dokumentation und Durchführung der studienbezogenen Therapieschemata erleichtern.

Das Definitionsprogramm hat die Aufgabe, dem Studiensekretariat und dem Anwender die Neuaufnahme und die Verwaltung von Studien und anderen Therapieschemata zu erleichtern. Mit Hilfe des Definitionsprogrammes können hier z. B. Therapien definiert werden, welche beim Anwender in der onkologischen Praxis bzw. in der Klinik „Standard“ sind.

### **InKA**

Das Internet-Kommunikations-Applikationsmodul InKA wurde im Rahmen einer Seminararbeit<sup>32</sup> am IMISE mit dem Ziel der Erhöhung der Datenqualität, der schnelleren Datenerfassung von Studiendaten sowie der Verbesserung der Kommunikationsstruktur entwickelt. Da im Studienkontext sensible Daten verarbeitet werden, müssen an die Datensicherheit besonders hohe Anforderungen gestellt werden. Die Datensicherheit bei dieser Anwendung wird durch ein mehrstufiges Sicherheitskonzept aus asymmetrischer Verschlüsselung, „Punkt zu Punkt“-Verbindungen, Benutzerauthentifikation und Transaktionsnummern realisiert.

Neben den Universitätskliniken und großen Krankenhäusern sollen insbesondere onkologischen Fachpraxen und Kliniken ohne permanenten Internetzugang eine Mitarbeit innerhalb multizentrischer Studien ermöglicht werden. Aus diesem Grund wurde es nötig, InKA als sogenannte „Offline“-Kommunikationslösung<sup>33</sup> umzusetzen. Durch die Umsetzung eines solchen Verfahrens können die Kosten für die Kommunikation der Praxen und Kliniken mit den Studienzentralen reduziert werden. Dem Anwender wird es durch dieses Konzept möglich, offline auf der lokalen Datenbank zu arbeiten und nur für die Zeit der Datenübertragung zwischen der lokalen Anwenderdatenbank und der zentralen Studiendatenbank bzw. des Datenbankgleiches mit der zentralen Studiendatenbank eine Online-Verbindung über das Internet aufzubauen. Die Abb. 1 soll den Aufbau von InKA verdeutlichen.

---

<sup>32</sup>Vgl. [Forster & Krüger 1998]

<sup>33</sup>Der Begriff der „Offline“-Kommunikation ist hier nicht wörtlich zu verstehen. Eine Kommunikation setzt immer eine Verbindungen zwischen zwei oder mehr beteiligten Kommunikationspartnern voraus. Der Begriff „Offline“-Kommunikation ist so zu verstehen, daß der Anwender lokal ohne ständigen Kontakt zum Server bzw. der Serverdatenbank arbeiten kann. Eine Verbindung zwischen Client und Server ist nur zum Abgleich und der Aktualisierung der Datenbanken notwendig. Um diese Strategie umzusetzen, wurde das Tool InKA entwickelt.

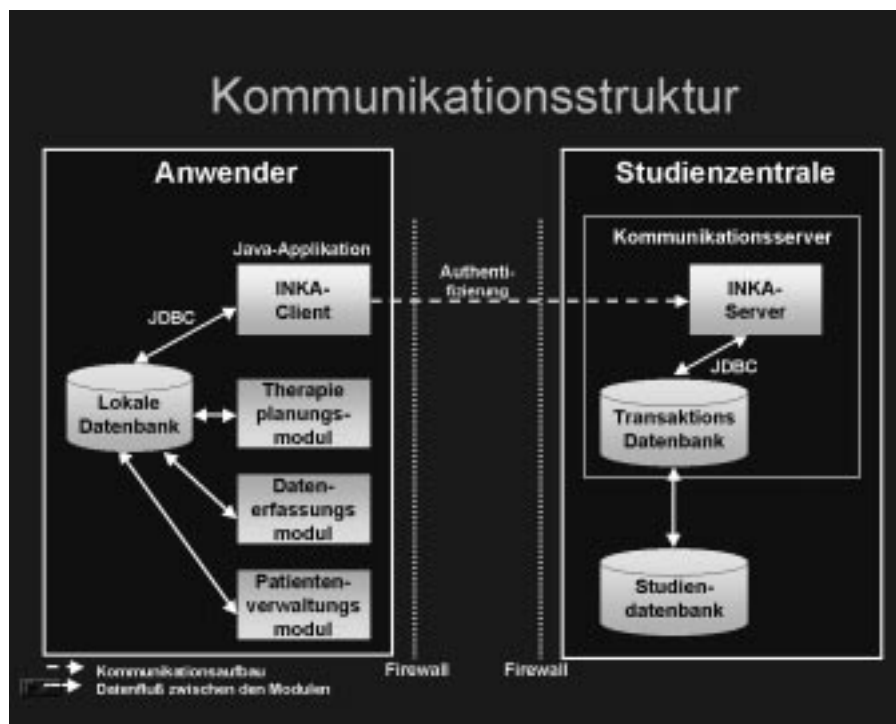


Abb. 1: Aufbau von InKA (Quelle: [Forster & Krüger 1998])

In Hinblick auf Plattform- und Datenbankunabhängigkeit wurde InKA in der Programmiersprache JAVA unter Verwendung der auf Standards basierenden Schnittstelle JDBC<sup>34</sup> entwickelt. Nähere Ausführungen hierzu sind in [Forster & Krüger 1998] zu finden.

#### 2.4.2 Einsatzszenario

Für die Konzeption und den Aufbau eines medizinischen Kompetenznetzes innerhalb des MedNet-Verbundes sollen rechnergestützte Werkzeuge für die Unterstützung der Daten- und Informationsflüsse zwischen den einzelnen Teilnehmern entwickelt und diesen zur Verfügung gestellt werden. Die Entwicklung der rechnergestützten Werkzeuge wird innerhalb eines Gemeinschaftsprojektes des IMISE (Studienzentrale für hochmaligne Lymphome in Leipzig), der Universität Köln (Studienzentrale für Morbus-Hodgkin-Lymphome) und der Universität München (Studienzentrale für niedrigmaligne Lymphome) erfolgen. Als Kommunikationsmedium wird neben Punkt-zu-Punkt-Verbindungen hauptsächlich das Internet eingesetzt.

Als einen ersten Schritt in diese Richtung sind die bisher entwickelten Softwarewerkzeuge ThEO und InKA zu sehen. Das zu erstellende Therapieauswahlssystem wird nach einer Beta-Testphase zu diesen beiden Werkzeugen hinzukommen. Das Therapieplanungssystem ThEO

<sup>34</sup>JAVA Database Connectivity

und das Studienauswahlmodul SAM sollen auf die gleichen Datenbanken aufsetzen, damit die Anwender insbesondere bei der Dateneingabe entlastet werden. Zudem wird der Verwaltungsaufwand zur Administration der Datenbanken so gering wie möglich gehalten. Ein weiterer positiver Effekt gemeinsamer Datenquellen besteht in einer Verminderung von inkonsistenten Daten in der Datenbasis, welche durch Fehleingaben entstehen können.

Das Softwarewerkzeug InKA ermöglicht die Kommunikation zwischen den lokalen Datenbanken und der zentralen Studiendatenbank. Auf diesem Weg ist die Arbeit mit den Softwarewerkzeugen auf der lokalen Datenbank offline möglich. Geplant ist der Einsatz dieser drei Softwarewerkzeuge in Universitätskliniken, onkologischen Fachpraxen und Kliniken.

Ein weiterer Schritt besteht in der standardisierten Aufbereitung der Studienprotokolle, sowie deren Repräsentation auf einem Informationsserver. Mit Hilfe dieses Informationsservers soll ein Informationsdienst für die Teilnehmer des MedNet-Verbundes realisiert werden, der auch für die Studienrepräsentation innerhalb der einzelnen Module eingesetzt werden wird.

Innerhalb des MedNet-Projektes ist die Entwicklung einer Reihe modularer Softwarewerkzeuge geplant, welche die Planung, Durchführung und Auswertung klinischer Studien in der Hämatookologie unterstützen. Eine Hauptaufgabe des Projektes liegt in der Optimierung des Ablaufs, des Informationsflusses und des Datenflusses. Geplant ist die Umsetzung der in Abb. 2 dargestellten Kommunikations- und Informationsstruktur.

Neben dieser Kommunikations- und Infrastruktur sollen leistungsfähige Konzepte für den Datenschutz und die Datensicherheit erarbeitet werden. Desweiteren sollen im Rahmen dieses Projektes Konzepte und Anwendungen für ein umfassendes Qualitäts-, Dokumenten- und Workflowmanagement im Studienkontext erarbeitet und umgesetzt werden.

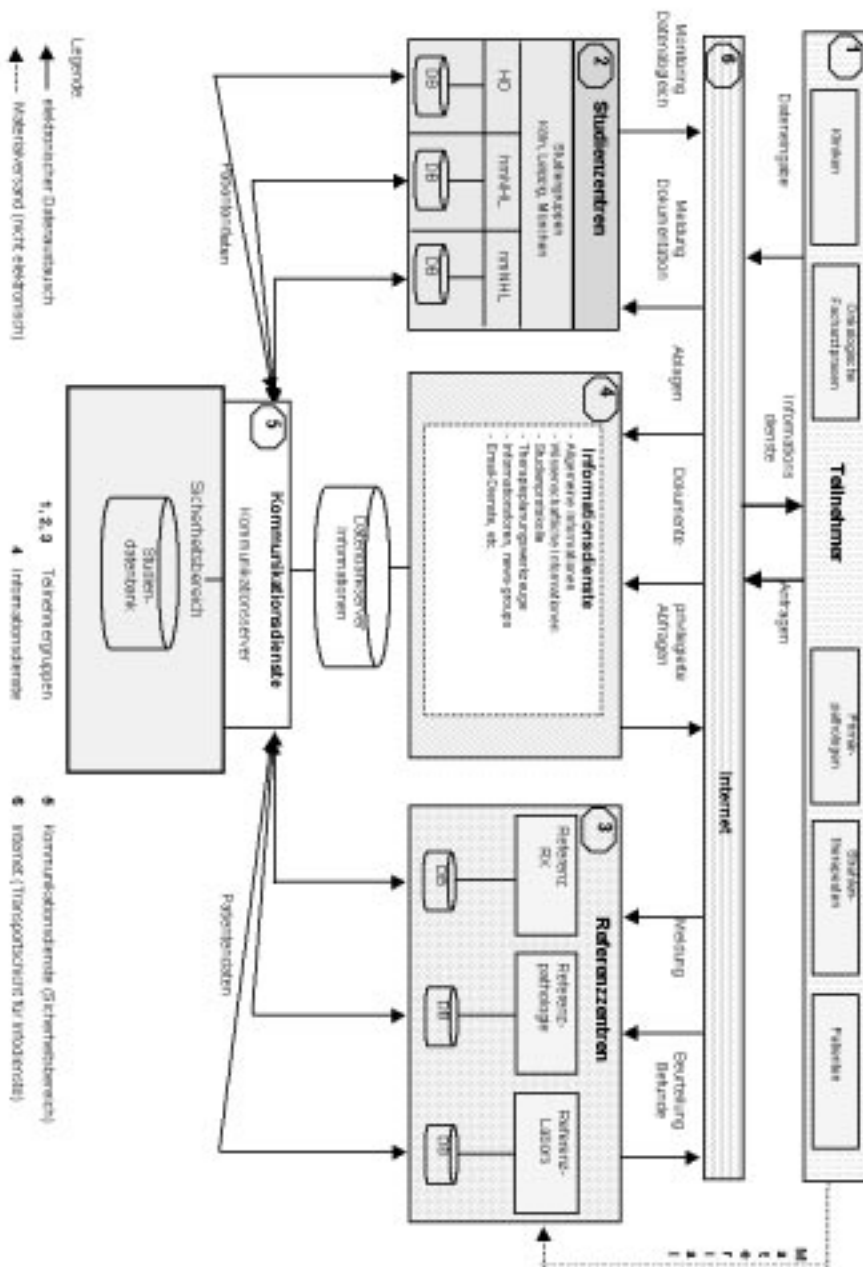


Abb. 2: Geplante Kommunikations- und Informationsstruktur im MedNet-Projekt (Quelle: [Heller & Löffler 1998], S.19)

## 3 Stand der aktuellen Forschung

### 3.1 Anforderungen an ein Therapieauswahlssystem

Wie jedes andere Anwendungssystem hat auch ein Therapieauswahlssystem allgemeine Anforderungen zu erfüllen. Zuerst wären hier allgemeine Akzeptanzkriterien zu nennen. Darunter fallen hauptsächlich die Forderung nach einer leicht erlernbaren und intuitiven Bedienbarkeit, einem leistungsfähigen Online-Hilfesystem, die Vermeidung von Mehraufwand durch Doppelingaben, Stabilität des Anwendungssystems, sowie kurze Antwortzeiten.

Neben den allgemeinen Anforderungen an ein Wissensbasiertes Therapieauswahlssystem, sind auch die spezifischen Anforderungen der Anwender zu berücksichtigen. Viele Anwender akzeptieren ein Wissensbasiertes System nur dann, wenn ihnen die Möglichkeit gegeben wird, die Ergebnisse des Schlußfolgerungsprozesse nachvollziehen zu können. Hieraus ergibt sich die Notwendigkeit des Einsatzes einer Erklärungskomponente. Wichtig für die Akzeptanz des Systems ist auch die Bedingung, daß das System nur entscheidungsunterstützend arbeitet, das bedeutet, daß die endgültige Entscheidung über die Therapie trifft der behandelnde Arzt bzw. der Studienarzt allein trifft. Wünschenswert ist zur besseren Entscheidungsunterstützung im Hinblick auf die Entwicklung eines Therapieauswahlsystems auch eine Abstufung des Eignungsgrades einer konkreten Studie für den Patienten.

Weitere Anforderungen ergeben sich aus der Sicht des Entwicklers, wie die Forderung nach Modularität, Erweiterbarkeit, leichte Wartbarkeit und Wiederverwendbarkeit.

Für die Studienauswahl sollte das System einen leistungsfähigen Algorithmus besitzen. Dieser Auswahlalgorithmus ist so zu konzipieren und zu implementieren, daß fehlende Daten entsprechend ihrer Wichtigkeit für den Auswahlprozeß toleriert werden können, d. h. die Eingabe aller Patientendaten ist nicht zwingend notwendig. Die Notwendigkeit der Toleranz gegenüber fehlenden Daten ergibt sich im wesentlichen daraus, daß einige Studien sehr spezifische Patientendaten verlangen. Zur Erhebung dieser Daten sind aber aufwendige und z. T. sehr invasive Untersuchungen notwendig. Zur Schonung des Patienten und auch aus Kostengründen empfiehlt es sich, diese Untersuchungen erst dann durchzuführen, wenn eine endgültige Eignung der Studie für den Patienten von diesen Untersuchungen abhängt.

Eine weitere Anforderung an das System ergibt sich aus der Betrachtung der Zielgruppe für ein solches System. Ein Therapieauswahlssystem kommt sowohl in Universitätskliniken und großen Krankenhäusern, als auch in Kliniken und onkologischen Fachpraxen zum Einsatz. Die onkologischen Fachpraxen und Kliniken verfügen im Gegensatz zu Universitäten und Universitätskliniken oft über keine permanente Kommunikationsverbindung mit dem Internet. Daher sollte ein solches System auch die Möglichkeit bieten, ohne eine ständige Verbindung zu einer Studienzentrale zu arbeiten. Der Verbindungsaufbau zum Studiensekretariat wäre

dann lediglich zur Aktualisierung und zum Abgleich der Studiendatenbank nötig<sup>35</sup>.

Die Zielgruppe für das System verfügt neben unterschiedlichen Zugang zum Internet auch über eine heterogene Rechnerumgebung. Diese Umgebung verlangt auch nach der Verfügbarkeit eines Therapieauswahlmoduls auf mehreren Plattformen bzw. Betriebssystemen.

Von diesen Punkten ausgehend lassen sich folgende Anforderungen an ein Therapieauswahlmodul formulieren:

- Intuitive und leicht erlernbare Bedienbarkeit
- Anbindung einer Patientendatenbank zur Vermeidung von Mehrfacheingaben
- Einfache Wartbarkeit insbesondere der Studien- bzw. Therapiedatenbank
- Toleranz gegenüber fehlenden Werten
- Generierung bewerteter Therapieempfehlungen
- Möglichkeit zur Offline-Bearbeitung
- Erklärungsfähigkeit der Ergebnisfindung und der Ergebnisse
- Kurze Antwortzeiten
- Einsatz unter verschiedenen Betriebssystemen (WinNT, Win9x, Unix, MacOS)

Mit Hilfe dieser Anforderungen sollen im folgenden Abschnitt einige bestehende Systeme zur Therapieauswahl bewertet werden.

## 3.2 Bestehende Systeme

### 3.2.1 Das Expertensystem ONCO-HELP

Das Expertensystem ONCO-HELP ist Teil des Projektes ONCO-CONS und wurde am Zentrum für angewandte Informationstechnologien (ZAIT) der Universität Bremen entwickelt<sup>36</sup>. ONCO-HELP soll eine individuelle und prognoseorientierte Behandlung von Tumorkranken ermöglichen und ist für mehrere Teilgebiete der Onkologie (z. B. Mammakarzinome, Bronchialkarzinome, Ovarialkarzinome, Melanome und Lymphome) verfügbar.

Durch die Erfassung individueller Patientendaten von Tumorart, Tumorhistologie, Metastasierungstyp, Metastasenlokalisierung und ihrer Anzahl sowie relevanter Labordaten wird mit Hilfe einer entsprechenden Wissensbasis eine individuelle Prognose erstellt und ein Therapieempfehlungsmodell entworfen. Zum kompletten Therapieempfehlungsmodell gehört auch die Erstellung einer

---

<sup>35</sup>Vgl. Abschnitt 1.2

<sup>36</sup>Vgl. [www.zait.uni-bremen.de/kilabor](http://www.zait.uni-bremen.de/kilabor)

effizienten Schmerztherapie und die Entwicklung eines Nachsorgeschemas. ONCO-HELP unterstützt weiterhin die Kontrolle des Therapiekonzeptes hinsichtlich Tumorprogression bzw. Tumorregression und Nebenwirkungen der Therapie. Gegebenenfalls wird eine Modifizierung des Konzeptes oder ein Wechsel der Therapie vorgeschlagen, das bedeutet, daß ONCO-HELP auch therapiebegleitend eingesetzt werden kann.

ONCO-HELP verfügt über pathophysiologisches und pharmakologisches Wissen, sowie über eine Plausibilitätskontrolle der gemessenen Werten hinsichtlich der Normwerte und der Werte im Zeitvergleich. Insgesamt bietet es folgende Funktionalität:

- Patientendatenerfassung
- Vorschlagsgenerierung für die Therapie, insbesondere für die adjunktive und palliative Therapie
- Therapiedokumentation
- Dosismodifikation
- Protokollgenerierung
- Individuelle Anpassung von Therapieschemata und Therapieabfolgen

Ein weiterer Vorteil von ONCO-HELP liegt bei den geringen Ansprüchen an die Hardware und das verwendete Betriebssystem. Als Systemvoraussetzung für den Einsatz von ONCO-HELP wird ein IBM-kompatibler PC mit einem Prozessor ab 80386 mit 33 MHz sowie mindestens 4 MByte Hauptspeicher angegeben. Als Betriebssystem wird mindestens MS-DOS 3.3 und MS-Windows 3.1 verlangt.

Der Nachteil dieses Systems besteht neben der Beschränkung auf die Windows-Betriebssysteme und der Hardwareplattform der PCs hauptsächlich darin, daß bei der Generierung der Therapievorschlüge keine Bewertung vorgenommen wird. Weiterhin fehlen zentrale Verwaltungsmechanismen für die Datenbasis, so daß jeder Nutzer für die Datenhaltung selbst verantwortlich ist. Ein einheitlicher Informationsstand kann daher nicht gewährleistet werden, da die einzelnen Nutzer auf einer unterschiedlichen Datenbasis operieren und kein Mechanismus vorgesehen ist, mit dessen Hilfe ein Datenaustausch möglich ist. Ursache hierfür liegt in der Konzeption des Systems als Stand-alone-System. Wird der Funktionsumfang betrachtet, liegt das Hauptaugenmerk von ONCO-HELP mehr auf der Therapiekontrolle und Therapieüberwachung als auf der Erstellung von Therapievorschlügen. Ein weiterer Schwerpunkt des Systems liegt in der Unterstützung der Verwaltung und Definition eigener Therapien. Für diese können Ein- und Ausschlußkriterien definiert werden, deren Einhaltung das System im speziellen Fall überprüft.

### 3.2.2 Die PDQ Clinical Trails Database

Die PDQ<sup>37</sup> Clinical Trails Database ist eine vom National Cancer Institute (NCI) der Vereinigten Staaten erstellte Datenbank mit Informationen über maligne Tumorerkrankungen für Patienten, Angehörige von Patienten, sowie für Ärzte und Krankenschwestern. Nach Angaben des NCI umfaßt die PDQ Datenbank Informationen von über 1700 klinischen Studien aus dem Bereich der Onkologie. Eine Aktualisierung der Daten erfolgt monatlich<sup>38</sup>.

Der Zugriff auf die Daten dieser Datenbank erfolgt über das World Wide Web<sup>39</sup>. Neben Informationen über die klinischen Studien sind weiterhin Informationen über die Tumörprävention, Therapiemöglichkeiten und supportive Maßnahmen abrufbar. Alle angeforderten Informationen erhält der Anwender in zwei Ausführungen. Die erste Version ist für Nutzer mit medizinischem Wissen, insbesondere für Beschäftigte im Gesundheitswesen bestimmt. Die Zielgruppe der zweiten Version sind medizinische Laien, also in der Regel die Patienten und deren Angehörige.

Die Suche nach Informationen in der Datenbank erfolgt mit Hilfe von zwei Arten von Informationen:

- Medizinische Informationen: Hierunter fallen Daten wie die Diagnose, der Typ der Studie, die Art der Therapie und der individuelle Zustand.
- Allgemeine Suchkriterien: PDQ stellt hier die Möglichkeit bereit, mit Hilfe von verschiedenen Kriterien wie die Phase der Studie, den Sponsor oder die geographische Lage den Suchraum zu beschränken.

