# **Refereed papers**

ion and similar papers at <u>core.ac.uk</u>

bre

# primary care diseases in the *medrapid.info* knowledge base

Pascal René van Quekelberghe

Tobias Jakob Dip Inform

Dirk Hoffmann Dip Med Inform

**Thomas Wetter** 

**Ekkehard Finkeissen** 

Institute of Medical Biometry and Informatics, Department of Medical Informatics, University of Heidelberg, Germany

#### ABSTRACT

**Background** Communication media commonly used in medicine today no longer meet the needs brought on by the present knowledge explosion. The Heidelberg *medrapid* project has been developed to quickly communicate high-quality clinical knowledge to physicians.

**Methods** In this paper, *medrapid* is introduced as an online clinical knowledge resource, and the methods used by the 'knowledge entry' function for the minimalist representation of clinical knowledge in the knowledge base are discussed.

**Results** On average, fewer than 1.4 problems per disease arose during the input of the formal representation of clinical knowledge using the 'knowledge entry' function. However, representation of

disease time processes, descriptions, warnings and graphics with the 'knowledge entry' function remains problematic.

**Conclusions** The 'knowledge entry' function allows fast formal representation of clinical knowledge (<14 minutes per disease) and testing using the integrated quality management system. In the near future, new measures must be found to improve the problematic representation of disease time processes, descriptions, warnings and graphics to formally represent clinical knowledge using the *medrapid* 'knowledge entry' function.

**Keywords:** knowledge communication, knowledge entry, *medrapid*, quality management

## Introduction

### Background

Structuring and standardising the representation of medical knowledge is an important step in developing an expert system for supporting medical decision making. This paper discusses how the *medrapid* authoring system formally represents relevant medical knowledge in its knowledge base in the domain of pulmonary diseases. In this area, decision support software systems have been developed, for example for occupational lung disease and distinction between benign and malignant tumours.<sup>1,2</sup> However, these systems do not provide a knowledge base in terms of 'one model for 20 000' diseases. Moreover, commonly used med-ical communication channels hardly fulfil Cimino's desiderata for a controlled clinical vocabulary.<sup>3</sup> This paper discusses how the *medrapid* knowledge base can help to fulfil Cimino's desiderata.

# Present medical knowledge communication

240

Presently, medical knowledge is most often communicated via classical resource channels such as books, specialty magazines or CD-ROMs. These media offer the advantage of covering wide sectors of specific medical specialty areas. However, these forms of knowledge communication are limited by long publication intervals, complicated publication procedures, a non-uniform presentation of content and varying media during knowledge communication.<sup>4</sup>

In commonly used media, language also presents a serious hurdle for knowledge communication, since medical terminology is poorly standardised and contains a large variety of synonymous specialty terms. Ambiguous medical terms also lead to communication problems.<sup>3</sup> The practicality of the knowledge contained in common media resources poses yet another problem: accessing the content is time-consuming and seriously hindered by tedious manual searches for the desired information. For example, an efficient search for possible differential diagnoses using common resources is barely possible due to their sequential knowledge structure. However, the search for and differentiation between similar disease patterns is urgently required for ensuring comprehensive patient treatment.

Some approaches have been tested for adaptive knowledge navigation, and have concentrated their efforts on web technologies or the usage of extensible mark-up language (XML) and XML-Schema.<sup>5,6</sup> Others have concentrated on domain knowledge but have not covered the whole breadth of the medical domain.<sup>7</sup> However, the inclusion of a large range of diseases is especially important for diagnostic purposes and the support of practitioners outside their own medical specialty.

Many researchers emphasise that knowledge has to be 'represented in the form of rules, constraints, calculations, guidelines, and other logical/algorithmic formats' to be executable.<sup>8</sup> Here, we will show a simpler way of knowledge representation for the compact transfer of clinical knowledge into routine.

# Need for new knowledge communication solutions

The need for a comprehensive solution for the communication of medical knowledge arises from the above mentioned problems posed by commonly used research media. Such a solution should meet the following requirements:

- integration of all sub-processes of medical knowledge communication
- widest possible availability
- highest possible level of currency
- unambiguous terms and vocabulary
- avoidance of varying media in knowledge communication
- high level of completeness, consistency, content accuracy and freedom from redundancies
- integrated quality management
- largely automated knowledge input and output processes
- high speed of knowledge communication
- high level of sensitivity and specificity for research functions.

The online resource *medrapid.info* (URL: public. medrapid.info/expert/cgi-bin/start.cgi [in German]) is presented here as a possible approach to such a solution.

### Materials and methods

# *medrapid* – a quick medical online reference

*medrapid* is a web-based system designed to communicate medical knowledge. Its functions range from knowledge representation via the authoring system to quality management and knowledge entry.

Medical knowledge is entered into the system's knowledge base by the author using an internet-based authoring tool (*medrapid* 'knowledge entry'). An important contrast to writing medical articles or books offered by *medrapid* is the formal representation of medical content:

- medical content is entered into the knowledge base by the author using a highly structured entry template (see Figure 1)
- the vocabulary used is also standardised and managed using a special tool.

