Pro-active medical information retrieval

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can
be important differences between the submitted version and the official published version of record.
People interested in the research are advised to contact the author for the final version of the publication,
or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page
numbers.

Link to publication

Pro-Active Medical Information Retrieval

Pro-Active Medical Information Retrieval

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus, prof. mr. G.P.M.F. Mols,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op woensdag 29 oktober 2008 om 14.00 uur

door

Loes Margaretha Marie Braun

UM
UNIVERSITAIRE
PERS MAASTRICHT

Promotores

Prof. dr. H.J. van den Herik (Universiteit van Tilburg / Universiteit Maastricht)
Prof. dr. ir. A. Hasman (Universiteit van Amsterdam)

Copromotor

Dr. F.J. Wiesman (Universiteit van Amsterdam)

Beoordelingscommissie

Prof. dr. E.O. Postma (Voorz., Universiteit van Tilburg / Universiteit Maastricht)
Prof. dr. H.H.M. Korsten (Technische Universiteit Eindhoven)
Prof. dr. G.G. van Merode
Prof. dr. ir. J.A. La Poutré (Technische Universiteit Eindhoven)
Dr. J.L. Talmon

Netherlands Organisation for Scientific Research

Cover design: Ilja Maas

Preface

This thesis is on pro-active medical information retrieval. In my research period, I have investigated how to support physicians in their information-retrieval process. I focused on two specific obstacles in medical information retrieval: the inadequate expression of information needs, and the time-consuming nature of the information-retrieval task. I implemented my solutions to these obstacles in a medical information-retrieval agent which was evaluated by physicians. My main findings are that the obstacles can be overcome by formulating a physician's information needs automatically (even if the information needs are implicit or suppressed) and by using these information needs as a starting point for the retrieval of relevant and patient-related literature. Over the years, I learned many things, which made the period an invaluable and particularly interesting experience. Yet, this experience would not have been possible without the support of many other people, which I would like to acknowledge below.

First and foremost, I would like to express my sincere gratitude towards my supervisors Jaap van den Herik and Arie Hasman. To Jaap I am grateful for his guidance, his encouragements, and his everlasting optimism. He has always been able to motivate me, and the many useful lessons he taught me extend far beyond the scope of this thesis. I am especially grateful for his support and for the independence he allowed me over the past year, which enabled me to combine my research with my medical education. To Arie I am grateful for his support, his enthusiasm, and his constructive criticism, particularly in the medical domain. The feedback and the many useful suggestions he provided, among others with respect to my experiments, have greatly improved the quality of the thesis. Furthermore, I would like to thank him for arranging valuable contacts with physicians and other experts in the medical field.

I am greatly indebted to my daily advisor Floris Wiesman. Despite the physical distance between Maastricht and Amsterdam, I could always contact him with my questions concerning research or writing. He never failed to answer them or to provide highly useful advice. Furthermore, I have enjoyed our many conversations on research and beyond. Without him, my thesis would not exist in its present form.

Moreover, I would like to thank my colleagues from the MICC-IKAT department and from the other departments. I would like to single out Femke de Jonge, Agnieszka Latoszek-Berendsen, Ivan Vermeulen, Linda Peelen, and Stefan Visscher. Then, I have had a pleasant cooperation with Pieter Spronck, Evgueni Smirnov,

Nico Roos, Marius Nap, and Barend Mons. I would also like to thank the physicians who participated in the evaluation of our system MIRA.

Special words of thanks go to the people from the secretariat: Joke Hellemons, Tons van den Bosch, Liesbeth Nederlands, Els van Aernsbergen, Hazel den Hoed, and Marlies van der Mee, and to our system administrator Peter Geurtz. Moreover, I am grateful to my *paranimfs*, Maaike Rietrae and Wesley Jongen.

I would like to express my gratitude towards my parents for their unconditional support and everlasting faith in me. They have taught me self-confidence and perseverance, which were indispensable when writing this thesis. But most of all, they have always encouraged me to pursue my goals in life. I can only aspire to be as supportive to them as they have always been to me. It fills me with great pride to be able to call them my parents.

My final words of gratitude are for the person who deserves them the most: Ron. His everlasting encouragement, understanding, and reassurance are unequalled. He made me realize the value and the importance of my work, and even while finishing his own thesis he was continuously focusing on mine as well. The period in which we both wrote our doctoral dissertations was extremely busy, but he succeeded in making it a particularly enjoyable and memorable experience at the same time. I will always cherish his confidence, dedication, and unconditional support. I am exceptionally grateful and proud to have him at my side.

Loes Braun
Leiden, Fall 2008

Acknowledgements

Contents

List of Abbreviations

The list below contains all abbreviations used in the thesis. Normal lexical abbreviations, for instance 'e.g.' and 'i.e.', are not listed. Abbreviations used only in tables or figures are explained in the corresponding table or figure.

AMR . . . . Automated Medical Record
AUI . . . . Atom Unique Identifier
CICS . . . . Clinical Informatics Consult Service
CWI . . . . Centre for Mathematics and Computer Science
CMR . . . . Computerized Medical Record
CUI . . . . Concept Unique Identifier
EBM . . . . Evidence-Based Medicine
EHR . . . . Electronic Health Record
EMR . . . . Electronic Medical Record
EPR . . . . Electronic Patient Record
ICD . . . . International Classification of Diseases
ICD-CM . . . . International Classification of Diseases, Clinical Modification
ICIS . . . . Intensive Care Information System
ICU . . . . Intensive Care Unit
IR . . . . Information Retrieval
LUI . . . . Term Unique Identifier
MESH . . . . Medical Subject Headings
MIA . . . . Medical Information Agents
MICC . . . . Maastricht ICT Competence Centre
MIRA . . . . Medical Information Retrieval Agent
MPAEX . . Multi-agent Pareto-Improvement Appointment Exchanging Algorithm
NCBI . . . . National Center for Biotechnology Information
NLM . . . . National Library of Medicine
NWO . . . . Netherlands Organisation for Scientific Research
PS . . . . Problem Statement
PSY . . . . Psychological Index Terms
RQ . . . . Research question
SNOMED . . . . Systematised Nomenclature of Medicine
SQL . . . . Structured Query Language
SUI . . . . String Unique Identifier

List of Tables

List of Figures

Chapter 1

Introduction

In this thesis we investigate how to support physicians in the information retrieval (IR) process. It is our position that tailored IR support provides physicians with point-of-care information and improves the quality of care.

This introductory chapter provides an overview of our research. In Section 1.1 we introduce the research area of IR and in particular medical IR. Section 1.2 discusses the Electronic Medical Record (EMR) which is an essential component of our approach. In Section 1.3 we present our problem statement and list five research questions. Section 1.4 describes the research methodology adopted in our research. In Section 1.5 we introduce the requirements of MIRA, the IR system we developed. Section 1.6 discusses the project of which our research constitutes a part. In Section 1.7 we provide the structure of the thesis. We remark that we number the definitions, examples, information needs, templates, etc. per chapter (i.e. not per item) in an increasing order.

1.1 Medical Information Retrieval

In the early 1940s, the field of IR started to gain interest from scientific researchers. Automated IR emerged about 20 years later, in the 1960s (Swanson, 1960; Swets, 1963). One of the domains in which IR gained particular interest was the medical domain (Ledley and Lusted, 1960; Nanus, 1960). Consequently, the field of medical IR was born. We define medical IR as follows.

Definition 1.1 (Medical IR) *Medical IR is IR that is completely focused on and applied to the medical domain.*

A schematic representation of an IR system is shown in Figure 1.1. The figure shows that the incentive of the user to employ the IR system is the existence of an information need. In this thesis, we define an information need as follows.
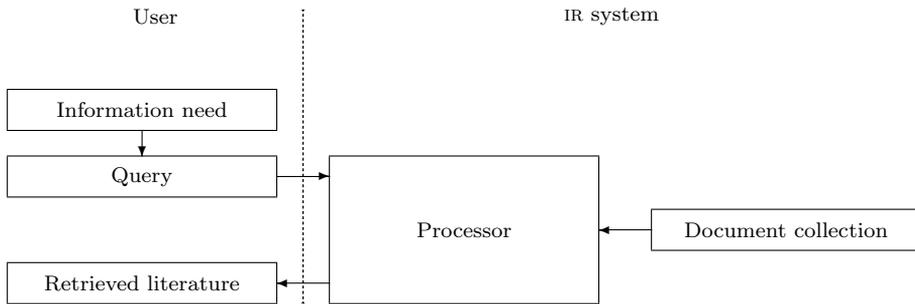
Figure 1.1: Schematic representation of an IR system. The figure is adopted from Van Rijsbergen (1979, p. 4) and slightly adapted.

Definition 1.2 (Information need) *An information need is a statement or question, the answer to which*
(a) is needed to perform a particular task, and
(b) is missing.

When applied to the medical domain, an information need always represents a need for medical information (e.g., the cause of a certain disease or the side effects of a certain medication).

In a conventional IR system, the user has to interact actively with the system by formulating an information need. However, for reasons elaborated upon in Subsection 1.1.2, in the medical domain the interaction between the physician and the IR system should be kept to a minimum. This thesis provides an approach to ensure minimal interaction by *automatic* formulation of information needs.

Furthermore, Figure 1.1 shows that the information need has to be transformed into a query. We define a query as follows.

Definition 1.3 (Query) *A query is an information need that is translated into the language that is understood by the processor of the IR system.*

When using a conventional IR system, the user himself has to translate the information need into a query. However, as with the formulation of information needs, this is not suitable in the medical domain (see Subsection 1.1.2). In the approach presented in this thesis, information needs are translated into queries automatically.

Moreover, we see that a document collection is needed to provide a user with appropriate literature. In medical IR the employed document collection always consists of medical literature (e.g., medical journal articles and medical books). In the medical domain, the most-frequently used collection is MEDLINE (McCarn and Leiter, 1973), which contains literature abstracts and for each abstract a reference to the corresponding document. MEDLINE is part of PubMed, an archive that is provided by the U.S. National Library of Medicine (NLM). At the time of writing, PubMed contains over 17 million citations from medical literature, mostly from

Table 1.1: Quantification of medical errors.

	NL	EU	USA
Preventable errors	30,000	1,000,000	1,000,000
Deaths	3000	100,000	98,000
Handicaps	750	25,000	NA
Reversible handicaps	26,000	900,000	NA

Note: Data concerning the Netherlands and the European Union are based on NeVeMeDis (2004). Data concerning the United States of America are based on Kohn, Corrigan, and Donaldson (2000). Data on handicaps in the USA are not available (NA).

MEDLINE.[1] For each literature citation, PUBMED stores a record containing metadata concerning the citation. The metadata comprises, for instance, the author, the title, the publication date, and the abstract of the cited document. The literature citations in PUBMED are indexed by Medical Subject Headings (MeSH)[2] (NLM, 2007), which are also part of the metadata. MeSH is a hierarchical vocabulary comprising over 24,000 biomedical terms. It was developed by the NLM for use with PUBMED. The PUBMED archive is searchable via ENTREZ,[3] the IR system used at the National Center for Biotechnology Information (NCBI) to search various biomedical databases.

In the remainder of this section, we discuss the importance of medical IR (Subsection 1.1.1) and the problems in medical IR (Subsection 1.1.2). Furthermore, we elaborate on evidence-based medicine (Subsection 1.1.3) and we survey various existing medical IR systems (Subsection 1.1.4).

1.1.1 The Importance of Medical Information Retrieval

Over the past years, the number of medical errors has reached alarming heights (Table 1.1). In the Netherlands alone, the number of preventable medical errors is $30,000$ per year. Of these errors, $26,000$ (87%) result in reversible handicaps, 750 (2.5%) result in irreversible handicaps, and 3000 (10%) results in death (NeVeMeDis, 2004). In the European Union and the United States of America, the numbers are approximately 30 times higher (Kohn *et al.*, 2000). Over the years to come, these numbers are expected to climb.

There is evidence that the use of information technology may reduce the number of medical errors (Koshy, 2005; Belmont and Waller, 2003; Kopec *et al.*, 2003). The information technology may be employed in a variety of ways to improve the quality of care. One direction that has received particular attention is the reduction of medication errors and adverse drug effects (Anderson *et al.*, 2003; Anderson, 2004). A second direction is the improvement in communication between physicians (Singh *et al.*, 2008). The focus direction of this thesis is to provide physicians with digital access to professional medical literature.

---

[1] http://www.ncbi.nlm.nih.gov/pubmed/
[2] http://www.nlm.nih.gov/mesh/
[3] http://www.ncbi.nlm.nih.gov/sites/entrez

Various studies have proven that the retrieval of relevant, patient-related literature is vital to the quality of care. Literature is considered *patient related* if the information in the literature is directly applicable to the specific patient case at hand. According to Barnett (2006), the use of IR systems will lessen the burden of routine medical tasks and grant physicians more time for patient care, thus greatly improving the quality of care. According to Gamble (1996), the use of medical literature improves the quality of care and reduces the patient costs. According to Klein *et al.* (1994), the employment of medical literature in specific patient cases leads to (a) lower costs, (b) lower charges, and (c) shorter length of stay.

To illustrate the importance of medical IR we provide an example.

Example 1.4 *An 84-year-old woman was brought into the emergency department of a hospital, suffering from dyspnea (shortness of breath) and loss of consciousness. According to her daughter she had been feeling ill for seven days and she visited her general practitioner five days ago. The general practitioner diagnosed her with suspected respiratory infection and described a drug called Clarithromycin. However, instead of improving, her condition worsened. Upon arrival at the hospital several tests were performed and finally the diagnosis considered was pneumonia. Other possible diagnoses were cardiac failure, pulmonary embolism, or a recent stroke. Upon her family's request, the patient was not admitted to the intensive care unit and so she died one day after she was admitted to the hospital. An autopsy was performed and revealed that the cause of death was neither pneumonia nor cardiac failure, but a case of severe acute pancreatitis. The autopsy also revealed that the most plausible cause for the pancreatitis was the use of Clarithromycin, since pancreatitis is a (rare) side effect of the use of Clarithromycin (Schouwenberg and Deinum, 2003).*

Since the incidence of Clarithromycin-induced pancreatitis is quite low, it is understandable that the physician in the example above was not aware of this possible side effect. However, if the physician had performed a literature search on the side effects of Clarithromycin, he[4] probably would have found an article by Leibovitch, Levy, and Shoenfeld (1996), in which another case of Clarithromycin-induced pancreatitis is discussed. If he had read this article and had acted accordingly, i.e., ceasing the use of Clarithromycin, it is probable that the patient would have survived.

A second possible way to avoid the situation in Example 1.4 is the use of an *alert system* (Raschke *et al.*, 1998; Oppenheim *et al.*, 2000; Degnan, Merryfield, and Hultgren, 2004; Furman, 2005). Alert systems forwarn physicians with respect to harmful situations (e.g., drug interactions and side effects). To this end, alert systems employ well-established medical knowledge. However, according to Balas and Boren (2000), it takes 15 to 20 years for new medical discoveries to become established medical knowledge and to be incorporated into routine patient care. Consequently, since alert systems rely on well-established medical knowledge, they are always lagging behind. As a result, alert systems cannot handle situations in which up to date knowledge is required. Since IR systems rely on medical literature,

---

[4]For brevity we will use the pronoun 'he' ('his') where 'he or she' ('his or her') is meant.

P$_1$: Unawareness of the
information needs

P$_2$:   Large amount of
medical literature

↓

↓

P$_3$: Inadequate formulation of
information needs

P$_4$:   Time-consuming nature
of the retrieval task

P$_5$: Inadequate information retrieval

↓

P$_6$: Lack of relevant information

↓

P$_7$: Medical errors
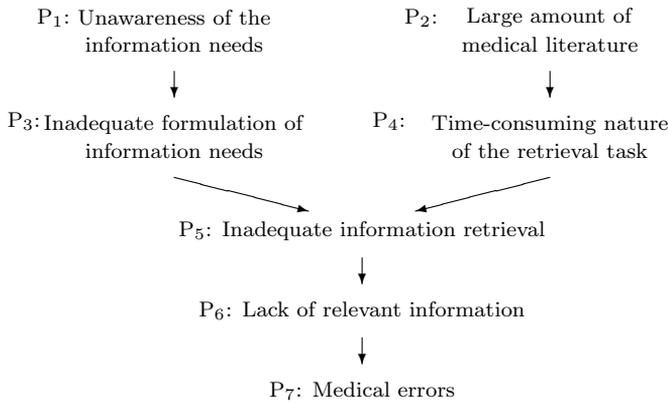
Figure 1.2: The problem bundle of our research domain. The arrows indicate causal relationships between the phenomena.

which is up to date, IR systems are useful in situations where up to date knowledge is required.

A new technique in information technology that has recently received much attention is *question answering* (Westbrook, Coiera, and Gosling, 2005; Lee *et al.*, 2006a). In this field, users are provided with direct answers to their questions, instead of documents in which answers may be found. However, in this thesis, we concentrate on the first step of the IR process: the retrieval of relevant and patient-related documents. Filtering these documents for direct answers to a specific question may be an interesting direction for future research.

1.1.2   Problems in Medical Information Retrieval

Example 1.4 clarifies the importance of medical IR. Moreover, it implicitly illustrates several phenomena in the medical domain that obstruct medical IR. Seven important phenomena are discussed below.

In Wieringa and Heerkens (2003), the notion of *problem bundle* is introduced and we would like to adopt this idea (see Figure 1.2). We note in passing that instead of the notion *problem*, we prefer the use of the notion *phenomenon*, since not every phenomenon needs to be a problem. The reader is invited to help us examining when a phenomenon (or a combination of phenomena) is a problem and when it is not a problem. The problem bundle contains seven phenomena ($P_1 - P_7$). An arrow from P_x to P_y indicates a causal relationship between P_x and P_y; P_x causes P_y $(x, y = 1, ..., 7)$.

P$_1$ Unawareness of the information needs

In many cases physicians are unaware of their information needs. According to Koopmans and Offringa (2003), physicians claim to have one information need per week, whereas they actually have seven to eight questions per day.

P₂ Large amount of medical literature

The quantity of medical literature is immense. ENTREZ alone publishes references to over $34,000$ journals each year (i.e., over one million journal articles).[5] Due to the large amount of medical literature, it is hard to find specific information.

P₃ Inadequate expression of information needs

Due to the physicians' unawareness of their information needs, many information needs cannot be appropriately expressed. In example 1.4, the physician could not express his information need because he was not aware of it. The fact that, in most cases, information needs have to be expressed in a specific query language further hampers the appropriate expression of information needs (Ter Hofstede, Proper, and van der Weide, 1996).

P₄ Time-consuming nature of the IR task

Due to the large amount of medical literature, the time needed to find relevant, patient-related literature is greatly enlarged. In most cases, the retrieval process takes more time than the physician can spare. According to Merry (1997), a retrieval time of more than 30 seconds is not acceptable for physicians.

P₅ Inadequate information retrieval

Due to the unawareness of their information needs and the time constraints, in many cases physicians do not perform IR. In example 1.4 no IR was performed because the physician could not explicitly express his information need concerning the side effects of Clarithromycin.