Der Anwender hat mit Hilfe eines Formulars (Abb. 3) die Möglichkeit, die Daten einzugeben und die Abfrage zu starten. Anschließend erhält der Anwender das Ergebnis, welches aus Informationen für den Patienten und Informationen für den Arzt besteht.

---

<sup>37</sup>Physician Data Query

<sup>38</sup>Vgl. <http://cancernet.nci.nih.gov/trails/users.html>

<sup>39</sup>Der Aufruf eines Formulars erfolgt über die URL: <http://cancernet.nci.nih.gov/prot/patsrch.shtml>. Siehe auch Abb. 3

PDQ® Clinical Trials Search Form - Patients - Netscape

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Stop

Bookmarks Location: <http://cancernet.nci.nih.gov/prot/patsrch.shtml> What's Related

**NATIONAL CANCER INSTITUTE** PDQ® Clinical Trials Search Form • Patients

To search for clinical trials in the PDQ database, fill in the boxes to select the appropriate criteria and click on the Search button at the bottom of the page. (See [PDQ Clinical Trials User's Guide](#) for additional assistance). If your browser can't handle forms, try our [text menu](#).

Clinical Trials New to PDQ (for trials entered in the last month).

**Diagnosis:**

**Phase:**

**Modality:**

**Trial Type:**

**Drug:**

**Protocol ID:**

**Sponsorship:**

**City:**

**State:**

**Country:**

NCI Protocols at NIH Clinical Center only

Search for PDQ clinical trials at Veterans Affairs (VA) and Military Treatment Facilities:

**VA Hospitals/Medical Centers:**

**Military Treatment Facilities:**

For information about veteran and military personnel/dependent participation in clinical trials, see [cancerTrials™](#).

Last updated: 09/99.

Document: Done

Abb. 3: Eingabemaske für die PDQ Clinical Trails Database (Quelle: <http://cancernet.nci.nih.gov/prot.patsrch.shtml>)

Die Informationen für den Patienten bestehen aus:

- **Rationale:** Die Rationale begründet, warum die Studie durchgeführt wird.
- **Ziel der Studie:** Das Ziel der Studie beschreibt, welche Annahmen mit dieser Studie begründet werden sollen.
- **Ein- und Ausschlußkriterien:** Diese Kriterien beschreiben die Patientenzielgruppe der Studie. Allerdings wird hier auf das fehlende medizinische Hintergrundwissen Rücksicht genommen, so daß die Liste der Kriterien vereinfacht und nicht vollständig ist.
- **Die Therapie:** Dieser Abschnitt beschreibt den Ablauf, die Zeitdauer und die Art der Therapie.
- **Ansprechpartner**

Im Gegensatz zu den Informationen für den Patienten werden dem Arzt folgende Informationen bereitgestellt:

- **Der Status der Studie,** welcher dem Arzt mitteilt, ob in die Studie im Moment Patienten aufgenommen werden oder nicht.
- **Die Ziele der Studie,** welche dem Arzt die Annahmen, die mit dieser Studie begründet werden sollen, darlegen.
- **Die Ein- und Ausschlußkriterien,** welche die Patientenzielgruppe der Studie beschreiben. Die Liste der Kriterien wird an dieser Stelle vollständig wiedergegeben und läßt sich in drei Bereiche unterteilen: die Krankheitscharakteristika, derzeitige und frühere Therapien, sowie patientenspezifische Daten.
- **Die Protokollbeschreibung,** welche eine detaillierte Beschreibung der Therapie und des zeitlichen Ablaufes der Studie enthält.

Die PDQ Clinical Trails Database ist ein Werkzeug, mit dessen Hilfe sich der Patient oder der Arzt über Studien informieren kann. Mit Hilfe des zur Verfügung gestellten Eingabeformulars ist eine gezielte Suche nach Studien für eine Tumorart möglich. Die Suchanfrage kann durch Angabe weiterer Patientendaten und weiterer Parametern, wie den Status der Studie und geographische Daten, eingeschränkt werden. Aufgrund der Verwendung eines Internetbrowser ist eine leichte Bedienbarkeit und die Verfügbarkeit dieses Werkzeuges auf relativ vielen Plattformen gewährleistet. Mit Hilfe dieser Datenbank und der verwendeten Formulare kann ein großer Personenkreis erreicht werden. Für das Erreichen einer möglichst hohen

Akzeptanz bei einem großen Personenkreis wurde dem unterschiedlichen Bildungsstand der potentiellen Anwender durch die Einrichtung von zwei Nutzergruppen Rechnung getragen.

Die Verwendung eines Browsers hat den Nachteil, daß die Daten und Suchanfragen nicht verschlüsselt werden. Dies liegt daran, daß die Verwendung einer Client/Server-Architektur nicht im Fokus dieser Anwendung liegt. Das Hauptziel dieser Anwendung ist eher in der Informationsbereitstellung und Informationsverteilung zu sehen. Ein weiterer Nachteil besteht darin, daß die Informationen nur in englischer Sprache zur Verfügung stehen. Dieser Fakt ist durch den hauptsächlichen Einsatz in den USA bedingt. Eine Bewertung der Studien hinsichtlich des Grades der Eignung erfolgt nicht durch die PDQ Datenbank, so daß der Arzt eine Einschätzung hinsichtlich der Eignung für den Patienten selbst vornehmen muß.

### 3.2.3 Das Anwendungssystem TheMPO

TheMPO<sup>40</sup> ist ein wissensbasiertes, protokollbezogenes Therapieplanungssystem für die pädiatrische Onkologie<sup>41</sup> und wurde am Universitätsklinikum Mainz erstellt<sup>42</sup>.

Das wissensbasierte Anwendungssystem TheMPO unterstützt die Therapieauswahl, Therapieplanung und die Therapiedokumentation in der pädiatrischen Onkologie. Das System verfügt neben einer Patientendatenbank über einen Therapieeditor zur Eingabe von Therapien, einen Protokolleditor zur Eingabe und Festlegung von zeitlichen Zusammenhängen und Abläufen sowie einen Regeleditor zur Festlegung von End-, Abbruch-, Verzweigungs- und Ausschlußbedingungen.

Ein Auswahlmodul generiert aufgrund vom Patientendaten Therapievorschlüge. Die Therapievorschlüge können vom Arzt angenommen, verworfen oder geändert werden. Mit Hilfe des Systems TheMPO kann desweiteren eine elektronische Patientenakte zur Patientenverfolgung geführt werden. Weitere Funktionen von TheMPO umfassen die Planung von Dosismodifikationen innerhalb von Therapien und das Erstellen von individuellen Therapieplänen. Aufgrund des vorgestellten Funktionsumfang wird deutlich, daß der Hauptschwerpunkt des Anwendungssystems TheMPO in der Therapieplanung und der Dokumentation der Therapie liegt. Obwohl das System für die pädiatrische Onkologie konzipiert wurde, ist es flexibel genug, um auch Therapieschemata der Erwachsenen-Onkologie aufzunehmen und zu verarbeiten.

Das System TheMPO wurde in der Programmiersprache C++ erstellt und nutzt die objektorientierte Datenbank POET. Implementiert wurde das System in einer NeXTStep-Umgebung. Die hieraus resultierenden Hard- und Softwareanforderungen schränken den Einsatzbereich des Systems TheMPO stark ein, da eine NeXTStep-Umgebung relativ teure Hard-

---

<sup>40</sup>TheMPO steht für Therapy Management in Pediatric Oncology

<sup>41</sup>nach [Pschyrembel 1997]: Kinder-Onkologie

<sup>42</sup>Vgl. [Müller 1997]

ware (Workstation) voraussetzt und damit der Verbreitungsgrad relativ gering ist. Aufgrund des geringen Verbreitungsgrades der Hardware kann das Anwendungssystem TheMPO auch keinen großen Verbreitungsgrad erreichen.

### **3.3 Vergleichende Beurteilung**

In diesem Abschnitt soll eine vergleichende Beurteilung der im vorangegangenen Abschnitt beschriebenen Systeme unter Beachtung der im Abschnitt 3.1 gestellten Anforderungen an ein Therapieauswahlsystem erfolgen.

Die Forderung nach Toleranz gegenüber fehlenden Patientendaten ist ebenso wie die Forderung nach relativ kurzen Antwortzeiten durch alle drei Systeme erfüllt. Von den drei vorgestellten Systemen verfügen jedoch nur ONCO-HELP und TheMPO über eine Patientendatenbank. Diese beiden Systeme bieten auch die Möglichkeit zur Offline-Bearbeitung und können dem Anwender die Ergebnisse sowie die Ergebnisfindung erklären. Die beiden Systeme ONCO-HELP und TheMPO sind nur für eine Plattform (WinNT/Win9x bzw. NeXTStep) verfügbar. Eine einfache Wartbarkeit der Studien- bzw. Therapiedatenbanken der Systeme, insbesondere der Abgleich bzw. ein Update der Datenbanken durch eine zentrale Stelle ist durch die Aufnahme von anwenderspezifischen Therapien in die Datenbasis der jeweiligen Anwendungssysteme nicht mehr gegeben.

Die Forderung nach einer einfachen Wartbarkeit wird nur durch die PDQ Clinical Trails Database erfüllt. Dieses System ist durch die Verwendung eines Browsers als Oberfläche als einziges auf mehreren Plattformen verfügbar.

## 4 Anforderungen an das Studienauswahlmodul SAM

### 4.1 Die Vorgehensweise des Arztes bei der Therapieauswahl

Für die Entwicklung effizienter und korrekter Strategien für die Therapieauswahl ist zunächst eine detaillierte Betrachtung der Vorgehensweise des Arztes notwendig.

Die Auswahl des Arztes erfolgt meist nach dem Prinzip: „Letztens hatte ich schon mal einen Patienten mit dieser Diagnose, den habe ich den nach dem Protokoll der Studie X behandelt. Da muß ich doch mal nachschauen, ob dieser Patient auch in die Studie X paßt.“ Diese Beschreibung erinnert sehr an ein fallbasiertes System, ist aber für die Modellierung einer geeigneten Auswahlstrategie zu ungenau. An dieser Stelle wird daher eine genauere Analyse der Vorgehensweise des Arztes notwendig.

Der Arzt wählt im ersten Schritt aus den ihm bekannten Studien bzw. Therapien<sup>43</sup>, die Studien heraus, welche anhand der Diagnose und evtl. auch nach dem Stadium der Krankheit und dem Alter zu dem speziellen Patienten passen. Daraufhin unterzieht der Arzt die so ausgewählten Studien einer genaueren Prüfung. Dabei vergleicht der Arzt die ihm bekannten Patientendaten mit den Ein- und Ausschlußkriterien der einzelnen Studien. Innerhalb dieses Prozesses unterscheidet der Arzt zwischen harten und weichen Ein- und Ausschlußkriterien. Harte Kriterien sind solche Kriterien, gegen die nicht verstoßen werden darf. Die weichen Kriterien dürfen hingegen mehr oder weniger stark verletzt werden. Eine stärkere Abweichung eines Patientendatums von den Ein- und Ausschlußkriterien kann der Arzt tolerieren, wenn bestimmte andere Werte in Ordnung sind. Diese Einschätzung kann der Arzt aufgrund seiner Erfahrung und seines medizinischen Wissens treffen.

Bei fehlenden Patientendaten hat der Arzt die Möglichkeit, eine Entscheidung über die Eignung einer Studie für einen speziellen Patienten zurückzustellen. Wichtig für die Entscheidung über eine Zurückstellung sind Faktoren, wie der Aufwand, der betrieben werden muß, um die notwendigen Daten zu erhalten, der Grad der Eignung, der ohne dieses Datum erreicht wurde bzw. erreicht werden kann und der Einfluß, den das betreffende Datum auf die Eignung hat. Diese Faktoren sind auch dafür entscheidend, ob der Arzt letztlich das Datum bzw. die Daten erhebt, oder sich für eine der anderen Studien entscheidet. Weitere Faktoren, die mit in die Bewertung der Studie einfließen können, sind z. B. persönliche Präferenzen und Erfahrungen, die der Arzt gemacht hat, sowie die Einschätzung der Persönlichkeit des Patienten. Die Einschätzung der Persönlichkeit des Patienten ist besonders wichtig bei Hochdosisstudien, bei denen z. T. relativ viele und schwere Nebenwirkungen auftreten. In der Praxis besteht auch die Möglichkeit, daß der Arzt keine geeignete Studie für den Patienten

---

<sup>43</sup>Zur Vereinfachung soll im Folgenden nicht mehr zwischen Studien und anderen Therapien unterschieden werden. Studie und Therapie werden unter dem Begriff der Studie zusammengefaßt werden.

kennt. An dieser Stelle muß der Selektionsvorgang unterbrochen werden. Für den Arzt besteht nun die Möglichkeit, sich entweder Informationen über andere Studien zu beschaffen und diese dann in die Auswahl einzubeziehen oder den Patienten an einen anderen Arzt bzw. eine andere Klinik zu überweisen.

Am Ende dieses Vorganges hat der Arzt dann eine Studie ausgewählt, die für den Patienten geeignet ist und nach welcher er den Patienten letztlich behandeln will.

## 4.2 Anforderungen an das zu entwickelnde Studienauswahlmodul SAM

Bei der Evaluation der bereits vorhandenen Systeme wurde deutlich, daß keins dieser Anwendungssysteme den im Abschnitt 3.1 gestellten Anforderungen genügt. Ausgehend von den Anforderungen im Abschnitt 3.1 und der im vorangegangenen Abschnitt beschriebenen Vorgehensweise des Arztes sollen die Anforderungen an das zu entwickelnde Studienauswahlmodul SAM aufgestellt werden.

Neben den unter Abschnitt 3.1 aufgeführten Anforderungen an ein Therapieauswahlsystem sind, vor dem Hintergrund eines wissensbasierten, speziell eines fallvergleichenden Therapieauswahlsystems, folgende weitere Anforderungen zu erfüllen:

- Leichte Erweiterbarkeit der Fallbasis
- Integrierbarkeit in das Gesamtprojekt<sup>44</sup>
- Toleranz gegenüber fehlenden Werten
- Einbeziehung synonymmer Begriffsstrukturen in den Prozeß der Therapieauswahl

In den Abschnitten 5 und 6 wird die Konzeption und Realisierung eines wissensbasierten Therapieauswahlsystems, welches den obigen Anforderungen genügt, vorgestellt.

---

<sup>44</sup>Vgl. Abschnitt 2.4 bzw. 2.4.2. Das zu entwickelnde System soll im Rahmen des MedNet-Projektes mit anderen Komponenten zusammenarbeiten. Eine dahingehende Integrierbarkeit ist demzufolge zwingend notwendig.

## 5 Konzeption der wissensbasierten Therapieauswahl

### 5.1 Abstraktion und Modellierung der Auswahlstrategien

Im Abschnitt 4.1 wurde die Vorgehensweise des Arztes bei der Therapieauswahl beschrieben. Bei der Betrachtung des Vorgehens fällt auf, daß dieser Vorgang in zwei Teilprozesse aufgeteilt werden kann. Im ersten Teilprozeß, der Vorauswahl, wird anhand relevanter und möglichst trennender Merkmale versucht, die Menge der zu untersuchenden Studien schnell einzuengen. Unter trennenden Merkmalen sind hier solche Merkmale zu verstehen, die gut zwischen potentiell geeigneten und potentiell ungeeigneten Studien unterscheiden. Diese Restriktion des Lösungsraums wird mit Hilfe einiger Regeln vorgenommen.

Der zweite Teilprozeß besteht aus einer detaillierteren Betrachtung der zuvor ausgewählten Studien. In diesem zweiten Teilprozeß werden Patientendaten und Ein- und Ausschlußkriterien in einem detaillierten Vergleich auf Übereinstimmung geprüft. Dieses Vorgehen entspricht im wesentlichen dem Vorgang des fallbasierten Schließens. Der Arzt unterscheidet bei der Auswahl zwischen harten und weichen Kriterien. Diese Unterscheidung erfordert besondere Beachtung bei der Modellierung der Auswahlstrategie für den detaillierten Fallvergleich in der zweiten Stufe des Auswahlprozesses.

In den folgenden beiden Abschnitten sollen diese Teilprozesse näher untersucht und Strategien für eine rechnergestützte Umsetzung entwickelt werden.

#### 5.1.1 Das Modell der Vorauswahl

Für die Entwicklung einer effizienten Strategie für die Therapieauswahl ist es notwendig, ein Modell für die Vorauswahl analog zu der vom Arzt verwendeten Vorgehensweise zu entwickeln. Das Ziel der Vorauswahl besteht in der Verminderung des Suchraumes für das zu lösende Problem. Dabei sollen alle potentiell zur Lösung des Problems geeigneten Fälle innerhalb der Fallbasis gefunden werden. Es gibt eine Reihe verschiedener Ansätze, das Problem der Vorauswahl zu lösen. Goos unterscheidet drei verschiedene Klassen von Ansätzen zur Lösung dieses Problems<sup>45</sup>:

- Auswahl auf Basis von Verfeinerungen: Verfahren dieser Art basieren auf hierarchischen Speicherstrukturen. Der Start erfolgt jeweils beim allgemeinsten Konzept, welches schrittweise verfeinert wird. Das Ergebnis besteht aus allen Fällen, welche durch den gefundenen Knoten bzw. das gefundene Blatt repräsentiert werden.
- Auswahlverfahren auf Basis von Partitionierungen: Bei diesen Verfahren werden die einzelnen Fälle nicht zerlegt, sondern als Ganzes betrachtet. Die Fälle werden dynamisch

---

<sup>45</sup>Vgl. [Goos 1996], S. 20

zu Fallklassen zusammengefaßt. Mit Hilfe dieser dynamischen Klassenbildung ist es möglich, von Fall zu Fall die Entscheidung für die Notwendigkeit einer detaillierteren Untersuchung zu treffen.

- Parallele Auswahlverfahren: Diese Verfahren gehen davon aus, daß für jeden bzw. für eine große Anzahl von Fällen der Fallbasis ein eigener Prozessor zur Verfügung steht. Somit hängt der Aufwand für die Vorauswahl nur noch vom Sortiermuster und der Zahl der zu untersuchenden Merkmale und nicht mehr von der Zahl der Fälle in der Fallbasis ab. Hier wird für jeden Fall ein grobes Ähnlichkeitsmaß berechnet und nur bei Überschreitung eines Schwellenwertes wird der Fall in den Problemlösungsprozeß einbezogen.

Wird das im Abschnitt 4.1.1 beschriebene Auswahlverfahren des Arztes zugrunde gelegt, scheint der zweite Ansatz das Verhalten des Arztes relativ gut zu beschreiben. Wichtig für die Realisierung einer solchen Strategie ist die Definition eines geeigneten Klassifikationskriteriums. Bei der Auswahl der Kriterien ist das vom Arzt verwendete Klassifikationskriterium zu berücksichtigen. Das Klassifikationskriterium muß so gewählt werden, daß alle potentiell geeigneten Fälle ausgewählt und alle ungeeigneten Fälle verworfen werden.

Die hierbei auftretenden Fehler werden in zwei Fehlerarten unterteilt. Das Verwerfen eines potentiell relevanten Falles kann Auswirkungen auf die Qualität der Lösung haben, da dieser verworfene Fall das aktuelle Problem unter Umständen lösen kann. Ein solcher Fehler wird auch als  $\alpha$ -Fehler bezeichnet. Die Auswahl eines ungeeigneten Falles hat keine Auswirkungen auf die Qualität der Lösung. Hier verschlechtert sich lediglich die Effizienz des Verfahrens, da die Präferenzrelation  $\succeq_Q$  für mehr als die unbedingt notwendigen ähnlichsten Fälle berechnet werden muß. Diese Art von Fehlern wird als  $\beta$ -Fehler bezeichnet.

Bei der Wahl der Klassifikationskriterien ist weiterhin darauf zu achten, daß der Unterschied zwischen dem Aufwand, der durch die Vorauswahl entsteht und dem Aufwand, der für den detaillierten Fallvergleich der verworfenen Fälle entstehen würde, möglichst groß ist. Je größer dieser Unterschied ist, desto größer die Effizienz der Vorauswahl und desto größer die Beschleunigung des Auswahlprozesses durch die Vorauswahl.

Es ist demnach ein Klassifikationskriterium zu wählen, welches die Menge der Fälle möglichst gut trennt, um die Zahl der Fehlklassifikationen zu minimieren. Weiterhin sollte der Aufwand für die Klassifikation eines Falles möglichst gering sein. Neben dem Klassifikationskriterium ist auch das verwendete Ähnlichkeitsmaß für die Qualität der Vorauswahl wichtig.

In der Literatur werden verschiedene, meist geometrisch motivierte Ähnlichkeitsmaße verwendet. Goos<sup>46</sup> schlägt für die Vorauswahl vor, um jeden relevanten Merkmalswert  $m_i$  ein Ähnlichkeitsintervall  $[m_u, m_o]$  zu legen, so daß für eine Mindestähnlichkeit  $M$  gilt:

$$(1) \quad \forall m_k \in [m_i - m_u, m_i + m_o] : \text{sim}(m_k, m_i) \geq M$$

Das folgende Beispiel soll die Anwendung eines solchen Ähnlichkeitsintervalls für die Mindestähnlichkeit verdeutlichen. Als Klassifikationskriterium für die Vorauswahl soll der Wert der Thrombozyten im Blut angenommen werden<sup>47</sup>. Der Wert  $m_i$  für die Thrombozyten betrage im Beispiel 100.000. Für die Mindestähnlichkeit sei eine Abweichung nach oben und unten von 5.000 akzeptabel, so daß  $m_o = m_u = 5.000$  gilt. Bei einer einfachen Division zur Berechnung der Ähnlichkeit ergibt sich bei einer maximalen Abweichung von 5.000 ein Wert von 0,95 für die Mindestähnlichkeit  $M$ . In diesem Beispiel werden alle Fälle ausgewählt, deren Wert für die Ähnlichkeit mindestens 0,95 ergibt bzw. deren Thrombozytenwerte nicht mehr als um 5.000 von  $m_i$  abweichen.