Knowledge output takes place using an intelligent search engine (*medrapid expert*), which delivers knowledge about specific diseases in a highly

241

🗿 Willkommen bei MedRapid -		_ = X		
Datei Bearbeiten Ansicht Favoriten Extras ?		A)		
🕒 Zurück 🝷 🕥 - 💌 🗟 🏠 🔎 Suchen 👷 Favo	riten 🜒 Medien 🤣 😥 🕏 😨 🔹 🛄 🎉 💈			
Adresse 🔊 http://www.medrapid.de/	✓ →	Wechseln zu		
Google -	😪 🐯 Web-Suche 👻 🕼 Ez Site-Popups zulässig 🔣 Optionen			
medrapid) Die medizinische Wissensbank für Praxis, Forschung & Bildung	BETA - TEST	out		
medrapid expert medrapid tutor medrapid forur	n äutor admin überluns hilfe			
wissensei	ngabe   externe quellen   fortschrittsanzeige			
Autor 🗸 Allergisches Lungenödem	👻 Editieren 👻 Inhalt kopieren von			
Status [Logbuch zum Krankheitsbild anzeigen]	EINGABE ABGESCHLOSSEN Da kein Coautor für das Krankheitsbild bestimmt wurde, kann der Freigabeprozess nicht fortgesetzt werdenl	<b>^</b>		
Definition				
Name des Krankheitsbildes	Allergisches Lungenödem			
Das Krankheitsbild wird definiert als	Massiver Austritt von Flüssigkeit aus den Lungenkapillaren in das Interstitium und den Alveolarraum, bedingt durch allergische Einflüsse			
ICD-Code				
Zugehörige Fachgebiete	Ľ			
Übergeordnete Krankheitsbilder, die diesem Krankheitsbild nicht zugeordnet sind	• Atopie d • ជំ			
Untergeordnete Krankheitsbilder, die diesem Krankheitsbild nicht zugeordnet sind	£			
Symptomatik				
Lokalisation des Primärgeschehens	● Lungenkapillaren d ● d			
Lokalisation des Sekundärgeschehens	■ Pulmonale Alveolen 的	Y		
E Fertig	🖉 Internet			

#### Figure 1 Entry template of the medrapid authoring system

structured manner, for example by result, disease enhancement, therapy, and so on (see Figure 2).

*medrapid* is currently in a testing phase. The system contains disease patterns for over 4000 diseases. Of these, 3000 have passed the quality management measures. The total number of diseases in western medicine is estimated as approximately 20 000.<sup>9</sup>

The medical knowledge contained in the *medrapid* online reference refers to individual diseases. All information necessary for diagnosis and therapy should be linked to a specific disease. *medrapid* spares the user the trouble of searching through several passages of various specialty books to find the relevant information.

Refinement of disease patterns, as presented in *medrapid*, differs greatly from common reference works: *medrapid* differentiates between diseases based on a very high level of refinement, for example 'allergic bronchial asthma, stage 2'. This level of refinement is more granular than that presented by the International Classification of Diseases (ICD-10), which was developed for billing purposes.<sup>10</sup>

*medrapid expert* helps the user to quickly find relevant information about a desired disease. The result of the search is a list of all possibly related diseases. Should the correct disease not be found directly, or if *medrapid* finds several matching diseases, the user receives an automatically generated list of related terms with which the search can be narrowed down or redirected (see Figure 3). By adding keywords to the search, the amount of results can easily be refined via a mouse click.

# Preparing medical knowledge for use by *medrapid*

Before the *medrapid* 'knowledge entry' can take action, a *medrapid* 'synopsis' must be performed: the author structures and groups the content of a desired specialty area using a standardised procedure.

This serves to create a hierarchical structure that allows precise mapping of each disease to specific lower categories within the hierarchy, for instance: 'cancer > bronchial cancer > small cell bronchial cancer > small cell bronchial cancer limited disease'. The goal of this structuring task is to attain the greatest possible coverage of knowledge of a specialty area. The initial structure chosen was taken from various German standard medical reference works.