P₆ Lack of relevant information

Due to the inadequate formulation of information needs, physicians have insufficient literature at their disposal (cf. O'Leary, 2000; Zhang, Patel, and Johnson, 2002). In example 1.4, the fatal side effect pancreatitis is not recognized because (a) the physician was not familiar with the side effect and (b) no literature on the subject was available to the physician.

P₇ Medical errors occur too often

Due to insufficient use of the medical literature, medical errors occur. As stated in Table 1.1, in the Netherlands, $30,000$ preventable medical errors are made each year, whereas in the European Union and in the United States of America, the number of medical errors amounts to $1,000,000$. It is not known exactly how many of these errors are caused by a lack of information (*P6*). An instance of a medical error due to a lack of information is provided in Example 1.4, where a patient dies because a fatal side effect which is not recognized.

The problem bundle can be used to select our *core problems*, i.e., those phenomena that we are going to investigate in our research (cf. Wieringa and Heerkens, 2003). According to Wieringa and Heerkens there are four criteria to select proper core problems.

---

[5]ftp://ftp.ncbi.nih.gov/pubmed/J_Entrez.txt

1. Choose phenomena close to the root,[6] preferably the root phenomena them-selves.

2. Choose phenomena which actually can be changed.

3. Choose phenomena which require actual research.

4. Choose a reasonable number of phenomena.

According to criterion (i), P1 and P2 should be our core problems, as they are the root phenomena. However, P1 and P2 are pre-defined characteristics of our research domain, that cannot be changed by our research. Consequently, P1 and P2 do not comply with criterion (ii). The phenomena closest to P1 and P2 are P3 and P4. They are now closest to the root phenomena, they can be changed, and they need actual research. Therefore, these phenomena comply with criteria (i), (ii), and (iii). Since we selected only two phenomena, we also comply with criterion (iv). As a result, we selected P3 and P4 as our core problems.

1.1.3   Evidence-Based Medicine

As stated above, the quantity of medical literature is immense. We mentioned already that each year, some 34,000 medical journals (i.e., over one million journal articles) are estimated to appear. The quality of these publications differs considerably. Nevertheless, physicians are expected to be familiar with all relevant information, which is an arduous task. A direct consequence is that the principle of evidence-based medicine (EBM) has become increasingly important over the past decades. According to this principle, a patient's medical information is combined with data obtained from clinical research in order to determine the best treatment for the specific patient (Offringa, Assendelft, and Scholten, 2003).

The application of evidence-based medicine comprises five steps (Offringa *et al.*, 2003).

1. Translating the clinical case into an answerable question.

2. Searching efficiently for the best evidence.

3. Weighing the evidence found, based on methodological quality results and on its applicability in the current situation.

4. Making a decision based on the available evidence.

5. Evaluating the quality of the process regularly.

As Offringa *et al.* (2003) mention, specific skills are required to apply these steps successfully. In this respect, two types of skills can be distinguished: IR skills (steps 1 and 2) and clinical skills (steps 3 to 5). Consequently, it is clear that IR plays an important role within the application of evidence-based medicine.

---

[6]The root is the origin of the sequence of causes.

Table 1.2: Types of information needs.

	Recognized	Not recognized
IR is performed	Explicit	–
IR is not performed	Suppressed	Implicit

According to Koopmans and Offringa (2003) a physician's information needs have two specific features: (a) whether they are recognized (i.e., a physician is aware of his information needs) and (b) whether IR is performed. Based on these two features, three kinds of information needs can be distinguished, viz. explicit, suppressed, and implicit information needs (see Table 1.2) (Ely *et al.*, 2002; Koopmans and Offringa, 2003). Obviously, information needs that are not recognized cannot be used as a starting point for IR. Consequently, the combination 'not recognized' and 'IR is performed' does not apply.[7] Below we provide our definitions of the three notions introduced above.

Definition 1.5 (Explicit information need) *An explicit information need is an information need that is recognized and for which IR is performed.*

An explicit information need in itself does not constitute a problem within medical IR. Since they are recognized and used as a starting point for IR, the physician is able to employ the newly obtained information directly in his patient cases.

Definition 1.6 (Suppressed information need) *A suppressed information need is an information need that is recognized but not used as a starting point for IR. This is due to time pressures, embarrassment, personal characteristics, or characteristics of the clinical setting (based on Ely* et al.*, 2002).*

According to Ramos, Linscheid, and Schafer (2003), residents have a larger tendency to suppress their information needs than, for instance, faculty physicians. Suppressed information needs are potentially problematic, since their existence implies a knowledge gap. They are mainly caused by the large amount of medical literature and the time-consuming nature of the retrieval task (cf. P1 and P3 in Figure 1.2). Consequently, an IR system may turn them into explicit information needs by reducing the time needed to perform IR.

Definition 1.7 (Implicit information need) *An implicit information need is an information need that is not recognized and consequently is not used as a starting point for IR.*

Implicit information needs are the most problematic type of information need. They are mainly caused by the physician's unawareness of his information needs and the

---

[7]In some cases, it might be possible that an unrecognized information need is answered by documents retrieved in response to another information need. Such an information need does not pose a problem to a physician.

resulting lack of formulation of information needs (cf. P2 and P4 in Figure 1.2). Consequently, an IR system may turn implicit information needs into explicit ones by both improving the formulation of information needs and by reducing the time needed to perform IR. If an IR system had been available in Example 1.4, the fatal outcome might have been avoided.

1.1.4   Existing Medical IR Systems

In the past two decades, various medical IR systems have been developed to overcome the IR problems in the medical domain (e.g., Miller *et al.*, 1991; Rada *et al.*, 1992; Van Mulligen, 1999; Wiesman, van den Herik, and Hasman, 2004). Systems resembling our approach are the so-called info buttons. According to Maviglia *et al.* (2006), info buttons are 'message-based content search and retrieval functions embedded within other applications that dynamically return information relevant to the clinical task at hand'.

Maviglia *et al.* (2006) mention a number of info-button systems. The first two are HEPATOPIX (Powsner *et al.*, 1989) and its extension PSYCHTOPIX (Powsner and Miller, 1992). HEPATOPIX enables a physician to use specific electronic patient data as a starting point for literature retrieval. For a specific patient case, PSYCHTOPIX automatically generates a number of topics that might be interesting to the case. However, the physician using the system still has to indicate which topics are to be used for a literature search. PSYCHTOPIX uses the same approach, but extends HEPATOPIX into the psychiatric domain.

Furthermore, Maviglia *et al.* (2006) mention various versions of MEDLINE BUTTON, a system developed by Cimino *et al.* (1992). The first version of MEDLINE BUTTON translates patient data from ICD9-CM (a certain medical terminology) into MeSH. The MeSH terms are used as a starting point for literature retrieval (Cimino *et al.*, 1992). However, a physician has to indicate explicitly which terms are to be used for literature retrieval. In INFOBUTTON, the second version of MEDLINE BUTTON, their own Medical Entities Dictionary is used to combine patient data, integrate them into information needs, and enhance them with terminological knowledge (Cimino, Elhanan, and Zheng, 1997). As with the previous version of MEDLINE BUTTON, a physician has to indicate which information needs he wants to use as a starting point for literature retrieval. INFOBUTTON MANAGER, the third version of MEDLINE BUTTON, anticipates a physician's information needs and constructs an button for each information need (Cimino *et al.*, 2002). Still, the physician has to indicate whether he would like to be provided with literature concerning a specific information need.

From a recent study by Cimino (2006), it may be concluded that the frequency of use of the MEDLINE BUTTON depends greatly on the context of the questions, and is rather low in some contexts. We argue that this is due to implicit information needs. In each of the three versions of MEDLINE BUTTON, the incentive for literature retrieval lies with the physician using the system. However, if the physician has implicit information needs, he will not recognize them, and consequently, he will not acknowledge the need for literature retrieval. Therefore, we claim that literature

should be retrieved by the system autonomously and pro-actively, without explicit action from the user.

1.2   Electronic Medical Record

According to Haas Binder *et al.* (1997) there are five different types of computer-ized medical records. These types can be distinguished based on three features, viz. (a) legality, (b) accessibility, and (c) the editor. The legality indicates to what ex-tent a computerized medical record complies with the legal requirements for medical records (NICTIZ, 2007). The legal requirements primarily concern the privacy of the patient data. The accessibility indicates to what extent the medical record is accessible. For instance, some medical records are accessible by physicians from one specific medical department only, whereas others are accessible by health care work-ers from various medical institutions (e.g., physicians, dentists, and pharmacists). The editor indicates who has a right to edit the information in the medical record. The five types of medical records discussed by Haas Binder *et al.* are presented be-low and summarized in Table 1.3 together with a particular instance, the Intensive Care Information System (ICIS). For an explanation of the latter, see below.

Automated Medical Record   Health-care workers within one department use an Automated Medical Record (AMR) as a working file. However, since the AMR is stored on a standard personal computer, the privacy requirements are hard to fulfil. Since patient privacy constitutes a major part of the legal requirements of a medical record, an AMR does not fulfil the legal requirements for medical records. Additionally, the documents in the AMR are still printed and filed on paper.

Computerized Medical Record   A Computerized Medical Record (CMR) is used by health-care workers to store scanned versions of paper documents. Analo-gously to the AMR, a CMR is used within one specific department. In contrast to the AMR, the CMR complies to some of the legal requirements for medical records.

Electronic Medical Record   An Electronic Medical Record (EMR) stores medical data from various medical departments. There are two distinctions (1) be-

Table 1.3: Six types of the medical record and their properties.

Record type	Legality	Accessibility	Editor
AMR	None	Medical department	Health care workers
CMR	Some	Medical department	Health care workers
EMR	Full	Medical enterprise	Health care workers
EPR	Full	All medical enterprises	Health care workers
EHR	Full	Everywhere	Health care workers / Patient
ICIS	Full	Medical department (ICU)	Health care workers

tween an EMR and an AMR and (2) between an EMR and a CMR. The first distinction is that an EMR is accessible from multiple departments within a medical enterprise. The second distinction is the fact that an EMR fulfils all legal requirements for medical records. At the time of writing (2008), this type of medical record is the most advanced medical record in existence.

Electronic Patient Record An Electronic Patient Record (EPR) is in essence similar to an EMR. The major difference is the presence in the EPR of information from various medical institutions (e.g., hospitals and dentists), accessible from each of them.

Electronic Health Record An Electronic Health Record (EHR) stores *all* medical information concerning a specific patient, even information provided by the patient himself. In contrast to the other types of medical records, the patient is in charge of part of this information, so the health-care workers are not the only persons responsible.

In our research we used a specific EMR system: the Intensive Care Information System (ICIS), used at the Intensive Care Unit (ICU) of the Catharina-ziekenhuis in Eindhoven.[8] In the field of intensive care medicine and anaesthesiology, the use of information retrieval has proven beneficial to the quality of care (Berkenstadt *et al.*, 2006; Catchpole, Bell, and Johnson, 2008). This system complies with all legal requirements for medical records, can be accessed only from the ICU, and is maintained by the hospital workers at the ICU (see Table 1.3). Based on these properties, the ICIS fits in none of the five categories discussed above. However, since the ICIS has most properties in common with the EMR, we will address it as an EMR in the remainder of the thesis. The ICIS contains a multitude of data for each patient. The most important data types and their occurrence are summarized in Table 1.4.

---

[8]Intensive Care Informatie Systeem, Version 2.8. INAD Computers & Software B.V. Eindhoven, Werkgroep ICIS Afd. Intensive Care, Dienst Informatie Voorziening, Catharina-ziekenhuis Eindhoven.

Table 1.4: Most important data types in the ICIS.

Data type	# Possible values	# Entries
Activities	138	729
Allergies	21	1155
Treatment	157	996
Complications	184	81
Indication for admission	476	181
Medication	475	1564
Tests	139	458
Comorbidity	51	116

1.3   Problem Statement and Research Questions

In Subsection 1.1.2 we chose two medical problems as our core problems: (a) the inadequate formulation of information needs (P3) and (b) the time-consuming nature of the retrieval process (P4). In this section, we present our problem statement which we inferred from the abovementioned problems. Furthermore, we derive five research questions (RQs) from the problem statement.

Our goal is to develop an IR system that solves the core problems mentioned above. Various articles have been written in which medical IR systems are discussed (Subsection 1.1.4). Our research concurs with these articles. However, in our opinion the overall shortcoming of the systems mentioned in the articles is that the degree of necessary interaction with the systems is too high. This is particularly true in the area of formulating implicit and suppressed information needs (cf. P3). Therefore, our first research objective is to investigate whether this step can be performed automatically. Our second research objective is to investigate whether relevant, patient-related literature can be retrieved based on the automatically formulated information needs (cf. P4). Consequently, our twofold problem statement (PS) is stated as follows.

Problem statement (PS):

PS1: *To what extent can a physician's information needs – implicit and suppressed – be formulated automatically?*

PS2: *To what extent can the automatically formulated information needs be used as a starting point for the retrieval of relevant and patient-related literature?*

PS1 gives rise to three research questions (RQ1, RQ2, and RQ3), PS2 gives rise to two research questions (RQ4 and RQ5). All five research questions are discussed below.

To formulate a physician's information needs automatically, a model of his information needs is a necessary starting point. However, since such a model does not exist, we have to compose a model. We do so as part of our research. The construction of such a model leads to our first research question.

Research question 1 (RQ1): *How can the information needs of a physician be modelled?*

Once a model of a physician's information needs has been constructed, the model may be used to formulate appropriate information needs. An information need is only appropriate if it is related to a specific patient and relevant to the physician treating the patient. The formulation of patient-related information needs gives rise to our second research question.

Research question 2 (RQ2): *How can patient-related information needs be formulated?*

The automatic formulation of patient-related information needs is based on the use of patient data. However, in the medical domain there is only little consensus on terminologies. Therefore, the patient data used have to be mapped onto the terminology (or terminologies) employed in the literature collection on which retrieval is performed. The terminology mapping results in our third research question.

Research question 3 (RQ3): *How can different medical terminologies be mapped?*

Once patient-related and relevant information needs have been formulated they may be used as a starting point for literature retrieval. However, care must be taken: a physician should not be overloaded with literature. Otherwise, the physician would have to spend a large amount of time on reading the retrieved literature. As a result we would not succeed in solving core problem P4 (time-consuming literature retrieval). The prevention of a literature overload leads to our fourth research question.

Research question 4 (RQ4): *How can a literature overload be prevented?*

After the relevant and patient-related literature has been retrieved, the literature has to be presented to the physician. However, the manner of presentation is crucial to the usability of our IR system. If our system is to be used in practice, it must be unobtrusive. Otherwise, it impedes the physician's daily workflow. The unobtrusiveness requirement raises our fifth research question.

Research question 5 (RQ5): *How can the IR system be designed to be unobtrusive?*

In order to evaluate our approach physicians would have to assess the information needs generated. However, since the information needs are implicit, physicians will probably not be aware of the fact that they have these information needs. A more appropriate method to evaluate our approach is to let physicians assess the information which is retrieved based on the generated information needs.

1.4   Research Methodology

A research methodology is an essential element of every research project (Weigand *et al.*, 2004). Within our research we use various methodologies. We discuss the adopted methodology per research question and do so too for the evaluation (summarized in Table 1.5).

To answer RQ1, we will identify what physicians' potential information needs are. We note that at this stage, the identified information needs are related to the

Table 1.5: Research methodologies used for answering the research questions and design questions and for evaluating MIRA.

	RQ1	RQ2	RQ3	RQ4	RQ5	Evaluation
Literature survey	●		●		●	
Interviews	●					
Analysis	●					
Abstraction	●					
Refinement	●					
Software experiments		●	●	●	●	●
User experiments						●

physician but not necessarily to the patient. To identify the information needs we will use two methods: a literature survey and interviews with physicians. Then, we will build a model based on the outcome of the identification. To model the information needs, we will use three methods: analysis, abstraction, and refinement. Eventually, answering the first research question will result in a model of physicians' information needs. The methods used for answering RQ1 correspond to the second column in Table 1.5.

To answer RQ2, we will determine how to use the EMR to check whether an information need applies to a specific patient. To investigate this, we will perform software experiments. Then, we will examine how to check whether the context of the information in the EMR is correct. To this end, we will perform software experiments as well. The assessment will result in a method to check whether an information need applies to a patient. The methods used for answering RQ2 correspond to the third column in Table 1.5.

RQ3 has already been thoroughly investigated over the past years. Since the emphasis of our research is on the automatic formulation of information needs instead of terminology mapping, we will mainly use literature surveys to find existing solutions to answer RQ3. Additionally, we will use software experiments to fine-tune MIRA to the specific application domain. The methods used for answering RQ3 correspond to the fourth column in Table 1.5.

To answer RQ4, we will design a variety of methods and use software experiments to assess them. The best method will be used as our final method to prevent a literature overload. The methods used for answering RQ4 correspond to the fifth column in Table 1.5.

To answer RQ5 we will employ the same course of action as with research question RQ3. We will develop a results-presentation method based on insights obtained from literature and we will test the method using software experiments. The methods used for answering RQ5 correspond to the sixth column in Table 1.5.

To evaluate the performance of our IR system , we will conduct two types of experiments. First, we will conduct a number of software experiments. These experiments will mainly be used to choose between various possible techniques, or to fine-tune a specific technique. Second, we will conduct user experiments. Here, we will provide users with a patient case and the literature retrieved by MIRA with

respect to that patient case. Then, we will measure various formal performance criteria as well as the physicians' satisfaction with MIRA. The methods used for evaluating our approach correspond to the last column in Table 1.5.

1.5   MIRA: Medical IR Agent

The IR techniques answering our research questions are incorporated in an *agent system*, since it had to meet two specific requirements, viz. (a) autonomy and (b) pro-activeness (Jennings, 2000).

The first requirement, autonomy, is met by the ability of the system to retrieve information without interacting directly with the user. The trigger for the system to start information retrieval is the entry of patient data into the EMR by the physician. When performing this task, no additional effort is required of the physician beyond his regular data entry activity, ensuring complete autonomy of the IR system. The only direct interaction of the physician with the system occurs when the physician reads the literature retrieved by the system.

The second requirement, pro-activeness, is met by the fact that MIRA is able to provide the physician with literature even before the physician can acknowledge his information needs. In that way, MIRA is always one step ahead of the physician's information needs.

1.6   The MIA Project

The MIA project[9] is part of the research programme ToKeN, which is funded by the Netherlands Organisation for Scientific Research (NWO). The project was started in 2002. The goal of the project is to investigate how physicians can be supported in their daily work by relieving them from routine tasks and providing them with relevant information from structured and unstructured medical sources. The project comprises three separate parts performed at three distinct institutions.

The first part concerns automatic feedback to physicians if they perform actions that are not in line with medical practice guidelines. The research is carried out at the Department of Medical Informatics of Maastricht University in Maastricht, The Netherlands. Several computer systems have been developed that alert physicians when they perform actions that are not in line with the current medical practice guidelines. However, most of these systems are rather inflexible, alerting physicians unnecessarily if they stray only little from the guideline. Latoszek-Berendsen, Talmon, and Hasman (2006) have developed a system that assesses a physician's actions by its intentions, rather than by the actions itself. The proposed methodology was implemented and evaluated for a heart failure guideline (Latoszek-Berendsen *et al.*, 2007). From the evaluation it transpired that the approach makes a decision support system more flexible.