Die in dieser Arbeit betrachteten Therapieoptimierungsstudien sind diagnoseorientiert. Aus diesem Grund liegt es nahe, die Diagnose als erstes Kriterium für die Beschränkung des Suchraumes zu verwenden. Auch der Arzt verwendet bei der Vorauswahl dieses Kriterium. Mit der Verwendung der Diagnose werden die besten Klassifikationsergebnisse erreicht. Andere Kriterien, welche der Arzt zur weiteren Eingrenzung verwendet, sollen an dieser Stelle vorerst unberücksichtigt bleiben. Die Eingrenzung soll die Modellierung der Strategie zur Vorauswahl zunächst vereinfachen.

Nach der Identifikation des Hauptklassifikationskriteriums muß überprüft werden, ob sich die Vorgehensweise und die Verwendung des Hauptklassifikationskriteriums des Arztes adäquat abbilden läßt und den oben genannten Anforderungen genügt. Das von Goos vorgeschlagene und in der Gleichung (1) formulierte Ähnlichkeitsmaß ist geometrisch motiviert. Für das Therapieauswahlssystem wurde jedoch die Diagnose als Hauptklassifikationskriterium identifiziert. Die Ähnlichkeit von Diagnosen kann aber nicht geometrisch formuliert werden. Hieraus ergibt sich die Notwendigkeit für eine andere Formulierung einer Mindestähnlichkeit für die Vorauswahl.

Im ersten Schritt für die Neuformulierung der Mindestähnlichkeit soll der Aufwand betrachtet werden, der für die Überprüfung der Diagnose notwendig ist. Der alleinige Vergleich der Diagnose erscheint zwar auf dem ersten Blick sehr effizient und verspricht einen großen

---

<sup>46</sup>Vgl. [Goos 1996], S. 66

<sup>47</sup>Der Wert der Thrombozyten im Blut soll nur die Anwendung eines Ähnlichkeitsintervalls für die Bestimmung der Mindestähnlichkeit verdeutlichen. Die Güte des verwendeten Merkmals zur Klassifikation der Fälle im Anwendungsgebiet wird an dieser Stelle nicht berücksichtigt.

Zeitvorteil gegenüber dem detaillierten Vergleich der einzelnen Fälle bzw. Studien. Dieses Vorgehen bringt aber den Nachteil mit sich, daß Diagnosen, welche mit einem anderen Detaillierungsgrad, nach einer anderen Klassifikation oder in einer anderen Sprache<sup>48</sup> gestellt wurden, nicht unbedingt berücksichtigt werden. Diese Vorgehensweise führt vermutlich zu einer hohen Anzahl von  $\alpha$ -Fehlern. Hierdurch kann es zu einer starken Qualitätsverschlechterung bei der Problemlösung kommen. Daher ist es zur Sicherung der Qualität der Lösung notwendig, Wissen über die verschiedenen Klassifikationen, die Diagnosehierarchien und evtl. auch die entsprechenden Diagnosen in anderen Sprachen, insbesondere in englischer Sprache, in den Prozeß der Vorauswahl einzubeziehen. Das notwendige Wissen über die verschiedenen Begriffs- und Klassifikationshierarchien, sowie die Synonyme wird durch eine Synonymverwaltung bereitgestellt. Die Beschreibung der Synonymverwaltung ist in den Abschnitten 5.3 zu finden. Mit Hilfe der Synonymverwaltung ist es möglich, alle ähnlichen Diagnosen zu bestimmen. Der Begriff der „ähnlichen Diagnosen“ wird weiter unten beschrieben. Statt nach einer einzelnen Diagnose wird nun nach einer Menge ähnlicher Diagnosen gesucht. Durch die Einbeziehung dieses Wissens verringert sich jedoch gegenüber dem einfachen Vergleich der Diagnosen der Zeitvorteil, der durch die Vorauswahl gewonnen wird. Dieser Umstand wird allerdings durch die hieraus zu erwartende Verbesserung der Lösungsqualität kompensiert.

Für das zu entwickelnde Anwendungssystem kann das von Goos vorgeschlagene und in Gleichung (1) beschriebene Mindestähnlichkeitsmaß, wie bereits erwähnt, nicht verwendet werden. Die Mindestähnlichkeit für das Therapieauswahlssystem SAM wird daher neu formuliert. Zur Realisierung der oben beschriebenen Strategie muß für eine Diagnose  $d$  eine Menge  $D$  aller ähnlichen Diagnosen gebildet werden. Hierzu ist es notwendig zu definieren, wann eine Diagnose  $d$  ähnlich bezüglich einer Diagnose  $d'$  ist:

**Definition 1** *Eine Diagnose  $d$  heißt ähnlich bezüglich einer Diagnose  $d'$ , wenn  $d'$*

- (1) *der Diagnose  $d$  bezüglich einer anderen Klassifikation entspricht,*
- (2) *der Diagnose  $d$  hinsichtlich der Klassifikationshierarchie der gleichen oder einer anderen Klassifikation übergeordnet ist oder*
- (3) *der Diagnose  $d$  hinsichtlich der Klassifikationshierarchie der gleichen oder einer anderen Klassifikation untergeordnet ist.*

Aus der Definition 1 kann geschlußfolgert werden, daß ein Fall ausgewählt wird, wenn sich die Diagnosen ähnlich sind. Das für diese Vorgehensweise erforderliche Wissen insbesondere

---

<sup>48</sup>Die Berücksichtigung der Sprache wird bei einer Einbeziehung internationaler Studien wichtig.

für die Abbildung der Synonyme und der Klassifikationshierarchien ist in einem eigenen Wissensmodul abzulegen. Für jede auf der Basis der Patientendiagnose vorausgewählte Studie wird anhand der Stellung der Diagnose in der Klassifikationshierarchie ein Ähnlichkeitswert  $\sigma_{0s}$  festgelegt. Dieser Wert soll den Ähnlichkeitswert der Diagnosen repräsentieren. Mit diesem Wert  $\sigma_{0s}$  startet jeweils der detaillierte Fallvergleich. Mit Hilfe von  $\sigma_{0s}$  ist es möglich, die Ähnlichkeit von Diagnosen bzw. auftretende Unsicherheiten abzubilden. Unsicherheiten treten insbesondere dann auf, wenn zu einer Diagnose ein Subtyp der Diagnose als ähnliche Diagnose angenommen wird.

Beispielsweise sind nach Definition 1 die Diagnosen zentrozytisches Lymphom und das Mantelzell-Lymphom ähnlich, da sie zwar unterschiedlichen Klassifikationshierarchien (Kiel- bzw. REAL-Klassifikation) angehören, sich aber einander entsprechen. Ebenso sind nach Definition 1 die noduläre Sklerose und das Hodgkin-Lymphom einander ähnlich, da die noduläre Sklerose einen Subtyp des Hodgkin-Lymphoms bildet. Die Ähnlichkeit der Diagnosen gilt in beide Richtungen.

Ein wesentlicher Vorteil dieser Vorgehensweise besteht darin, daß die Angabe der Klassifikation für die Diagnosen zur Zeit der Entwicklung nicht notwendig ist. Die Klassifikationshierarchie wird erst zur Laufzeit dynamisch generiert und kann den Anforderungen des konkreten Falles entsprechend aufgebaut werden. Mit diesem Verfahren ist es auch möglich, abgestufte Zugehörigkeitsfunktionen zu definieren und die auftretende Unsicherheit abzubilden.

Nähere Ausführungen zur Repräsentation und Verwendung des während der Vorauswahl verwendeten Wissens ist in den Abschnitten 5.2 und 6.2.1 zu finden.

### 5.1.2 Der detaillierte Fallvergleich

Im vorangegangenen Abschnitt wurde eine Vorgehensweise zur Bestimmung eines Ähnlichkeitsmaßes beschrieben, mit deren Hilfe es möglich ist, die Anzahl der näher zu untersuchenden Fälle zu reduzieren. In diesem Abschnitt soll eine Vorgehensweise beschrieben werden, mit deren Hilfe aus dieser Teilmenge der ähnlichste (geeignetste) Fall bzw. die ähnlichsten (geeignetsten) Fälle gefunden werden können. Hierzu ist es notwendig, ein Maß für die Ähnlichkeit der Fälle zu definieren.

In der Literatur werden verschiedene Ähnlichkeitsmaße und Strategien zur Bestimmung der Ähnlichkeit diskutiert. In diesem Zusammenhang werden auch verschiedene Anforderungen an ein Ähnlichkeitsmaß gestellt. Goos stellt in seiner Arbeit<sup>49</sup> zehn Anforderungen an ein Ähnlichkeitsmaß auf, die im Anschluß diskutiert werden sollen:

- Berücksichtigung von domänenspezifischem Hintergrundwissen

---

<sup>49</sup>Vgl. [Goos 1996], S. 86

- Verhinderung der Mehrfachbewertung von Merkmalen beim Ähnlichkeitsvergleich
- Berücksichtigung von domänenspezifischem Ausschlußwissen
- Unterscheidung von generell relevanten und generell irrelevanten Merkmalen für den Ähnlichkeitsvergleich
- Möglichkeit der Berücksichtigung der wechselnden Relevanz von Merkmalen in verschiedenen Kontexten
- Unterschiedliche Bedeutung von normalen und abnormalen Merkmalsausprägungen
- Behandlung von nicht erfaßten und unbekanntem Merkmalen
- Leichte Adaptierbarkeit zum Einsatz maschineller Lernverfahren
- Berücksichtigung von partiellen Ähnlichkeiten zwischen den Merkmalsausprägungen
- Nachvollziehbarkeit der Bewertungen durch den Nutzer

Domänenspezifisches Wissen ist Wissen, welches in einem mehr oder weniger abgeschlossenen Gebiet bzw. in einem bestimmten Kontext benötigt wird, um die Bedeutung von Merkmalsausprägungen zu verstehen. Dieses Wissen wird benötigt, da einige Merkmalsausprägungen erst in Zusammenhang mit anderen Merkmalen an Aussagekraft gewinnen.

Eine ungewollte Mehrfachbewertung von Merkmalen kann im behandelten Kontext nicht auftreten, da die Ein- und Ausschlußkriterien der Studie über die Ähnlichkeit entscheiden und damit jedes Kriterium nur einmal betrachtet wird. Ebenso braucht nicht zwischen generell relevanten und generell nicht relevanten Merkmalen explizit unterschieden werden, da für die Aufnahme eines Patienten in eine Studie alle Ein- und Ausschlußkriterien erfüllt sein müssen. Die zu untersuchende Studie gibt jeweils die zu vergleichenden Merkmale vor.

Die Einbeziehung partieller Ähnlichkeiten zwischen Merkmalsausprägungen soll durch die Verwendung unscharfer Ähnlichkeitsbewertungen berücksichtigt werden. Mit Hilfe der unscharfen Ähnlichkeitsbewertungen der Merkmale ist es auch möglich, eine Berücksichtigung von physiologischen<sup>50</sup>, pathophysiologischen<sup>51</sup> und pathologischen<sup>52</sup> Merkmalsausprägungen vorzunehmen.

Das mögliche Auftreten von nicht erfaßten und unbekanntem Werten soll beim detaillierten Fallvergleich mit berücksichtigt und toleriert werden. Allerdings darf die Toleranz des

---

<sup>50</sup>Physiologisch bezeichnet normale Werte, Werte die innerhalb einer Norm liegen. [Pschyrembel 1997]

<sup>51</sup>Pathophysiologisch bezeichnet Werte in einem Grenzbereich, der leicht ausserhalb einer Norm liegt. [Pschyrembel 1997]

<sup>52</sup>Pathologisch bezeichnet Werte, die klar ausserhalb eines Normbereiches liegen und krankhaft sind. [Pschyrembel 1997]

Systems gegenüber nicht erfaßten bzw. unbekanntem Werten unbeschränkt sein. Hier scheint die Unterscheidung zwischen unbedingt notwendigen und weniger notwendigen Patientendaten sinnvoll. Bei der Auswertung wird der Anwender auf das Fehlen dieser Werte hingewiesen. Eine Erklärung der Auswirkungen des Fehlens dieser Werte wird durch die Erklärungskomponente vorgenommen. Ebenso soll dem Anwender die Möglichkeit gegeben werden, mit Hilfe der Erklärungskomponente die Arbeitsweise des Fallvergleiches und die Wirkung des Ähnlichkeitsmaßes zu verstehen.

Eine leichte automatische Adaptierbarkeit des Ähnlichkeitsmaßes ist nicht notwendig, da der Einsatz maschineller Lernverfahren für das zu realisierende Therapieauswahlmodul nicht vorgesehen ist. Das Ähnlichkeitsmaß ist allerdings so zu konzipieren, daß durch die Aufnahme neuer Studien keine Änderung des Ähnlichkeitsmaßes notwendig wird.

Ausgehend von den beschriebenen Anforderungen soll nun ein entsprechendes Ähnlichkeitsmaß zur Bestimmung der Ähnlichkeit zweier Fälle entwickelt werden. Wie im Abschnitt 5.1.1 wird an dieser Stelle erneut die Vorgehensweise des Arztes herangezogen. Als erster Schritt muß eine Unterscheidung der Kriterien in

- harte Kriterien
- weiche Kriterien

vorgenommen werden.

Harte Kriterien sind Kriterien, bei denen eine Unterscheidung zwischen „erfüllt“ bzw. „nicht erfüllt“ getroffen werden muß. Solche Kriterien sind beispielsweise HIV-Positivität, vorhandene Schwangerschaft und der Leukozytenwert im Blut des Patienten. Hingegen kann bei weichen Kriterien, wie z. B. dem Gesamteiweiß im Blut, der Zahl der Thrombozyten im Blut und dem Alter des Patienten, eine Abweichung in gewissen Grenzen toleriert werden.

Der Abgleich der harten Kriterien ist relativ einfach, da diese nur dahingehend überprüft werden, ob die Patientendaten mit dem Ein- und Ausschlußkriterien der Studie übereinstimmen bzw. ob die Patientendaten im geforderten Intervall liegen. Bei Verstoß gegen ein hartes Kriterium kann der Vergleich des speziellen Falles abgeschlossen, der Fall als ungeeignet markiert und der Fallvergleich mit dem nächsten vorausgewählten Fall fortgesetzt werden. Aufgrund der Tatsache, daß die harten Kriterien einfacher zu überprüfen sind und bei fehlender Übereinstimmung sofort zum Ausschluß des Falles führen, empfiehlt es sich, zunächst die harten Kriterien zu untersuchen. Das Kriterium, welches zum Ausschluß geführt hat, ist jeweils entsprechend zu kennzeichnen.

Im Gegensatz zu den harten Kriterien ist die Untersuchung der weichen Kriterien komplexer und erfordert einige Vorüberlegungen. Hierzu ist wiederum die Vorgehensweise des Arztes heranzuziehen. Eine geringe Abweichung des speziellen Patientendatums von den Ein-

und Ausschlußkriterien scheint der Arzt ohne Abstriche zu tolerieren. Es existiert also eine Abweichung  $\varepsilon$  für ein weiches Ein- bzw. Ausschlußkriterium  $k_i$  mit deren Hilfe sich ein Intervall  $(k_i \dots k_i + \varepsilon)$  bzw. für eine negative Abweichung  $(k_i \dots k_i - \varepsilon)$  formulieren läßt. Bei einer etwas größeren Abweichung  $\delta$  mit  $\delta > \varepsilon$ , das Patientendatum liegt also etwas außerhalb des obigen Intervalls, bewertet der Arzt diesen Wert mit Hilfe eines zweiten Intervalls  $(k_i + \varepsilon \dots k_i + \delta)$  bzw. analog für die negative Abweichung  $(k_i - \varepsilon \dots k_i - \delta)$ . Liegt das Patientendatum in diesem zweiten Intervall beurteilt der Arzt im Zusammenhang mit verschiedenen anderen Werten, ob es möglich ist, den Patienten trotz dieses leichten Verstoßes gegen die Ein- und Ausschlußkriterien in dieser Studie zu behandeln.

Die Anwendung der soeben beschriebenen Intervalle sollen die folgenden Beispiele verdeutlichen. Innerhalb einer Studie gefordert, daß die Zahl der Thrombozyten im Blut größer 100.000 ist. Der tatsächlich Thrombozytenwert eines Patienten liegt jedoch bei 99.500. Der Wert für  $\varepsilon$  könnte in diesem Fall bei 1.000 liegen. In diesem Fall würde der Patient trotz der kleinen Abweichung vom geforderten Wert das Kriterium noch erfüllen.

Bei einem anderen Patienten liegt der Wert der Thrombozyten im Blut bei 97.500 und damit ist die Abweichung des Wertes größer als  $\varepsilon$ . An dieser Stelle muß geprüft werden, ob das Patientendatum auch außerhalb des zweiten Intervalls liegt. Der Wert für  $\delta$  soll in diesem Beispiel bei 3.000 liegen. Da die Abweichung des Patientendatums mit 2.500 kleiner als  $\delta$  ist, könnte der Patient dennoch in die Studie aufgenommen werden, wenn andere Patientendaten eine solche Entscheidung zulassen. Die Entscheidung über die Aufnahme muß jedoch der Arzt treffen.

Für die weichen Kriterien ist genau zu evaluieren, wie groß diese beiden Abweichungen sind und wie die lokale Ähnlichkeitsfunktion  $\mu_{(p_i, s_i)}$  im speziellen Fall aussieht, wobei  $p_i$  das  $i$ -te Merkmal des Patienten und  $s_i$  das  $i$ -te Merkmal der Studie ist. Insbesondere ist der Verlauf der Funktion innerhalb des zweiten Intervalls zu ermitteln. Desweiteren ist zu hinterfragen, wie ein leichter Verstoß gegen alle weichen Kriterien behandelt wird.

Die Vorgehensweise bei der Ermittlung der lokalen Ähnlichkeit wurde soeben beschrieben. Neben der lokalen Ähnlichkeit ist ein Modell zur Bestimmung der globalen Ähnlichkeit notwendig, um die Fälle richtig einordnen zu können. Im folgenden soll das Modell der globalen Ähnlichkeitsbestimmung beschrieben werden. In der Literatur werden eine Reihe von Modellen zur globalen Ähnlichkeitsbestimmung definiert. Das globale Ähnlichkeitsmaß  $sim$  wird in Anlehnung an Weisbrod<sup>53</sup> definiert, so das gilt:

$$(2) \quad sim_{(p,s)} : \mathcal{P} \times \mathcal{S} \rightarrow [0, 1]$$

Wobei  $p \in \mathcal{P}$  die Merkmale eines speziellen Falles, d. h. eines speziellen Patienten  $p$ , aus der

---

<sup>53</sup>Vgl. [Weisbrod 1996], S. 22

Menge aller Fälle (Patienten)  $\mathcal{P}$  und  $s \in \mathcal{S}$  für die Merkmale einer speziellen Studie  $s$  aus der Menge der Studien  $\mathcal{S}$  der Fallbasis stehen. Desweiteren ist ein Schwellenwert  $\omega_g \in (0, 1)$  zu definieren, der den Ähnlichkeitsgrad angibt, bis zu dem, die Fälle einander hinreichend ähnlich sind bzw. eine Studie für den Patienten geeignet ist.

Neben diesem Schwellenwert  $\omega_g$  ist ein zweiter Schwellenwert  $\omega_u \in (0, 1)$  mit  $\omega_g > \omega_u$  festzulegen. Dieser zweite Schwellenwert  $\omega_u$  ist das Abbruchkriterium für den Fallvergleich. Wird der Schwellenwert  $\omega_u$  unterschritten, hat dies den Abbruch des detaillierten Fallvergleichs zur Folge und der Fall bzw. die Studie wird als zu wenig ähnlich bzw. nicht geeignet eingestuft.

Bei Beginn eines detaillierten Fallvergleiches ist davon auszugehen, daß alle vorausgewählten Fälle ähnlich sind, d. h. der Patient erfüllt die entsprechenden Ein- und Ausschlußkriterien. Diese Annahme gilt es für alle vorausgewählten Fälle zu prüfen. Der detaillierte Fallvergleich startet mit dem durch die Vorauswahl festgelegten Ähnlichkeitswert  $\sigma_{0s}$ . Dieser Ähnlichkeitswert kann im vorliegenden Fall auch als Grad der Eignung verstanden werden. Für alle Werte  $\sigma_{is}$  soll hierbei gelten:

$$(3) \quad \sigma_{is} \in [0, 1] \subset \mathcal{R} \quad \text{mit } i = 0, 1, \dots, n$$

Zu Beginn sollen die harten Kriterien überprüft werden. Für den Vergleich der einzelnen Kriterien wird hierzu eine Funktion  $f$  definiert mit:

$$(4) \quad f_{(p_i, s_i)} = \begin{cases} 0 & \text{wenn } p_i \neq s_i \\ 1 & \text{sonst} \end{cases}$$

Mit Hilfe dieser Funktion  $f$  kann der Wert für die globale Ähnlichkeit  $\sigma_{is}$  schrittweise wie folgt ermittelt werden:

$$(5) \quad \sigma_{is} = \sigma_{i-1s} \times f_{(p_i, s_i)} = \begin{cases} 0 & \text{wenn } f_{(p_i, s_i)} = 0 \\ \sigma_{i-1s} & \text{sonst} \end{cases}$$

Wobei  $p_i$  und  $s_i$  jeweils das Merkmal des Patienten bzw. der Studie ist, welches im  $i$ -ten Schritt bzw. beim  $i$ -ten Merkmalsvergleich betrachtet wird.

Die Gleichungen (3), (4) und (5) verdeutlichen, daß ein erfülltes Kriterium den bisher erreichten Ähnlichkeitswert nicht verändert. Eine Studie hat beispielsweise einen Ähnlichkeitswert  $\sigma$  von 0,95. Die Falldaten aus der Fallbasis verlangen, daß ein Patient nicht HIV positiv ist<sup>54</sup>. Das konkrete Patientendatum ist: Der Patient ist nicht HIV positiv<sup>55</sup>. Aus dem Vergleich der Merkmale ergibt sich eine Übereinstimmung, es gilt also  $p_1 = s_1$  und damit

<sup>54</sup>Dieses Merkmal des Falles (der Studie) wird mit der Variable  $s_1$  bezeichnet.