PR van Quekelberghe, T Jakob, D Hoffmann et al

242

atel Bearbeiten Ansicht Favoriten Extra	- 2.2 ·					
ator bearbeiterr Arbitrit Payonterr Extre						
3 Zurück 🔹 🕥 - 💌 😰 🕼 🖉	🖓 Suchen 👷 Favoriten	Medien 🛞 🖉	3• 🎍 🖸 • 🗌			
resse 🔊 http://www.medrapid.de/						👻 🛃 Wechselr
ioogle -	*	😚 Web-Suche 🔹 🕷	El Site-Popups zulà	ssig 🛛 💽 Optionen		
nedrapid) e medizinische Wissensbank für Praxis, Fo	rschung & Bildung	BETA - TEST			Pascal van Quel	kelberghe, Editor Logout
medrapid expert medrapid tutor	medrapid forum	í	autor	admin	über uns	hilfe
		ratschlag   favorite	en			
Link zum Forum Krankheitsbil	ld den persönlichen Favori	iten hinzufügen		Ihre Him	weise zur Anfrage	
		[	• • • • • • • • • • • • • • • • • • •		a - In contract statements	
Kleinzelliges Bronchialkarzi	nom zentral Limited	d disease	Name and de	efinition of the	e disease	
		-				
<ul> <li>Lungenkrebs vom kleinzelliger</li> </ul>	n Typ, zentral gelegen, Tun	nor ist auf eine Thoraxse	ite begrenzt, kein Pleu	raerguss, keine Eint	flussstauung	
				raerguss, keine Eint	flussstauung	
Autor(en):	Correspondi	ng details of the	e respective			
	Correspondi		e respective			
Autor(en):	Correspondi	ng details of the	e respective			
Autor(en):	Correspondi	ng details of the	e respective			
Autor(en): Herr van Quekelberghe, Pasca	Correspondi	ng details of the ered by sympto	e respective m, diagnostics			
Autor(en): • Herr van Quekelberghe, Pasca Symptomatik	Correspondi	ng details of the ered by sympto	e respective m, diagnostics	s, therapy, el		2
Autor(en): • Herr van Quekelberghe, Pasca Symptomatik Lokalisation des Primärgeschehens	Correspondi	ng details of the ered by sympto	e respective im, diagnostics agnostik agnostische Maßnahm	s, therapy, el		
Autor(en): • Herr van Quekelberghe, Pasca Symptomatik	Correspondi	ng details of the ered by sympto	e respective m, diagnostics agnostik agnostische Maßnahm • Anamnese	s, therapy, et	tc.)	
Autor(en): • Herr van Quekelberghe, Pasca Symptomatik Lokalisation des Primärgeschehens • Bronchien	Correspondi disease (ord	ng details of the ered by sympto	e respective m, diagnostics agnostik agnostische Maßnahm - Anamnese - Allgemeine Körg	s, therapy, et	tc.)	
Autor(en): • Herr van Quekelberghe, Pasca Symptomatik Lokalisation des Primärgeschehens • Bronchien Lokalisation des Sekundärgeschehens	Correspondi disease (ord	ng details of the ered by sympto	e respective m, diagnostics agnostik agnostische Maßnahm	s, therapy, et	tc.)	
Autor(en): • Herr van Quekelberghe, Pasca Symptomatik Lokalisation des Primärgeschehens • Bronchien Lokalisation des Sekundärgeschehens • Atmungssystem	Correspondin disease (ord	ng details of the ered by sympto	e respective m, diagnostics agnostik agnostische Maßnahm Anamnese Allgemeine Körg Berufsanamnes Inspektion	s, therapy, et	tc.)	
Autor(en): • Herr van Quekelberghe, Pasca Symptomatik Lokalisation des Primärgeschehens • Bronchien Lokalisation des Sekundärgeschehens	Correspondin disease (ord	ng details of the ered by sympto	e respective m, diagnostics agnostik agnostische Maßnahm - Allgemeine Körg - Berufsanamnes - Inspektion - Palpation	s, therapy, et	tc.)	
Autor(en): • Herr van Quekelberghe, Pasca Symptomatik Lokalisation des Primärgeschehens • Bronchien Lokalisation des Sekundärgeschehens • Atmungssystem • Den ganzen Menschen betreffe	Correspondin disease (ord	ng details of the ered by sympto	e respective m, diagnostics agnostik agnostische Maßnahm Anamnese Allgemeine Körr Berufsanamnes Inspektion Palpation Auskultation	s, therapy, et	tc.)	
Autor(en): • Herr van Quekelberghe, Pasca Symptomatik Lokalisation des Primärgeschehens • Bronchien Lokalisation des Sekundärgeschehens • Almungssystem • Den ganzen Menschen betreffe Zustandsbeschreibung des Krankheits	Correspondin disease (ord	ng details of the ered by sympto	e respective m, diagnostice agnostik agnostische Maßnahm - Allgemeine Körg - Berufsanannes - Inspektion - Palpation - Auskultation - Perkussion	s, therapy, ef nen perliche Untersuchun e	t <b>c .)</b> ng durchführen	
Autor(en): • Herr van Quekelberghe, Pasca Symptomatik Lokalisation des Primärgeschehens • Bronchien Lokalisation des Sekundärgeschehens • Atmungasystem • Den ganzen Menschen betreffe Zustandsbeschreibung des Krankheits • Husten	Correspondin disease (ord	ng details of the ered by sympto	e respective m, diagnostics agnostische Maßnahm Anamnese Allgemeine Körg Berufsanamnes Inspektion Palpation Auskuitation Perkussion Routinelabor mi	s, therapy, el	t <b>c .)</b> ng durchführen	