---

[9]NWO grant number 634.000.021.

The second part concerns automatical patient scheduling in order to improve patient flow through a hospital. The research is performed at the Center for Mathematics and Computer Science (CWI) in Amsterdam, The Netherlands. According to Vermeulen *et al.* (2006), the waiting lists for medical procedures in Dutch hospitals are exceptionally long. A promising solution to this problem is automatic patient scheduling. Because of the decentralized nature of hospital departments, a multi-agent approach to patient scheduling performs better than a more traditional and centralized approach (Vermeulen, Bohte, and La Poutré, 2007a). Consequently, Vermeulen *et al.* developed MPAEX: a multi-agent Pareto-improvement appointment exchanging algorithm. MPAEX is a dynamic and adaptive approach to patient scheduling and resource calendar management which proved to be both effective and efficient (Vermeulen *et al.*, 2007b; 2007c).

The third part concerns the research presented in this thesis. The research was carried out at the Maastricht ICT Competence Centre (MICC) in Maastricht, The Netherlands.
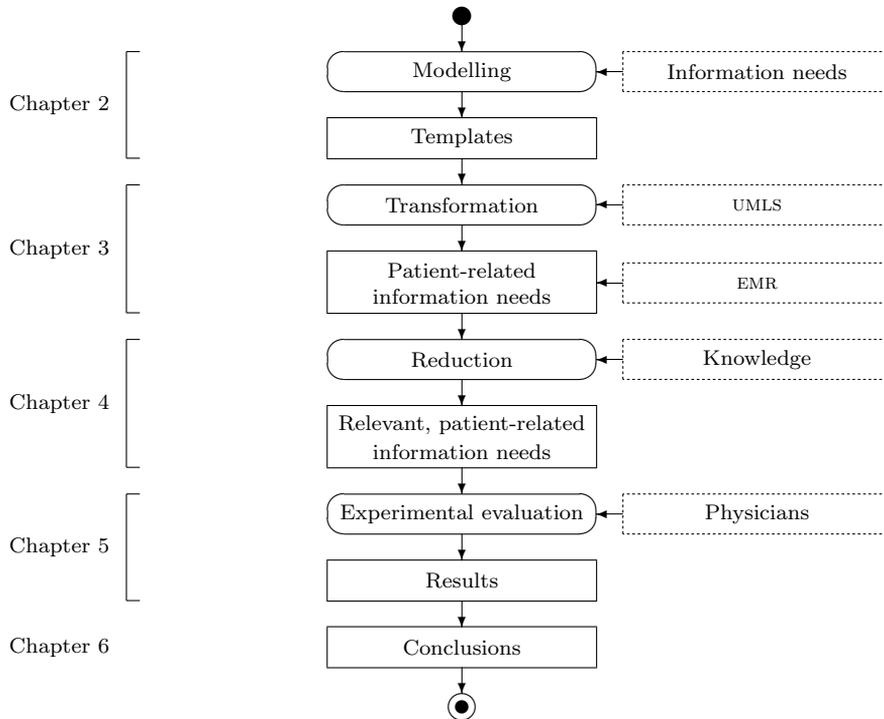
Figure 1.3: Schematic overview of the thesis structure.

1.7 Structure of the Thesis

The structure of the thesis is as follows. In Chapter 1, the problem statement and research questions are formulated. Chapter 2 describes our efforts to build a model of a physician's information needs (RQ1). Chapter 3 describes our method of formulating information needs based on patient data (RQ2) and our method for retrieving literature based on the formulated information needs (RQ3). Chapter 4 presents our method for tuning the formulated information needs and the retrieved literature to the physician using MIRA (RQ4) and our way of ensuring unobtrusiveness (RQ5). Chapter 5 provides the results of the evaluation of MIRA. Finally, Chapter 6 provides our conclusions, discusses our results, and indicates future research.

In Appendix A we present the information needs that we identified by means of a literature survey and interviews with physicians. Appendix B provides the templates that constitute our model of physicians' information needs. In Appendix C we list the heuristics that are used to translate patient data. Appendix D provides the materials used during the evaluation of MIRA. In Appendix E we present the evaluation results of MIRA. For a schematic overview of the thesis structure, we refer to Figure 1.3.

Chapter 2

Modelling a Physician's Information Needs[*]

A prerequisite for making the implicit information needs of a physician explicit is to have a model of the physician's information needs. This chapter focuses on the process of modelling a physician's information needs (see Figure 2.1).

Section 2.1 establishes the importance of information needs. Then, Section 2.2 describes the Unified Medical Language System, which was an essential part of the modelling process. Subsequently, Section 2.3 describes how we identified physicians' information needs. Thereafter, Section 2.4 discusses how we use the identified information needs to develop our model INMOD. Finally, Section 2.5 provides the chapter conclusions.

2.1   The Importance of Information Needs

To make information needs explicit it is essential to be able to anticipate a physician's information needs. Therefore, knowledge about a physician's potential information needs is required. This knowledge has to satisfy two conditions, viz. (a) the knowledge has to be complete and (b) the knowledge has to be accurate.

An ideal solution would be to construct a set of all potential information needs. However, such a set cannot satisfy the two aforementioned requirements for two reasons. First, such a set is never complete, since it is impossible to capture all information needs that a particular physician may have in the future, let alone a large number of physicians. Second, it is difficult to ensure the accuracy of such a set. Physicians generate new information needs over time, which should be added

---

[*]This chapter was based on Braun, L.M.M., Wiesman, F., van den Herik, H.J., Hasman, A., and Korsten, E. (2005). Towards Automatic Formulation of a Physicians Information Needs. *Journal of Digital Information Management*, 3(1): pp. 40-47. ISSN 0972-7272. Also published in *Proceedings of the 5th Dutch-Belgian Information Retrieval Workshop (DIR 2005)*, (ed. R. van Zwol), pp. 25-32. Centre for Content and Knowledge Engineering, Utrecht, The Netherlands

Section 2.1

Section 2.2

Section 2.3

Section 2.4

Section 2.5
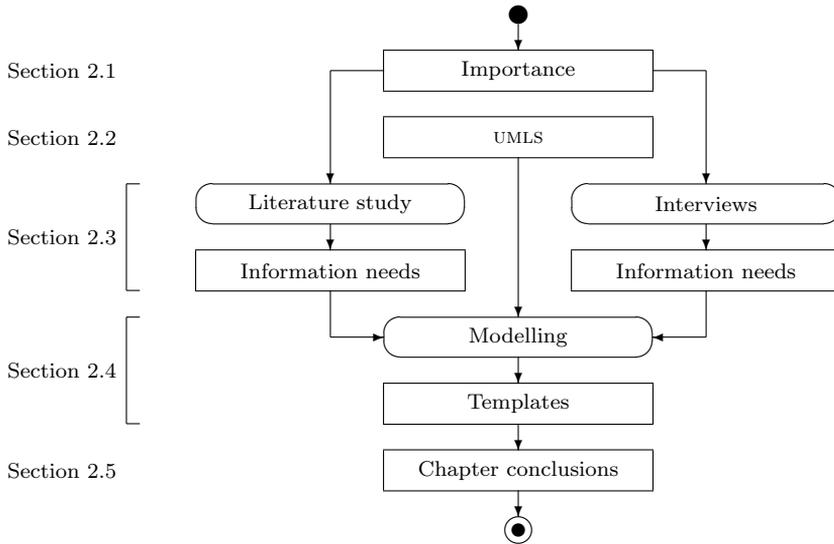
Figure 2.1: Schematic representation of the modelling of physicians' information needs.

to the set and this is hard to facilitate.

Our solution is to build a model of a physician's information needs. Such a model may satisfy the aforementioned requirements, since it represents information needs on a higher level and can also anticipate future information needs.

2.2   The Unified Medical Language System

The Unified Medical Language System (UMLS) is a product of the U.S. National Library of Medicine (NLM).[1] It is a collection of knowledge sources "to facilitate the development of computer applications that behave as if they 'understand' the meaning of the language of biomedicine and health" (NLM, 2006d). The UMLS comprises several knowledge sources and software tools, which are described below.

The UMLS Knowledge Sources[2] consist of the Metathesaurus (Subsection 2.2.1), the Semantic Network (Subsection 2.2.2), and the SPECIALIST Lexicon and Lexical Tools (Subsection 2.2.3). The descriptions below are taken from the UMLS Knowledge Sources Documentation Set (NLM, 2008).

2.2.1   The Metathesaurus

The Metathesaurus is a multi-purpose, multi-lingual vocabulary database that contains information about biomedical and health-related concepts, their various names,

---

[1]http://www.nlm.nih.gov/research/umls/
[2]http://umlsks.nlm.nih.gov/

and the relationships between them (NLM, 2006a). The scope of the Metathesaurus is determined by the combined scope of its source vocabularies. The meanings, concept names, and relationships from the source vocabularies are incorporated and preserved. The Metathesaurus contains mappings between concepts from various source vocabularies (e.g., ICD and SNOMED). Additionally, the Metathesaurus facilitates translation of its source vocabularies into specific languages (e.g., English and Dutch). The Metathesaurus is linked to both the Semantic Network (Subsection 2.2.2) and the SPECIALIST Lexicon (Subsection 2.2.3).

The Metathesaurus is organized by concept or meaning. In essence, its purpose is to link alternative names and views of the same concept together and to identify relationships between different concepts. Each concept within the Metathesaurus is identified by a *Concept Unique Identifier* (CUI). For example, the concept of *Disorder of cardiac rhythm characterized by rapid, irregular atrial impulses and ineffective atrial contractions* is uniquely identified by the CUI C0004238. The concept has four different concept names (see Table 2.1). The concept name *Atrial Fibrillation* is the preferred concept name. For the English language, the concept names of a concept are divided into groups of lexical variants. Each such a group is called a 'term' and is uniquely identified by a *Term Unique Identifier* (note the abbreviation: LUI, not TUI). For the concept C0004238, there are two groups of lexical variants, with the

Table 2.1: A concept with corresponding terms, strings, and atoms.

Concept (CUI)	Terms (LUIs)	String (SUIs)	Atoms (AUIs)
C0004238 Atrial Fibrillation *(preferred)* Atrial Fibrillations Auricular Fibrillation Auricular Fibrillations	**L0004238** Atrial Fibrillation *(preferred)* Atrial Fibrillations	**S0016668** Atrial Fibrillation *(preferred)*	**A0027665** Atrial Fibrillation *(from MeSH)*
			A0027667 Atrial Fibrillation *(from PSY)*
		S0016669 Atrial Fibrillations	**A0027668** Atrial Fibrillations *(from MeSH)*
	L0004327 *(synonym)* Auricular Fibrillation Auricular Fibrillations	**S0016899** Auricular Fibrillation *(preferred)*	**A0027930** Auricular Fibrillation *(from PSY)*
		S0016900 *(plural variant)* Auricular Fibrillations	**A0027932** Auricular Fibrillations *(from MeSH)*

Note: Source of the table is the UMLS *Knowledge Sources Documentation (NLM, 2008)*

LUIs L0004238 and L0004327 (see Table 2.1). Each concept name within a term is represented by a separate string which can be identified by a *String Unique Identifier* (SUI). Table 2.1 shows that the concept C0004238 refers to four different SUIs, viz. S0016668, S0016669, S0016899, and S0016900. Each occurrence of a concept name in a specific source vocabulary is represented by an atom, which is identified by an *Atom Unique Identifier (*AUI*)*. As can be seen from Table 2.1, the string *Atrial Fibrillation* (S0016668) has two different atoms (A0027665 and A0027667) because it occurs in two different source vocabularies, viz. Medical Subject Headings (MeSH) and Psychological Index Terms (PSY).

2.2.2   The Semantic Network

The purpose of the UMLS Semantic Network is to provide a consistent categorization of all concepts represented in the Metathesaurus (NLM, 2006b). To achieve this goal, the Semantic Network comprises two parts. The first part is a set of 135 broad subject categories that provide a categorization of all concepts within the Metathesaurus (see Subsection 2.2.1). These subject categories are called *semantic types*. The semantic types are ordered hierarchically, according to the 'is-a' relationship. Each concept within the Metathesaurus is assigned exactly one semantic type, viz. the most specific semantic type available in the hierarchy. The concept automatically inherits all ancestors of its semantic type.

To illustrate how a semantic type is assigned to a concept, we provide an example based on the concept *pancreatitis* (see Figure 2.2).

- Initially, the concept *pancreatitis* is assigned the most general semantic type *Event*.

- The semantic type *Event* has two subtypes, viz. *Activity* and *Phenomenon or Process*. The latter is a more specific description of the concept, so the concept is assigned the type *Phenomenon or Process*.

- The semantic type *Phenomenon or Process* has three subtypes, viz. *Human-caused Phenomenon or Process*, *Natural Phenomenon or Process*, and *Injury or Poisoning*. Since the second subtype describes the concept more accurately, the concept is assigned the type *Natural Phenomenon or Process*.

- The semantic type *Natural Phenomenon or Process* has only one subtype, *Biologic Function*, which is a more specific description of the concept. Therefore, the concept is assigned the type *Biologic Function*.

- The semantic type *Biologic Function* has two subtypes, viz. *Physiologic Function* and *Pathologic Function*. Since the latter subtype describes the concept more accurately, the concept is assigned the subtype *Pathologic Function*.

- The semantic type *Pathologic Function* has three subtypes, viz. *Cell or Molecular Dysfunction*, *Disease or Syndrome*, and *Experimental Model of Disease*. Since the second subtype is a more specific description of the concept, the concept is assigned the type *Disease or Syndrome*.
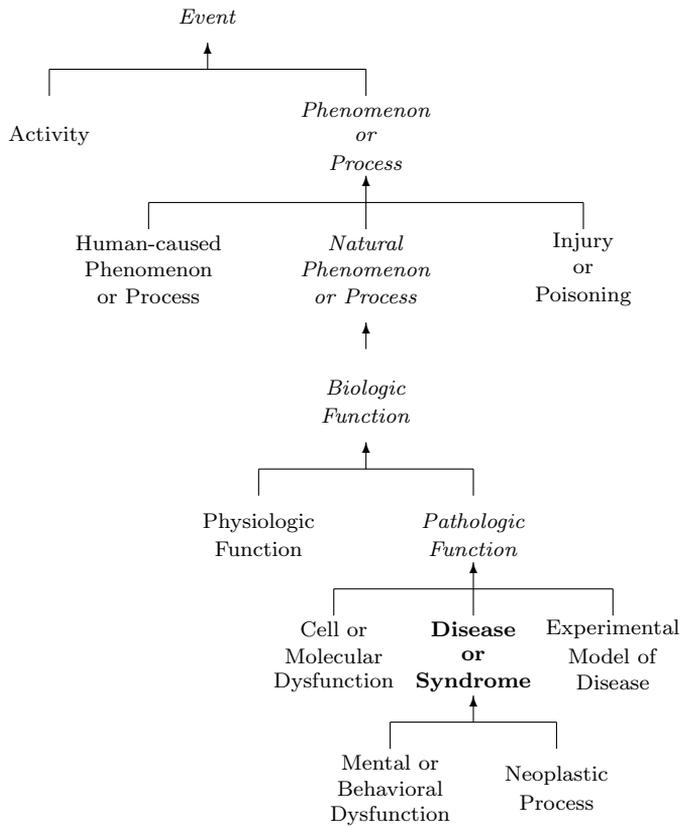
Figure 2.2: Part of the hierarchy of the Semantic Network. An arrow represents an 'is-a' relationship.

- The semantic type *Disease or Syndrome* has two subtypes, viz. *Mental or Behavioral Dysfunction* and *Neoplastic Process*. Since neither describes the concept more accurately, the concept *pancreatitis* is assigned the most specific semantic type found, viz. *Disease or Syndrome*. All ancestors of the semantic type (*Pathologic Function, Biologic Function, Natural Phenomenon or Process, Phenomenon or Process, and Event*) are automatically inherited by the concept.

The second part of the Semantic Network is a set of 54 relationships between the semantic types. These relationships are called *semantic relations*. The primary relationship in the Semantic Network is the 'is-a' relationship, because it is used to order the semantic types hierarchically. Furthermore, the Semantic Network uses a set of non-hierarchical relationships, which are grouped into five main categories, viz. 'physically related to', 'spatially related to', 'temporally related to', 'functionally related to', and 'conceptually related to'. Relationships are inherited downwards along the is-a hierarchy of semantic types.

2.2.3 The SPECIALIST Lexicon and Lexical Tools

The SPECIALIST lexicon is an English lexicon. In contrast to the Metathesaurus (Subsection 2.2.1) the SPECIALIST Lexicon does not provide links between various medical vocabularies. Instead it contains commonly used English words, English biomedical terms, and the lexical variations of both. As such it uses information from English dictionaries (e.g., Longman's Dictionary of Contemporary English) as well as medical dictionaries (e.g., Webster's MedicalDesk Dictionary). The SPECIALIST Lexicon is a part of the SPECIALIST Natural Language Processing System and provides lexical information to this system (NLM, 2006c).

2.3 Identification

To identify physicians' information needs, several methods may be used. Subsection 2.3.1 mentions three identification approaches and assesses them all useful for our purpose. The three identification approaches are discussed separately in Subsections 2.3.2–2.3.4, respectively. Subsection 2.3.5 presents the final results of the identification process.

2.3.1 Identification Approaches

In the past 25 years, extensive research has been performed on identifying users' information needs in general and physicians' information needs in particular. In Table 2.2 we mention eight publications in this field. They are indicative for the research performed and are given in chronological order.

De Vries Robbé, Beckers, and Zanstra (1988) identified physicians' information needs in two different ways. The first way was by registering physicians' real-life information needs. This was done by registering consultations performed by physicians. The consultations were conducted in writing and over the telephone. The second way was by interviewing physicians about their information-seeking behaviour. The interviews were conducted according to a specific interview scheme.

Gorman (1995) used interviews as well as registration in a two-step approach. First, the physicians were interviewed to collect the information needs arising during patient consultations. Then, the physicians recorded their perception of these information needs after each consultation during two to five days.

Smith (1996) performed a literature survey to identify physicians' information needs.

Ely, Osheroff, and Ebell (1999) collected information needs by registering them. During consultation with patients the questions a physician asked were registered manually by one of the researchers.

Strasberg, Hubbs, Rindfleisch, and Melmon (1999) registered the questions asked by physicians by keeping a record of physicians' search actions.

Zeng and Cimino (2000) identified information needs by registering physicians' information needs during their patient consults. Before the registration took place, a literature survey was used to establish a general framework of information needs.

Jerome, Giuse, Gish, Sathe, and Dietrich (2001) identified physicians' information needs by registering the questions received by the librarians of the Clinical Informatics Consult Service (CICS) at Vanderbilt University Medical Center.