<sup>55</sup>Das konkrete Patientenmerkmal wird mit  $p_1$  bezeichnet.

nach Gleichung (4)  $f_{p_1, s_1} = 1$ . Zur Berechnung der globalen Ähnlichkeit aller bisher verglichenen Merkmale wird die Gleichung (5) herangezogen und es ergibt sich aus  $0,95 \times 1$  ein neuer Wert für  $\sigma$  von 0,95.

Diese Vorgehensweise ist korrekt, da beim Start des detaillierten Fallvergleiches davon ausgegangen werden kann, daß die Fälle zueinander ähnlich sind. Somit ist ein erfülltes Kriterium eine Unterstützung der gestellten These. Bei einem Verstoß hingegen erfolgt der Ausschluß des Falls vom weiteren Fallvergleich, da der Verstoß gegen ein hartes Kriterium einer Nichteignung des Patienten für diese Studie entspricht. Dies wird auch deutlich, wenn der oben definierte Schwellenwert  $\omega_u$  betrachtet wird. Der aktuelle Wert 0, der aus einem Verstoß gegen die Ein- und Ausschlußkriterien resultieren würde, unterschreitet laut Definition immer den Wert von  $\omega_u$ , was zu einem Ausschluß vom weiteren Fallvergleich führen und einer Einordnung als nicht geeignete Studie entsprechen würde.

Nach dem Vergleich der harten Kriterien soll der Vergleich der weichen Kriterien folgen. Es ist zu definieren, wie die partielle Ähnlichkeit des einzelnen Kriterium die globale Ähnlichkeit beeinflusst. Hierzu sind einige Vorüberlegungen notwendig, da der Vergleich nicht nur wahr oder falsch als Ergebnis liefern soll, sondern auch abgestufte Wahrheitswerte.

An dieser Stelle werden einige Anleihen aus dem Bereich der Fuzzy-Set-Theorie bzw. der Aussagenlogik gemacht. In diesen Gebieten werden unter anderem Aussagen über  $t$ -Normen und  $s$ -Normen getroffen. Bei den  $s$ -Normen handelt es sich um Funktionen zur Berechnung des Zugehörigkeitsgrades von alternativen Aussagen. Die  $t$ -Normen sind hingegen Funktionen zur Berechnung der Zugehörigkeit von konjunktiven Aussagen<sup>56</sup>. Zur Ermittlung der globalen Ähnlichkeit können diese Funktionen herangezogen werden.

Da es sich bei dem Vergleich um eine Konjunktion handelt, der Patient soll alle Ein- und Ausschlußkriterien erfüllen, sollen im folgenden nur die  $t$ -Normen näher betrachtet werden. Weisbrod<sup>57</sup> diskutiert in seiner Arbeit mehrere dieser  $t$ -Normen. Diese  $t$ -Normen haben jedoch den Nachteil, daß sie entweder keine Möglichkeit zur Gewichtung bieten oder deren Wirkung nur schwer nachvollziehbar ist. Weisbrod stellt fest, daß im allgemeinen Fall nur der Einsatz der Maximum- bzw. Minimum-Norm gerechtfertigt scheint<sup>58</sup>. Jedes andere Normpaar  $(s, t) \neq (max, min)$  führt zu einer „unnötig großen“ Vereinigungsmenge und einer „unnötig kleinen“ Schnittmenge. Der Einsatz anderer Normpaare ist jedoch durch den Einsatz von Zusatzwissen über das zu lösende Problem zu rechtfertigen. Dieses Zusatzwissen muß global sein, da es über die allgemeinen Operatoren unabhängig von den konkreten Parametern bestimmt. Desweiteren sollte es nicht nur implizit in den Aufbau der Normen fließen, sondern auch ex-

---

<sup>56</sup>Vgl. [Weisbrod 1996], S. 13 f. und [Biewer 1997], S. 83 ff.

<sup>57</sup>Vgl. [Weisbrod 1996], S. 16

<sup>58</sup>Vgl. [Weisbrod 1996], S. 19 f.

plizit dokumentiert werden. Dieses Wissen kann dann auch durch die Erklärungskomponente zur Generierung von Erklärungen genutzt werden.

Ein Nachteil der Maximum- und Minimum-Normen liegt darin, daß der Einfluß der graduellen Zugehörigkeitsgrade nicht berücksichtigt wird. Bei der Konjunktion schwächen sich zwei graduelle Zugehörigkeitsgrade gegenseitig ab, so daß z. B. ein Kunde, der ein Paar preiswerte und schöne Schuhe sucht, sich nicht für ein Paar Schuhe entscheiden wird, welches er für einigermaßen preiswert und einigermaßen schön hält. Gerade eine solche Abhängigkeit soll jedoch im vorliegenden Fall modelliert werden.

Aufgrund der Tatsache, daß mit Hilfe der Maximum- und Minimum-Normen weder der Einfluß der graduellen Unterschiede untereinander abgebildet noch eine Gewichtung der Faktoren vorgenommen werden kann, ist die Verwendung einer anderen  $t$ -Norm notwendig. Hier empfiehlt sich eine  $t$ -Norm der Form  $a * b$ . Wird diese Form mit Hilfe der Gewichte  $n_1$  bzw.  $n_2$  erweitert, kann auch eine Gewichtung der Eigenschaften vorgenommen werden. Die Gewichtung kann mit Hilfe von Exponenten erfolgen, so daß  $n_i < 1$  eine Abschwächung und  $n_i > 1$  eine Verstärkung auf den Wert der Ähnlichkeit bedeutet.<sup>59</sup> Es ergibt sich dann eine  $t$ -Norm der Form  $a^{n_1} * b^{n_2}$ . Für das globale Ähnlichkeitsmaß im  $i$ -ten Schritt kann dann geschrieben werden:

$$(6) \quad \sigma_{is} = \sigma_{i-1s} * \mu_i^n(p_i, s_i)$$

Wobei für  $n_1, n_2 \in (0, \infty) \subset \mathcal{R}$  gilt. Die beiden Variablen  $n_1$  und  $n_2$  geben die Gewichtung bzw. den Einfluß des  $i$ -ten Merkmals auf die Gesamtfunktion an. Die genaue Gewichtung der einzelnen Merkmale ist noch zu evaluieren. Bei der Modellierung der partiellen Ähnlichkeitsfunktion ist zu beachten, daß stets gelten soll:

$$(7) \quad \mu_i^n(p_i, s_i) \rightarrow [0, 1]$$

Alle Ein- und Ausschlußkriterien der Studien werden mit den Patientendaten verglichen und mit Hilfe der Funktionen (4), (5), (6) und (7) wird ein entsprechender Wert für die Ähnlichkeit bzw. die Eignung der Studie für den Patienten ermittelt. Zur Verdeutlichung der Vorgehensweise soll nachfolgendes Beispiel dienen: Der geforderte Wert der Thrombozyten im Blut in einem Fall aus der Fallbasis betrage 100.000. Der Patient hat in diesem Beispiel einen Thrombozytenwert von 98.500. Die lokale Ähnlichkeitsfunktion für das Merkmal liefert in diesem Fall 0,95.<sup>60</sup> Für das Gewicht  $n$  wird ein Wert von 2 angenommen, so daß sich nach Gleichung (7) ein gewichteter lokaler Ähnlichkeitswert von 0,9025 ergibt. Die betrachtete

<sup>59</sup>Vgl. [Biewer 1997], S. 310

<sup>60</sup>Auf die konkrete lokale Ähnlichkeitsfunktion wird an dieser Stelle nicht näher eingegangen. Nähere Ausführungen zur Modellierung der lokalen Ähnlichkeitsfunktionen sind im Abschnitt 6.2.3 zu finden.

Studie wurde bisher mit einem Ähnlichkeitswert von 0,85 bewertet. Nach der Gleichung (6) ergibt sich durch den berechneten Wert von 0,9025 ein neuer globaler Ähnlichkeitswert  $\sigma$  von rund 0,767.

Der so errechnete Wert für den Grad der Eignung wird mit den Schwellenwerten  $\omega_g$  und  $\omega_u$  verglichen und die Studie wird entsprechend kategorisiert. Der Vergleich mit  $\omega_u$  sollte allerdings nach jeder Neubewertung des Ähnlichkeitsmaßes erfolgen, um den Fallvergleich bei Unterschreiten von  $\omega_g$  abbrechen zu können.

## 5.2 Synonymverwaltung

Mit Hilfe der Synonymverwaltung soll das im Abschnitt 4.1.2 beschriebene Wissen über die verschiedenen Klassifikationen und Diagnosehierarchien, sowie deren sprachliche Abweichungen abgebildet werden. Dieses Wissen hat im wesentlichen die Aufgabe, den Ausschluß potentiell relevanter Fälle während der Phase der Vorauswahl zu vermeiden. Die Einbeziehung dieses Wissens hat das Ziel, die Qualität der Lösung zu verbessern. Ein großer Nachteil der Einbeziehung der Synonymverwaltung in die Phase der Vorauswahl besteht in der Erhöhung des Aufwandes für die Vorauswahl. Dies erscheint jedoch in Hinblick auf die zu erwartende Verbesserung der Qualität der Lösung akzeptabel.

Für die Konzeption der Synonymverwaltung müssen zunächst erst einmal ihre Aufgaben identifiziert werden. Als Aufgaben sind hier zu nennen:

- Abbildung der verschiedenen Klassifikationen
- Abbildung des Aufbaus der Klassifikation
- Abbildung der Beziehungen zwischen den Klassifikationen
- Abbildung der Hierarchieebene zur Navigation und Steuerung

Die Abbildung der verschiedenen Klassifikationen, wie z.B. der Kiel- und der REAL-Klassifikation der Non-Hodgkin-Lymphome, ist Grundgedanke der Synonymverwaltung. Daher bedarf dieses Ziel wohl keiner näheren Erläuterung.

Für die Navigation innerhalb einer Klassifikation ist es notwendig, den Aufbau der Klassifikation abzubilden. Es ist also notwendig, daß ein Diagnosetyp innerhalb einer Klassifikation seine übergeordneten Typen und/oder seine Subtypen kennt. Für die Abbildung dieser Beziehungen wird eine Relation  $isA(c_i, c)$  definiert, welche aussagt, daß die Klasse  $c_i$  eine Unterklasse der Klasse  $c$  ist.

Mit der Abbildung der Beziehungen zwischen den Klassifikationen wird die Synonymverwaltung erst sinnvoll. Für eine effiziente und leicht wartbare Abbildung muß eine Referenz-

klassifikation  $C_{ref}$  definiert bzw. festgelegt werden. Weiterhin ist die Vereinbarung einer Relation der Form  $relatedTo(c_i, c_{ref})$  notwendig. Sie legt fest, daß die Klasse  $c_i$  der Klasse  $c_{ref}$  der Referenzklassifikation entspricht.

Die Abbildung der Hierarchieebene für eine Klasse  $c$  innerhalb ihrer Klassifikation soll neben der Navigation auch den Suchvorgang steuern und effizienter gestalten. Bei der Suche nach Synonymen sollen neben den entsprechenden Begriffsstrukturen in den anderen Klassifikationen auch über- und untergeordnete Begriffsstrukturen mitberücksichtigt werden. Die Berücksichtigung der über- und untergeordneten Begriffsstrukturen darf nicht über die gesamte Klassifikation ausgedehnt werden, um einer unnötigen Vergrößerung des Suchvektors für die Vorauswahl entgegenzuwirken. Weiterhin kann mit Hilfe der Hierarchieebene dynamisch entschieden werden, ob weiter verfeinert bzw. generalisiert werden soll.

Das Ähnlichkeitsmaß ist so zu gestalten, daß die Berücksichtigung der unterschiedlichen Ähnlichkeit von zwei sich entsprechenden Diagnosen verschiedener Klassifikationen der gleichen Hierarchiestufe und von zwei Diagnosen, die auf verschiedenen Hierarchiestufen stehen, möglich ist.

### 5.3 Wissensrepräsentation und Aufbau der Fallbasis

In den vorangegangenen Abschnitten wurde die Vorgehensweise zur Bestimmung einer geeigneten Studie, sowie die Struktur der Synonymverwaltung beschrieben. In diesem Abschnitt wird auf die Struktur und die verwendeten Modelle zur Repräsentation des Wissens der Fallbasis eingegangen.

Die Wissensrepräsentation legt in einem Wissensbasierten System die Formalismen fest, mit denen das Wissen, welches dem System zur Problemlösung zur Verfügung steht, formuliert werden kann. Für die Wissensrepräsentation wird in der Literatur zwischen drei verschiedenen Arten von Wissen unterschieden<sup>61</sup>:

- Domänenwissen ist Wissen über die spezielle Domäne, in welcher das wissensbasierte System eingesetzt wird. In der vorliegenden Arbeit wird das Domänenwissen in Form der Fallbasis und der Begriffshierarchien in der Synonymverwaltung abgelegt.
- Inferenzwissen beinhaltet Regeln und Ableitungsschritte, mit deren Hilfe sich neues Wissen aus den bereits vorhandenen Daten ableiten läßt. Dieses Wissen wird hier im wesentlichen durch die verschiedenen Ähnlichkeitsmaße repräsentiert.
- Kontrollwissen besteht im wesentlichen aus Wissen über den allgemeinen Ablauf und

---

<sup>61</sup>Vgl. [Hunang, Jennings & Fox 1994], S. 221

Ausschlußwissen bzw. Wissen über Wertebereiche, sofern es domänenübergreifend verwendet werden kann.

Diese Dreiteilung verdeutlicht, daß eine strikte Trennung von domänenspezifischen und domänenunabhängigen Wissen angestrebt wird. Die Trennung von domänenspezifischen und domänenunabhängigen Wissen soll die Wiederverwendung von Wissen in anderen System ermöglichen, den Aufwand für die Wartung des Wissens verringern sowie eine Erweiterung des Systems erleichtern. Die unabhängige Repräsentation des Kontrollwissens soll die Anwendung des gleichen Wissen in verschiedenen Teilen des wissensbasierten Anwendungssystems ermöglichen und damit die Redundanz im System verringern. Weiterhin besteht die Möglichkeit, daß das Domänenwissen unabhängig von Kontroll- und Inferenzwissen akquiriert und verändert werden kann<sup>62</sup>. Die strikte Trennung von domänenspezifischen und domänenunabhängigen Wissen hat sich in den letzten Jahren bei der Entwicklung von Wissenbasierten systemen durchgesetzt und wird als State of the Art betrachtet.

Im Mittelpunkt des Modells zur Wissensrepräsentation steht das Domänenwissen, dessen wesentlicher Bestandteil in einem fallvergleichenden System die Fallbasis ist. Die Fallbasis besteht aus einer Ansammlung von Fällen und ist als Erweiterung einer Datenbank zu verstehen. Dies ist leicht einzusehen, da mit einer Anfrage an die Fallbasis nicht nur alle gleichen Fälle, wie bei einer Datenbank, sondern auch ähnliche Fälle gefunden werden sollen. Hierbei muß sichergestellt werden, daß ein gespeicherter Fall auch wiedergefunden werden kann. Daher sollte das System in der Lage sein, identisch wiederkehrende Fälle auch als solche wiederzuerkennen und die gleiche Lösung liefern, sofern die Fallbasis nicht verändert wurde<sup>63</sup>.

Ein Fall der Fallbasis wird hierbei meist als Beschreibung eines Problems zusammen mit der Beschreibung der Lösung dieses Problems verstanden. Zusätzlich zu Problembeschreibung und -lösung können mit dem speziellen Fall noch Zusatzinformationen gespeichert werden<sup>64</sup>. Für das Studienauswahlmodul sind die Ein- und Ausschlußkriterien als Problembeschreibung, die Beschreibung des Therapieablaufs als Problemlösung und alle anderen Informationen als Zusatzinformationen zu betrachten. Fälle können in Fallbasen mit Hilfe von verschiedenen Repräsentationsformalismen abgebildet werden<sup>65</sup>:

- Attribut-Wert-Repräsentation
- Repräsentation durch strukturierte Mengen von prädikatenlogischen Formeln
- Objektorientierte Fallrepräsentation

---

<sup>62</sup>Vgl. [Huang 1994], S. 223

<sup>63</sup>Vgl. [Weß 1996], S. 125

<sup>64</sup>Vgl. [Kolodner 1993], S. 147

<sup>65</sup>Vgl. [Bergmann 1996], S. 38

Die Attribut-Wert-Repräsentation ist die einfachste Form zur Abbildung von Falldaten. Sie findet vor allem bei einfach strukturierten, analytischen Problembereichen Anwendung. Ein Teil der Attribute wird hierbei zur Repräsentation der Merkmale der Problemsituation benutzt, während der andere Teil bei der Repräsentation der Lösung verwendet wird.

In fallbasierten Systemen für synthetische Aufgaben werden die Falldaten oft mit Hilfe von strukturierten Mengen prädikatenlogischer Formeln dargestellt. Diese Form der Repräsentation der Falldaten findet insbesondere bei der Plangenerierung, wie z. B. der Aktionsplanung, Verwendung. In komplexeren Domänen wird mittlerweile oft die objektorientierte Repräsentation der Falldaten verwendet.

In der vorliegenden Domäne besteht die Fallbasis aus einer Ansammlung von Studien. Damit enthält die Fallbasis relativ wenige, aber sehr komplexe Fälle. Der komplexe Aufbau der Fälle legt die Verwendung der objektorientierten Falldatenrepräsentation nahe. Ein großer Vorteil dieser Repräsentationsform ist, daß die Struktur der Domäne ohne große Probleme bei der Repräsentation der Falldaten erhalten und abgebildet werden kann.

Bei der Modellbildung für die Wissensrepräsentation eines fallbasierten Systems sind hinsichtlich der Struktur der Fallbasis folgende Ebenen zu unterscheiden<sup>66</sup>:

- Strukturelle Beschreibung der Begriffswelt
- Funktionale Beschreibung der Abhängigkeiten
- Strukturelle Beschreibung der Falldaten
- Beschreibung der Ähnlichkeiten von Fallbeispielen

Durch die strukturelle Beschreibung der Begriffe innerhalb der Domäne wird die Grundstruktur des Modells der Domäne festgelegt. Die funktionale Beschreibung der Abhängigkeiten formuliert die Relationen zwischen den einzelnen Domänenbegriffen. In den so entstandenen Rahmen werden die speziellen Domänenobjekte eingeordnet. Die Struktur des Domänenmodells ist so zu formulieren, daß die funktionalen Abhängigkeiten abgebildet werden können. Weiterhin läßt sich aus der strukturellen Beschreibung der Domänenbegriffe auch die Struktur ableiten, in der die zu bearbeitenden Fallbeispiele abgebildet werden. Die Summe dieser Beschreibungen wird auch als Domänenmodell bezeichnet, da diese den Rahmen für die konkrete Entwicklung der Fallbasis und des fallvergleichenden Systems vorgeben<sup>67</sup>:

Grundlage für die Beschreibung des Domänenmodells bilden sogenannte Primitive. Als Primitive sollen alle Grundtypen bzw. alle Grundbausteine bezeichnet werden, mit deren

---

<sup>66</sup>Vgl. [Weß 1996], S. 111

<sup>67</sup>Vgl. [Weß 1996], S. 111

Hilfe es möglich ist, komplexere Strukturen aufzubauen. Solche primitiven Typen sind beispielsweise boolesche Variablen, ganze und gebrochene Zahlen und alphanumerische Zeichen. Diese Primitive bilden die Grundlage, aus der alle anderen, komplexeren Typen, Klassen, Regeln und Constraints aufgebaut sind.

Klassen sind als Beschreibung von Entitätstypen der betrachteten Domäne zu verstehen. Die konkreten Entitäten werden als Objekte bzw. Instanzen von Klassen bezeichnet. Objekte bzw. Instanzen von Klassen beschreiben demzufolge eine konkrete Entität mit den zugehörigen Merkmalsausprägungen entsprechend der im Domänenmodell festgelegten Struktur. Merkmale von Objekten werden als Attribute bezeichnet und können wiederum mit Hilfe von Objekten oder mit Hilfe von Primitiven beschrieben werden. Relationen beschreiben die Beziehungen zwischen Objekten, Regeln und Constraints die Abhängigkeiten zwischen den einzelnen Objekten. Mit Hilfe der Regeln und Constraints ist es auch möglich, Ausschlußwissen zu formulieren und den Klassen als Methoden zuzuordnen.

Bei der Konzeption und Entwicklung der Fallbasis ist darauf zu achten, daß die Fallbasis, welche durch die Studiendatenbank repräsentiert wird, strikt von den Patientendaten zu trennen ist. Die Trennung ist notwendig, um die Wartbarkeit der Patientendaten und insbesondere der Studiendaten zu gewährleisten. Diese Trennung ist auch in Hinblick auf die Erweiterbarkeit und die Wartbarkeit der Fallbasis notwendig. Weiterhin muß die Konsistenz der Daten der Synonymverwaltung mit den Daten der Studiendatenbank gewährleistet sein. Hieraus folgt, daß die Studiendatenbank keine Diagnosen bzw. Tumorbezeichnungen enthalten soll, die die Synonymverwaltung nicht kennt.

## 5.4 Strategisches Auswahlwissen

Eine zentrale Rolle innerhalb eines fallbasierten Ansatzes spielt neben der Repräsentation der Fälle die Repräsentation der Ähnlichkeit. Das Wissen eines fallvergleichenden Systems verteilt sich demnach im wesentlichen auf die Fallbasis und das verwendete Auswahlwissen, welches seinen Ausdruck im globalen Ähnlichkeitsmaß  $sim$  und im lokalen Ähnlichkeitsmaß  $\mu$  findet. Das globale und das lokale Ähnlichkeitsmaß ist so zu wählen, daß die verwendete Strategie zur Bestimmung der Ähnlichkeit unterstützt wird. Die Effizienz eines fallvergleichenden Systems hängt somit im starken Maße vom verwendeten Ähnlichkeitsbegriff ab<sup>68</sup>.