Autor(en): • Herr van Quekelberghe, Pasca Symptomatik Lokalisation des Primärgeschehens • Bronchien Lokalisation des Sekundärgeschehens • Atmungssystem • Den ganzen Menschen betreffe Zustandsbeschreibung des Krankheits • Hamoptyse	Correspondin disease (ord	ng details of the ered by sympto	e respective m, diagnostics agnostik agnostische Maßnahm Anamnese Algemeine Körr Berufsanamnes Inspektion Palpation Auskultation Perkussion Routinelabor mi Diagnostik von 1	s, therapy, effectives of the second se	t <b>c .)</b> ng durchführen	
Autor(en): • Herr van Quekelberghe, Pasca Symptomatik Lokalisation des Primärgeschehens • Bronchien Lokalisation des Sekundärgeschehens • Atmungssystem • Den ganzen Menschen betreffe Zustandsbeschreibung des Krankheits • Husten • Hausten • Hanoptyse • Dyspnoe	Correspondin disease (ord	ng details of the ered by sympto	e respective m, diagnostice agnostische Maßnahm Anamnese Allgemeine Körg Berufsanamnes Inspektion Palpaton Auskultation Perkussion Foutinelabor mi Diagnostik von 1 - FI-Karz	s, therapy, el	t <b>c .)</b> ng durchführen	
Autor(en): • Herr van Quekelberghe, Pasca Symptomatik Lokalisation des Primärgeschehens • Bronchien Lokalisation des Sekundärgeschehens • Atmungssystem • Den ganzen Menschen betreffe Zustandsbeschreibung des Krankheits • Husten • Hamoptyse • Dyspnoe • Fieber	Correspondin disease (ord	ng details of the ered by sympto	e respective m, diagnostics agnostische Maßnahm Anamnese Altgemeine Körr Berufsanamnes Inspektion Palpation Auskuitation Perkussion Routinelabor mi Diagnostik von E HKarz CYFRA	s, therapy, el	t <b>c .)</b> ng durchführen	
Autor(en): • Herr van Quekelberghe, Pasca Symptomatik Lokalisation des Primärgeschehens • Bronchien Lokalisation des Sekundärgeschehens • Atmungssystem • Den ganzen Menschen betreffe Zustandsbeschreibung des Krankheits • Husten • Hamoptyse • Dyspnoe • Fieber • Lokalisierte Schmerzen	Correspondin disease (ord	ng details of the ered by sympto	e respective m, diagnostics agnostik agnostische Maßnahm Anamnese Algemeine Körr Berufsanamnes Inspektion Palpation Auskultation Perkussion Routinelabor mi Diagnostik von 1 EFL-Kar CYFRA Blutbild	s, therapy, el nen berliche Untersuchuu e t CRP-Bestimmung fumormarkern innoembryonäre Antig 21-1	t <b>c .)</b> ng durchführen	
Autor(en): • Herr van Quekelberghe, Pasca Symptomatik Lokalisation des Primärgeschehens • Bronchien Lokalisation des Sekundärgeschehens • Almungssystem • Den ganzen Menschen betreffe Zustandsbeschreibung des Krankheits • Husten • Hamoptyse • Dyspnoe • Fieber • Lokalisierte Schmerzen • Thorax	Correspondin disease (ord	ng details of the ered by sympto	e respective m, diagnostics agnostische Maßnahm Anamnese Allgemeine Körg Berufsanamnes inspektion Palpation Auskultation Perkussion Routinelabor mi Olagnostik von 1 +Fi-Karz CYFRA Blutbild Diagnostik zur E	s, therapy, el nen berliche Untersuchuu e t CRP-Bestimmung fumormarkern innoembryonäre Antig 21-1	t <b>c .)</b> ng durchführen	
Autor(en): • Herr van Quekelberghe, Pasca Symptomatik Lokalisation des Primärgeschehens • Bronchien Lokalisation des Sekundärgeschehens • Atmungssystem • Den ganzen Menschen betreffe Zustandsbeschreibung des Krankheits • Husten • Hämoptyse • Dyspnoe • Fieber • Lokalisierte Schmerzen • Thorax • Schultern	Correspondin disease (ord	ng details of the ered by sympto	e respective m, diagnostics agnostik agnostische Maßnahm Anamnese Altgemeine Körr Berufsanamnes Inspektion Palpation Auskultation Perkussion Perkussion Berufeabor mi Diagnostik von T FIKar CYFRA Blutbild Diagnostik zur E Röttigen	s, therapy, el nen berliche Untersuchuu e t CRP-Bestimmung fumormarkern innoembryonäre Antig 21-1	t <b>c .)</b> ng durchführen	
Autor(en): • Herr van Quekelberghe, Pasca Symptomatik Lokalisation des Primärgeschehens • Bronchien Lokalisation des Sekundärgeschehens • Almungssystem • Den ganzen Menschen betreffe Zustandsbeschreibung des Krankheits • Husten • Hamoptyse • Dyspnoe • Fieber • Lokalisierte Schmerzen • Thorax	Correspondin disease (ord	ng details of the ered by sympto	e respective m, diagnostice agnostische Maßnahm Anamnese Allgemeine Körg Berufsanamnes Inspektion Palpation Auskultation Perkussion Routinelabor mi Diagnostik von T • Fi-Karz • CYFRA Blutbild Diagnostik zur E Röntgen • Thorax	s, therapy, el nen berliche Untersuchuu e t CRP-Bestimmung fumormarkern innoembryonäre Antig 21-1	t <b>c .)</b> ng durchführen	