Reddy, Pratt, Dourish, and Shabot (2002) registered the information needs by recording questions asked by physicians during their morning rounds.

As can be seen from Table 2.2 we observed three approaches to identify physicians' information needs. We chose all three approaches for our research.

Table 2.2: Publications on the identification of physicians' information needs.

Author(s)	Identification approach(es)	Identification domain(s)
De Vries Robbé, Beckers, and Zanstra (1988)	Registration Interviews	General practice Cardiology Pulmonology Allergology
Gorman (1995)	Interviews Registration	Primary care
Smith (1996)	Literature survey	Primary care
Ely, Osheroff, and Ebell (1999)	Registration	Family care
Strasberg, Hubbs, Rindfleisch, and Melmon (1999)	Registration	Diverse
Zeng and Cimino (2000)	Registration Literature survey	Diverse
Jerome, Giuse, Gish, Sathe, and Dietrich (2001)	Registration	Diverse
Reddy, Pratt, Dourish, and Shabot (2002)	Registration	Surgical care

2.3.2   Literature Study

Several publications have been written presenting a physician's information needs
(see Table 2.3). Below, we mention eight of these publications. The identification
domain and the number of identified information needs per publication are listed
in Table 2.3. Please note that the publications in Table 2.3 differ slightly from
the ones mentioned in Table 2.2. In Table 2.3, we removed the publications by
Strasberg *et al.* (1999) and Zeng and Cimino (2000), since they did not specify any
identified information needs. In Table 2.3 we added publications by Cucina *et al.*
(2001) and Grundmeijer, Reenders, and Rutten (1999). These publications are not
mentioned in Table 2.2, since they do not discuss their identification approach.

De Vries Robbé, Beckers, and Zanstra (1988) identified 76 generic questions,
which were collected from general practitioners, cardiology, pulmonology, and
allergology.

Gorman (1995) identified 16 questions asked by primary care physicians.

Smith (1996) provided 8 examples of clinical questions asked by primary care
doctors.

Ely, Osheroff, and Ebell (1999) collected 1101 questions asked by 103 family

Table 2.3: Number of information needs identified by various studies.

Author(s)	Identification domain(s)	# INs identified
De Vries Robbé, Beckers, and Zanstra (1988)	General practice Cardiology Pulmonology Allergology	76
Gorman (1995)	Primary care	16
Smith (1996)	Primary care	8
Ely, Osheroff, and Ebell (1999)	Family care	10
Grundmeijer, Reenders, and Rutten (1999)	*Not explicit*	34
Cucina, Shah, Berrios, and Fagan (2001)	Outpatient care Inpatient care Internal medicine	16
Jerome, Giuse, Gish, Sathe, and Dietrich (2001)	Diverse	10
Reddy, Pratt, Dourish, and Shabot (2002)	Surgical care	2

doctors. The 10 most common generic queries from these queries are presented in the publication.

Grundmeijer, Reenders, and Rutten (1999) described all stages of the medical process. From this publication we were able to identify 34 information needs.

Cucina, Shah, Berrios, and Fagan (2001) collected 110 questions in inpatient, outpatient, and critical-care academic internal medicine. From these questions they identified 16 generic queries.

Jerome, Giuse, Gish, Sathe, and Dietrich (2001) provided 10 distinct examples of questions.

Reddy, Pratt, Dourish, and Shabot (2002) provided 2 questions.

The identified information needs are listed in Appendix A.1.

2.3.3   Interviews

To identify a set of information needs which is as diverse as possible, we interviewed five physicians in five different fields, viz. neurology, anaesthesiology, surgery, pulmonology, and cardiology. The physicians were asked (a) to what kind of medical questions they need an answer on a regular basis and (b) what kind of questions prompt them to conduct a literature search. From the interviews, we identified nine information needs (see Appendix A.2).

2.3.4   Registration

Finally, we identified the real-life information needs of physicians by registering their search actions. To this end we built a web application, through which the physicians could register each of their search actions separately. In order to register a search action, physicians had to fill out a web form, stating their search question, the reason for the search action and the collection in which they conducted their search. The registered search actions were stored in a database. After three months, the database was analyzed and the information needs were derived from the registered search questions.

However, the registration approach resulted in only a small number of information needs. Furthermore, no information needs were registered that we had not already identified. Therefore, the approach will be left out of consideration in the remainder of this section.

2.3.5   Identification Results

Table 2.4 lists the number of information needs identified in various medical specialisms by the above-mentioned identification approaches, viz. literature study and interviews. The specific numbers of information needs identified per specialism with

each specific identification approach are listed in the second and third column. The
fourth column lists the total number of information needs identified per specialism.
The final row lists the total number of information needs identified per identification
approach. In total, we identified 181 information needs. (See Appendix A).

2.4   Modelling

In this section we discuss how we modelled a physician's information needs. We
started analyzing the identified information needs (Subsection 2.4.1). Since a model
is an abstract representation of the real world, we abstracted the information needs
based on the results of the analysis (Subsection 2.4.2). We call the model INMOD.
Finally, we refined INMOD (Subsection 2.4.3).

2.4.1   Analysis

When analyzing the identified information needs, we noted that the information
needs have a certain structure, viz. they consist of *medical concepts* and *text ele-
ments*. Let us consider the information need below.

Information Need 2.1 *Does **Clarithromycin** cause **pancreatitis**?*

Information need 2.1 contains two text elements: 'Does' and 'cause' (the plain italic
parts), and two medical concepts: 'clarithromycin' and 'pancreatitis' (the bold italic
parts). All identified information needs have a similar structure, i.e., they consist of
text elements and medical concepts. The number and order of the various elements
differ per information need.

2.4.2   Abstraction

We abstracted the information needs by employing their underlying structure. The
abstraction process is based on Ely *et al.* (1999). In their study, Ely *et al.* abstract
an information need by replacing each of its medical concepts by a variable. For

Table 2.4: The number of information needs identified in various specialisms with
various identification approaches.

Specialism	Literature survey	Interviews	Total
Anaesthesiology	0	2	2
Cardiology	0	1	1
Primary care	24	0	24
Pulmonology	0	3	3
Surgery	0	3	3
Various	148	0	148
Total	172	9	181

Table 2.5: Information needs and their corresponding templates.

Information need	Templates
What are the side effects of Amoxicillin?	What are the side effects of <ORGANIC CHEMICAL> ?
What are the treatment options for hypereosinophilic syndrome?	What are the treatment options for <DISEASE OR SYNDROME>?
For diagnosis of deep vein thrombosis, how good is ultrasound?	For diagnosis of <FINDING>, how good is <DIAGNOSTIC PROCEDURE>?
Is there a connection between asthma and migraine?	Is there a connection between <DISEASE OR SYNDROME> and <DISEASE OR SYNDROME>?
Does Norpace cause fatigue?	Does <ORGANIC CHEMICAL> cause <SIGN OR SYMPTOM>?

example, according to Ely *et al.*, Information need 2.1 would be abstracted into *Does drug X cause disease Y?*

We choose to abstract a medical concept more formally by representing it by its semantic type instead of a variable. Such a semantic type is obtained via the UMLS Semantic Network (see Subsection 2.2.2). The process of abstracting an information need consists of replacing each medical concept in the information need by its semantic type. We applied the abstraction process to each identified information need, resulting in a set of information needs, called *information-need templates*.

Definition 2.2 (Information-need template) *An information-need template is an abstract representation of an information need, that can be obtained by replacing each medical concept in an information need by its semantic type.*

In the remainder of this thesis, the term information-need template will be abbreviated into 'template'. An information need is represented by exactly one template. For example, the template corresponding to Information need 2.1 is shown below.

Template 2.3 *Does* `<Chemical>` *cause* `<Disease or Syndrome>`*?*

In general, a specific template represents a group of information needs. For example, Template 2.3 represents both Information need 2.1 and 2.4.

Information Need 2.4 *Does Amoxicillin cause hepatitis?*

Our abstraction process resulted in 181 templates; one for each information need (see Table 2.4). Table 2.5 shows a selection of information needs and their corresponding templates.

2.4.3 Refinements

We refined the set of 181 templates according to three perspectives, viz. (a) syntax, (b) semantics, and (c) context.

Syntactic Refinement

Our set of templates contained multiple templates that were syntactically equal, i.e., they were exactly similar. This was caused by the fact that some information needs were represented by the same template. Since our set of templates contained one template for each information need, some templates occurred multiple times. To obtain a proper *set* of templates, we removed all doubles. During the syntactic refinement 14 templates were discarded, reducing the set to 167 templates.

Semantic Refinement

Our set of templates also contained multiple templates that were semantically equal, i.e., they were syntactically distinct, but they conveyed the same meaning. For example, Templates 2.5 and 2.6 are semantically similar.

Template 2.5 *What does* <Health Care Activity> *cost?*

Template 2.6 *What are the costs of* <Health Care Activity>*?*

To obtain a proper set of templates from the semantic perspective, we reduced each group of semantically similar templates to one template. It was of no particular consequence which specific template was chosen for this purpose, since the text elements of the templates are standardized as well (see Subsection 3.5.2). During the semantic refinement 25 templates were discarded, reducing our set of templates to 142 templates.

Contextual Refinement

According to Koopmans and Offringa (2003), physicians have information needs in four distinct contexts.

- Diagnosis

- Prognosis

- Therapy (including prevention)

- Side effects (including etiology)

Consequently, the templates in our set can also be categorized according to their subjects. Our set contained 40 templates related to diagnosis, 8 templates related to prognosis, 71 templates related to therapy (including prevention), and 23 templates related to side effects (including etiology). It is difficult to anticipate a physician's information needs during the *diagnostic process*, since a physician's thoughts during this process are hard to predict (cf. Norman, Young, and Brooks, 2007; Patel, Arocha, and Zhang, 2004; Lemieux and Bordage, 1992; Gale, 1982). Therefore, we choose to concentrate INMOD on the remaining three contexts. Consequently, during contextual refinement 40 templates were discarded, reducing our set to 102 templates. The complete set of templates in INMOD is listed in Appendix B.

2.5   Chapter Conclusions

In this chapter we described how we modelled information needs. From the results of the modelling process, i.e., the development of INMOD, we may draw three conclusions.

First, we may conclude that modelling of information needs is an important step in the process of automatically formulating information needs (Subsection 2.1). Without a model, it is impossible to handle all possible information needs that physicians may have.

Second, we may conclude that we identified 181 information needs (Subsection 2.3). These information needs were identified by a literature study and interviews with physicians.

Third, we may conclude that we developed a model of a physician's information needs, consisting of 102 templates (Subsection 2.4). The templates were constructed by analysis, abstraction, and refinement of the identified information needs.

INMOD will be used to formulate a physician's information needs in a specific patient situation and to retrieve literature based on the formulated information needs. This will be the topic of the next chapter.

Chapter 3

From Patient Data to Literature*

As mentioned in Chapter 1 quality of care can be improved by providing physicians with patient-related literature. In Chapter 2 we explained how we modelled the physicians' information needs into templates. By combining the templates with patient data, patient-related information needs can be formulated and subsequently, patient-related literature may be retrieved. The process to get from patient data to literature is the topic of this chapter. The chapter consists of two parts. Part 1 is the process of automatically retrieving patient-related literature (Sections 3.1 – 3.5). Part 2 consists of experiments, evaluation, and chapter conclusions (Sections 3.6 – 3.8).

The outline of Part 1 is as follows (see Figure 3.1). The process of automatically retrieving literature from patient data comprises five steps (indicated as rounded rectangles in the top part of Figure 3.1). In the first step, templates are selected to be transformed into patient-related information needs (Section 3.1). In the second step, the appropriate patient data are extracted from the EMR (Section 3.2). In the third step, the patient data resulting from the second step are translated into English (Section 3.3). In the fourth step, the translated patient data resulting from the third step are used to instantiate the templates, resulting into patient-related information needs (Section 3.4). In the fifth step, the formulated information needs are used as a starting point for literature retrieval, resulting in a set of articles (Section 3.5).

The outline of Part 2 is already given briefly above. For completeness we mention the following. Experiments concerning the data translation are discussed in Section 3.6. The evaluation of our approach of retrieving literature based on patient data is discussed in Section 3.7. Finally, the chapter conclusions are provided in Section 3.8.
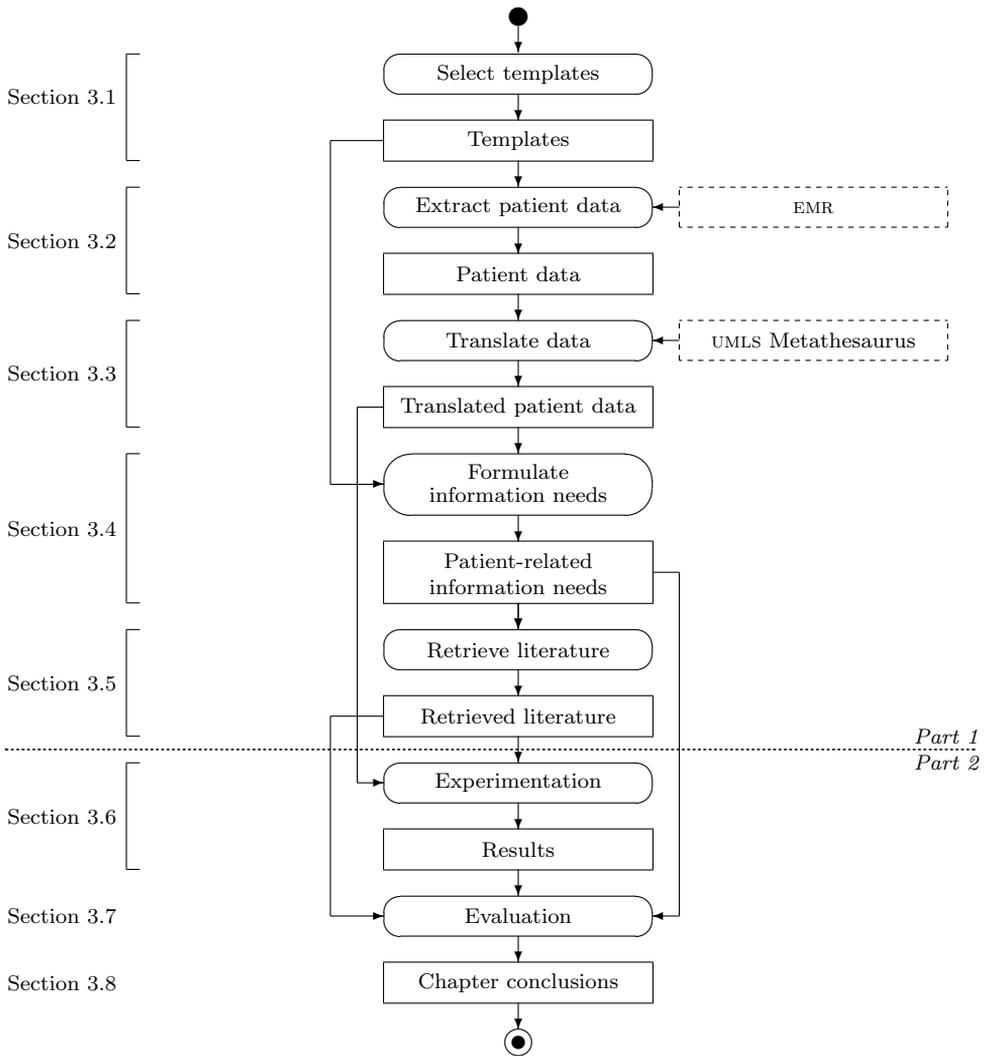
Figure 3.1: Schematic representation of the automatic retrieval of patient-related literature (Part 1), followed by experiment, evaluation, and chapter conclusions (Part 2).

3.1   Selecting Information-Need Templates

To ensure efficient selection of templates (see Appendix B), and thus, efficient formulation of information needs, we grouped the 102 templates into a hierarchy of semantic clusters, based on the semantic types they contain.
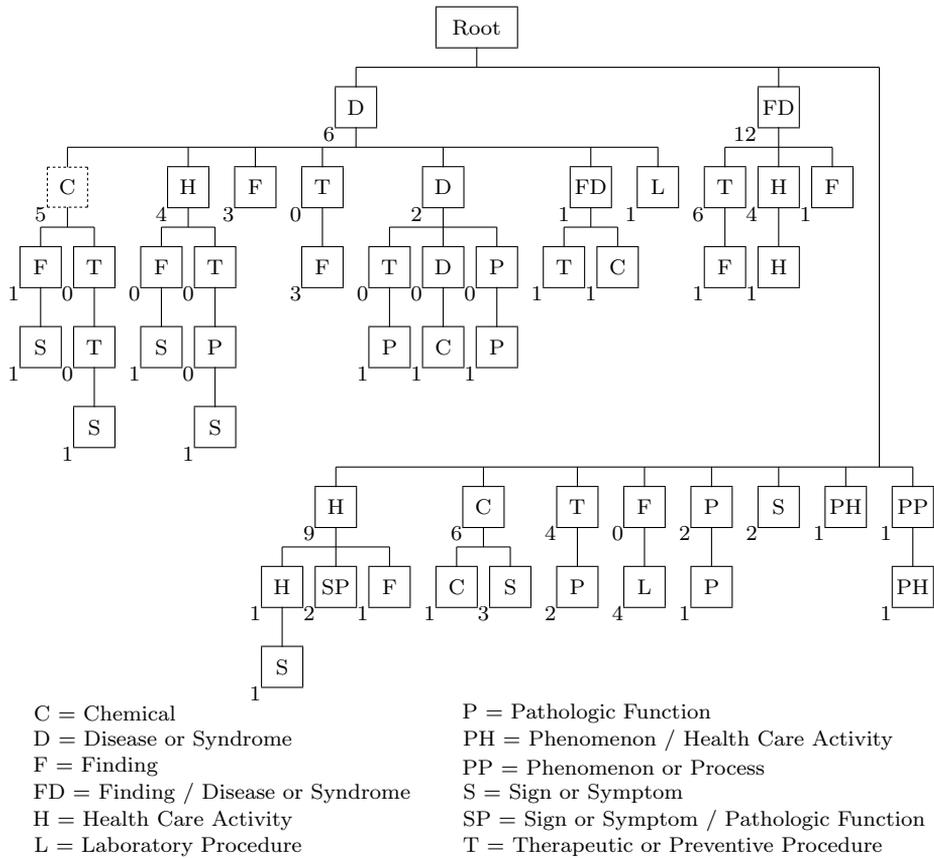
Figure 3.2: Semantic clusters, structured as a tree based on the semantic types they contain. The number under the lower left corner of each node indicates the number of templates that the semantic cluster contains.

Definition 3.1 (Semantic cluster) *A semantic cluster is a set of templates, all containing the same bag[1] of semantic types.*

Figure 3.2 provides an overview of the clusters we derived from the templates discussed in Chapter 2. In the following discussion, we distinguish between a general node n and a specific node, called node N. Each node n represents a semantic type. Assume node N is the node represented by the dashed box (see the upper left corner of Figure 3.2). A semantic cluster S_n is denoted by the path starting at the root R and ending in node n. Consequently, the semantic cluster S_N is represented by the path starting at the root R and ending in node N. A semantic cluster S_n comprises all templates containing exactly those semantic types that are encountered on the path from the root R to node n. Therefore, the semantic cluster S_N comprises all

---

[1]A bag is defined as a set that may contain multiple instances of the same element.