### 5.4.1 Modell der Ähnlichkeitsbestimmung

Im Abschnitt 4.2.2 wurde bereits die Vorgehensweise bei der Bestimmung der Ähnlichkeit beschrieben. Diese Beschreibung verdeutlicht, daß das verwendete Modell zur Bestimmung

---

<sup>68</sup>Vgl. [Weß 1996], S. 124 f.

der Ähnlichkeit berechnungsorientiert ist und damit die Auswahl der Fälle erst zur Laufzeit erfolgt. Eine Anfrage  $Q$  wird entsprechend der bereits beschriebenen Strategie mit den in der Fallbasis  $CB$  gespeicherten Fällen verglichen. Entsprechend der Anfrage  $Q$  wird eine spezifische Präferenzrelation über der Fallbasis  $CB$  gebildet:

$$(8) \quad y \succeq_Q z \quad \text{mit } y, z \in CB$$

Mit Hilfe der Präferenzrelation  $\succeq_Q$  liefert das System entsprechend der in einer Anfrage enthaltenen Patientendaten eine nach der Eignung der Studie für den Patienten geordnete Liste von Studien. Das Sortierungskriterium innerhalb der Liste wird durch die Präferenzrelation  $\succeq_Q$  festgelegt.

In Fallvergleichenden Systemen wird jeweils zwischen der lokalen und der globalen Ähnlichkeit unterschieden. Die lokale Ähnlichkeit wird auf der Ebene der Merkmale festgestellt. Sie beruht also auf der partiellen Ähnlichkeit spezieller Merkmalsausprägungen eines Merkmals zueinander. Zur Bestimmung der lokalen Ähnlichkeit wird eine Präferenzrelation auf Ebene der Merkmale in der folgenden Form gebildet:

$$(9) \quad u \succeq_v^i w \quad \text{mit } u, v, w \in W_i$$

Die Ausprägung von  $v_{ij}$  eines Merkmals  $M_i$  legt hierbei fest, in welchem Kontext welche Merkmalsausprägung vorzuziehen ist.

Beispielsweise soll für einen Patienten mit einem Alter von 62 Jahren eine geeignete Therapie gefunden werden. Eine Studie  $S_1$  verlangt, daß die Patienten nicht älter als 60 Jahre sind. Eine zweite Studie  $S_2$  hingegen verlangt, daß die Patienten zwischen 60 und 75 Jahre sind. Mit Hilfe der in (9) beschriebene Präferenzrelation lassen sich diese Studien hinsichtlich der Ähnlichkeit zum Patientendatum ordnen. In diesem Beispiel ist die Studie  $S_2$  der Studie  $S_1$  hinsichtlich des Alters vorzuziehen, wenn das Alter als alleiniges Vergleichsmerkmal zugrunde gelegt wird.

Ein Objekt bündelt eine Menge von Merkmalen. Daher kann der Ähnlichkeitsvergleich von Objekten als eine Erweiterung des Vergleichs von Merkmalsausprägungen verstanden werden und zur Ermittlung der globalen Ähnlichkeit herangezogen werden. Zum Vergleich von Objektinstanzen soll die globale Präferenzrelation  $y \succeq_Q z$  bzw. das im Abschnitt 5.2.2 definierte Ähnlichkeitsmaß  $sim$  dienen.

#### 5.4.2 Ähnlichkeit in der Vorauswahl

Das Wissen für die Bestimmung der Ähnlichkeit während der Vorauswahl wird im wesentlichen durch die Synonyme, die Klassifikationshierarchien und der Strategie zur Vorauswahl repräsentiert. Die Vorauswahl ist in der vorliegenden Arbeit als ein Spezialfall des lokalen

Ähnlichkeitsvergleiches zu sehen. Diese Sonderstellung ergibt sich aus der Tatsache, daß weder ein harter Vergleich der Merkmalsausprägungen der beiden Fälle, noch eine Berechnung der Ähnlichkeit, wie bei den weichen Kriterien, durchgeführt wird. Aufgrund dieser Sonderstellung und der Tatsache, daß die anderen Teilprozesse das für die Vorauswahl genutzte Wissen nicht benötigen, wurde für dieses Wissen ein eigenes Modul gebildet. Das Wissen über die Ähnlichkeit der Begriffe wurde in der Synonymverwaltung abgelegt<sup>69</sup>.

### 5.4.3 Die lokale Ähnlichkeitsbestimmung

Grundlage für die Bestimmung der lokalen Ähnlichkeit bildet der Vergleich von Basistypen. Unter dem Begriff der Basistypen sind die schon zuvor beschriebenen Primitiven und elementaren Eigenschaften der Objekte zusammengefaßt. Für diese Basistypen müssen innerhalb des Systems bestimmte Vorgaben zur Ähnlichkeitsbestimmung gemacht werden. Diese Vorgaben sind der folgenden Tabelle zu entnehmen<sup>70</sup>:

Basistyp	Vordefiniertes Ähnlichkeitsmaß
<i>Integer, Real, ...</i>	bewertete Differenz oder Gleichheit
<i>String</i>	Transformationskosten oder Gleichheit
<i>Boolean, Symbol</i>	Gleichheit
<i>Ordered Symbol</i>	bewerteter Abstand in der definierten Ordnung
<i>Taxonomy</i>	Abstand in der Taxonomy

Die hier definierten Maße lassen sich über die Typenhierarchie auch auf abgeleitete Typen übertragen, so daß diese Maße sich auch auf abgeleitete und zusammengesetzte Objekte übertragen lassen. Für den Merkmalstyp *Taxonomy* und den Merkmalstyp *Ordered Symbol* müssen die Ähnlichkeitsbeziehungen explizit definiert werden. Im Anwendungssystem SAM wird dieses Wissen in der Synonymverwaltung abgebildet. Die benötigten Referenzrelationen werden bei Bedarf dynamisch erzeugt. Die Formulierung ist dabei als Distanz, Präferenzrelation  $\succ$  bzw.  $\succeq$  oder direkt als Ähnlichkeit möglich. Der Vergleich der Merkmalsausprägungen kann intern nach drei Prinzipien, entweder als Liste, Tabelle oder als Funktion erfolgen:

- Die Ähnlichkeit wird anhand der Nähe von Objekten in der Ordnung über die Liste ausgedrückt. Die Liste repräsentiert direkt die Präferenzrelation  $x \succeq z$ . Voraussetzung für die Verwendung einer Liste zur Repräsentation der Ähnlichkeit ist ein endlicher Wertebereich.

<sup>69</sup>Vgl. Abschnitt 5.3

<sup>70</sup>Die an dieser Stelle aufgeführten Maße sind in Anlehnung an die im System *Inreca* verwendeten Ähnlichkeitsmaße definiert. Nähere Ausführungen hierzu sind [Weß 1996], S. 127 f. zu entnehmen.

- Für endliche Wertebereiche kann eine Tabelle zur Repräsentation der Ähnlichkeit definiert werden. Die Vergleichswerte der entsprechenden Merkmalsausprägungen entsprechen dann den, in den Zeilen und Spalten eingetragenen Werte. Jede Zeile  $z$  der Tabelle  $T$  definiert somit eine spezifische Präferenzordnung der Form:

$$T_{i(x,y)} \geq T_{i(x,z)} \Leftrightarrow y \succeq_x z.$$

- Eine weitere Möglichkeit bietet die Definition einer Vergleichsfunktion  $f$ . Diese Funktion liefert einen entsprechenden Ähnlichkeitswert zurück. Die entsprechende Präferenzrelation  $y \succeq_x z$  wird mit Hilfe der Funktionswerte über  $f_{(x,y)} \geq f_{(x,z)}$  induziert.

Mit Hilfe dieser Strategien ist es möglich, lokale Präferenzrelationen zu definieren bzw. die lokalen Ähnlichkeitswerte zu bestimmen. Durch die Anwendung der objektorientierten Programmierung kann der lokale Ähnlichkeitsbegriff bzw. der Ähnlichkeitsbegriff für die Primitiven auch auf komplexere Typen und Strukturen übertragen werden.

#### 5.4.4 Die globale Ähnlichkeitsbestimmung

Durch die Verwendung globaler Ähnlichkeitsmaße ist es möglich, nicht nur die einzelnen Merkmalsausprägungen, sondern auch ganze Objekte zu vergleichen. Zur Bestimmung der globalen Ähnlichkeit wurde bereits im Abschnitt 4.2 die Funktion  $sim$  definiert. Mit Hilfe dieser Funktion  $sim$  ist es möglich, für ein bestimmtes Anfrageobjekt  $x$  eine spezifische Präferenzrelation der Form:

$$(10) \quad sim_{(x,y)} \geq sim_{(x,z)} \Leftrightarrow y \succeq_x z$$

zu bestimmen. Die entsprechende Präferenzrelation kann demnach auch anstatt des spezifischen Maßes direkt im Domänenmodell repräsentiert werden.

Neben dem Vergleich von Objekten gleicher Klassenzugehörigkeit ist es auch möglich, Objekte unterschiedlicher Klassenzugehörigkeit zu vergleichen. Für die vorliegende Arbeit ist der Vergleich von Objekten unterschiedlicher Klassenzugehörigkeit zwingend notwendig, da einem Patienten eine Menge von potentiell geeigneten Studien zugeordnet werden sollen. Der Vergleich von Objekten unterschiedlicher Klassen kann nach drei grundsätzlichen Strategien erfolgen:

- Vergleich von Merkmalen, die in in beiden Objekten vorhandenen sind
- Vergleich aller vorhandenen Merkmale aus der Vereinigung von zwei Objekten<sup>71</sup> oder
- Vergleich von Merkmalen anhand einer zuvor spezifizierten Präferenzrelation<sup>72</sup>

---

<sup>71</sup>Vgl. [Weß 1996], S. 130

<sup>72</sup>Vgl. [Bergmann & Stahl 1998], S. 31

Die Verwendung aller Merkmale macht jedoch wenig Sinn. Ebenso ist die Verwendung einer zuvor spezifizierten Präferenzrelation zu aufwendig und scheidet aufgrund des zu erwartenden relativ hohen Wartungsaufwandes aus. Hieraus wird deutlich, daß für diese Arbeit nur der erste Ansatz in Frage kommt. An dieser Stelle spielt auch die Behandlung nicht erhobener Daten eine Rolle. Diese Daten sollen wie Merkmale behandelt werden, die nicht in beiden Klassen vorkommen. Allerdings muß deren Existenz für die richtige Kategorisierung der Fälle bzw. Studien vermerkt werden.

Im Unterschied zu dem Vergleich einzelner Merkmale kann beim Objektvergleich der Fall auftreten, daß einige Merkmale für die Bewertung der Ähnlichkeit wichtiger sind als andere. Dieser Unterschied im Einfluß auf die Bewertung der Ähnlichkeit kann auch als eine Art von Relevanz für die Problemlösung interpretiert werden. Unter dem Begriff der Relevanz können jedoch je nach Kontext verschiedene Dinge verstanden werden. Eine gängige Art und Weise zur Modellierung der Relevanz bildet die Verwendung von Gewichten, insbesondere wenn ein geometrisch motiviertes Ähnlichkeitsmodell Verwendung findet.

## 6 Das Studienauswahlmodul SAM

### 6.1 Architektur des wissensbasierten Studienauswahlmoduls SAM

In diesem Abschnitt wird näher auf die Architektur des Studienauswahlmoduls SAM eingegangen.

Die wichtigsten Komponenten des Studienauswahlmoduls SAM sind:

- Fallbasis
- Synonymverwaltung
- Patientendatenbank
- Benutzeroberfläche
- Ablaufsteuerung des Auswahlprozesses

In enger Verbindung zur Fallbasis, der Synonymverwaltung und der Patientendatenbank stehen die Datenbankschnittstellen und die Datenbankengine.

#### 6.1.1 Funktionsumfang des Anwendungssystems

Das Studienauswahlmodul SAM hat in erster Linie die Aufgabe, dem Anwender aus einer Menge von Studien anhand von relevanten Patientendaten eine oder mehrere geeignete Studien vorzuschlagen. Bei der Generierung eines Therapievorschlages wird zunächst eine Kategorisierung der Studien auf Basis des Grades der Eignung für den Patienten vorgenommen. Bei der Einordnung der Studien in die Kategorien wird zwischen potentiell geeigneten, bedingt geeigneten und ungeeigneten Studien unterschieden. Innerhalb der einzelnen Kategorien werden die Studien entsprechend ihrem Grad der Eignung für den Patienten sortiert. Näheres zur Ergebnispräsentation ist im Abschnitt 6.5.1 beschrieben.

Innerhalb von SAM wird eine Unterscheidung in studienrelevante Patientendaten (Pflichtdaten) und weniger relevanten Patientendaten, deren Eingabe nicht zwingend notwendig ist, vorgenommen. Fehlende Patientendaten können so teilweise toleriert werden. Bei der Studienauswahl verwendet SAM das in der Synonymverwaltung abgelegte Wissen, um die Lösungsqualität zu verbessern. Näheres zur Synonymverwaltung und der Unterstützung des Auswahlprozesses durch die Synonymverwaltung ist in den Abschnitten 5.2 und 6.2.2 zu finden.

Für die Verwaltung von Patientendaten und zur Vermeidung von Mehrfacheingaben wurde in SAM neben der Studiendatenbank eine Datenbankanbindung für die Patientendaten realisiert, deren Aufbau und Funktionsweise im Abschnitt 6.4 erklärt ist. Weiterhin ist es

möglich, sich mit Hilfe der Erklärungskomponente den Prozeß der Therapieauswahl und das Ergebnis erklären zu lassen. Die Aufgaben und der Aufbau der Erklärungskomponente sind im Abschnitt 6.3 beschrieben.

### 6.1.2 Der Therapieauswahlprozeß

Im Abschnitt 5.2 wurde der Algorithmus für die Auswahl einer geeigneten Studie erläutert. An dieser Stelle soll der programmtechnische Ablauf der Studiauswahl beschrieben werden.

Nach dem Start von SAM erfolgt die Anmeldung bzw. Autorisierung und der Verbindungsaufbau zu den Datenbanken. Anschließend können nun Patientendaten schrittweise eingegeben werden. In SAM wird zu Beginn einer Auswahl potentiell geeigneter Studien zwischen studienrelevanten und weniger studienrelevanten Daten unterschieden. Die studienrelevanten Daten, die sehr wichtig für die Auswahl einer geeigneten Studie sind, sind Pflichtdaten und müssen eingegeben werden. Pflichtdaten sind beispielsweise die gestellte Diagnose, das Alter und der WHO-Status des Patienten, sowie Labordaten wie der Leukozytenwert im Blut. Nach der Eingabe der Pflichtdaten wird der Suchalgorithmus gestartet. Für die Auswahl wird im ersten Schritt ein Suchvektor bestehend aus der gestellten Diagnose und den Synonymen für die gestellte Diagnose generiert. Mit Hilfe dieses Suchvektors wird die Vorauswahl durchgeführt. Eine detaillierte Beschreibung der Vorgehensweise bei der Vorauswahl ist im Abschnitt 5.1.1 bzw. 6.2.2 zu finden. Im nächsten Schritt wird mit der durch die Vorauswahl erhaltenen Liste der eigentliche Fallvergleich gemäß der in den Abschnitten 5.1.2 und 6.2.3 beschriebenen Vorgehensweise durchgeführt und der Grad der Eignung der Studie für den Patienten anhand eines Ähnlichkeitsmaßes errechnet.

Im Anschluß an den Auswahlprozeß werden alle vorausgewählten Studien anhand des ermittelten Eignungsgrades für den Patienten in drei Kategorien eingeteilt. Jede Kategorie wird dem Anwender einzeln präsentiert. Die erste Kategorie besteht aus Studien, die hinsichtlich der Einhaltung der Ein- und Ausschlußkriterien am geeignetsten erscheinen. Die Ein- und Ausschlußkriterien der Studie werden vom Patienten entweder vollständig oder nur mit geringen Abweichungen erfüllt. In der zweiten Kategorie befinden sich die Studien, welche für den Patienten bedingt geeignet sind. Eine bedingte Eignung einer Studie für den Patienten liegt dann vor, wenn entweder die eingegebenen Patientendaten nicht vollständig sind und/oder die Patientendaten eine größere Abweichung von den im Studienprotokoll angegebenen Ein- und Ausschlußkriterien aufweisen. Im ersten Fall, die Patientendaten sind nicht vollständig, ist es möglich, den Durchlauf mit vervollständigten Patientendaten zu wiederholen, um damit evtl. ein besseres Ergebnis zu erzielen. Im zweiten Fall, die Patientendaten weichen stärker von einem bzw. mehreren Ein- und Ausschlußkriterien ab, liegt es in der Entscheidung des Arztes, ob die Studie trotz der Abweichungen von den Ein- und Ausschlußkriterien für den Patienten

geeignet erscheint. Die dritte Kategorie beinhaltet Studien, welche für den Patienten aufgrund von großen Abweichungen und/oder Verstößen gegen die Ein- und Ausschlußkriterien nicht geeignet sind.

Für die Nachvollziehbarkeit der Problemlösung durch das System, insbesondere bei der Kategorisierung von Studien nach dem Grad der Eignung für den Patienten, ist es notwendig, daß das System über eine Erklärungskomponente verfügt.

## 6.2 Die Fallbasis

### 6.2.1 Struktur der Fallbasis

Die Anzahl der Fälle der zugrundeliegenden Domäne ist im Vergleich zu Fallbasen anderer fallvergleichender Systeme gering. Dennoch stellt die strukturelle Aufbereitung der einzelnen Fälle besondere Anforderungen an die Modellierung der Fallbasis. Die Struktur der Fallbasis muß so aufgebaut werden, daß die auftretenden Fallvarianzen adäquat abgebildet werden können. Der Aufbau der Struktur der Fallbasis muß demzufolge sehr flexibel sein.

Im Abschnitt 4.4 wurde bereits dargestellt, daß ein Fall sich aus einer Zusammenfassung von Problembeschreibung und Problemlösung verstehen läßt, wobei dem Fall noch Zusatzinformationen mitgegeben werden können. Für die vorliegende Arbeit sind die Fälle die einzelnen Studien. Die Ein- und Ausschlußkriterien bilden die Problembeschreibung und die Lösung des Problems ist die Darstellung der Therapie. Informationen zu Studienleitung, Studienzentrale, sowie zur Kontaktaufnahme stellen wichtige Zusatzinformationen dar, welche für den Arzt von primärem Interesse sein könnten. Insbesondere Informationen über den Meldevorgang, das Einschleusen des Patienten in die Studie, sowie Informationen über notwendige Voruntersuchungen für die Studie stellen wichtige Zusatzinformationen dar. Daraus läßt sich schließen, daß die eigentliche Problemlösung für den Arzt weniger in der Darstellung der jeweiligen Therapieschemata liegt, sondern vielmehr aus der Bereitstellung der benötigten Zusatzinformationen besteht. Ausgehend von diesen Betrachtungen läßt sich für die Repräsentation der Studien folgende allgemeine Klassenbeschreibung ableiten:

```
public class Studie
{
    Studienname;
    Einschlußkriterien;
    Studieninformation;
}
```

Das Attribut `Studienname` dient zur Identifikation der konkreten Studie. Im Gegensatz zu

**Studiename** ist das Attribut **Studieninformationen** selbst ein Objekt, daß die benötigten Informationen für den Arzt bereitstellt. Die Objektstruktur ist so gestaltet, daß eine Erweiterung ohne großen Aufwand vorgenommen werden kann. Die Ein- und Ausschlußkriterien einer Studie beschreiben die Patientenzielgruppe<sup>73</sup>. Wie bereits erwähnt, ist es möglich, zu jedem Ausschlußkriterium ein äquivalentes Einschlußkriterium zu formulieren. Diese Umformulierung wurde bei der Modellierung der Fallbasis von SAM vorgenommen. Der Vorteil der Umformulierung der Ausschlußkriterien in Einschlußkriterien besteht darin, daß nur noch Einschlußkriterien betrachtet werden müssen. Die Einschlußkriterien bestehen aus patientenbezogenen Merkmalen, Angaben über Vor- und Begleiterkrankungen sowie Angaben über den therapeutischen Verlauf. Die patientenbezogenen Merkmale beziehen sich größtenteils auf Patientenmerkmale wie Alter, Diagnose oder Geschlecht. Das Hauptproblem bei der Modellierung der Einschlußkriterien besteht in der hohen Varianz der für die Beschreibung der Zielgruppe verwendeten Merkmale. Weiterhin besteht die Möglichkeit, daß Merkmale oderverknüpft sind und/oder an zusätzliche Bedingungen gebunden sein können. Ein Beispiel hierfür sind die Einschlußkriterien der NHL-B Studie<sup>74</sup>. Innerhalb der NHL-B-Studie werden Patienten, welche zwischen 18 und 60 Jahre alt sind und einen LDH-Wert größer dem oberen Normwert aufweisen bzw. Patienten, die zwischen 61 und 75 Jahre alt sind, berücksichtigt. Für die Abbildung der Studien in der Fallbasis ist eine Struktur zu finden, welche es ermöglicht, die auftretende Varianz abzubilden.

Die Programmiersprache JAVA bietet hierfür mehrere Konzepte für die Abbildung und Behandlung dynamischer und flexibler Strukturen. Hier ist neben Arrays und Listen, die auch in anderen Programmiersprachen unterstützt werden, vor allem das Konzept der Vektoren zu nennen. Mit dem Konzept der Vektoren lassen sich beliebige Objekte zu einer Struktur innerhalb der Programmiersprache JAVA zusammenfassen. Vektoren lassen sich demnach als verkettete Listen von Objekten verstehen, wobei die Objekte des Vektors unterschiedliche Klassenzugehörigkeiten aufweisen können. In einen Vektor sollten jedoch nicht wahllos Objekte geschrieben werden, da für das Auslesen der Elemente aus einem solchen Vektor aufwendige Castingoperationen (Typenumwandlungen) und Programmverzweigungen notwendig wären. Auch die Verwendung von Vektoren macht demzufolge eine Strukturierung der Einschlußkriterien nicht überflüssig.