**Figure 2** Example of knowledge output by *medrapid expert* (an excerpt): the header lists name and definition of the disease, followed by the corresponding details of the respective disease (ordered by symptom, diagnostics, therapy, and so on)

Willkommen bei MedRapid - Datei Bearbeiten Ansicht Favoriten Extras 7		_ 2 ×
	Favoriten 🜒 🔗 🎍 🔯 • 🛄 🕻	10
resse 🗃 http://www.medrapid.de/		Vechseln zu
	V 📸 Web-Suche 🔹 🚳 🔎 Site-Popuos zulassia	P Optionen
Google -	👻 😚 Web-Suche 🔹 💞 🗗 Site-Popups zulässig	Pag Optionen
medrapid) e medizinische Wissensbank für Praxis, Forschung & Bil	BETA - TEST	Pascal van Quekelberghe, Editor Logout
medrapid expert medrapid tutor medrap	d forum autor	admin über uns hilfe
	ratschlag   favoriten	
inge adipositas	e starten 0.38 sec Ihre Hinweise zur Anfrage	Ähnliches bei Google - Ähnliche Bücher
kalisationiArt: atmungssystem larynx 6 fregei (female, fat, forty, fertile, fair, flatulent dyspepsia) chronische obstipation chronische obstipation chronische obstipation chronische obstipation immundefekte keimbesiedlung der gallenwege obstipation <b>mptom:</b> ansteg der entzundungsparameter big beschleunigung cholezystilis cop anstieg diffuser schmerz mit ausstrahlung zwischen schulterblätter entfarbung des utnis entfarbung des stuhls heiserkeit heise	Respiratorische Addose Akute obstruktive Langatis (Krupp) Kehlkopfstanose Gallenblasenstein mit akuter Cholezystitis ohne Gallenw Gallenblasenstein ohne Cholezystitis ohne Gallenwegso Gallenblasenstein ohne Cholezystitis mit Gallenwegsobs Gallengangsstein mit Cholangitis ohne Gallenwegsobs Gallengangsstein mit Cholezystitis ohne Gallenwegsobs Gallengangsstein mit Cholezystitis mit Gallenwegsobs Cholezystolithiasis Cholezystolithiasis Cholezystolithiasis	egsobstruktion obstruktion bstruktion struktion struktion Tutktion Thine Gallenwegsobstruktion Allenwegsobstruktion

Figure 3 Example of the *medrapid* search function: upon entering a term, the user can select further terms from an offered list of related terms to refine the search

As far as possible, all diseases listed in medical reference works are sorted and linked to specific disease groups. This structure then allows creation of a heterarchy structure for all diseases and disease groups. The end nodes of the structure are later filled by the *medrapid* 'knowledge entry' function.

The disease structure created is integrated into the *medrapid* system. Given basic structures must, in other words, be redesigned to match the created disease structure described above. The ICD-10 disease structure is one of the basic structures contained within the *medrapid* system.<sup>10</sup>

#### Knowledge entry in medrapid

The classified knowledge is not entered as freestyle text, but rather by selecting terms from the standardised vocabulary. For this purpose, *medrapid* offers two further lists in addition to the disease list: the specification list and the list of processes. The terms contained in these lists are hierarchically structured, that is, they are classified in terms of subject groups.

The specification list contains terms with which the search term can be refined. Time, location and general specifications are possible; for example, location: thorax, lungs, bronchial tubes; time: fast, slow; general: severe, light, and so on.

The list of processes contains 'action' terms describing diagnostics, therapies, and so on concerning a specific disease, such as cooling, x-ray, surgery, radiation.

The entry fields of the authoring system restrict the use of terms to specific lists. For example, the entry field 'disease condition' only allows the use of terms from the disease list. This serves to limit entry errors.

In summary, three lists of terms exist from which the medical content is taken by the authoring system: the list of diseases, the specification list and the process list. The lists can be processed and edited by an author at any time. Over 20 different disease characteristics serve to better describe the knowledge entered using the authoring system (see Figure 2).

# *medrapid* 'knowledge entry': evaluation methods

The *medrapid* 'knowledge entry' function has been evaluated qualitatively and quantitatively. The qualitative evaluation examined the problems occurring during knowledge entry: why they occurred and how they were solved.<sup>11</sup> For this purpose, a logbook was kept during knowledge entry using the electronic database system Microsoft Office Access 2003<sup>TM</sup>, whereby the problems arising were described subjectively in freestyle text form. After completing the knowledge entry process, the problems were categorised by problem areas and examined.

Complete representation of the knowledge source and the time required for this 'synopsis' process were the goals of the quantitative evaluation. The evaluation targeted completeness of representation and was conducted according to the 20 disease characteristics. The representation of a characteristic was considered complete if the synopsis knowledge could be represented by *medrapid*. The completeness was measured binarily – a graded assessment was specifically rejected because of the numerous methodological problems involved.

### Results

#### Qualitative study results

Some problems occurred while mapping the synopsis knowledge with the *medrapid* system. Four specific areas were identified: mapping of disease time processes, descriptions, warnings and graphics. However, a learning effect could be observed during use of the authoring system. The number of initial problems decreased as the author's understanding increased during the time the authoring system was in use. Each of the four identified problem areas will now be described and detailed by examples.

Time processes represented a recurring problem during knowledge representation with the *medrapid* system. Noticeably, many of the synopsis sources contained a large quantity of information containing disease time processes. A fundamental difference in understanding of the disease concept could be identified between *medrapid* and commonly used medical knowledge resources, which is reflected in this recurring problem. Commonly used sources, in contrast to *medrapid*, do not consider disease as a timeless (current) state, but rather as just the opposite, as a condition that defines itself by its progression in time.