Table 3.1: Five templates in the specific semantic cluster S_n.

- Does <CHEMICAL> cause <DISEASE OR SYNDROME> ?

- Is <CHEMICAL> effective for <DISEASE OR SYNDROME> ?

- Give a description of the farmacodynamics of <CHEMICAL> in <DISEASE OR SYNDROME>.

- After a course of <CHEMICAL> in a patient with <DISEASE OR SYNDROME>, what treatment is appropriate for persistent symptoms?

- Is <CHEMICAL> indicated in a patient with <DISEASE OR SYNDROME>?

---

templates containing the semantic types *Disease or Syndrome* and *Chemical*. The templates belonging to the semantic cluster S_N are listed in Table 3.1. The number beneath a node n indicates the number of templates belonging to the semantic cluster S_n. The number 5 beneath node N indicates that the semantic cluster S_N contains five templates (cf. Table 3.1). In total, there are 51 distinct semantic clusters, represented by 51 distinct paths in our tree. However, 10 of these paths represent semantic clusters containing no templates (paths ending in a node with the number 0 beneath it). These 10 semantic clusters are not included in the set of semantic clusters. In total, we distinguished 41 semantic clusters.

We continue our general discussion. MIRA selects templates by traversing the tree in a depth-first manner. For every node n that is encountered, MIRA queries the EMR for information of the semantic type represented by node n (see Section 3.2). If patient data are found, all templates of the semantic cluster S_n can be selected and the search continues with the next node. If no patient data are found, the templates in semantic cluster S_n are not selected. The tree is constructed in such a way that non-selection of a certain semantic cluster has consequences for the selection of all clusters in the corresponding subtree. If the templates in the semantic cluster S_n are not selected, the templates in the clusters of the subtree of node n are not selected either. Consequently, the whole subtree will be pruned. The tree is constructed based on the occurrence of the various semantic types in the semantic clusters. This ensures an efficient selection of templates.

For instance, starting from the root of the tree, the first semantic cluster selected contains templates with the semantic type 'Disease or Syndrome' (D). The next semantic cluster (when using depth-first search) contains templates with the semantic types 'Disease of Syndrome' (D) and 'Chemical' (C). The third semantic cluster contains templates with the semantic types 'Disease of Syndrome' (D), 'Chemical' (C), and 'Finding' (F). Assume that the EMR contains no data of the semantic type 'Finding'. Then the templates in the third semantic cluster are not selected. Consequently, the fourth semantic cluster (containing templates with the semantic types 'Disease of Syndrome' (D), 'Chemical' (C), 'Finding' (F), and 'Sign or Symptom' (S)) will not be selected either. In this way, subtrees can be pruned.

3.2 Extracting Appropriate Patient Data

As mentioned in the chapter introduction, patient data are essential in the process
of converting templates into patient-related information needs. However, not all
patient data are equally important with respect to each template. Therefore, the
second step in the conversion process is to extract only the appropriate data from
the EMR. In order to do this, MIRA (a) formulates EMR queries and (b) executes the
formulated queries on the EMR. Both steps are discussed separately below.

3.2.1 Formulating EMR Queries

MIRA starts to determine which patient data are appropriate with respect to a se-
lected template. This can be achieved by looking at the semantic types of the data
slots in the template. For each data slot in a template, patient data of the corre-
sponding semantic type should be extracted from the EMR.

In order to extract the appropriate data from the EMR we have to indicate where
the specific data can be found. To this end we define the notion of EMR *query*.

Definition 3.2 (EMR query) *An EMR query is an expression indicating where pa-
tient data of a particular semantic type can be found in the EMR.*

For our research, we have formulated a number of EMR queries, based on the structure
of the EMR. Since the EMR is stored in a database, EMR queries are database queries.
In our investigations we use SQL as the query language.

As discussed in Subsection 2.2.2, the semantic types used in the templates are
arranged hierarchically (see Figure 3.3). The hierarchy of semantic types was used
in the formulation of the EMR queries in a bottom-up fashion. An EMR query for a
semantic type that is a 'leaf' in the hierarchy simply indicates where information of
this type can be found in the EMR.

For a semantic type that is not a leaf, the situation is different. For most semantic
types with subtypes, the corresponding subtypes do not completely cover all con-
cepts comprised by the semantic type itself. For example, the semantic type 'Disease
or Syndrome' has two subtypes (see Figure 3.3): (1) 'Mental or Behavioral Dysfunc-
tion' and (2) 'Neoplastic Process'. The medical concept 'Dementia' is covered by the
semantic type 'Mental or Behavioral Dysfunction' and the medical concept 'Brain
Neoplasms' is covered by the semantic type 'Neoplastic Process'. Both concepts are
therefore also covered by the semantic type 'Disease or Syndrome'. However, the
medical concept 'Pneumonia' is only covered by the semantic type 'Disease or Syn-
drome', but not by one of its subtypes. If only the queries of the two subtypes are
used to find information of the type 'Disease or Syndrome', the concepts 'Demen-
tia' and 'Brain Neoplasms' will be found. However, the concept 'Pneumonia' would
not be found, despite the fact that it is covered by the semantic type 'Disease or
Syndrome'. Consequently, a separate EMR query is required for the semantic type
'Disease or Syndrome'. Therefore, in these cases, the compound EMR query is com-
posed of the EMR queries corresponding to the subtypes *and* the EMR query of the
semantic type itself. Since MIRA uses SQL to query the EMR, we can use the UNION

```
Event
├─── Activity
│      ├─── Behavior
│      │      ├─── Social Behavior
│      │      ├─── Individual Behavior
│      ├─── Daily or Recreational Activity
│      ├─── Occupational Activity
│      │      ├─── Health Care Activity
│      │      │      ├─── Laboratory Procedure
│      │      │      ├─── Diagnostic Procedure
│      │      │      ├─── Therapeutic or Preventive Procedure
│      │      ├─── Research Activity
│      │      │      ├─── Molecular Biology Research Technique
│      │      ├─── Governmental or Regulatory Activity
│      │      ├─── Educational Activity
│      ├─── Machine Activity
├─── Phenomenon or Process
       ├─── Injury or Poisoning
       ├─── Human-caused Phenomenon or Process
       │      ├─── Environmental Effect of Humans
       ├─── Natural Phenomenon or Process
              ├─── Biologic Function
              │      ├─── Physiologic Function
              │      │      ├─── Organism Function
              │      │      │      ├─── Mental Process
              │      │      ├─── Organ or Tissue Function
              │      │      ├─── Cell Function
              │      │      ├─── Molecular Function
              │      │      │      ├─── Genetic Function
              │      ├─── Pathologic Function
              │      │      ├─── Disease or Syndrome
              │      │      │      ├─── Mental or Behavioral Dysfunction
              │      │      │      ├─── Neoplastic Process
              │      │      ├─── Cell or Molecular Dysfunction
              ├─── Experimental Model of Disease
```

Figure 3.3: Hierarchy of UMLS semantic types which are subtypes of the type *Event*.

keyword here. The compound query can be assembled by the following recursive
algorithm.

```
compoundQuery(T):
    result := query(T)
    for each subtype S of T
        result := result + "UNION" + compoundQuery(S)
```

For example, assume that a semantic type *A* has two subtypes, *B* and *C*, corre-
sponding to EMR queries *QueryB* and *QueryC*, respectively. The semantic types *B*
and *C* have no further subtypes. In addition, the semantic type has its own query:
QueryA. Then the compound query corresponding to A would be *QueryA* UNION
QueryB UNION *QueryC*.

To illustrate the EMR query selection, we provide an example. Normally, EMR query selection is performed for all templates, but for clarity, we take only a single template into account.

Template 3.3 *Is* <Chemical> *effective for* <Disease or Syndrome>*?*

Template 3.3 contains two semantic types: (1) Chemical and (2) Disease or Syndrome. Consequently, two types of patient data have to be obtained from the EMR: data concerning the patient's medication and data concerning the patient's diseases. Therefore, two EMR queries are selected: one for the semantic type 'Chemical' (EMR query 3.4) and one for the semantic type 'Disease or Syndrome' (EMR query 3.5). For reasons of conformity with the ICIS we provide questions and answers in Dutch as given to and by MIRA.

EMR **query 3.4** `SELECT Medicijn`
`FROM Medicatie`
`WHERE PatientNummer=X`

EMR **query 3.5** `SELECT Indicatie`
`FROM OpnameIndicatie`
`WHERE PatientNummer=X`

Both semantic types have subtypes. Consequently, both queries should be extended by queries for these subtypes. However, our EMR does not store information of the subtypes of 'Chemical' or 'Disease or Syndrome'. Therefore, the subtypes of both semantic types have no specific queries. As a result, the queries above are not extended.

As can be seen in EMR queries 3.4 and 3.5, the WHERE clause is generic. To make a query executable, the 'X' has to be replaced by the patient number of the patient under consideration. Based on the template mentioned above and the patient number of our specific patient,[2] the EMR queries 3.6 and 3.7 are formulated.

EMR **query 3.6** `SELECT Medicijn`
`FROM Medicatie`
`WHERE PatientNummer=246897531`

EMR **query 3.7** `SELECT Indicatie`
`FROM OpnameIndicatie`
`WHERE PatientNummer=246897531`

3.2.2 Executing EMR Queries

To obtain the patient data, EMR queries have to be executed on the patient's EMR. The results of an EMR query are called *active concepts* of the corresponding semantic type.

Definition 3.8 (Active concept) *An active concept of a particular semantic type is a result of the EMR query corresponding to this semantic type.*

---

[2]All patient data throughout this thesis are fictitious.

Table 3.2: Two semantic types and their active concepts.

Semantic type	Active concept
Chemical	Clarithromycine
	Amoxi/Clavulaan
	Furosemide-iv
Disease or Syndrome	Pneumonie infectieus
	COPD

Assume we execute EMR queries 3.6 and 3.7 (Subsection 3.2.1) on the EMR of our patient (with patient number 246897531). A part of the particular EMR is shown in Figure 3.4. The specific part of the EMR indicates that the patient is using three different medications ('Clarithromycine', 'Amoxi/Clavulaan', and 'Furosemide-iv') and is suffering from two different diseases ('Pneumonie Infectieus' and 'COPD'). Based on these data EMR query 3.6 has three results and EMR query 3.7 has two results. Consequently, the semantic type 'Chemical' has three active concepts and the semantic type 'Disease or Syndrome' has two active concepts (see Table 3.2).

3.3   Translating Patient Data

The active concepts obtained in Subsection 3.2.2 introduce an obstacle. They are in Dutch, whereas the information needs to be formulated have to be in English. The latter requirement is imposed since MIRA searches English literature databases. Consequently, the terms have to be translated. Moreover, our EMR does not solely use standardized terms; this makes translation of the terms more difficult. To overcome the obstacle, we investigated two approaches for translating our EMR, viz. manual and automatic translation. For both approaches, the use of the UMLS Metathesaurus (see Subsection 2.2.1) is essential. To make a comparative assessment between the two approaches, we constructed two translation mechanisms, one based on the manual approach (Subsection 3.3.1) and the other based on the automatic approach (Subsection 3.3.2). Subsection 3.3.3 motivates our choice for one of the mechanisms as the translation mechanism to be used.

3.3.1   Manual Translation

The manual translation mechanism consists of manually mapping all terms from the EMR (a collection of 1583 different terms) onto UMLS concepts. In the experiments we were able to map 72.5% of the terms. The remaining 27.5% of the terms (e.g., 'geen plaats afd' meaning 'no free bed in unit') could not be mapped because they had no corresponding UMLS concepts. The mapping was stored in a database. To translate a Dutch term, the corresponding UMLS concept is looked up in the database and translated into English via the UMLS Metathesaurus. For example, the term

Figure 3.4: Interface of the ICIS, showing a patient's particulars, diagnoses, and medication.

'Clarithromycine per os' (per os = oral) is mapped onto the UMLS concept *C0055856* and translated into 'Clarithromycin'.

3.3.2   Automatic Translation

Our automatic translation mechanism is based on Tran *et al.* (2004) and Deville (2001). Tran *et al.* discussed a five-step method for preprocessing French terms so that they can be translated rather easily by means of a medical thesaurus (e.g., the UMLS Metathesaurus). Deville presented rules for translating suffixes of medical terms into English to improve translation performance.

We combined the methods by Tran *et al.* (2004) and Deville (2001) into a three-step procedure for preprocessing.

1. Normalization

2. Lexical transformation

3. Morphosyntactical transformation

These three steps have to be performed in the order in which they are listed above.

Normalization comprises the removal of punctuation marks and the conversion of uppercase characters to lower case. For example, the term *Clarithromycine-po* is transformed into *clarithromycine po*.

Lexical transformation comprises the application of several domain-specific heuristics. These heuristics remove unnecessary extensions and write abbreviations in their complete form. For example, *clarithromycine po* is transformed into *clarithromycine per os*. In total we constructed 14 lexical heuristics (see Appendix C).

Morphosyntactical transformation comprises the application of domain-specific heuristics to transform suffixes of terms.[3] For example, *clarithromycine per os* is transformed into *clarithromycin per os*. The construction of our set of morphosyntactical heuristics is based on Jensen (2004). In total we constructed 102 morphological heuristics (see Appendix C).

After the medical terms were preprocessed by the above mentioned procedure, they were translated by the UMLS Metathesaurus. To reiterate our example, *clarithromcyin per os* is translated into *clarithromycin*. The suffix *per os* is removed by UMLS. If a term, after preprocessing, can still not be translated by UMLS, we have two options: either to leave the term unchanged or to remove the term from our list of active concepts. However, removing a term from the list of active concepts means that this term cannot be used in the formulation of information needs, despite the fact that the term is present in the patient data. This is an undesirable consequence, therefore the terms that cannot be translated are left unchanged.

The automatic translation mechanism can be used in several different configurations. Based on an experimental analysis, we chose the configuration in which terms are preprocessed by all three preprocessing steps and translated by using the *complete* Metathesaurus. The experimental basis for this decision can be found in Subsection 3.6.1.

3.3.3   Choosing a Translation Mechanism

To make a comparative assessment between the use of the manual and the automatic translation mechanism, we performed an experiment. Details of this experiment can be found in Subsection 3.6.2. The results of this experiment indicated that the manual translation mechanism is more effective but less scalable than the automatic translation mechanism. Since we favour effectiveness over scalability, we prefer to

---

[3]We have successfully used a similar approach in a different domain (Braun, Wiesman, and Sprinkhuizen-Kuyper, 2002).

Table 3.3: Translation of five medical terms from Dutch into English via UMLS concepts.

ICIS (Dutch)	UMLS concept (CUI)	UMLS (English)
Clarithromycine	C0055856	Clarithromycin
Amoxi/Clavulaan	C0054066	Amoxicillin clavulanic acid
Furosemide-iv	C0016860	Furosemide
Pneumonie Infectieus	C0339973	Infectious mononucleosis pneumonia
COPD	C0024117	Chronic Obstructive Airway Disease

use the manual translation mechanism to translate the medical concepts from Dutch into English.

By use of our manual translation mechanism, the active concepts obtained in Section 3.2 are translated according to Table 3.3.

3.4   Formulating Information Needs

To formulate an information need from a template, MIRA replaces each semantic type within the template by one of its active concepts. We recall from Subsection 3.2.2 that *active concepts* of a semantic type are the results of its corresponding database query (definition 3.8). Consequently, an information need can only be formulated from a template if each of the semantic types in the template has one or more active concepts. In that case, we call the template *applicable*.

Definition 3.9 (Applicable template) *An applicable information-need template is a template in which each semantic type has one or more active concepts.*

By systematically replacing all semantic types by all their active concepts, all information needs are formulated.

The total number of possible information needs which can be formulated from a template is the product of the numbers of instances of all semantic types in the template. We recall from Subsection 3.2.1 that Template 3.3 contains two semantic types: 'Chemical' and 'Disease or Syndrome'. These active concepts have three and two active concepts, respectively (Table 3.2). As a result, six information needs can be formulated from Template 3.3. The resulting templates are listed below.

- Is **Clarithromycin** effective for **Infectious mononucleosis pneumonia?**

- Is **Clarithromycin** effective for **Chronic Obstructive Airway Disease?**

- Is **Amoxicillin clavulanic acid** effective for **Infectious mononucleosis pneumonia?**

- Is **Amoxicillin clavulanic acid** effective for **Chronic Obstructive Airway Disease?**

- Is **Furosemide** effective for **Infectious mononucleosis pneumonia?**

- Is **Furosemide** effective for **Chronic Obstructive Airway Disease?**

3.5 Retrieving Literature

To perform literature retrieval, we used a specific retrieval tool discussed in Subsection 3.5.1. Initially, MIRA formulates a query for each information need formulated in the previous section (Subsection 3.5.2). Subsequently, the queries are executed on our document collection by the retrieval tool, resulting in a number of documents (Subsection 3.5.3). Finally, the retrieved documents are presented to the user (Subsection 3.5.4).

3.5.1 Retrieval Tool

To search the document collection, we used the APACHE LUCENE search engine library.[4] LUCENE is capable of searching full-text collections in any domain. The advantage of LUCENE over common search engines for MEDLINE (e.g., ENTREZ, the standard engine for searching MEDLINE) is LUCENE's ability to sort the retrieved documents with respect to their relevance before they are presented to the user.

To enable LUCENE to search a document collection (e.g., MEDLINE), the collection has to be indexed. To this end, each document in the collection has to be structured into a number of fields according to which it will be indexed (e.g., title, abstract, and keywords). For MEDLINE records, this is quite straightforward, since they are already structured by means of their metadata. Consequently, the format of MEDLINE records renders them highly useful for indexing by LUCENE. Using the appropriately formated documents as input, LUCENE indexes the document collection automatically. When searching the document collection, LUCENE actually searches its self-created index, speeding up the search process considerably.

To determine the relevance of a document—required for sorting retrieved documents with respect to their relevance—LUCENE combines the Boolean model (see e.g., Grossman and Frieder, 2004; Manning, Raghavan, and Schütze, 2007, Chapter 1) and the Vector Space Model (see e.g., Grossman and Frieder, 2004; Manning *et al.*, 2007, Chapter 6). The Boolean model is used to refine the document collection to those documents that match the query. Subsequently, the Vector Space Model is used to determine the relevance of each document (i.e., the extent to which the document matches the query), expressed by a number between 0 and 1. In addition LUCENE also uses some fuzzy retrieval techniques.

3.5.2 Building Queries

Building a query for an information need is done in four steps. Each of these steps is discussed separately below.

---

[4]LUCENE 2.0. The Apache Software Foundation. http://lucene.apache.org/

Table 3.4: MeSH qualifiers.