### 6.2.2 Vorauswahl beim Fallvergleich

In diesem Abschnitt soll die verwendete Strategie bei der Vorauswahl von Fällen (Studien) näher beschrieben werden. Die Vorauswahl von Fällen ist notwendig, um die Zahl der Fälle,

---

<sup>73</sup>Siehe Begriffsdefinitionen im Abschnitt 2.1

<sup>74</sup>Vgl. [Löffler et. al. 1994], S. 12

welche später durch den Fallvergleich betrachtet werden müssen, zu reduzieren. Diese Reduktion ist insbesondere in Hinblick auf die angestrebte Erweiterung des Systems auf die gesamte Hämatookologie von Bedeutung. Neben der Reduktion der Anzahl der zu untersuchenden Studien soll der Prozeß der Vorauswahl durch die Verwendung von Wissen über Synonyme und Begriffsstrukturen sicherstellen, daß möglichst keine potentiell geeignete Studie vom detaillierten Fallvergleich ausgeschlossen wird. Durch diese Vorgehensweise soll neben der Steigerung der Effizienz des Auswahlprozesses auch eine Sicherung der Qualität der Lösung erzielt werden.

Bevor der Prozeß der Vorauswahl starten kann, ist die Eingabe verschiedener Patientendaten, insbesondere der Diagnose notwendig. Bei der Eingabe der Patientendaten wird dabei zwischen Pflichtdaten und für die Studienauswahl weniger notwendigen Daten unterschieden<sup>75</sup>. Im folgenden wird anhand der gestellten Diagnose mit Hilfe der Synonymverwaltung der Suchvektor bestimmt.

Durch eine Anfrage an die Synonymverwaltung wird zur gestellten Diagnose  $d$  die Referenzdiagnose  $d_{ref}$  anhand der Referenzklassifikation ermittelt. Beim Aufbau der Synonymverwaltung wird eine Diagnoseklassifikation zur Referenzklassifikation bestimmt, d. h. alle Diagnosen sind der entsprechenden Diagnose in der Referenzklassifikation zugeordnet. Mit Hilfe der Referenzdiagnose  $d_{ref}$  werden alle relevanten Diagnosen ermittelt. Weiterhin ist es möglich, die entsprechenden über- bzw. untergeordneten Diagnosen zu finden, um diese in den Suchprozeß mit einzubeziehen. Der Anwender wird bei der Eingabe der Patientendiagnose ebenfalls durch die Berücksichtigung von Diagnosen auf unterschiedlichen Granularitätsstufen und Diagnosen in verschiedenen Klassifikationssystemen unterstützt. Dies führt dazu, daß alle potentiell relevanten Studien durch die Vorauswahl gefunden werden. Allerdings sollte die Berücksichtigung der Diagnosen hinsichtlich ihres Abstraktionsgrades nicht uneingeschränkt erfolgen, um einer Aufblähung des Suchvektor durch eine zu starken Verallgemeinerung der Diagnose entgegenzuwirken.

Eine Berücksichtigung der Abweichung der Diagnosen hinsichtlich der Klassifikationshierarchien erfolgt durch die Einbeziehung eines entsprechenden Ähnlichkeitswertes für die einzelnen Diagnosen. Die Zuordnung der Ähnlichkeitswerte zu den Diagnosen erfolgt anhand der im Abschnitt 5.1.1 beschriebenen Vorgehensweise. Die Diagnosen werden in einem Vektor  $v_d$  mit der Form  $\langle d_{ref}, d_i \rangle$  abgelegt. Wobei  $d_i$  für eine zu  $d$  ähnlichen Diagnosen steht. Für die Auswahl ähnlicher Diagnosen wurde der in Definition 1 festgelegte Ähnlichkeitsbegriff verwendet. Neben den Diagnosen ist für jedes  $d_i$  ebenfalls der Grad der Ähnlichkeit festzuhalten. Grundlage für die Abbildung des Ähnlichkeitsgrades bilden die Hierarchieebenen.

Bei der Bestimmung der Abstufung der Ähnlichkeit sind einige Besonderheiten zu be-

---

<sup>75</sup>Nähere Ausführungen hierzu sind im Abschnitt 6.1.2 zu finden.

achten. Die Ähnlichkeit der Diagnosen ist nicht geometrisch motiviert, d. h. der Abstand zwischen zwei Merkmalsausprägungen kann nicht durch deren Differenz ausgedrückt werden. Für die Berechnung der Ähnlichkeit muß eine entsprechende Abbildung realisiert werden. Wird eine Diagnose  $d$  in mehrere Unterdiagnosen  $d_u$  aufgeteilt, so bildet jede Menge von Unterdiagnosen eine echte Teilmenge der Menge der Diagnose  $D$ , so daß gilt:  $D_u \subset D$ . Die Mengen der Unterdiagnosen bilden hierbei disjunkte Mengen. Demzufolge besteht an dieser Stelle eine gewisse Unsicherheit darüber, ob ein Patient mit der Diagnose  $d$  auch mit einem Studienprotokoll für maligne Tumorerkrankungen der Art  $d_u$  behandelt werden kann.

Zum besseren Verständnis dieser Problematik soll ein Beispiel dienen: Bei einem Patienten wurde ein T-Zell-Lymphom niedriger Malignität festgestellt. Das T-Zonen-Lymphom und das Sezary-Syndrom werden innerhalb der Kiel-Klassifikation dem T-Zell-Lymphom niedriger Malignität zugeordnet. Der Patient hat in diesem Fall tatsächlich das T-Zonen-Lymphom. Aufgrund der beschriebenen Vorgehensweise zur Erstellung des Suchvektors wird allerdings auch das Sezary-Syndrom aufgenommen, da an dieser Stelle eine Unsicherheit bezüglich der genauen Diagnose vorliegt. Die hier auftretende Unsicherheit soll bei der Bestimmung der Ähnlichkeit mit berücksichtigt werden.

Für die Bestimmung einer Therapie ist in vielen Fällen schon eine grobe Diagnose ausreichend, obwohl eine Feindiagnose gefordert wird. Die Feindiagnose ist für diese Fälle oft nur für die Auswertung der Studien bzw. für die Prognose des Patienten relevant. Daher muß die Synonymverwaltung von SAM neben einer Spezialisierung auch die Generalisierung der Diagnosen unterstützen. Bei der Generalisierung wird einer Diagnose  $d$  eine übergeordnete Diagnose  $d_o$  zugeordnet. Für die Diagnose  $d$  gilt hier:  $d \subset d_o$ . Die Diagnose  $d$  bildet hierbei also eine echte Teilmenge von  $d_o$ . Bezogen auf den Studienkontext bedeutet dies, daß alle potentiell geeigneten Studien für die Erkrankung  $d_o$  auch für die Erkrankung  $d$  potentiell geeignet sind. Eine Abstufung der Ähnlichkeitswerte, wie im Fall der Diagnosespezifizierung, erscheint hier nicht notwendig.

Mit Hilfe der oben beschriebenen Vorgehensweise und den bereits im Abschnitt 4.3 angestellten Überlegungen läßt sich folgende Objektstruktur für die Speicherung der Synonyme und Begriffshierarchien herleiten<sup>76</sup>:

---

<sup>76</sup>Die hier verwendete Struktur wurde in Anlehnung an [Chen 93] modelliert.

```

class Diagnose
{
    String      : DiagnoseName;
    Diagnose   : ReferenzDiagnose;
    Diagnose   : HigherDiagnose;
    Integer    : HierarchieEbene;
}

```

Diese Objektstruktur wird während der Laufzeit des Programms dynamisch um einen Wert erweitert, der den Grad der Ähnlichkeit der aktuellen Diagnose zu der gestellten Diagnose repräsentiert.

Mit Hilfe des Vektors  $v_d$  wird die eigentliche Vorauswahl gestartet. Die gesamte Fallbasis wird für jede Diagnose des Vektors  $v_d$  durchsucht. Stimmt die für eine Studie relevante Diagnose mit der aktuell betrachteten Diagnose aus  $v_d$  überein, wird die Studie vorausgewählt und der Grad der Ähnlichkeit entsprechend der relativen Stellung der Diagnose in der Klassifikationshierarchie bestimmt. Das hierzu notwendige Durchsuchen der Datenbank ist mit Hilfe von Datenbankabfragen realisiert. Nähere Ausführungen sind in Abschnitt 6.4 beschrieben. Die Datenbankabfragen liefern schließlich die potentiell relevanten Studien, welche in einem Vektor gespeichert und für den anschließenden detaillierten Fallvergleich herangezogen werden.

### 6.2.3 Detaillierter Fallvergleich

Der detaillierte Fallvergleich hat die Aufgabe, alle ausgewählten Studien näher zu untersuchen und die Studien in drei Kategorien (potentiell geeignet, bedingt geeignet und ungeeignet) einzuordnen. Die Grundlage für den detaillierten Fallvergleich im System SAM bildet die im Abschnitt 5.1.2 beschriebene Vorgehensweise. In diesem Abschnitt soll die Umsetzung des detaillierten Fallvergleiches näher erläutert werden.

Wie bereits erwähnt, startet der detaillierte Fallvergleich mit einer bewerteten Liste von Studien, d. h. zu jeder vorausgewählten Studie wird gemäß der im vorhergehenden Abschnitt beschriebenen Vorgehensweise ein Wert für die Ähnlichkeit bestimmt. Dieser Wert wird zusammen mit dem Fall in eine Liste aufgenommen. Aufgrund der Tatsache, daß sich ein Ausschlußkriterium stets so umformulieren läßt, daß es als Einschlußkriterium aufgefaßt werden kann, erfolgte diese Transformation bei der Erstellung der Fallbasis. Damit soll eine einfachere Handhabbarkeit der Studien und eine Vereinfachung des Verfahrens für den detaillierten Fallvergleich erreicht werden.

Für jede Studie in der Liste wird nun der detaillierte Vergleich beginnend mit den har-

ten Kriterien durchgeführt. Jedes harte Kriterium wird nun dahingehend überprüft, ob die Ein- bzw. Ausschlußkriterien der Studie durch die Patientenmerkmale erfüllt sind. Dieser Vergleich erfolgt je nach Merkmal durch Prüfung von Gleichheit oder durch Überprüfung der entsprechenden Intervallgrenzen. Hierbei wird das Maß für die Eignung bzw. Ähnlichkeit  $\sigma_{is} = \sigma_{i-1s}$  entweder beibehalten, das Kriterium ist erfüllt, oder es gilt  $\sigma_{is} = 0$ , das Kriterium ist nicht erfüllt. Ist bei einem Patienten ein hartes Kriterium nicht erfüllt, so wird die Studie für diesen Patienten als ungeeignet markiert und vom weiteren Fallvergleich ausgeschlossen.

Die Überprüfung der weichen Patientenmerkmale wird aufgrund des verwendeten Modells zur Ähnlichkeitsbestimmung etwas aufwendiger. Im ersten Schritt sind für die weichen Merkmale lokale Ähnlichkeitsfunktionen zu definieren. In der Gleichung (7) im Abschnitt 5.1.2 wird für die lokalen Ähnlichkeitsfunktionen der Wertebereich festgelegt. Es werden aber keine Aussagen über den konkreten Verlauf dieser Funktionen getroffen. Aufgrund der im Abschnitt 5.1.2 getroffenen Aussagen lassen sich jedoch vier verschiedene Funktionsbereiche bzw. Fallunterscheidungen für die lokalen Ähnlichkeitsfunktionen ableiten. Im ersten Fall wird das Kriterium erfüllt, die Patientenmerkmale stimmen mit den Ein- bzw. Ausschlußkriterium der Studie genau überein bzw. die Patientenmerkmale liegen im geforderten Intervall. Demzufolge muß für das lokale Ähnlichkeitsmaß gelten:

$$(11) \quad \mu_{i(p_i, s_i)} = 1$$

Im zweiten Fall liegt das Patientenmerkmal knapp außerhalb des geforderten Bereiches bzw. das Merkmal weicht geringfügig vom geforderten Wert ab. Daher kann für die lokale Ähnlichkeit nicht mehr  $\mu_{i(p_i, s_i)} = 1$  gelten. Für diesen, zweiten Fall muß der Wert für die lokale Ähnlichkeit modifiziert werden. Aufgrund der Tatsache, daß diese Abweichung nur relativ gering ausfällt, ist die Modellierung einer lineare Funktion für die Ähnlichkeit ausreichend. Für die lokale Ähnlichkeitsfunktion im zweiten Fall gilt demnach<sup>77</sup>:

$$(12) \quad \mu_{i(p_i, s_i)} = 1 - \frac{\Delta_i}{\delta_i} \quad \text{mit } \Delta_i \leq \varepsilon_i \quad \text{und } \varepsilon_i \leq \delta_i$$

Für  $\varepsilon_i$  und  $\delta_i$  gelten die im Abschnitt 4.2.2 formulierten Festlegungen, wobei der Index  $i$  für das  $i$ -te Merkmal steht.  $\Delta_i$  soll hier für die tatsächliche Abweichung des Patientenmerkmals vom geforderten Wert stehen. Das folgende Beispiel soll die Anwendung der beschriebenen Ähnlichkeitsfunktion verdeutlichen. In einer Studie  $s$  soll der Wert der Thrombozyten im Blut größer als 100.000 sein. Der tatsächliche Wert für die Thrombozyten eines Patienten

<sup>77</sup>Diese Funktion wurde in Anlehnung an die von Weiß (Siehe [Weiß 1996], S. 53) beschriebene Ähnlichkeitsfunktion modelliert.

beträgt 99.500, woraus sich für  $\Delta_i$  ein Wert von 500 ergibt. Für  $\varepsilon$  wird ein Wert von 1.000 angenommen und für  $\delta$  ein Wert von 2.500. Unter Verwendung der Gleichung (12) ergibt sich ein Wert für  $\mu$  von 0,8.

Im dritten liegt der Wert des Patientenmerkmals im zweiten Intervall<sup>78</sup>. In diesem Fall werden die Studien maximal in die Kategorie der „bedingt geeigneten Studien“ eingeordnet, d. h. der Anwender muß über die Eignung letztlich selbst entscheiden. Wenn sich der Wert von  $\Delta_i$  dem Wert von  $\delta_i$  nähert, fällt die Ähnlichkeit mehr als linear ab, während der Ähnlichkeitswert in der Nähe von  $\varepsilon$  weniger stark ansteigt. In diesem Fall kann also eine Ähnlichkeitsfunktion der folgenden Form präferiert werden:

$$(13) \quad \mu_{i(p_i, s_i)} = 1 - \frac{\varepsilon_i}{\delta_i} - \left( \frac{\Delta_i - \varepsilon_i}{\delta_i} \right)^{n_i} \quad \text{mit } \Delta_i \leq \delta_i$$

Für den Exponenten  $n_i$  gilt  $n_i \geq 1 \in \mathcal{R}$ . Auch die Anwendung dieser Ähnlichkeitsfunktion soll durch ein Beispiel verdeutlicht werden. Es gelten die oben beschriebenen Werte. Der Wert der Thrombozyten im Blut für den Patienten beträgt in diesem Beispiel allerdings 98.500. Für den Exponenten  $n$  wird der Wert 1 angenommen. Unter Verwendung der Ähnlichkeitsfunktion in Gleichung (13) ergibt sich ein Wert für  $\mu$  von 0,4.

Mit Hilfe dieser Ähnlichkeitsfunktion wird einerseits der zweite Fall und die, für den dritten Fall getroffenen Feststellungen mitberücksichtigt und andererseits bleibt so die Monotonie der Ähnlichkeitsfunktion  $\mu_{i(p_i, s_i)}$  gewahrt<sup>79</sup>. Für die Patienten  $p$  und  $p'$  muß demnach bezüglich der Präferenzrelation  $\succeq_{s_i}$  der Studie  $s$  gelten:

$$(14) \quad \mu_{i(p_i, s_i)} \geq \mu_{i(p'_i, s_i)} \Leftrightarrow p_i \succeq_{s_i} p'_i$$

Im vierten Fall liegt der Patientenwert außerhalb des zweiten Intervalls. Daraus folgt, daß das Kriterium nicht erfüllt wird und somit gilt:

$$(15) \quad \mu_{i(p_i, s_i)} = 0$$

Nach der Ermittlung der lokalen Ähnlichkeit mit Hilfe der eben beschriebenen Teilfunktionen wird in einem nachfolgenden Schritt die globale Ähnlichkeit ermittelt. Diese wird auf

<sup>78</sup>Vgl. Abschnitt 4.2.2

<sup>79</sup>Die Monotonie ist eine Eigenschaft, welche bei Ähnlichkeitsmaßen im Allgemeinen nicht unbedingt gefordert wird. In vorliegenden Fall ist jedoch leicht einzusehen, daß eine Abnahme des Abstandes des Patientenmerkmals vom geforderten Wert eine Zunahme der partiellen Ähnlichkeit zufolge hat. Die Ähnlichkeitsfunktion sollte deshalb diese Abhängigkeiten mit berücksichtigen. Weitere Ausführungen zu Ähnlichkeit und Monotonie sind in [Weß 1996], S. 46 ff. zu finden.

Basis der in der Gleichung (6) beschriebenen globalen Ähnlichkeitsfunktion ermittelt. Anschließend ist der Ähnlichkeitswert  $\sigma_{is}$  mit dem Schellenwert  $\omega_u$  zu vergleichen. Wird der Schwellenwert  $\omega_u$  vom Ähnlichkeitswert  $\sigma_{is}$  unterschritten, wird die Studie vom weiteren Vergleich ausgeschlossen und den nicht geeigneten Studien zugeordnet.

Der Vergleich von Studien- und Patientendaten erfolgt fallweise, d. h. eine Studie wird solange mit den Patientendaten verglichen, bis die Kategorisierung der Studie feststeht. Diese Vorgehensweise hat den Vorteil, daß jede vorausgewählte Studie nur einmal aus dem Vektor bzw. der Liste der vorausgewählten Studien geholt werden muß. Damit läßt sich diese Vorgehensweise effizienter implementieren.

### 6.3 Die Erklärungskomponente

Obwohl die Erklärungskomponente einen großen Einfluß auf die Akzeptanz des Gesamtsystems hat, kommt ihr in vielen Arbeiten aus dem Bereich der Wissensbasierten Anwendungssysteme eine eher stiefmütterliche Behandlung zu. Die Möglichkeit, den Problemlösungsprozeß in Wissensbasierten Systemen für den Anwender nachvollziehbar zu machen, ist eine der Besonderheiten und ein Vorteil dieser Systeme gegenüber konventionellen Anwendungsprogrammen. Hierbei wird die strikte Trennung des Domänenwissens von der Programmlogik ausgenutzt<sup>80</sup>. Im Anschluß an eine kurze Einführung, in der die allgemeinen Anforderungen an eine Erklärungskomponente beschrieben werden, wird im folgenden auf die Realisierung der Erklärungskomponente in SAM näher eingegangen.

#### 6.3.1 Allgemeine Anforderungen an eine Erklärungskomponente

Die Hauptaufgabe einer Erklärungskomponente besteht darin, dem Anwender in einer verständlichen Form darzustellen, warum und auf welche Art und Weise eine bestimmte Lösung vom System gefunden wurde<sup>81</sup>. Damit der Lösungsweg für den Anwender transparent gestaltet werden kann, sind die einzelnen Schritte bei der Lösungsfindung begründet darzulegen. Durch die Darlegung der einzelnen Lösungsschritte mit ihren jeweiligen Begründungen werden evtl. Fehler bzw. Inkonsistenzen im Lösungsverfahren und/oder in der Fall- bzw. Wissensbasis aufgedeckt. Die Erklärungskomponente dient demnach auch der Nachvollziehbarkeit der Problemlösung, Plausibilitätskontrolle, der Prüfung der Korrektheit und des Konsistenzchecks der Wissensbasis. Diese Aufgaben kann die Erklärungskomponente um so besser erfüllen, je genauer sie die Lösungsfindung und den Lösungsweg beschreibt.

In [Puppe 1991] werden zwei Arten von Erklärungen unterschieden, die eine gute Erklärungskomponente verwenden sollte:

---

<sup>80</sup>Vgl. [Puppe 1991], S.132

<sup>81</sup>Vgl. [Helbig 1996], S.248

- Direkte Erklärungen, welche die Teile des Programmes identifizieren, die zur Herleitung des Ergebnisses eingesetzt wurden und welche Rolle diese Programmteile gespielt haben. Die Möglichkeit zur Generierung von direkten Erklärungen bildet den eigentlichen Vorteil der Wissenbasierten Systeme im Bereich der Erklärungstechniken gegenüber konventionellen Systemen.
- Indirekte Erklärungen dienen der Begründung des Einsatzes von bestimmten Methoden und Komponenten und sollen die Nachvollziehbarkeit der allgemeinen Programmlogik ermöglichen.

Wichtig für die Generierung guter Erklärungen ist das richtige Abstraktionsniveau. Bei guten direkten Erklärungen sollten alle unnötigen Details, insbesondere die programmtechnischen Einzelheiten, strikt unterdrückt werden. Hierbei sind drei relevante Wissenstypen zu unterscheiden:

- Strategisches Wissen: Wissen, welches zur Rechtfertigung der Reihenfolge von Aktionen des wissensbasierten Systems verwendet wird
- Strukturelles Wissen: Wissen über die Objekte und ihre Beziehungen untereinander
- Unterstützendes Wissen: Wissen zur Generierung der indirekten Erklärungen

Auf Basis dieser drei Wissenstypen ist es nach [Puppe 1991] möglich, rechnergestützte Erklärungen zu generieren, wobei das strategische Wissen mehr den Prozeß der Lösungsfindung erklärt, die anderen beiden Wissenstypen hingegen die Erklärungsgenerierung unterstützen.

### 6.3.2 Inhalte der Erklärungskomponente von SAM

Die Erklärungskomponente von SAM hat im wesentlichen die Aufgabe, die vorgenommene rechnergestützte Kategorisierung einer Studie anhand des Eignungsgrades für einen konkreten Patienten dem Anwender zu erklären.