Since these basic concepts are congruent, mapping problems were inevitable. The time information contained in the synopsis sources refers to all aspects of the particular disease, for instance pathogenesis, characteristics, therapy and prognosis. Acute tonsillitis caused by streptococci, for example, is described by medical textbooks as follows: 'Acute tonsillitis is characterised by an *incubation time* of approx 72 hours, a *sudden* onset of complaints, increased pain in the *evening hours, quick* response to antibiotic treatment, and an *infectious period* of approx 24 hours following antibiotic intake. Therapeutically, penicillin is taken over a *period* of 10 days each *morning* and *evening*'.<sup>12</sup>

243

244

During the authoring process using medrapid, it became apparent that common media often take a descriptive approach to disease pattern definition. This is also a fundamentally different approach in comparison to the medrapid system. Common reference media describe disease in freestyle text form. This allows a detailed description of the disease. The medrapid system, however, uses a highly structured vocabulary. Freestyle text entries are only possible during disease definition. The diagnosis polyposis nasi, for example, is described by common medical resources as: 'Rhinoscopic examination shows glassy, grey-coloured swollen polyps that can be encircled with the probe, and whose stem most often can be followed into the intermediary nasal duct'.<sup>12</sup> This type of information cannot be represented by the authoring system due to its formal knowledge representation form. In other words, the content cannot be described, but rather, it must be exactly defined.

Because of the serious nature of medical decisions, commonly used medical resources contain explicit warnings to make the physician aware of special conditions or therapies. These warnings are contained in the synopsis sources in various forms, such as using bold print, coloured markings or graphical accentuations. Dangerous contraindications are an example of how common medical reference sources make use of warnings, for example: 'caution, nephrotoxicity' or 'caution, liver toxicity'. These types of warnings are not yet used in the *medrapid* authoring system.

Graphics also presented a recurring problem during the attempt to adequately represent synopsis knowledge from common reference media using the authoring system. Such sources contain various amounts of graphical material, which serves to efficiently and effectively pass on knowledge to the user and to offer assistance during the search for diagnostic and differential diagnostic information. Photographs of mucous membrane lesions, for example, are often shown to depict infections of the upper respiratory pathways, or schematic diagrams are presented to better describe therapeutic procedures or pathological processes.

Most initial problems that occurred during the authoring process arose while attempting an approach to representing the medical knowledge of the synopsis sources with the *medrapid* authoring system that was as text-oriented as possible. For example, the attempt to represent 'The discrepancy between the severity of the x-ray and the clinical symptoms was noticeable' for the disease emphysema failed. The reason lies within the concept on which the lists used by the *medrapid* system are based: the information formulated in sentences cannot be mapped 1:1. The non-text-oriented, but meaningful, synopsis (knowledge acquisition) allows an easier process for acquiring knowledge using the authoring system.

A learning effect was observed while mapping the synopsis knowledge. An increasingly routine handling of the lists in terms of less time required to represent a specific disease reflects this effect.

A further initial problem occurred during the authoring process when representing symptoms, such as 'swallowing disorders', 'bloody sputum', or 'pounding pain', since each can, in turn, be seen as separate diseases, each possessing its own symptoms and findings. Therefore, they could not be mapped using the specification or process lists. Adding the difficult terms to the list of clinical conditions solved this problem. Hereby, the initial list of diseases was transformed to a list of clinical conditions containing diseases, symptoms and results.

The increased use of specifications during authoring decreased the workload and enabled a more precise representation of the diseases. For example, when mapping information such as 'radiation level 60 Gy', the phrase 'radiation level 60 Gy' was not included in the process list, but rather the information was put together from information taken from the process and specification lists: the term 'radiation' was taken from the process list, and the value '60 Gy' from the specification list. This procedure helped to keep a better overview of the process list.

Of the problems initially encountered, the following remain to be solved: representation of disease time processes, descriptions, warnings and graphics.

#### Quantitative study results

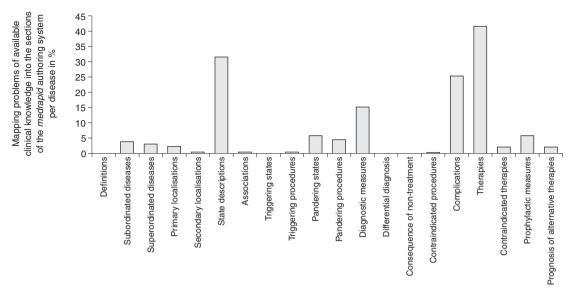
The number of times synopsis content could not be mapped to a specific field using the *medrapid* authoring system entry template was evaluated in the quantitative study. The results can be seen for all diseases in Figure 4.

Mapping speeds for knowledge representation using the authoring system are shown in Table 1.

### Discussion

As mentioned above, of the problems initially encountered, the following remain to be solved: representing disease time processes, descriptions, warnings and graphics.