AB	abnormalities	EM	embryology	PP	physiopathology
AD	administration &	EN	enzymology	PO	poisoning
	dosage	EP	epidemiology	PC	prevention & control
AE	adverse effects	ES	ethics	PX	psychology
AG	agonists	EH	ethnology	RE	radiation effects
AA	analogs & derivatives	ET	etiology	RA	radiography
AN	analysis	GE	genetics	RI	radionuclide imaging
AH	anatomy & histology	GD	growth & development	RT	radiotherapy
AI	antagonists &	HI	history	RH	rehabilitation
	inhibitors	IM	immunology	SC	secondary
BI	biosythesis	IN	injuries	SE	secretion
BL	blood	IR	innervation	ST	standards
BS	blood supply	IS	instrumentation	SN	statistics &
CF	cerebrospinal fluid	IP	isolation & purification		numerical data
CS	chemical sythesis	LJ	legislation &	SD	supply &
CI	chemically induced		jurisprudence		distribution
CH	chemistry	MA	manpower	SU	surgery
CL	classification	ME	metabolism	TU	therapeutic use
CO	complications	MT	methods	TH	therapy
CN	congenital	MI	microbiology	TO	toxicity
CT	contraindications	MO	mortality	TM	transmission
CY	cytology	NU	nursing	TR	transplantation
DF	deficiency	OG	organization &	TD	trends
DI	diagnosis		administration	US	ultrasonography
DU	diagnostic use	PS	parasitology	UL	ultrastructure
DH	diet therapy	PY	pathogenicity	UR	urine
DE	drug effects	PA	pathology	UT	utilization
DT	drug therapy	PK	pharmacokinetics	VE	veterinary
EC	economics	PD	pharmacology	VI	virology
ED	education	PH	physiology		

Note: Source of the table is http://www.nlm.nih.gov/mesh/topsubscope2008.html

First, stop words are removed from the information need. Stop words are words that occur very often in the document collection. The removal of stop words is quite common practice in the field of IR (Van Rijsbergen, 1979). Stop-word removal prevents queries from containing words without any discriminating power (Luhn, 1958). If we perform this step on the first of the six information needs mentioned in the previous section, the resulting query looks like this.

```
clarithromycin effective infectious mononucleosis pneumonia
```

Second, each medical term in the query is annotated with a prefix indicating in which index field of the document the term must occur for the document to be retrieved. The information concerning the appropriate annotations is provided in the template corresponding to the information need. Furthermore, part of the query may be annotated by a plus-sign, indicating that a document must contain the query terms in parentheses in order to be retrieved. After the second step the query looks like this.

```
+(meshterms:"clarithromycin"
meshterms:"infectious mononucleosis pneumonia")
```

The prefix *meshterms* indicates that the query terms have to be found in the field *meshterms*. The query terms are mentioned in double quotes, meaning that the terms have to be found exactly as stated.

Third, the medical terms in the query are annotated with appropriate MeSH qualifiers (NLM, 2007). The qualifiers to be used are provided by the template. There are 83 different MeSH qualifiers (see Table 3.4). They are used to indicate the context in which a specific term is used. In this way, a single term may be used within different contexts. For instance, when the term *Clarithromycin* is annotated with the qualifier AD, it indicates that the term is used within the context of *administration and dosage*. However, if the same term is annotated with the qualifier AE, it indicates that the term is used within the context of *adverse effects*. For each query, we specify which qualifiers should be used. The appropriate qualifier is determined by the semantics of the query. In our example, the term *clarithromycin* is annotated with the qualifier PD, since this qualifier concerns effectiveness of drugs. The term *infectious mononucleosis pneumonia* is annotated with the qualifier DT, since this qualifier concerns drug therapy (DT). After the third step the query looks like this.

```
+(meshterms:"clarithromycin/PD"
meshterms:"infectious mononucleosis pneumonia/DT")
```

As mentioned above, the plus sign at the start of the query indicates that the document must contain the query terms in parentheses in order to be retrieved.

Fourth, the remaining terms and synonyms of all medical terms in the information need are added to the query. Synonyms are added to decrease the influence of the specific terminology used in the literature collection. The synonyms are obtained from the UMLS Metathesaurus. Currently, the query term *clarithromycin* has no useful synonyms. The query term *infectious mononucleosis pneumonia* has one synonym: *glandular fever pneumonia*. Furthermore, the remaining non-medical terms from the original information need are added to the query as well. They may be annotated by a '*'-sign, indicating a wildcard that matches any sequence of characters.

```
+(meshterms:"clarithromycin/PD"
"infectious mononucleosis pneumonia/DT")
title:"clarithromycin"
abstract:"clarithromycin"
title:"infectious mononucleosis pneumonia"
abstract:"infectious mononucleosis pneumonia"
title:"glandular fever pneumonia"
abstract:"glandular fever pneumonia"
title:"effective*"
abstract:"effective*"
```

Information needs formulated

Retrieved articles

Lipid-soluble and water-soluble beta-blockers. Comparison of the central nervous system depressant effect.
Drug-induced acute interstitial nephritis: report of 10 cases.
Use and abuse of allopurinol [editorial] [published erratum appears in Br Med J (Clin Res Ed) 1987 Aug 8;295(6594):350]
In vitro suppression of HTLV-III/LAV infectivity by a combination of acyclovir and suramin.

Abstract

Drug-induced acute interstitial nephritis: report of 10 cases.(JOURNAL ARTICLE.)

Handa SP.

Can Med Assoc J 8703; 135(11):1278-81

Between January 1979 and June 1985, 10 patients with acute allergic interstitial nephritis were seen in a clinical nephrology service at a large regional hospital. The onset of renal failure was temporally related to the use of a drug: a nonsteroidal anti-inflammatory agent (NSAID) (in four patients), cimetidine (in three), antibiotics (in two) or allopurinol (in one). The onset of renal failure was acute in three patients and insidious in seven. Two patients also exhibited marked proteinuria. Clinical features such as fever, rash, hematuria, pyuria with or without eosinophiluria, and mild to marked proteinuria had led to suspicion of the disease. The diagnosis was confirmed by renal biopsy findings of inflammatory cells, predominantly lymphocytes, plasma cells and eosinophils. Three patients required hemodialysis; two of them received steroids as well. Steroid therapy was also used in two patients with NSAID-induced proteinuria. Renal function improved in nine patients by 35 days, but one patient continued to have slow but progressive deterioration of renal function. Acute interstitial nephritis can be distinguished from other forms of acute renal failure by heavy renal uptake of gallium 67, maximal 48 hours or more after injection. The improvement in renal function after discontinuation of the implicated drug, the characteristic histopathological findings of allergic interstitial nephritis, and the presence of

Figure 3.5: MIRA's minimalist interface, presenting formulated information needs and retrieved literature.

After the fourth step, the query is complete and can be executed on the document collection to retrieve the appropriate literature.

3.5.3 Executing Queries and Retrieving Documents

The execution of the queries and the retrieval of the appropriate documents is quite a trivial step in the retrieval process. This step is completely handled by LUCENE. When all documents matching the query are found, they are sorted according to their relevance. MIRA only returns the single most relevant document to the physician.

3.5.4 Presenting Documents

After the documents are retrieved they are presented to the user, i.e., the physician. To avoid impeding the physician's workflow, we ensured that MIRA presents the documents in one batch at the moment that the physician indicates that he has finished entering patient data (instead of after each data entry). Moreover, MIRA provides information only via a minimized window in the task bar. So, the user is never obliged to use the information provided by MIRA. When clicking the window, the retrieved documents are presented to the user. Initially, the user only sees the titles of the documents. By clicking one of the titles, the document is selected and shown in the text area (cf. Figure 3.5).

3.6   Experiments and Results

As mentioned in Section 3.3 we performed two experiments with respect to data translation. The purpose of the first experiment was to tune our automatic translation mechanism (Subsection 3.6.1). In the second experiment the automatic translation mechanism was compared to the manual translation mechanisms (Subsection 3.6.2).

   To build an appropriate test set for each experiment, we extracted a collection of 3349 terms from the EMRs of 82 patients. All terms had one of the three semantic types which are most used in our templates, viz. (a) *Chemical*, (b) *Therapeutic or Preventive Procedure*, and (c) *Disease or Syndrome*. Since many terms occurred in the collection more than once, the doubles were removed to obtain a proper set of terms. The resulting set consisted of 347 distinct terms. One half of the set, the *tuning set* (173 terms), was used for tuning the automatic translation mechanism, whereas the other half, the *test set* (174 terms), was used to compare the two mechanisms. Both experiments and their results are discussed below.

3.6.1 Tuning the Automatic Translation Mechanism

As mentioned in Subsection 3.3.2, automatic translation comprises *preprocessing* and *translation* by means of the UMLS Metathesaurus. As a consequence, tuning the automatic translation mechanism consisted of (a) choosing the best combination of preprocessing steps, and (b) choosing the best UMLS Metathesaurus subset.

   In Subsection 3.3.2, we discussed three preprocessing steps, viz. (i) normalization, (ii) lexical transformation, and (iii) morphosyntactical transformation. Since the three steps have to be executed in a predefined order (i.e., i, ii, iii), four possible combinations were considered in our experiment.

- No preprocessing + UMLS

- Normalization + UMLS

- Normalization + lexical transformation + UMLS

Table 3.5: Configuration of the test runs.

	Test runs											
	1	2	3	4	5	6	7	8	9	10	11	12
UMLS *Metathesaurus subset*												
Dutch Metathesaurus	•	•	•	•								
English Metathesaurus					•	•	•	•				
Complete Metathesaurus									•	•	•	•
Preprocessing steps												
UMLS	•	•	•	•	•	•	•	•	•	•	•	•
Normalization		•	•	•		•	•	•		•	•	•
Lexical transformation			•	•			•	•			•	•
Morphosyntactical transformation				•				•				•

- Normalization + lexical transformation + morphosyntactical transformation + UMLS

With respect to the UMLS Metathesaurus, we may distinguish three distinct subsets that are appropriate for our purpose.

- The Dutch Metathesaurus

- The English Metathesaurus

- The complete (multi-language) Metathesaurus[5]

During the experiment, we applied each of the four possible combinations of preprocessing steps to the three different subsets of the UMLS Metathesaurus. Combining the four combinations of preprocessing steps with the three subsets yielded 12 test runs (see Table 3.5).

We used two different matching methods incorporated in the UMLS software tools: EXACTMATCH, returning the UMLS concept matching the query term exactly, and APPROXMATCH, returning all UMLS concepts matching the query term to a certain degree. During each test run we first attempted to match each term (from the tuning set) by using EXACTMATCH. Then, the remaining terms (that were not matched) were matched by using APPROXMATCH (returning only the best-matching concept). The resulting translations were checked manually for correctness.

The results of the test runs were assessed according to three performance criteria.

- **Percentage of correctly translated terms**
  To obtain a positive assessment, this percentage should be as high as possible.

- **Percentage of incorrectly translated terms**
  To obtain a positive assessment, this percentage should be as low as possible.

---

[5]Basque, Czech, Danish, Dutch, Finnish, French, Hebrew, Hungarian, German, Italian, Japanese, Norwegian, Portuguese, Russian, Spanish, and Swedish

■ Correctly translated ■ Incorrectly translated □ Not translated
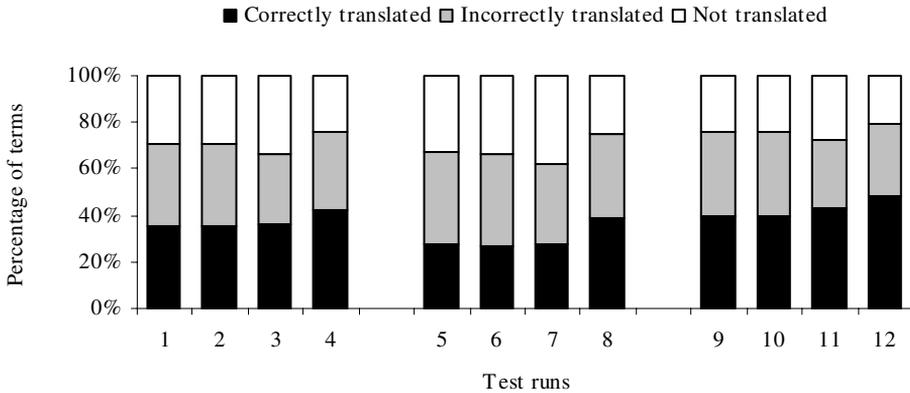
Figure 3.6: Percentage of correctly, incorrectly, and non-translated terms with respect to four different translation mechanisms (numbers on the x-axis refer to the test runs in Table 3.5).

- **Percentage of non-translated terms**
  To obtain a positive assessment, this percentage should be as low as possible. However, since we prefer a non-translated term in favour of an incorrectly translated term, the percentage of non-translated terms may be higher than the percentage of incorrectly translated terms.

The results of the experiments are presented in Figure 3.6. When comparing the three groups of test runs applied to the three different subsets of the UMLS Metathesaurus with respect to the aforementioned performance criteria, four observations stand out.

- Test run 9 outperforms test runs 1 and 5 on two of the criteria.

- Test run 10 outperforms test runs 2 and 6 on all three criteria.

- Test run 11 outperforms test runs 3 and 7 on all three criteria.

- Test run 12 outperforms test runs 4 and 8 on all three criteria.

From these observations it is clear that a test run using the complete UMLS Metathesaurus outperforms its single-language counterparts using only the Dutch or the English subsets of the UMLS Metathesaurus. Consequently, we may conclude that the use of the complete UMLS Metathesaurus is more appropriate for our purpose than the Dutch or English subsets.

When comparing the four groups of test runs using the four distinct combinations of preprocessing steps with respect to the aforementioned performance criteria, three observations stand out.

- The test runs using all preprocessing steps (i.e., 4, 8, and 12) have the highest percentage of correctly classified terms.

- The test runs using two preprocessing steps (i.e., 3, 7, and 11) have the lowest percentage of incorrectly classified terms.

- The test runs using all preprocessing steps (i.e., 4, 8, and 12) have the lowest percentage of non-classified terms.

From these observations it is clear that test runs using all preprocessing steps outperform test runs using only part of the preprocessing steps on two of the three performance criteria. Consequently, we may conclude that the use of all three preprocessing steps is most appropriate for our purpose.

Based on our conclusions that (a) the use of the complete UMLS Metathesaurus, and (b) the use of all three preprocessing steps serve our purpose best, we choose configuration 12 as our mechanism for translation of medical terms.

3.6.2 Comparing the Automatic and Manual Translation Mechanisms

For the comparison of the automatic and manual translation approaches, two criteria are important, viz. *effectiveness* and *scalability*. Even without any quantitative results it seems safe to assume that the automatic translation approach is more scalable than the manual translation approach. The preprocessing steps of the automatic approach are applicable to any EMR. In contrast, the manually constructed mapping is developed for our particular EMR, and consequently, a new mapping has to be constructed for use with another EMR. Therefore, we may state that, with respect to scalability, the automatic translation approach outperforms the manual translation approach.

With respect to effectiveness it may be expected that the manual translation approach outperforms the automatic translation approach. Since the manually constructed mapping is developed specifically for our particular EMR, we expect that the approach translates many terms correctly, resulting in high effectiveness. In contrast, automatic translation is expected to be less effective for small sets; it will classify fewer terms correctly.

To make a comparative assessment between the translation approaches, we have to choose between effectiveness and scalability. For our purpose, effectiveness is more important than scalability, so we are in favour of the manual translation approach. However, if an automatic translation mechanism could be developed that is sufficiently effective, automatic translation might be preferred, since the approach is scalable as well.

To determine whether the automatic translation mechanism is sufficiently effective in comparison to the manual translation mechanism, we let both mechanisms translate the terms in the test set. As mentioned in the introduction of this Subsection, the test set contains 174 terms of the types 'Chemical', 'Therapeutic or Preventive Procedure', and 'Disease or Syndrome'. We recall that the manual

Table 3.6: Evaluation results of AUTOTRANS and MANUTRANS.

| | | AUTOTRANS | | | |
		Correct	Incorrect	Non-translated	Total
MANUTRANS	Correct	42.5	18.4	12.7	73.6
	Incorrect	0	0	0	0
	Non-translated	4.6	15.5	6.3	26.4
	Total	47.1	33.9	19.0	

translation mechanism matches a term to the appropriate UMLS concept using our manually constructed mapping. Then, the UMLS concept is translated into English via the UMLS Metathesaurus (Subsection 3.3.1). Moreover, we recall that the automatic translation mechanism preprocesses the term by using all three preprocessing procedures. Then the term is translated via the complete UMLS Metathesaurus (Subsection 3.3.2).

For the comparison between the two translation mechanisms, the same performance criteria are used as for the tuning of the automatic translation approach, viz. percentage of (a) correctly, (b) incorrectly, and (c) non-translated terms.

The results of the experiment are shown in Table 3.6. From the results three observations can be made.

- The manual translation mechanism results in more correctly translated terms than the automatic translation mechanism (73.6% versus 47.1%).

- By definition, the manual translation mechanism translates no terms incorrectly, versus 33.9% for the automatic translation mechanism.

- The manual translation mechanism results in more non-translated terms than the automatic translation mechanism (26.4% versus 19.0%).

According to these results, we may state that the manual translation mechanism outperforms the automatic translation mechanism on two of the three performance criteria (correctly and incorrectly translated terms). However, to substantiate our statement properly, we continue our research below, by examining each of the performance criteria separately.

Correctly translated terms

When looking at the first performance criterion, percentage of correctly translated terms, we made three observations.

- Of the 47.1% of the terms correctly translated automatically, 90.2% is correctly translated manually as well; of the 73.6% of the terms correctly translated manually, only 57.7% is correctly translated automatically.

- Of the 47.1% of the terms correctly translated automatically, none are incorrectly translated manually by definition; of the 73.6% of the terms correctly translated manually, 25.0% is incorrectly translated automatically.

- Of the 47.1% of the terms correctly translated automatically, 9.8% is not translated manually; of the 73.6% of the terms correctly translated manually, 17.3% is not translated automatically.

From the first observation it is clear that most of the terms correctly translated automatically are correctly translated manually as well, whereas only half of the terms correctly translated manually are correctly translated automatically. The second observation indicates that some of the terms correctly translated manually are incorrectly translated automatically, whereas – by definition – no terms correctly translated automatically are incorrectly translated manually. The third observation shows that the number of terms that are correctly translated manually and cannot be translated automatically is larger than the number of terms that are correctly translated automatically and cannot be translated manually. Based on these three observations we may state that, with respect to the first performance criterion, the manual translation mechanism outperforms the automatic translation mechanism.

Incorrectly translated terms

When looking at the second performance criterion, percentage of incorrectly translated terms, we made three observations.

- Of the 33.9% of the terms incorrectly translated automatically, 54.3% is correctly translated manually.

- Of the 33.9% of the terms incorrectly translated automatically, none are incorrectly translated manually.

- Of the 33.9% of the terms incorrectly translated automatically, 45.7% is not translated manually.

The first observation shows that more than half of the terms incorrectly translated automatically is correctly translated manually. The second observation indicates that no terms incorrectly translated automatically are incorrectly translated manually. From the third observation it is clear that less than half of the terms incorrectly translated automatically are not translated manually. As mentioned before, we prefer non-translated terms over incorrectly translated terms. Based on this preference and the observations above we may state that, with respect to the second performance criterion, the manual translation mechanism outperforms the automatic translation mechanism.