Hierbei soll dem Anwender für jede vom System vorausgewählte Studie die Bewertung der einzelnen studienrelevanten Patientenmerkmale verdeutlicht werden. Für diese Art der Erklärungsfähigkeit ist es notwendig, daß die detaillierte Bewertung auch nach der abgeschlossenen Kategorisierung solange erhalten bleibt, bis die Patientendaten geändert werden.

Mit Hilfe der Erklärungskomponente soll insbesondere auf die relevanten Patientendaten, die den Ausschlag für die entsprechende Kategorisierung gegeben haben, hingewiesen werden. Dies gilt insbesondere für Studien, welche in die Kategorie „bedingt geeignete Studien“ eingestuft werden, da für diese Kategorie ein besonderer Erklärungsbedarf besteht. Dem Anwender ist insbesondere darzulegen, ob die betreffende Studie zum einen aufgrund fehlender

Patientendaten und/oder zum anderen durch die leichte Verletzung eines oder mehrerer weicher Ein- und Ausschlußkriterien zu der Einordnung geführt haben. Für den Fall, daß dem Anwender eine „bedingt geeignete“ Studie erfolgversprechender erscheint als eine „geeignete“ Studie, wird diesem die Möglichkeit gegeben, einen erneuten Durchlauf mit zusätzlichen Patientendaten durchzuführen. Bei stärkeren Abweichungen der Patientenmerkmale von den Ein- und Ausschlußkriterien einer Studie, liegt es in der Verantwortung des Anwenders, ob der Patient innerhalb der „bedingt geeigneten“ Studie behandelt werden kann. Oft toleriert ein Arzt die Verletzung eines Kriteriums oder einiger weniger wichtiger Kriterien, wenn die anderen relevanten Patientendaten die Kriterien erfüllen. SAM simuliert zwar dieses Verhalten mit Hilfe der Einbeziehung von unscharfen Bereichen, überläßt aber die endgültige Entscheidung dem Arzt, da die Abbildung der komplexen Regel- und Bedingungsstruktur bei einem solchen Vorgang den Rahmen des Systems sprengen würde, zumal diese Einschätzungen z. T. sehr subjektiv sind und komplexen Bedingungsstrukturen unterliegen.

Die Einordnung von Studien in die Kategorie der „geeigneten Studien“ verlangt wenig Erklärungsbedarf und soll im wesentlichen der Verifikation und der Prüfung der Plausibilität der Entscheidung dienen. Werden Studien in die Kategorie der „nicht geeigneten Studien“ eingeordnet, ist dem Arzt verständlich darzustellen, welcher Verstoß gegen die Ein- bzw. Ausschlußkriterien zu dieser Kategorisierung geführt hat bzw. an welcher Stelle der Vergleich abgebrochen wurde.

Neben der Erklärung des Lösungsweges soll dem Anwender darüberhinaus die Möglichkeit gegeben werden, sich über alle Studien, welche in den Fallvergleich eingeschlossen wurden, detailliert zu informieren. Hierzu soll es möglich sein, die Informationen zu den Studien direkt abzurufen. Im Prototypen von SAM wurden die Möglichkeiten zum Informationsabruf nur ansatzweise entwickelt.

#### 6.4 Die Datenbankanbindung

Für die Implementierung von SAM wurde zur Realisierung der Datenbanken das Datenbanksystem Personal Oracle Lite for JAVA in der Version 3.5 verwendet. Hierbei handelt es sich um ein objektrelationales Datenbankmanagementsystem. Dieses System hat den Vorteil, daß es über einen nativen JDBC-Treiber<sup>82</sup> verfügt und JAVA-Klassen sich mit Hilfe von sogenannten JAVA-Acess-Classes direkt speichern lassen. Dadurch werden aufwendige Mappingoperationen, welche bei der Verwendung von objektorientierter Programmierung und relationalen Datenbanksystemen auftreten, vermieden. Ein weiterer Vorteil dieses Datenbanksystems besteht darin, daß die Datenmanipulationssprache SQL für Datenbankoperationen verwendet

---

<sup>82</sup>JDBC steht für JAVA Database Connectivity. JDBC stellt eine herstellerunabhängige Schnittstelle für den Zugriff auf Datenbanken mittels der Datenmanipulationssprache SQL bereit.

werden kann. Diese Sprache wurde in Personal Oracle Lite dahingehend erweitert, daß neben Transformationen auf Tabellen und Relationen auch Operationen über Objekte möglich sind. Das für SAM verwendete objektrelationale Datenbanksystem Personal Oracle Lite for JAVA vereinigt dementsprechend die Vorzüge der objektorientierten Programmierung mit den Vorteilen relationaler Datenbanksysteme und bietet mit einer SQL-Schnittstelle ein mächtiges Werkzeug zur Datenmanipulation.

Für die Generierung, die Ausführung und die Auswertung sowie die Verwaltung der Datenbankverbindungen ist unter Verwendung des JAVA-Klassenkonzeptes ein Datenbankmodul, die Datenbankengine, implementiert worden. Hierdurch ist es möglich, den eigentlichen Anwendungscode vom Datenbankmodell und den Datenbankzugriffen zu trennen und damit besser zu modularisieren. Die Datenbankengine generiert mit Hilfe einiger Parameter eine oder mehrere Datenbankanfragen, welche vom Datenbankmanagementsystem ausgewertet werden. Das Ergebnis wird der Datenbankengine von der Datenbankmanagementsystem (DBMS) in Form von Resultsets übermittelt. Diese werden wiederum von der Datenbankengine zerlegt und ausgewertet. Die Methode, welche die Datenbankengine aufgerufen hat, erhält nun das entsprechende Ergebnis.

Aufgrund dieser Vorgehensweise wird deutlich, daß die Datenbankanbindung und das Datenbankmodell für den Rest der Anwendung vollkommen transparent sind und eine Änderung im Datenmodell oder die Verwendung eines neuen Datenbankmanagementsystems lediglich Änderungen in der Datenbankengine und den von der Datenbankengine verwendeten JAVA Access Klassen bedarf.

## 6.5 Die interaktive Benutzeroberfläche

Die Benutzeroberfläche von SAM wurde mit Hilfe der JFC<sup>83</sup> „Swing“ implementiert und an die Oberfläche des Therapieplanungsmoduls ThEO angelehnt. Zur Oberfläche gehören das Hauptfenster und mehrere Dialoge, welche im wesentlichen vom Hauptfenster aus aufgerufen werden können.

### 6.5.1 Das Hauptfenster

Das Hauptfenster (Abb. 4) beinhaltet die Menüleiste, die Toolbar und ein System von acht Karteikarten, auch als Aktenreiter bezeichnet, welche den Anwender durch die Studienauswahl geleiten. Die Toolbar besteht aus einer Reihe von Icons und befindet sich direkt unter

---

<sup>83</sup>JAVA Foundation Classes. Die JFC bilden eine Bibliothek, welche Klassen und Komponenten für die Oberflächenprogrammierung bereitstellt. Entsprechend dem JAVA-Konzept können die Oberflächenelemente dem „Look and feel“ des jeweiligen Betriebssystems ohne Änderung des Programmcodes angepaßt werden. Codename der JAVA Foundation Classes ist „Swing“.

**SAM Studienauswahlmodul**

Programm Benutzer Hilfe

Stammdaten Anamnese I Anamnese II Klinik Labor Geeignete Studien Bedingt geeignete Studien Ungeeignete Studien

Bitte ankreuzen!

1 Patient hat Lymphknoten tastbar.  ja  nein

2 Patient hat Schmerzen.  ja  nein

3 Patient hat mehr als 10% seines Gewichtes in den letzten 6 Monaten verloren.  ja  nein

4 Patient hat unklares Fieber über 38° C.  ja  nein

5 Patient hat Nachtschweiß.  ja  nein

6 Patient leidet unter Juckreiz.  ja  nein

7 Patient hat Husten.  ja  nein

8 Patient leidet unter Alkoholschmerz.  ja  nein

Auftreten der ersten Symptome:

Welche

Datum (Monat/Jahr)

Weiter Abbruch

Abb. 4: Die Oberfläche von SAM (Quelle: eigene Darstellung)

der Menüleiste. Mit Hilfe der Toolbar wird der Aufruf wichtiger Menüpunkte, wie „Laden“ und „Speichern“, erleichtert und beschleunigt.

Die einzelnen Karteikarten dienen zur Datenerhebung und zur Darstellung der Ergebnisse. Die Datenerhebung kann durch Dateneingabe von Seiten des Nutzers oder durch den Zugriff auf bereits in der Datenbank gespeicherte Patientendaten erfolgen. Dem Nutzer stehen für die Dateneingabe unterschiedliche Oberflächenkomponenten wie Textfelder, Comboboxen und Checkboxen, zur Verfügung. Die Textfelder dienen zur freien Eingabe von Texten oder Zahlen, Comboboxen zur Auswahl eines Elementes aus einer Menge von möglichen Elementen und die Checkboxen der Auswahl von keinem, einem oder mehreren Elementen aus einer Menge möglicher Elemente. Die Eingabe der Diagnose erfolgt dialoggesteuert. In einer alphabetisch geordneten Liste werden alle dem System bekannten Diagnosen aufgeführt. Der Anwender hat die Möglichkeit, eine der Diagnosen aus der Liste auszuwählen. Intern wird bei der Eingabe von Daten zwischen Daten, die zwingend eingegeben werden müssen, und Daten, die nicht zwingend notwendig sind, unterschieden. Im unteren Teil der Karteikarten befinden sich jeweils zwei Button. Ein Button soll den Abbruch und die Neueingabe ermöglichen, der

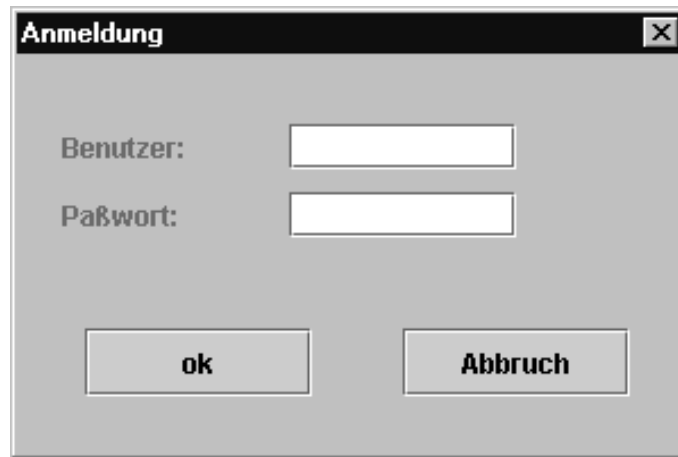


Abb. 5: Der Paßwortdialog (Quelle: eigene Darstellung)

andere die Eingabe weiterer Daten und damit die Freigabe der nächsten Karteikarte bzw. den Suchvorgang starten.

Zur Ergebnispräsentation werden ebenfalls die Karteikarten benutzt. Die Studien werden entsprechend ihrer Eignung getrennt präsentiert. Innerhalb der einzelnen Karteikarten werden die Studien entsprechend ihrer Eignung in einer Tabelle geordnet. An dieser Stelle können Informationen über die Studien abgerufen werden. Desweiteren ist hier eine Erklärung für die Bestimmung des Grades der Eignung und die Einordnung der Studie in diese Kategorie abrufbar. Durch die nach Eignung der Studien für den Patienten aufgeteilte Ergebnispräsentation soll dem Anwender der Einfluß fehlender Patientendaten verdeutlicht werden.

### 6.5.2 Die Dialoge

Die Interaktionen zwischen Nutzer und System werden durch den Einsatz verschiedener Dialoge unterstützt. Die einzelnen Dialoge werden nachfolgend beschrieben.

#### Der Paßwortdialog

Bei jedem Start von SAM ist die Anmeldung des Benutzers im System sowie an der lokalen Datenbanken erfolgen. Für die Abfrage von Nutzernamen und Paßwort dient der Paßwortdialog (vgl. Abb. 5).

Die paßwortgeschützte Anmeldung an der lokalen Datenbank soll zusammen mit der Verwaltung von Nutzerprofilen den nicht autorisierten Zugriff auf Studien- und Patientendaten verhindern. Bei fehlerhafter Eingabe des Paßwortes wird dies dem Nutzer in einer Fehlermeldung mitgeteilt. Die Anmeldung kann bis zu drei mal versucht werden. Kann eine Anmeldung an der Datenbank aus technischen Gründen, wie z. B. fehlerhafte Netzwerkverbindung, Hard-

The image shows a standard Windows-style dialog box with a title bar that reads 'Laden von Patientendaten' and a close button (X) on the right. The dialog has a light gray background. It contains three rows of labels and text boxes: 'Name : Schmidt', 'Vorname : Katja', and 'Datum : 05.05.1941'. At the bottom, there are two buttons: 'ok' on the left and 'Abbruch' on the right.

Abb. 6: Dialog zum Laden von Patientendaten (Quelle: eigene Darstellung)

waredefekt oder einer Beschädigung der Datenbank, nicht erfolgen, erscheint ebenfalls eine Fehlermeldung. Schlägt auch der dritte Versuch der Anmeldung fehl, wird das Programm beendet. Desweiteren kann durch die Betätigung des Abbruch-Buttons das Programms beendet werden.

### Das Laden von Patientendaten

Durch den Aufruf des Menüpunktes „Laden“ ist es möglich, gespeicherte Patientendaten aus der Personal Oracle Lite Datenbank zu laden. Damit wird dem Anwender die Möglichkeit gegeben, den Auswahlprozeß für den Patienten erneut zu durchlaufen, ohne alle Patientendaten erneut einzugeben.

Für die Identifikation eines, in der Datenbank gespeicherten Patienten ist die Eingabe des vollständigen Patientennamen, sowie des Geburtsdatums notwendig (Siehe Abb. 6). Nach dem Betätigen des OK-Buttons wird eine entsprechende Anfrage an die Datenbank generiert, vorhandene Patientendaten geladen und in den hierfür vorgesehenen Karteikarten visualisiert.

### Der Datenbankdialog

Unter dem Menüpunkt „Datenbankoptionen“ wird dem Anwender mit Hilfe eines Dialoges die Möglichkeit gegeben, die verschiedenen Quellen für die Datenbanken festzulegen.

Die Verzeichnisse für die Datenbanken können jeweils manuell eingetragen werden, oder es kann mit Hilfe des Buttons „Durchsuchen“ ein weiterer Dialog, ein Filedialog, geöffnet werden (vgl. Abb. 7).

Dieser Filedialog soll die Suche nach den Datenbanken erleichtern. Nach der Bestätigung des Buttons „Öffnen“ bzw. des Buttons „OK“ wird getestet, ob sich die erforderlichen Daten-

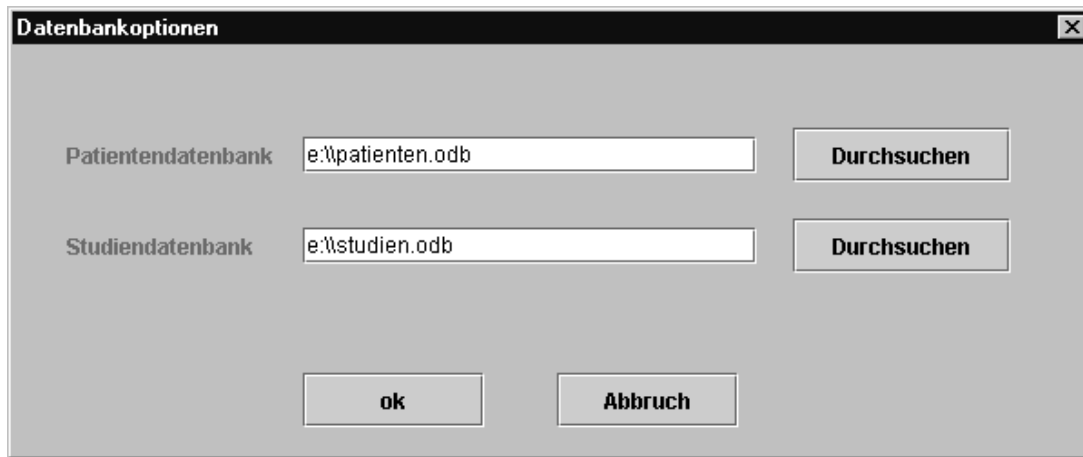


Abb. 7: Dialog zur Angabe der Datenbankpfade (Quelle: eigene Darstellung)

banken an der angegebenen Stelle befinden und ein Zugriff auf diese Datenbanken möglich ist. Auftretende Fehler werden analog der bei der passwortgeschützten Anmeldung beschriebenen Vorgehensweise behandelt.

### Das Erklären von Ergebnissen

Durch Betätigen der Schaltfläche „Erklären“ auf einer der Karteikarten zur Ergebnisinterpretation kann sich der Anwender erklären lassen, warum die Einordnung der Studie in diese Kategorie erfolgte und wie der Grad der Eignung ermittelt wurde. Zu diesem Zweck werden alle Einschlußkriterien zusammen mit der lokalen Ähnlichkeit und einer rechnergestützten qualitativen Bewertung der Patientendaten in einer Tabelle, wie z. B. „leichte Abweichung vom Einschlußkriterium“, aufgeführt. Am Ende der Tabelle wird die Vorgehensweise bei der Zusammenfassung der einzelnen Werte für die lokale Ähnlichkeit zu einem Gesamtmaß erläutert.

### Die Studieninformationen

Nähere Informationen zu der Studie können durch Betätigen der Schaltfläche „Info“ abgerufen werden. Die Studieninformationen beinhalten zur Zeit Daten über die Studienleitung, die Studiensekretariate, die unterschiedlichen Ansprechpartner und den Therapieablauf. Die Funktion soll im Laufe des MedNet-Projektes<sup>84</sup> mit den Informationsdiensten zusammengelegt werden, so daß die Daten des Informationsservers genutzt werden können, um Redundanz zu verringern. Aus diesem Grund ist diese Funktion in der bestehenden Version des Studienauswahlmoduls SAM nur im Ansatz realisiert.

<sup>84</sup>Vgl. Abschnitt 2.4.2 und [Heller & Löffler 1998], S. 12

### Der Informationsdialog

Der Menüpunkt „Info“ im Hilfe-Menü dient zur Anzeige allgemeiner Programminformationen über die Version und das Entwicklerteam von SAM. Dieser Menüpunkt ist nicht mit den allgemeinen medizinischen Informationen und dem Hilfesystem des System zu verwechseln.

## 6.6 Integration in das Gesamtsystem

SAM ist zunächst als ein Stand-alone-System entwickelt worden. Bei der Umsetzung der Konzepte wurde jedoch darauf geachtet, daß eine Integration in das Gesamtsystem ohne großen Aufwand möglich ist. Zur Zeit ist eine Integration des Studienauswahlmoduls SAM mit der Internetkommunikationsapplikation InKA<sup>85</sup> denkbar. Mit Hilfe von InKA kann eine Aktualisierung der Datenbestände, insbesondere der Studiendatenbank, welche die Fallbasis bildet, von SAM vorgenommen werden. Für eine individuelle Therapieplanung kann nach der Auswahl einer potentiell geeigneten Studie durch Studienauswahlmodul SAM das Therapieplanungsmodul ThEO eingesetzt werden.

Der hier beschriebene Workflow wird jedoch noch nicht vollständig informationstechnisch unterstützt, da nach der Auswahl der Studie im Regelfall noch die papiergestützte Bestätigung der Aufnahme, sowie die Randomisation durch das Studiensekretariat erfolgt. Für diese Schritte sind im Rahmen des MedNet-Projektes „Maligne Lymphome“ bereits entsprechende Softwarewerkzeuge geplant.

Im Laufe des MedNet-Projektes „Maligne Lymphome“ ist der Aufbau eines Baukastensystems, das alle Arbeitsschritte bei der Planung, der Durchführung und der Auswertung onkologischer Therapiestudien abdeckt, geplant. Hierzu sollen die modularen Softwareeinheiten unter einer gemeinsamen dynamischen Benutzeroberfläche integriert werden.

---

<sup>85</sup>Vgl. Abschnitt 2.4.1 und [Forster & Krüger 1998]

## 7 Diskussion des erstellten Studienauswahlmoduls SAM

In den vorangegangenen Abschnitten 5 und 6 wurde der Aufbau des Studienauswahlmoduls SAM, sowie die verwendeten Strategien bei der Studienauswahl beschrieben. In diesem Kapitel wird eine Bewertung des entwickelten Systems hinsichtlich der in den Abschnitten 3.1 und 4.2 gestellten Anforderungen vorgenommen. Diese Anforderungen sind nachfolgend noch einmal kurz aufgeführt:

- Intuitive und leicht erlernbare Bedienbarkeit
- Anbindung einer Patientendatenbank zur Vermeidung von Mehrfacheingaben
- Einfache Wartbarkeit insbesondere der Studien- bzw. Therapiedatenbank
- Toleranz gegenüber fehlenden Werten
- Generierung bewerteter Therapievorschlage
- Moglichkeit zur Offline-Bearbeitung
- Erklarfahigkeit der Ergebnisfindung und der Ergebnisse
- Kurze Antwortzeiten
- Einsatz auf verschiedene Plattformen (WinNT, Win9x, Unix, MacOS)
- Leichte Erweiterbarkeit der Fallbasis
- Integrierbarkeit in das Gesamtprojekt
- Toleranz gegenüber fehlenden Werten
- Einbeziehung synonymmer Begriffsstrukturen in den Proze der Therapieauswahl

Die Bewertung der gestellten Anforderungen im Folgenden getrennt nach den Anforderungen des Anwenders und nach Systemanforderungen werden, insbesondere hinsichtlich der Anforderungen an die Wissensbasis, diskutiert.

### 7.1 Akzeptanzkriterien des Anwenders

Wichtig fur die Verwendung eines Softwarewerkzeuges ist die Akzeptanz durch die Anwender. In Hinblick auf die Akzeptanz durch den Anwender wird das Studienauswahlmodul SAM unter den folgenden Kriterien naher betrachtet:

- Anwenderfreundlichkeit

- Stabilität
- Funktionalität
- Zeit- und Ressourcenverhalten

### 7.1.1 Anwenderfreundlichkeit

Der Begriff der Anwenderfreundlichkeit umfaßt alle Eigenschaften einer Software, die das Erlernen des Bedienens und den Umgang selbst mit einem Anwendungssystem erleichtern.