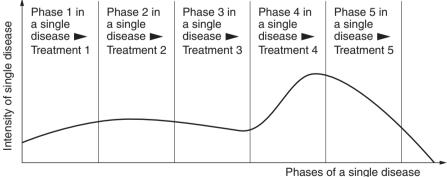
Mapping disease time processes was problematic, as shown by the quantitative analysis. Hereby, a difference is made between disease time processes and disease time specifications. *medrapid* 'knowledge entry' maps time processes using disaggregations (see Figure 5). The disease time process is thereby broken down into several diseases, each representing a different



Sections of the medrapid authoring system

#### Figure 4 Mapping problems per disease

#### Table 1 Mapping speed (averages in minutes) Duration of Total duration Duration of Disease group knowledge of knowledge knowledge representation using representation without representation copy function copy function Infectious diseases 18.26 39.00 18.95 Malignant diseases 5.84 61.00 6.47 Other diseases 14.72 47.33 17.84 Total 12.21 47.08 13.92 Phase 1 in Phase 2 in Phase 3 in Phase 4 in Phase 5 in



#### Figure 5 Disaggregation of the disease time pattern

stage of the disease. The dynamics of the process are then represented as a series of static diseases. This should serve to better describe the disease time processes. Depending on how many stages an individual disease is divided into, a higher or lower refinement of the representation can be achieved. A complex disease possessing a complex time process will, therefore, be broken into several disease stages, all held as sub-phases of the overall disease. However, limitations of this process were often met, so that highly complex diseases, such as malignant lung tumours, could not be broken down fully. The individual time

245

process of a disease cannot be handled comprehensively by *medrapid*, since an exponential growth in the number of diseases created by the breakdown is expected. These limitations in statistical mapping within the medical field were also mentioned by Fleck: 'It is not the *status praesens* but rather the *historia morbi* that leads to the diagnosis'.<sup>13</sup>

246

Disease time specifications, such as incubation period, infectious period, time of intake of medication, and so on, could hardly be mapped by the authoring system: the necessary concepts are missing. However, since the time specifications are of high relevance, a representation of a time-specific concept is greatly needed.

In order to meet the target of highly structured knowledge representation and effective retrieval, freestyle text entries are only permitted in a few sections by the system. The other sections required formal representation based on the controlled vocabulary. This is seen as a barrier for describing relationships between diseases. Although specifically listed terms can be used to describe diseases, explanations, comparisons or therapeutic options cannot be entered. The pathogenesis of a disease, as described in medical reference books, can hardly be mapped and represented by medrapid. The same applies for warnings: the *medrapid* entry template does not allow marking or highlighting of specific text passages. However, especially in an electronic knowledge communication system that aims to speed up knowledge transfer, such as medrapid, such a feature is essential. The problems encountered with descriptions and warnings will mostly affect inexperienced physicians. At least one case study was able to show that the detail-oriented (encyclopaedic) questions stated by experienced physicians could, for the most part, be answered by the system.

Since freestyle text entries are not possible in *medrapid*, descriptions cannot be entered for graphics. This feature, however, would be meaningful. Commonly used medical textbooks use graphics to explain or illustrate complex relationships or therapeutic methods. Especially in cases of malignant diseases, high-resolution x-rays can help physicians to better assess this (sadly) frequent diagnosis.

The author entering content into the system is involved in a constant learning process while working with the *medrapid* system. The initial set of problems encountered was reduced during work with the system. This learning process is good for the author, but poses an overall problem for the *medrapid* system. Initially, many authors try to recite the sources in detail, sometimes without doing a real synopsis of all sources. Therefore, a large number of varying terms have been entered into the authoring system, leading to redundant, semantically equivalent entries. Other authors, in turn, use these terms to describe identical content. This leads to imprecise and ambiguous content. In addition, many of the terms are included in the wrong lists, and are then used incorrectly. Subsequently, processes were added to the specification list and specifications to the list of diseases, and so on. Accordingly, additional features for restricting same, similar or synonym terms have been added.

During proofreading by other authors, different styles of authoring were also observed. Some authors made frequent use of the specification list, while others, for example, tried to integrate specifications into the process list.

Another aspect is the varying degrees of refinement with which the authors enter diseases into the system: authors coming from smaller specialty areas listed more diseases in specific disease groups than their colleagues working in larger specialty areas.

Of the diseases represented in *medrapid*, over 80% could be represented either completely or at least to 90%. The difficulties described arose disproportionately within the group of malignant diseases. This is due to the higher degree of complexity of these diseases as compared to other diseases. They, for example, on average, involve higher numbers of therapies, examinations, differential diagnoses, results, and so forth. In addition, it is very difficult to clearly separate individual disease stages from one another. In contrast, infectious diseases are much less difficult to represent, as they, on average, involve fewer numbers of therapies, examinations, and so on.

Finding solutions for the problems posed by disease time processes, graphics, descriptions and warnings will improve knowledge representation using the authoring system, and will surely help to considerably improve representation of complex medical knowledge content.

The copy function was of considerable importance in terms of the speed with which knowledge was mapped with the *medrapid* system. Initial editing of a disease pattern in the area of respiratory diseases without using the copy function took an average time of 47 minutes 5 seconds, as opposed to an average of 12 minutes 13 seconds using the copy function. Since the copy function can be used any time similar diseases are to be mapped, it is advisable to appoint the same author to homogenous specialty areas during knowledge acquisition.