Non-translated terms

When looking at the third performance criterion, percentage of non-translated terms, we made three observations.

- Of the 19.0% of the terms not translated automatically, 66.8% is correctly translated manually; of the 26.4% of the terms not translated manually, 17.4% are translated correctly automatically.

- Of the 19.0% of the terms not translated automatically, none are incorrectly translated manually; of the 26.4% of the terms not translated manually, 58.7% are incorrectly translated automatically.

- Of the 19.0% of the terms not translated automatically, 33.2% is not translated manually; of the 26.4% of the terms not translated manually, 23.9% are not translated automatically.

The first observation indicates that many of the terms not translated automatically are translated correctly manually, whereas only few of the terms not translated manually are translated correctly automatically. From the second observation it is clear that more than half of the terms not translated manually are incorrectly translated automatically, whereas no terms that are not translated automatically are incorrectly translated manually. We note once more that we prefer non-translated terms over incorrectly translated terms. The third observation indicates that there is not a great difference between the percentages of terms that are not translated by one mechanism and not translated by the other mechanism as well. Based on these observations we conclude that, even with respect to the third performance criterion, the manual translation mechanism outperforms the automatic translation mechanism.

Final remarks

The experiment indicated that the manual mechanism outperforms the automatic translation mechanism with respect to all three evaluation criteria. In summary, the automatic mechanism was not sufficiently effective in comparison with our manual approach. Despite the superior scalability of the automatic translation mechanism we chose to use our manual translation mechanism to map terms from the EMR to UMLS concepts, since we consider effectiveness more important than scalability with respect to answering our research questions.

If and when a Dutch standard terminology with a mapping to UMLS becomes available, translation of patient data is no longer necessary. In that case, Dutch terms may be mapped on a language-independent identifier, and from there, translated into any possible language, including English. However, since such a terminology is not yet available, manual translation is still necessary.

3.7 Evaluation

To determine whether our approach for retrieving literature based on patient data is feasible, we evaluated the two outputs of our approach: information needs and

literature. The evaluation criterion for the feasibility of the information-need formulation was the number of information needs. The evaluation criterion for the feasibility of the document retrieval was the number of retrieved documents.

In this section we discuss the patient data used in the evaluation (Subsection 3.7.1), we describe the document collection used (Subsection 3.7.2), we evaluate the number of formulated information needs (Subsection 3.7.3), and we evaluate the number of retrieved documents (Subsection 3.7.4).

3.7.1 Patient Data

In the evaluation we simulated the physicians' use of MIRA. To this end, we employed real patient data, concerning 82 patients and entered by 31 physicians over six months time. The physicians were engaged in four different specialties.

- Anaesthesiology (25 physicians)

- Internal medicine (4 physicians)

- Pulmonology (1 physician)

- Bacteriology (1 physician)

During the simulation all data entries were entered separately and automatically into our simulation EMR in their original order.

Definition 3.10 (Data entry) *A data entry is one single data element in the* EMR.

In total the collection of patient data consisted of 2462 data entries. Each data entry has a type. In our simulation, four types of data entries were used, viz. (a) chemical, (b) disease or syndrome, (c) therapeutic of preventive procedure, and (d) diagnostic procedure. The data entries can be grouped into interactions.

Definition 3.11 (Interaction) *An interaction is a complete series of data entries that were entered into the* EMR *in one session (i.e., from login to logout).*

In total, the 2462 data entries constituted 345 interactions. This leads to an average of 7 data entries per interaction.

After each simulated interaction MIRA formulated information needs and retrieved corresponding documents. For each interaction, we counted (a) the number of information needs formulated and (b) the number of documents retrieved. Based on these numbers we evaluated the number of information needs formulated (Subsection 3.7.3) and the number of retrieved documents (Subsection 3.7.4).

3.7.2   The OHSUMED Document Collection

During our experiments, the use of the complete, online MEDLINE collection did not scale well. To conduct our experiments more efficiently, we used a test collection: the OHSUMED document collection. This collection was obtained by Hersh and Hickam (1994) and Hersh *et al.* (1994). In 2000, the OHSUMED document collection was used for the TREC-9 Filtering Track. The remainder of this section is based on the description of the OHSUMED document collection on the TREC-9 website.[6]

The OHSUMED document collection is a set of references from MEDLINE. The document collection consists of 348,566 document references from 270 medical journals. The documents were collected over a time period of five years (1987–1991). Each document is represented as a collection of fields.

- Sequential identifier

- MEDLINE identifier

- Human-assigned MeSH terms

- Title

- Publication type

- Abstract

- Author (or authors)

- Document source

To evaluate MIRA's performance, we used the set of document references from 1987 (i.e., 54,710 references). The remaining document references are used to tune the approaches discussed in Chapter 4.

3.7.3   Formulated Information Needs

According to Koopmans and Offringa (2003), physicians have an average of 8 questions per day. Therefore, we assume that 10 formulated information needs per interaction (i.e., per day) is a manageable number. Consequently, we distinguish three categories with respect to the number of information needs formulated per interaction: (i) no information needs, (ii) a manageable number of information needs (1 to 10), and (iii) an unmanageable number of information needs (> 10).

The numbers of information needs formulated per interaction are shown in Figure 3.7. On the horizontal axis, the number of formulated information needs per interaction is divided into the three categories discussed above. The vertical axis indicates the number of interactions for each category on the horizontal axis. From Figure 3.7 we make three observations.

---

[6]http://trec.nist.gov/data/t9_filtering/README
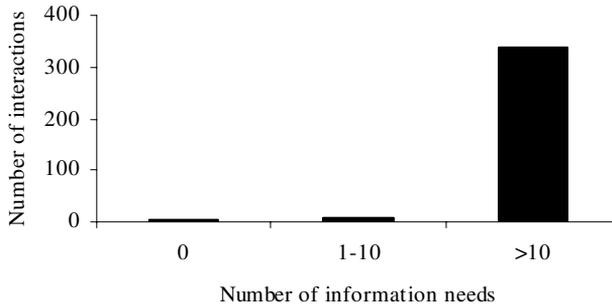
Figure 3.7: Number of interactions for which a specific number of information needs was formulated.

- For 1% of the interactions (5 out of 345), no information needs are formulated at all.

- For 2% of the interactions (7 out of 345), the number of formulated information needs is manageable (1 to 10).

- For 97% of the interactions (338 out of 345), the number of information needs is unmanageable (> 10).

Consequently, the number of information needs formulated per interaction is too high for adequate use.

3.7.4 Retrieved Literature

In the previous subsection, we mentioned that, in the ideal case, MIRA should formulate at most 10 information needs per interaction. Therefore, we assume that physicians prefer a maximum of 10 answers (i.e., relevant documents) per day. Sometimes, no relevant documents can be found with respect to a specific information need. In that case, no documents should be presented to the user. So, the ideal number of documents to be retrieved is 0 to 10. Consequently, we distinguish three categories with respect to the number of retrieved documents per interaction: (i) no documents, (ii) a manageable number of documents (1 to 10), and (iii) an unmanageable number of documents (> 10).

The numbers of documents retrieved per interaction are shown in Figure 3.8. On the horizontal axis, the number of documents retrieved per interaction is divided into the three categories. The vertical axis indicates the number of interactions for each category on the horizontal axis. From Figure 3.8, we make three observations.

- For 25% of the interactions (87 out of 345), no documents are retrieved. This is only useful if there are no relevant documents with respect to the information need (see Section 4.2.1).
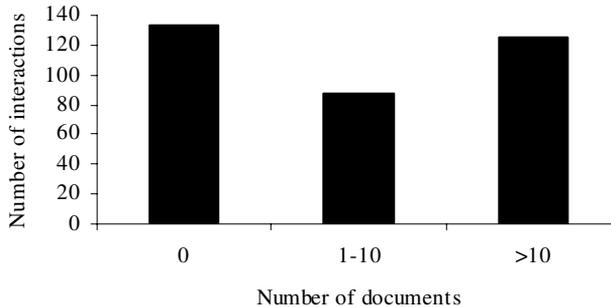
Figure 3.8: Number of interactions for which a specific number of documents was retrieved.

- For 39% of the interactions (133 out of 345), a manageable number of documents is retrieved.

- For 36% of the interactions (125 out of 345), an unmanageable number of documents is retrieved. This is undesirable, since we do not want MIRA to cause a literature overload (see Chapter 4).

Consequently, for only 39% of the interactions, the ideal number of documents (1 to 10) is retrieved. For the other 61% of the interactions, the number of documents retrieved is either too high or no documents are retrieved.

3.8   Chapter Conclusions

In this chapter we designed an approach to get from patient data to retrieved literature. The approach comprises five steps: (1) selecting templates (Section 3.1), (2) extracting appropriate patient data (Section 3.2), (3) translating patient data (Section 3.3), (4) formulating information needs (Section 3.4), and (5) retrieving literature (Section 3.5).

In Subsection 3.3 we mentioned that Dutch non-standardized terms in the EMR present an obstacle to our approach. To address this obstacle, we investigated two different translation mechanisms. An experiment showed that a translation mechanism employing a manually constructed mapping provides us with the best translation results (Section 3.6).

From the evaluation with respect to MIRA (Section 3.7) we may conclude that our approach is adequate and can be generalized to other EMRs, as long as they use a clear information structure that is programmatically accessible. If EMRs are standardized with respect to medical terminology, our approach is even more generalizable. However, the evaluation indicated that the number of formulated information needs per interaction and the number of retrieved documents per interaction is still

high (Subsection 3.7). In this way, physicians would be overloaded with information. So, our approach is not yet sufficiently adequate. To improve the approach, the number of formulated information needs and the number of retrieved documents should be reduced. That is the topic of Chapter 4.

Chapter 4

Preventing a Literature Overload*

In Chapter 3 we established that the number of information needs as formulated by MIRA developed so far is quite high. We remark that a high number of information needs is likely to result in an even higher number of documents retrieved, leading to a *literature overload*. Since we do not want the physicians to be overloaded by literature, MIRA does not yet perform adequately.

To prevent a literature overload, MIRA has to be enhanced in two respects. First, the number of information needs should be reduced. Second, the set of documents retrieved for the formulated information needs should contain as few non-relevant documents as possible (i.e., the precision is high). This chapter discusses both enhancements.

The outline of the chapter is as follows (see Figure 4.1). In Section 4.1 we simulate the formulation of information needs and analyze the formulated information needs to determine the cause underlying the multitude of information needs. Based on this cause Section 4.2 presents two approaches to reduce the number of information needs, viz. discarding information needs and ranking information needs. In Section 4.3, two approaches are discussed to improve on the precision of the set of retrieved documents, viz. patient-related document scoring and patient-related document clipping. Section 4.4 describes experiments conducted to tune the two reduction approaches and the two precision-improvement approaches and shows the results of these experiments. In Section 4.5, we outline the final form of a template as used in MIRA. Section 4.6 presents the final architecture of MIRA. In Section 4.7 we provide the chapter conclusions.
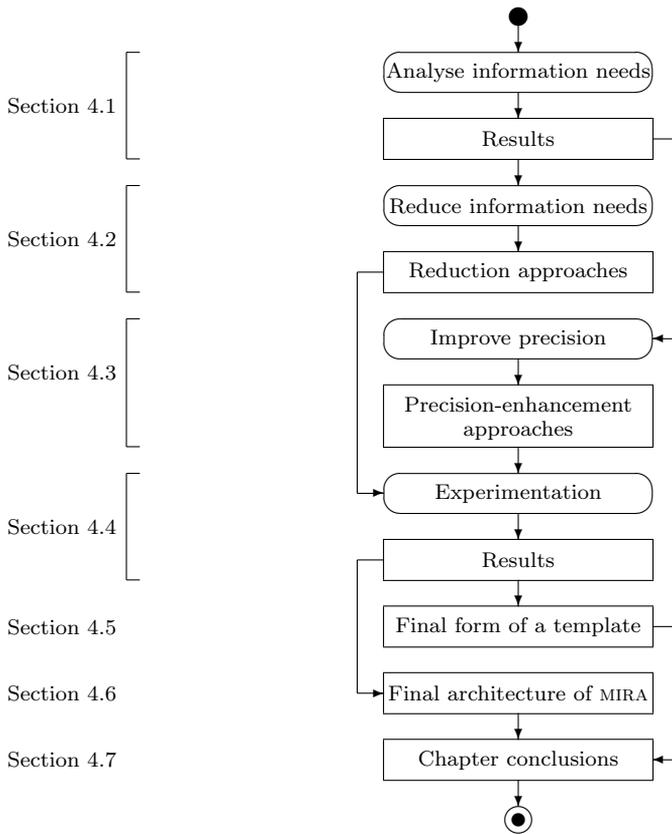
Figure 4.1: Outline of Chapter 4.

4.1   Analysis of Formulated Information Needs

In the previous chapter we concluded that the number of information needs formulated by MIRA is too high for adequate use. The reduction approach to be used depends on the specific cause underlying the multitude of information needs. To determine the nature of this cause, we performed an experimental analysis of the formulated information needs. The results of the analysis provide directions for determining which reduction approach is most appropriate for our purpose. Subsection 4.1.1 describes the set-up of the analysis. In Subsection 4.1.2 the results are presented and analyzed. Subsection 4.1.3 provides a discussion on the subject.

4.1.1   Experimental Set-up

In our experiment we simulated the use of the EMR by physicians. During the simulation all data entries were entered in our simulation EMR separately and automatically. During data entry, the order of the data entries was retained. We recall

from definition 3.10 that a data entry is one separate data element in the EMR. Information needs were formulated after each interaction. Moreover, we recall from definition 3.11 that an interaction is a series of data entries by a specific physician within one day.

The patient data used in the simulation were the same as the data used in Subsection 3.7.1, concerning 82 patients and entered by 31 physicians over six months of time. The patient data included data entries of four different semantic types, namely (a) 'Disease or Syndrome', (b) 'Laboratory Procedure', (c) 'Therapeutic or Preventive Procedure', and (d) 'Chemical'. We confined the analysis to these types, since they represent the most data entries in the EMRs.

During the simulation, two types of data were recorded for each physician separately: (a) the number of information needs generated by each semantic cluster per interaction and (b) the semantic type of each data entry.

4.1.2   Results and Analysis

From the first type of data recorded during the simulation we calculated the average number of information needs generated per interaction by each separate semantic cluster. This calculation was performed for each physician separately. From these results we calculated the number of information needs generated by each semantic cluster, averaged over all interactions and all physicians. The results of these calculations are shown in Figure 4.2.

The results indicate that there are five semantic clusters (13, 23, 26, 27, and 31) which clearly generate more information needs than the other semantic clusters. In Figure 4.3 (which is an adaptation of Figure 3.2) the five clusters are represented by a path ending in a dashed node. All five clusters comprise only one information need. These information needs are shown in Table 4.1. We note that three of the templates contain semantic types that were not considered explicitly in our analysis ('Health Care Activity' and 'Pathologic Function'). However, these semantic types are supertypes of the considered semantic types, and are therefore considered
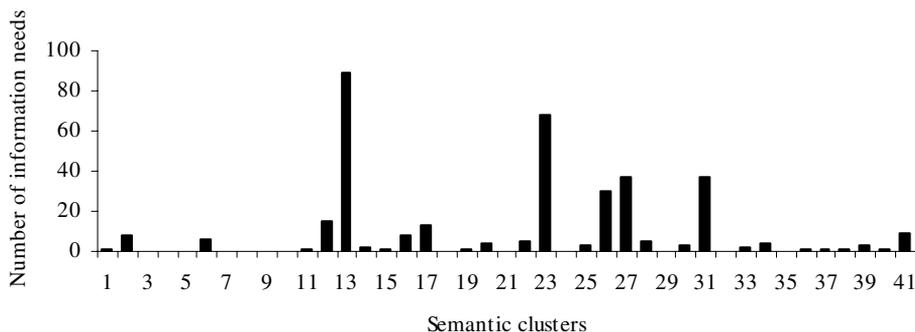
Figure 4.2: Number of templates generated by each semantic cluster.

Root

D 6   FD 12

C 5   H 4   F 3   T 0   D 2   FD 1   L 1   T 6   H 4   F 1

F 1   T 0   F 0   T 0   F 3   T 0   D 0   P 0   T 1   C 1   F 1   H 1

S 1   T 0   S 1   P 0   P 1   C 1   P 1

S 1   S 1

H 9   C 6   T 4   F 0   P 2   S 2   PH 1   PP 1

H 1   SP 2   F 1   C 1   S 3   P 2   L 4   P 1   PH 1

S 1

C = Chemical
D = Disease or Syndrome
F = Finding
FD = Finding / Disease or Syndrome
H = Health Care Activity
L = Laboratory Procedure

P = Pathologic Function
PH = Phenomenon / Health Care Activity
PP = Phenomenon or Process
S = Sign or Symptom
SP = Sign or Symptom / Pathologic Function
T = Therapeutic or Preventive Procedure

Figure 4.3: Semantic clusters, structured as a tree based on the semantic types they contain (cf. Figure 3.2). The five clusters represented by a path ending in a dashed node generate the majority of information needs.

implicitly.

To determine why these semantic clusters generate significantly more information needs, we used the information on the semantic type of each data entry to determine the relative frequency of data entries of the four semantic types. 62% of the data entries were of the semantic type 'Chemical', 25% of the semantic type 'Therapeutic or Preventive Procedure', 7% of the semantic type 'Disease or Syndrome' and 6% of the semantic type 'Laboratory Procedure'. From these results it is clear that data entries of the semantic types 'Therapeutic or Preventive Procedure' and 'Chemical' occur significantly more often than data entries of the the semantic types 'Disease or Syndrome' and 'Laboratory Procedure'. The reason for the differences in relative frequencies is the fact that EMRs contain information on all ICU admissions of the patients. In the course of a lifetime, patients do not generally develop many different diseases, keeping the number of data entries of the type 'Disease or Syndrome'

Table 4.1: Templates within the five semantic clusters (13, 23, 26, 27, and 31) that produce the majority of information needs.

#	Information Need
13	Can <CHEMICAL> be used as prophylaxis against <DISEASE OR SYNDROME> in a patient with <DISEASE OR SYNDROME> and no clinical risk factors for <DISEASE OR SYNDROME> ?
23	Give a description of the advantages and disadvantages of <HEALTH CARE ACTIVITY> in relation to <HEALTH CARE ACTIVITY> in case of <FINDING> /<DISEASE OR SYNDROME> .
26	How can <HEALTH CARE ACTIVITY> be used in the evaluation of <HEALTH CARE ACTIVITY> ?
27	How does <HEALTH CARE ACTIVITY> compare with <HEALTH CARE ACTIVITY> in the setting of <SIGN OR SYMPTOM> /<PATHOLOGIC FUNCTION> ?
31	What is the interaction of <CHEMICAL> with <CHEMICAL> ?

low. However, despite a low number of diseases, patients may receive different kinds of treatments and medications. The reason is that physicians are not always immediately sure what treatment or medication is best for a specific patient, causing treatment changes or medication changes in a later stage. This leads to many entries of the semantic type 'Therapeutic or Preventive Procedure' and 'Chemical'. The number of entries is even enlarged by the frequent occurrence of dosage adjustments, which also involve data entries of the semantic type 'Chemical'.