Eine zentrale Rolle bei der Erlernbarkeit einer Software spielt die Oberfläche, welche die Schnittstelle bei der Mensch-Maschine-Kommunikation bereitstellt. Die Oberfläche eines Anwendungssystem soll möglichst intuitiv bedienbar sein. Hierbei sind bei deren Entwicklung allgemeine Standards einzuhalten.

Die Oberfläche des Studienauswahlmoduls SAM wurde in Anlehnung an die Oberfläche von ThEO implementiert, um einerseits eine einheitliche Nutzerführung zu ermöglichen und andererseits bereits bekannte Kritikpunkte mit in den Entwicklungsprozeß von SAM einfließen zu lassen. Damit eine Einarbeitung der Nutzer in das Anwendungsprogramm zu erleichtern, wurden die Eingabemasken für die Patientendaten entsprechend der Gruppierung der Daten auf den Erhebungsbögen der Studien HD 10 - 12 und NHL-B<sup>86</sup> gruppiert. Ebenso wurde SAM so realisiert, daß nur die wichtigsten studienrelevanten Patientendaten, wie beispielsweise die Diagnose und das Stadium der Krankheit, eingegeben werden müssen. Bei der Dateneingabe, insbesondere bei der Eingabe der Diagnose, wird der Anwender durch die Verwendung synonymmer Begriffsstrukturen unterstützt.

Mit Hilfe des Studienauswahlmoduls SAM können nach dem Grad der Eignung abgestufte Listen von Therapievorschlügen generiert werden. Zur Verbesserung der Übersicht wurden die Studien anhand des Grades der Eignung für den konkreten Patienten in die drei Kategorien („geeignet“, „bedingt geeignet“ und „ungeeignet“) eingeteilt. Der Anwender kann sich mit Hilfe der Erklärungskomponente erläutern lassen, warum eine Studie einer konkreten Kategorie zugeordnet wurde. Somit ist einerseits der Anwender in der Lage, die Kategoriebildung nachzuvollziehen und andererseits können so mögliche Fehler bei der rechnergestützten Kategoriebildung aufgedeckt werden. Zur Erhöhung der Akzeptanz beim Nutzer ist es weiterhin möglich, mit SAM off-line zu arbeiten und nur zum Abgleich der Datenbank eine direkte Verbindung zur Studienzentrale via Internet mit Hilfe der Internetkommunikationsapplikation InKA aufzubauen<sup>87</sup>.

---

<sup>86</sup>Vgl. [Sieber et. al. 1999] und [Löffler et. al. 1994], S. 14

<sup>87</sup>Siehe auch Abschnitt 2.4.1 und 2.4.2

### 7.1.2 Stabilität und Zuverlässigkeit

Unter Stabilität und Zuverlässigkeit werden hier alle Eigenschaften eines Anwendungssystems gezählt, die dafür sorgen, daß das System die gestellten Anforderungen erfüllt. In erster Linie zählt zu diesen Eigenschaften die Unempfindlichkeit der Software gegenüber falschen Eingaben und unsachgemäßer Bedienung. Hierbei ist vor allem das Abfangen falscher Eingaben, z. B. die Eingabe von Zeichenketten an Stelle von Zahlen, bzw. falscher Eingabewerte, d. h. Werte außerhalb eines erlaubten Wertebereiches zu nennen. Mit dem Test der Eingabedaten auf Gültigkeit des Wertebereiches ist es teilweise möglich, inhaltliche Fehler abzufangen. Durch diesen Test können jedoch nicht alle inhaltlichen Fehler abgefangen werden. Die zusätzliche Überprüfung der Eingaben durch den Nutzer ist deshalb, gerade in Hinblick auf die Sensibilität der Patientendaten, unumgänglich.

Ein weiterer Punkt, der die Stabilität und Robustheit eines Anwendungssystems kennzeichnet, ist die Unempfindlichkeit gegenüber unerwarteten Ereignissen und Störungen. Die Programmiersprache JAVA bietet zur Behandlung von möglichen unerwarteten Ereignissen und Störungen das Konzept der Exceptions. Mit Hilfe dieses Konzeptes können auftretende Fehler abgefangen und anschließend beispielweise in einem Fehlerdialog behandelt werden. Nach der Fehlerbehandlung kann die Arbeit mit dem Programm je nach Art und Schwere des aufgetretenden Fehlers, normal fortgesetzt bzw. kontrolliert beendet werden.

Ein weiterer Aspekt der Zuverlässigkeit besteht im Schutz und in der Sicherheit der gespeicherten Daten vor unerlaubter Manipulation. Für den Schutz der Daten, ist es hierbei notwendig, den Zugriff auf die Datenbank zu kontrollieren. Die Zugriffskontrolle auf die Datenbank erfolgt jeweils nach dem Start des Programm mit Hilfe des Paßwortdialogs<sup>88</sup>.

### 7.1.3 Zeit- und Ressourcenverhalten

Wichtig für die Akzeptanz einer Software beim Anwender ist das Zeit- und Ressourcenverhalten. Unter dem Zeitverhalten eines Systems, Zeit zu verstehen, die das Anwendungssystem für die Abarbeitung eines Anwendungsfalles benötigt. Das Ressourcenverhalten beschreibt den Bedarf des Anwendungssystems an Betriebsmitteln<sup>89</sup>.

Der, für das Anwendungssystem SAM benötigte Platz auf der Festplatte wird weitestgehend durch die virtuelle Maschine von SUN zur Interpretation des Bytecodes und der Klassenbibliotheken bestimmt. Durch die virtuelle Maschine von SUN werden zusammen mit den benötigten Klassenbibliotheken ca. 9 MB belegt. Der durch SAM zusätzliche benötigte Speicherplatz beträgt zusammen mit den Hilfedateien ca. 1 MB.

---

<sup>88</sup>Siehe Abschnitt 6.5.2

<sup>89</sup>Insbesondere werden an dieser Stelle der Bedarf des Hauptspeichers und der, durch das System belegte Platz auf der Festplatte betrachtet.

Zur Laufzeit benötigt SAM zusammen mit der virtuellen Maschine ca. 14 MB im Hauptspeicher. Die Ursache für diesen relativ hohen Speicherbedarf liegt in der Verwendung von JAVA und den JFC „Swing“. Als Hardwarevoraussetzung für den Einsatz von SAM gelten somit Computersysteme mit mindestens 64 MB Hauptspeicher.

Das Anwendungssystem SAM benötigt beim Start eine relativ lange Zeit zur Initialisierung (ca. 10 Sekunden). Die Ursache hierfür liegt in der Verwendung des JDK<sup>90</sup> in der Version 1.1.7. Durch Verwendung einer neueren Version des JDK insbesondere der JFC „Swing“ kann diese Zeit allerdings noch verkürzt werden. Nach der Initialisierung treten, z. B. beim Aufruf der Dialoge, nur kurze Antwortzeiten (unter 1 Sekunde) auf. Über das Zeitverhalten bei der Studienauswahl kann noch keine Aussage getroffen werden, da die Fallbasis im Prototyp von SAM noch sehr klein ist.

## 7.2 Wissensbasis

Der zentrale Bestandteil eines Wissensbasierten Systems ist die Wissensbasis. Die Wissensbasis entscheidet zusammen mit den verwendeten Problemlösungsstrategien über die Kompetenz und die Effizienz des Wissensbasierten Systems. In diesem Abschnitt sollen die Wissensbasis und die verwendeten Problemlösungsstrategien in Hinblick auf ihre Erweiterbarkeit und ihre Effizienz hin bewertet werden.

### 7.2.1 Medizinisches Wissen

Das verwendete medizinische Wissen, insbesondere das studienspezifische Wissen, ist zum großen Teil domänenspezifisches Wissen und somit vom Anwendungsgebiet abhängig. Der Hauptbestandteil der Wissensbasis in einem Fallvergleichenden System ist die Fallbasis. Besonders wichtig, in Hinblick auf die Erweiterbarkeit einer Fallbasis, ist der unabhängige Aufbau der Fallbasis vom verwendeten Ähnlichkeitsmaß und der Auswahlstrategien bei der Repräsentation der Fälle.

Aus diesem Grund wurde beim Entwurf der Fallbasis von SAM darauf geachtet, daß das Wissen in der Fallbasis durch andere Softwarewerkzeuge, z. B. ThEO, wiederverwendet werden kann. Desweiteren wurde bei der bei der Repräsentation der einzelnen Fälle die Forderung nach einer unscharfen Bewertung einzelner Kriterien berücksichtigt. So ist es möglich, für einzelne Kriterien Abweichungen anzugeben, welche noch toleriert werden können. Eine singuläre Festlegung der tolerierbaren Grenzen ist somit für jeden Fall möglich.

Die Repräsentation des Wissens für die Synonymverwaltung wurde ebenfalls streng von der Fallrepräsentation getrennt. Durch die Trennung der Fallbasis und der Synonymverwal-

---

<sup>90</sup>JAVA Development Kit

tung ist es möglich, die Fallbasis unabhängig von der Synonymverwaltung zu erweitern bzw. gegen anderssprachige Module auszutauschen. Der modulare Aufbau des Systems ermöglicht darüberhinaus eine schnelle Erweiterung der Synonymverwaltung mit neuen Begriffshierarchien.

### 7.2.2 Erweiterbarkeit der Fallbasis

Eine der Hauptanforderungen an die Fallbasis eines fallvergleichenden Systems ist die leichte Wartbarkeit der Fallbasis. Voraussetzung für die leichte Wartbarkeit der Fälle ist ein modularer Aufbau und die strikte Trennung von Falldaten, Patientendaten und Auswahlwissen.

## 7.3 Erweiterbarkeit

Beim Entwurf des Prototypen für das Studienauswahlmodul SAM wurde besonders die Integrierbarkeit des Systems in das Gesamtprojekt berücksichtigt. Insbesondere wurde hierbei der Datenaustausch mit den bereits realisierten Softwarewerkzeugen ThEO und InKA gefordert<sup>91</sup>.

Ein Datenaustausch mit vom Studienauswahlmodul SAM mit den vorhandenen Softwarewerkzeugen ThEO und InKA ist bis dato noch nicht möglich, Allerdings werden zur Zeit Portierungen von ThEO auf JAVA sowie der TherapiepiDatenbank auf das Datenbankmanagementsystem ORACLE geplant. Aufgrund der geplanten Portierung der Datenbank von ThEO ist auch eine Anpassung der Internetkommunikationsapplikation InKA erforderlich. Das Studienauswahlmodul SAM wurde so realisiert, daß nach der Portierung von ThEO und der Anpassung von InKA ein direkter Datenaustausch mit diesen Softwarewerkzeugen möglich ist. Mit Hilfe von SAM wird es dem Arzt ermöglicht, nach einer, für den Patienten geeigneten Therapie zu suchen. Nachdem sich der Anwender für eine Therapie entschieden hat, kann sich der Anwender mit Hilfe von ThEO individuelle Therapiepläne erstellen lassen. Der Einsatz von InKA soll eine Arbeit mit beiden Softwarewerkzeugen offline ermöglichen. Durch die Verwendung von JAVA ist der Einsatz von SAM auf verschiedenen Plattformen möglich.

---

<sup>91</sup>Siehe Abschnitt 1.3 und 2.4.2

## 8 Zusammenfassung und Ausblick

Mit dem Studienauswahlmodul SAM steht dem Anwender ein Wissensbasiertes System zur Studienauswahl im Bereich der Hämatookologie für Erwachsene zur Verfügung. SAM ist ein fallvergleichendes System, mit dessen Hilfe auf der Grundlage studienrelevanter Patientendaten potentiell geeignete Studien für den Patienten ausgewählt werden können. Zur Kommunikation mit dem Anwender bietet SAM eine interaktive Benutzerschnittstelle, welche eine leicht erlernbare und intuitive Bedienung ermöglicht.

Bei der Suche nach potentiell relevanten Studien werden fallbasierte Strategien verwendet. Für die Repräsentation der Studien (Fälle) kommt die objektrelationale Datenbank Personal Oracle Lite for JAVA in der Version 3.5 zur Anwendung. Die einzelnen Fälle werden im System als Objekte abgebildet.

Das Studienauswahlmodul SAM wurde mit dem Ziel entwickelt, den Anwender bei der Suche nach potentiell geeigneten Studien zu unterstützen und den Auswahlprozeß zu beschleunigen. Für die medizinischen Anwender in den onkologischen Fachpraxen und Kliniken soll durch das System medizinisches Wissen aus dem Gebiet der Onkologie schnell und übersichtlich zur Verfügung gestellt werden. Dies gilt insbesondere dann, wenn die, innerhalb des Verbundprojektes MedNet „Maligne Lymphome“ geplanten Informationsdienste bereitgestellt und innerhalb von SAM genutzt werden können. Der Einsatz von SAM soll in beratender Weise erfolgen. Die endgültige Entscheidung für eine bestimmte Therapie bzw. Studie liegt alleine beim medizinischen Anwender. Jedoch soll die Behandlung der Patienten in Studien präferiert werden, um so einen gewissen Standard bei der Behandlung von onkologischen Patienten einhalten zu können.

Für den Einsatz von Entscheidungsunterstützenden Systemen in der Medizin ist jedoch das größte Problem in der geringen Akzeptanz solcher Systeme zu sehen. Um die Akzeptanzprobleme möglichst gering zu halten, wurde das System so entworfen, daß für ein breites Spektrum von Krankheiten eine Therapieauswahl mit diesem System vorgenommen werden kann. Desweiteren wurde die endgültige Entscheidung über die Therapie ausdrücklich dem Anwender überlassen, um Bedenken auszuräumen, daß das System den Anwender bevormunden oder ihm bestimmte Entscheidungen vorschreiben würde.

Für den praktischen Einsatz von SAM ist eine Erweiterung der Fallbasis um weitere onkologische Fälle notwendig, da SAM mit der vorhandenen, stark begrenzten Fallbasis keinen größeren praktischen Nutzen hat. Desweiteren ist eine ausführliche Testphase in Planung, um mögliche Schwachstellen und Fehler innerhalb des Systems aufzudecken. Neben der korrekten und fehlerfreien Arbeitsweise von SAM sollte insbesondere die Bedienbarkeit in Hinblick auf das richtige Verstehen der Oberfläche und die Qualität der Benutzerführung getestet und

bewertet werden.

Obwohl SAM den im Abschnitt 4.2 gestellten Anforderungen genügt, sind einige Weiterentwicklungen und Verbesserungen denkbar, welche im nachfolgenden aufgeführt sind:

- Ausbau der Online-Hilfe um Hilfestellungen bei medizinischen Fragen, welche beispielsweise im Zusammenhang mit der Stadieneinteilung und der Verwendung von Klassifikationsstrategien stehen.
- Erweiterung der abrufbaren Informationen über einzelne Studien durch die Kommunikation mit den geplanten Informationsdiensten innerhalb des MedNet-Projektes „Maligne Lymphome“.
- Erweiterung der Fallbasis (Studiendatenbank) von SAM um neue hämatoonkologische Studien In einem weiteren Schritt ist zu überlegen, inwieweit auch die Aufnahme von Studien, welche sich mit soliden Tumoren, wie z. B. Mammakarziom beschäftigen, möglich und sinnvoll ist.
- Entwicklung eines Managementtools zur Unterstützung der Eingabe neuer Studien und der Änderungen zu bereits vorhandenen Studien in Hinblick auf die bessere Wartbarkeit der Fallbasis.
- Entwicklung von SAM als Client/Serversystem auf Basis von CORBA; SAM als Client/Serversystem hat den Vorteil, daß innerhalb einer Klinik oder einer Fachpraxis mit mehreren, untereinander vernetzten Rechnern eine Arbeit auf der gleichen Datenbasis möglich wird. Die Daten können unter Verwendung einer solchen Lösung für die Klinik oder die Praxis an einer zentralen Stelle gehalten werden, was den Aufwand für die Pflege der Daten innerhalb der Klinik bzw. der Praxis wesentlich verringern könnte.
- Einführung hierarchisch geordneter Zugangsstrukturen insbesondere in Hinblick auf die Entwicklung von SAM als Client/Serversystem.

## A Literatur

### Literatur

- [Berger 1996] H. Berger: Konzeption und Implementierung eines Konsultationssystems zur Therapieplanung in der Erwachsenenonkologie; Universität Heidelberg 1997.
- [Bergmann 1996] R. Bergmann: Effizientes Problemlösen durch flexible Wiederverwendung von Fällen auf verschiedenen Abstraktionsebenen; Infix St. Augustin 1996.
- [Bergmann & Stahl 1998] R. Bergmann, A. Stahl: Similarity Measures for Object-Oriented Case Representations in B. Smyth, P. Cunningham (Hrsg.): Advances in Case-Based Reasoning; Springer Berlin 1998.
- [Biewer 1997] B. Biewer: Fuzzy-Methoden; Springer Berlin 1997.
- [Chen 1993] P. S. Chen: On Inference Rules of Logic-Based Information Retrieval Systems; Aachener Informatik-Berichte 1993.
- [Classen, Diehl & Kochsiek 1998] M. Classen, V. Diehl, K. Kochsiek: Innere Medizin; Urban & Schwarzenberg München 1998.
- [Diehl & Pfreundschuh 1997] V. Diehl, M. Pfreundschuh: Hodgkin-Lymphome in Ostendorf, Seeber (Hrsg.): Hämatologie, Onkologie; Urban & Schwarzenberg München 1997, S. 534 - 545.
- [Engelhard 1997] M. Engelhard et. al.: Non-Hodgkin-Lymphome in Ostendorf, Seeber (Hrsg.): Hämatologie, Onkologie; Urban & Schwarzenberg München 1997, S. 546 - 566.
- [Forster & Krüger 1998] M. Forster, M. Krüger: Internet-Kommunikationsapplikation zu onkologischen Konsultationssystem ThEO; Universität Leipzig 1998.
- [Goos 1996] K. Goos: Fallbasiertes Klassifizieren; Infix St. Augustin 1996.
- [Helbig 1996] Helbig: Künstliche Intelligenz und automatische Wissensverarbeitung; Verlag Technik, Berlin 1996.
- [Heller & Löffler 1998] B. Heller, M. Löffler: Telematik und rechnerbasiertes Qualitätsmanagement in einem Kommunikationsnetzwerk. Teilprojekt 2 MedNet MALIGNEN LYMPHOME. Antrag zur Förderung eines medizinischen Netzes bei malignen Lymphomen; Universität Leipzig 1998.

- [Hunang, Jennings & Fox 1994] J. Hunang, N. R. Jennings, J. Fox: An Agent Architecture for Distributed Medical Care in M. J. Wooldridge, N. R. Jennings: Intelligent Agents; Springer Verlag 1994.
- [Kolodner 1993] J. L. Kolodner: Case-Based Reasoning; Morgan Kaufmann Publishers; San Mateo CA, 1993.
- [Löffler et. al. 1994] M. Löffler, H. Köppler, M. Pfreundschuh, K. Havemann: Integratives Konzept zur Behandlung hochmaligner Non-Hodgkin-Lymphome; Universität Leipzig 1994.
- [Mainzer 1999] K. Mainzer: Komplexe Systeme, intelligente Computer und Selbstorganisation in Telepolis, [www.heise.de/deutsch/special/bio/2161/2.htm](http://www.heise.de/deutsch/special/bio/2161/2.htm), 1999.
- [Müller 1997] R. Müller: TheMPO: A Knowledge-Based System for Therapy Planning in Pediatric Oncology in Computers in Biology and Medicine Vol. 27(3), 1997.
- [Ostendorf & Seeber 1997] Ostendorf, Seeber (Hrsg.): Hämatologie, Onkologie; Urban & Schwarzenberg München 1997.
- [Petersen 1996] J. Petersen: Ein baukastenartiges Expertensystem-Architekturmodell mit fuzzy-fallbasierter Schlußfolgerungskomponente; VDI Verlag Düsseldorf 1996.
- [Pschyrembel 1997] Pschyrembel: Klinisches Wörterbuch; CD-Version, Walter de Gruyter & Co 1997.
- [Puppe 1990] F. Puppe: Problemlösungsmethoden mit Expertensystemen; Springer Verlag Berlin, 1990.
- [Puppe 1991] F. Puppe: Einführung in Expertensysteme; Springer Verlag Berlin, 1991.
- [Sieber et. al. 1999] M. Sieber et. al.: Studienprotokolle der Primärtherapie; Hodgkin Studiensekretariat Köln 1999.
- [Schulz 1998] R. Schulz: Fallbasierte Entscheidungsunterstützende Systeme - ein Ansatz zur Lösung betrieblicher Entscheidungsprobleme; Universität Leipzig, 1998.
- [Stolpmann und Weiß 1999] M. Stolpmann, S. Weiß: Optimierung der Kundenbeziehung mit CBR-Systemen; Addison-Wesley-Longmann, Bonn 1999.
- [Weisbrod 1995] J. Weisbrod: Unscharfes Schließen; Infix St. Augustin 1995.
- [Weiß 1996] S. Weiß: Fallbasiertes Problemlösen in wissensbasierten Systemen zur Entscheidungsunterstützung und Diagnostik; Infix, St. Augustin 1996.

[Winston 1989] . H. Winston: Künstliche Intelligenz; Addison-Wesley Publishing Company,  
Bonn 1989.

**B Abkürzungsverzeichnis**

CBR	Case-Based Reasoning
CORBA	Common Object Request Broker Architecture
DBMS	Datenbank-Managementsystem
IMISE	Institut für Medizinische Informatik, Statistik und Epidemiologie
InKA	Internetkommunikationsapplikation
JAC	JAVA Access Classes
JDBC	JAVA Database Connectivity
JDK	JAVA Development Kit
JFC	Java Foundation Classes
KI	Künstliche Intelligenz
KIS	Krankenhausinformationssysteme
PDQ	Physical Data Query
REAL	Revised European-American Classification of Lymphoid Neoplasms
SAM	Studienauswahlmodul
TheMPO	Therapy Management in Pediatric Onkology
ThEO	Therapieplanung in der Erwachsenenonkologie