Surprisingly, the area of malignant diseases was represented with above-average speed because of the high degree of refinement necessitated by their complexity. A large number of very similar diseases were, therefore, represented – an optimal situation for making use of the copy function.

Improving the *medrapid* authoring system to better allow representation of disease time processes, warnings, descriptions and graphics is an important task for the near future. As soon as these problems have been solved and medical knowledge has been broadly covered by *medrapid*, a comprehensive system evaluation is targeted. The evaluation will cover:

- knowledge acquisition using medrapid synopsis
- knowledge entry using medrapid knowledge entry
- knowledge processing
- knowledge output using medrapid expert
- knowledge practicality
- the effect of knowledge communication on medical treatment.

Future refinements of the *medrapid* system also include faster usability and linking external knowledge services (such as hospital information systems) to *medrapid*. This is fundamental for successful knowledge-based systems in terms of decision support.<sup>14</sup> Translation of content based on Unified Medical Language System (UMLS) of the NLM is also planned for the near future.<sup>4</sup> This will help to fulfil the goal of international use of *medrapid*.

Broad use of the *medrapid* system as an online medical reference resource can take place following a comprehensive and positive system evaluation.

### Conclusion

Common forms of medical communication no longer meet the needs brought on by the knowledge explosion of our time. The Heidelberg *medrapid* project serves to quickly communicate up-to-date and high-quality clinical knowledge to physicians in routine practice.

The online quick reference resource *medrapid* has been presented and the *medrapid knowledge entry* function was discussed as a means for entering knowledge into the knowledge base.

On average, fewer than 1.4 problems per disease arose during formal representation of the clinical knowledge using the 'knowledge entry' function. However, representation of disease time processes, descriptions, warnings and graphics using the 'knowledge entry' function remains problematic.

The 'knowledge entry' function allows fast formal representation of clinical knowledge (<14 minutes per disease) and testing using the integrated *medrapid* quality management system. In the near future, new measures must be found to improve the problems of representation of disease time processes, descriptions, warnings and graphics using *medrapid knowledge entry*.

#### REFERENCES

 Harber P, McCoy JM, Howard K, Greer D and Luo J. Artificial intelligence-assisted occupational lung disease diagnosis. *Chest* 1991;100:340–6.

247

- 2 Aoyama M, Li Q, Katsuragawa S, MacMahon H and Doi K. Automated computerized scheme for distinction between benign and malignant solitary pulmonary nodules on chest images. *Medical Physics* 2002;29:701–8.
- 3 Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods of Information in Medicine* 1998;37:394–403.
- Finkeissen E, Fuchs H, Jakob T and Wetter T. medrapid

   medical community & business intelligence system.

   <u>Studies in Health Technology and Informatics 2002;90:</u> 582–6.
- 5 Pagesy R, Soula G and Fieschi M. Improving knowledge navigation with adaptive hypermedia. *Medical Informatics and Internet Medicine* 2000;25:63–77.
- 6 Hoelzer S, Schweiger RK and Dudeck J. Knowledge representation in digital medical documents for efficient information access and retrieval. *Studies in Health Technology and Informatics* 2003;95:451–6.
- 7 Masarie FE Jr, Miller RA and Myers JD. INTERNIST-I properties: representing common sense and good medical practice in a computerized medical knowledge base. *Computers and Biomedical Research* 1985;18:458–79.
- 8 Greenes RA. Decision support at the point of care: challenges in knowledge representation, management, and patient-specific access. <u>Advances in Dental Research</u> 2003;17:69–73.
- 9 Finkeissen E, Fuchs H and Jakob T. medrapid aktuelles medizinisches wissen auf knopfdruck [up-to-date medical knowledge at your fingertips]. *Telemedizinführer Deutschland* [Telemedicine Guide Germany] 2003:2–5.
- 10 Deutsches Institut f
  ür Medizinische Dokumentation und Information [German Institute of Medical Documentation and Information]. Cologne: Deutscher Ärzte-Verlag, 2003.
- 11 Burkle T, Ammenwerth E, Prokosch HU and Dudeck J. Evaluation of clinical information systems. What can be evaluated and what cannot? *Journal of Evaluation in Clinical Practice* 2001;7:373–85.
- 12 Probst R, Grevers G and Iro H. Hals-Nasen-Ohren-Heilkunde. [Otorhinolaryngology]. Stuttgart, New York: Thieme, 2004.
- 13 Fleck L. Erfahrung und Tatsache: gesammelte aufsätze [Experiences and Facts: collected essays]. Frankfurt am Main: Suhrkamp, 1983.
- 14 Wetter T. Lessons learnt from bringing knowledgebased decision support into routine use. *Artificial Intelligence in Medicine* 2002;24:195–203.



#### CONFLICTS OF INTEREST

None.

#### ADDRESS FOR CORRESPONDENCE

Ekkehard Finkeissen Institute for Medical Biometry and Informatics Department of Medical Informatics University of Heidelberg Im Neuenheimer Feld 400 69120 Heidelberg Germany Email: Ekkehard.Finkeissen@medrapid.info

Accepted September 2005