When analyzing the semantic clusters generating most information needs, we discovered two factors influencing the number of formulated information needs. The first factor is the *number* of semantic types within the templates of a semantic cluster. In general, the larger the number of semantic types, (a) the larger the number of combinations which can be generated by their instances, and (b) the larger the number of information needs formulated from the semantic cluster. However, this is not always true. For instance, cluster 14 contains 4 semantic types (cf. Figure 4.3) but still it generates only 2 information needs (cf. Figure 4.2). The reason is that its semantic types have only a few instances, resulting in a small number of information needs. The second factor is the *kind* of semantic types within the templates of a semantic cluster. In general, semantic clusters involving the semantic types with most data entries ([Therapeutic or Preventive Procedure] and [Chemical]) result in the largest number of information needs. However, this is not always the case. For instance, clusters 30 and 33 both contain one of the semantic types [Therapeutic or Preventive Procedure] and [Chemical] (cf. Figure 4.3), but still they generate only 3 and 2 information needs, respectively (cf. Figure 4.2). The reason is that both clusters use only one semantic type, resulting in a small number of information needs. Our analysis indicated that the large number of information needs formulated

by the five semantic clusters shown in Table 4.1 can be explained by a *combination* of the two factors discussed above. Most information needs are formulated from (1) semantic clusters using a combination of the semantic types [Therapeutic or Preventive Procedure] and [Chemical], and (2) from semantic clusters using the semantic type [Chemical] in combination with at least three occurrences of other semantic types.

4.1.3   Discussion

When analyzing the information needs resulting from the five semantic clusters generating the majority of information needs, we discovered that not all information needs are useful to a physician. In this respect, we made four observations which will be elaborated upon in Subsection 4.2.1.

- Many information needs are formulated repeatedly for a certain physician. These information needs are considered *repeated* and should be discarded.

- Often, information needs contain data entries in the wrong order. For instance, the information need *Does Clarithromycin cause pancreatitis?* is formulated when pancreatitis occurred *before* the administration of Clarithromycin. These information needs are considered *non-sensical* in a specific situation and should be discarded.

- Some information needs formulated by MIRA can be answered by hospital alert systems (see e.g., Raschke *et al.*, 1998; Oppenheim *et al.*, 2000; Degnan *et al.*, 2004; Furman, 2005). These information needs are considered *redundant* and should be discarded.

- Many information needs do not result in any retrieved documents. These information needs are considered *superfluous* and should be discarded.

Consequently, we may conclude that one way to reduce the number of information needs is to *discard information needs* based on the four observations above.

Furthermore, we discovered that not all non-discarded information needs are equally relevant to a physician. In this respect, we made three observations which will be further discussed in Subsection 4.2.2.

- Information needs containing data entries that are rare are more likely to be relevant to a physician than information needs containing data entries that occur often.

- Information needs containing recently entered data are more likely to be relevant to a physician than information needs containing data entries that were entered some time ago.

- Information needs resulting in a small number of documents are more likely to be relevant to a physician than information needs returning many documents.

Based on these observations, information needs can be assigned a measure of relevance. This measure may then be used to rank the information needs. After ranking them, only the most-relevant information needs are selected. Consequently, we may conclude that a second way to reduce the number of information needs is to *rank and select information needs* according to their relevance.

4.2 Reducing the Number of Information Needs

As mentioned in the previous section there are two approaches to reduce the number of information needs: (1) by discarding information needs directly (Subsection 4.2.1) and (2) by ranking the information needs according to their relevance and selecting the most-interesting ones (Subsection 4.2.2). This section discusses both approaches more elaborately.

4.2.1 Discarding Information Needs

As mentioned in Subsection 4.1.3, there are four features that may cause an information need to be discarded. We have associated these features with knowledge types that may be used to determine whether an information need should be discarded.

- *Repeatedness*: formulation history

- *Non-sensicalness*: entry order

- *Redundancy*: domain knowledge

- *Superfluity*: number of retrieved documents

In the remainder of this section each knowledge type is discussed in more detail. To clarify the use of the knowledge types, examples will be provided. All examples are related to example 1.4 introduced in Chapter 1. For convenience, the patient data corresponding to example 1.4 are summarized in Table 4.2.

Formulation History

A physician's formulation history represents the information needs that have been formulated previously. This knowledge can be used to prevent the physician from being provided with information needs (and corresponding literature) he has already seen. These information needs are considered *redundant* and should not be presented again. Since physicians might forget information they have seen before, a time stamp may be associated with each information need. In this way, information needs can be presented anew after a specific time interval. Since the patient data used in our research were entered during only six months time, we deemed the interval too short to legitimate the use of time stamps. Consequently, the current version of MIRA does not incorporate time stamps.

For instance, consider the patient data in Table 4.2. One of the information needs formulated is *What are the side effects of Valsartan?* If the physician prescribes

Table 4.2: EMR related to example 1.4.

	Data entry	Entry time
Diagnosis	• Pancreatitis	09/06/2003 - 17:56:00
	• Pneumonia	07/06/2003 - 13:14:00
	• Heart failure NOS	07/06/2003 - 13:14:00
	• Pulmonary embolism	07/06/2003 - 13:14:00
	• Respiratory tract infections	02/06/2003 - 13:49:00
	• Transient Ischemic Attack	03/11/2002 - 15:24:00
	• Hypothyroidism	09/04/1992 - 16:19:00
	• Chronic Obstructive Airway Disease	15/08/1988 - 11:38:00
Diagnostic Procedure	• Radiography of leg	12/01/2003 - 14:31:00
Medication	• Furosemide	07/06/2003 - 15:41:00
	• Amoxicillin-Potassium Clavulanate Combination	07/06/2003 - 15:41:00
	• Clarithromycin	02/06/2003 - 13:49:00
	• Aspirin	03/11/2002 - 15:31:00
	• Valsartan	03/11/2002 - 15:31:00
	• Thyroxine	09/04/1992 - 16:23:00
	• Chlorthalidone	09/04/1992 - 16:23:00
Therapy	• Oxygen therapy care	07/06/2003 - 09:45:00

Valsartan to another patient, the same information need would be formulated. This information need is considered redundant and should therefore be discarded.

Entry Order

The entry order represents the order in which data are entered into the EMR. This type of knowledge may prevent physicians from being provided with information needs that contain data entries in the wrong order. These information needs are considered *non-sensical* and should be discarded. The entry order is only applicable for templates containing data slots that are temporally dependent.

For instance, consider the patient data in Table 4.2. One of the information needs formulated for the patient under concern is *Does Clarithromycin cause hypothyroidism?* In general, this might be an interesting information need. However, Table 4.2 shows that for this specific patient, the data entry *hypothyroidism* was entered before the data entry *Clarithromycin*. This implies that the disease hypothyroidism occurred before the patient was prescribed the medication Clarithromycin. Consequently, it is unlikely that the Clarithromycin caused the hypothyroidism. As a result, the information need *Does Clarithromycin cause hypothyroidism?* is considered non-sensical and should be discarded in this case. If the data *Clarithromycin*

were entered before the data entry *hypothyroidism*, the hypothyroidism might be a side effect of the Clarithromycin. In that case, the information need is not considered to be non-sensical and should not be discarded.

Domain Knowledge

Domain knowledge represents the standard medical knowledge (e.g., guidelines and systematic reviews). This type of knowledge may be used to prevent physicians from being provided with literature with respect to information needs that are already met by other information sources. For instance, some hospitals use alert systems that automatically forewarn physicians to potentially harmful actions, such as prescribing interacting medications (Raschke *et al.*, 1998; Oppenheim *et al.*, 2000; Degnan *et al.*, 2004; Furman, 2005). The warnings issued by an alert system match some of the information needs formulated by MIRA. Consequently, these information needs are redundant to the physician. According to the physicians we interviewed (Subsection 2.3.3), redundant information is considered rather impeding. Therefore, the redundant information needs should be discarded by MIRA.

For instance, consider the patient data in Table 4.2. In this case, the physician prescribes various medications to our patient. One of the information needs formulated is *Does Valsartan cause respiratory tract infections?* Since a respiratory tract infection is a common side effect, the physician will either be aware of it or be alerted to it by the hospital alert system. Consequently, the information need is redundant and should be discarded. A second information need formulated based on Table 4.2 is *Does Clarithromycin cause pancreatitis?* Despite the fact the pancreatitis is indeed a side effect of Clarithromycin it is not recorded in the domain knowledge (yet). Therefore, no warning will be issued by the alert system and the information need is not met. Consequently, the information need is not redundant and should not be discarded.

Since alert systems use medical domain knowledge as their knowledge source, the standard medical knowledge sources can be employed to predict which warnings will be issued. Based on this information, MIRA can determine whether an information need is redundant and should be discarded. In our research we used the domain-knowledge base G-STANDAARD, developed by Z-INDEX.[1] G-STANDAARD is a medical knowledge base (in text-format) containing information on various subjects, for instance, allergies, dosage, drug interactions, and contraindications. G-STANDAARD is employed by many medical institutions (e.g., general practitioners, pharmacies, and hospitals), mostly in the form of an alert system embedded into existing information systems. As such, the G-STANDAARD is a representative choice of medical domain-knowledge source for our research.

For use with MIRA we ported the G-STANDAARD from text format into a database. For each information need formulated, MIRA queries this database to determine whether the G-STANDAARD would issue an alert that matches the information need. In that case, the information need is discarded.

---

[1]http://www.z-index.nl

Number of retrieved documents

The number of documents resulting from an information need can be used to prevent physicians from being provided with information needs for which no literature can be retrieved. These information needs are considered non-relevant and should be discarded.

Below we give two examples as instructive instances. Consider the patient data in Table 4.2. As a first instance, we consider the information needs formulated from these patient data: *Does Aspirin cause pancreatitis?*. When a literature search is performed based on this information need, no results are retrieved. Therefore, this information need is non-relevant to the physician and should be discarded. A second information need formulated based on the patient data in Table 4.2 is *Is Clarithromycin effective for respiratory tract infections?* When this information need is used as a starting point for IR, several documents are retrieved. Consequently, this information need is relevant and should not be discarded.

4.2.2 Ranking Information Needs

As mentioned in Subsection 4.1.3 there are three factors determining the relevance of an information need. We have translated these reasons into knowledge types that may be used to determine the relevance of an information need.

- Knowledge about a physician's specialism

- Knowledge about data entry time

- Knowledge about the number of results of an information need

These knowledge types may be used to rank the formulated information needs. Only the 10 most relevant information needs are presented to the user.

In the remainder of this section each knowledge type is discussed in some detail. As in the previous subsection, all examples are related to example 1.4 introduced in Chapter 1 and refer to the data in Table 4.2.

Specialism

Each medical specialism has its own frequency distribution for diseases, treatments, prescribed medication, etc. Diseases encountered on a daily basis by a physician in a certain medical specialism may be very rare to a specialist in another domain. This distribution is reflected in the patient data and may be used to determine how interesting an information need is to a certain physician. In general we may state that data entries encountered often are less interesting than data entries encountered rarely. We described the relation between occurrence and interest of data entries in the formula below.

$$\text{interest}_{\text{specialism}} = 1 - \frac{\#\ \text{occurrences}}{\text{total data entries}} \tag{4.1}$$

Based on this equation the interest of a data entry can be calculated. The interest
of an information need is the average interest of all its data entries.

For instance, consider the patient data in Table 4.2. A potential information
need formulated from these data is *Is there a connection between pneumonia and
pulmonary embolism?* Assume the patient record system contains 3000 entries of
which the data entry *pneumonia* occurs 300 times and the data entry *pulmonary
embolism* occurs 100 times. The data entry *pneumonia* has an interest value of 0.9
and the data entry pulmonary embolism has an interest value of 0.97. Consequently,
the interest value of the information need is 0.93. A second information need for-
mulated is: *What is the prognosis in case of pancreatitis.* Assume the data entry
pancreatitis occurs only 5 times in 3000 data entries. Then the interest value of the
data entry (and the information need) is 0.998. Therefore, the second information
need is more relevant than the first one.

Entry Time

The entry time represents the time at which data are entered into the EMR. This type
of knowledge may be used to determine the relevance of specific data entries. Data
entered outside a certain time interval might be considered irrelevant and discarded
during the instantiation of the templates. We describe the relation between age and
relevance of a data entry according to the formula below.

$$\text{interest}_{\text{entry time}} = \frac{1}{\text{data age}} \tag{4.2}$$

We note that the data age is expressed in days. Based on this equation the interest
of a data entry can be calculated. The interest of an information need is the average
interest of all its data entries.

For instance, consider the patient data presented in Table 4.2. One of the in-
formation needs formulated is: *Does Chlorthalidone cause pneumonia?* Table 4.2
shows that the data entries were done 16 years ago (i.e., 5840 days) and two days
ago, respectively. Therefore, the interest value of the data entry *Chlorthalidone* is
0.00017 and the interest value of the data entry *pneumonia* is 0.5. Consequently,
the interest value of the information need is 0.25. A second information need that
can be formulated based on the data in Table 4.2 is: *Is there a connection between
pneumonia and pulmonary embolism.* Both data entries are done two days ago.
Consequently, the interest value of the data entries (and the information need as a
whole) is 0.5. Therefore, the second information need is more interesting than the
first information need.

Number of Retrieved Documents

In the previous subsection we mentioned that the number of retrieved documents
can be used to determine whether an information need is relevant. If the number of
retrieved documents is 0, the information need is labelled non-relevant and should

be discarded. However, if the number of retrieved documents is larger than 0, this number can be used to determine how interesting the information need is. In our interviews with physicians, it was pointed out repeatedly that physicians prefer as few retrieved documents as possible, provided that the documents that are retrieved are relevant. Consequently, we describe the relation between the number of retrieved documents and the relevance of an information need according to the formula below.

$$\text{interest}_{\text{documents}} = \frac{1}{\#\ \text{retrieved documents}} \qquad (4.3)$$

Based on this equation the interest of an information need can be calculated.

Again we present two instances in which the patient data presented in Table 4.2 are considered. A first information need is: *Is there a connection between pneumonia and pulmonary embolism?* When using this information need as a starting point for IR, 23 documents are retrieved. Consequently, the interest of the information need is 0.043. A second information need is: *Does Clarithromycin cause pancreatitis?* This information need results in only 2 documents. Therefore, the interest of this information need is 0.5. Consequently, the second information need is more interesting than the first one.

4.3   Improving Precision

As mentioned in the introduction of the chapter, a second way to prevent a literature overload is to ensure that the number of non-relevant documents that is retrieved is as small as possible. In other words, the precision of the set of retrieved documents should be as high as possible. The precision is calculated as follows.

$$\text{precision} = \frac{\#\ \text{relevant retrieved documents}}{\#\ \text{retrieved documents}} \qquad (4.4)$$

Enforcing a high precision usually has a negative influence on another important measure in information retrieval: recall, which is calculated as follows.

$$\text{recall} = \frac{\#\ \text{relevant retrieved documents}}{\#\ \text{relevant documents}} \qquad (4.5)$$

In our research we concentrate particularly on obtaining a high precision. The reason is that physicians prefer being provided with a few documents that are all relevant over being provided with a large number of documents, many of which are non-relevant. This preference was mentioned by the physicians we interviewed during our research (cf. Subsection 2.3.3).

In our research, we devised two approaches to ensure a high precision, viz. patient-related document scoring and patient-related document clipping. Both approaches will be discussed separately below. In addition to these two approaches, we improve precision by presenting to the user only the most relevant document for each information need.

4.3.1   Patient-Related Document Scoring

The relevance score of a document with respect to an information need is calculated by Lucene, as mentioned in Subsection 3.5.3. Lucene's calculation of a document's relevance score is solely based on the correspondence of the document to the information need. However, as mentioned above, the relevance score of a document should reflect its relevance with respect to the patient as well. To this end, patient-related document scoring can be used. The scoring entails that all patient data are taken into account when calculating the relevance score of a document. To accomplish this, all data entries in the EMR are added to the optional part of the query. This does not imply that only documents containing all added data entries satisfy the query. However, the more of these data entries occur in the document title and/or abstract, the higher the relevance score is. If none of the data entries occur, the document gets the minimal relevance score. If all the data entries are found, the maximal relevance score is granted. Since adding additional patient data to the query does not influence whether a document satisfies the query, patient-related document scoring does not influence the recall.

For instance, a potential information need for the EMR outlined in Table 4.2 may be *Does Clarithromycin cause pancreatitis?* This information need would be transformed into the following query.

```
+(meshterms:"clarithromycin"
meshTerms:"pancreatitis/CI"
meshTerms:"etiology")
title:"caus*"
abstract:"caus*"
title:"induc*"
abstract:"induc*"
```

To facilitate patient-related document scoring, a separate clause is added to the query for each data entry in the EMR (Table 4.2). For instance, for the data entry *Hypothyroidism*, the following clause would be added.

```
title:"hypothyroidism" abstract:"hypothyroidism"
```

We do the same for each synonym of each data entry. For instance, for the synonym *hypothyroid* of the term *hypothyroidism*, the following clause would be added.

```
title:"hypothyroid" abstract:"hypothyroid"
```

We note that not only documents containing all added terms are labelled relevant to the query. In that case, the number of documents retrieved would be quite low, most likely zero. A document is considered relevant to the query as soon as all terms in the clause preceded by the plus sign are found in the appropriate fields of the document. In this example, the terms *clarithromycin*, *pancreatitis*, and *etiology* must

be present in the field *meshTerms*. All the other terms are utilized to determine the extent to which the document is relevant with respect to the query. The more terms are found in the document, the more the document applies to the specific patient and consequently, the more relevant is it.

4.3.2   Patient-Related Document Clipping

In the previous subsection, we argued that the entries from the EMR should occur in the documents. Additionally, we now argue that the MeSH terms from the document must occur in the EMR.

Each document is indexed by two types of MeSH terms: major and minor terms. The major MeSH terms describe the main topics of the document. The minor MeSH terms describe additional topics briefly touched upon in the document. We decided to take into account only the major topics of the document. Furthermore, we only take into account major topics of semantic types that are actually stored in the EMR. All such major topics of the document should occur in the EMR to ensure that the document is not about topics unrelated to the patient. Otherwise, the document will be labelled non-relevant to the query and discarded.

For instance, in the previous subsection the information need *Does Clarithromycin cause pancreatitis?* was formulated based on our EMR (Table 4.2). When executing the corresponding query, MIRA might retrieve an article titled *Acute pancreatitis after clarithromycin and beta-methasone* (Rassiat *et al.*, 2003). However, one of the major topics of this article is *Betamethasone*, which is not present in the EMR of our patient. Consequently, the document is removed from the list of retrieved documents. Another article that might be retrieved by MIRA is titled *Pancreatitis induced by Clarithromycin* (Leibovitch *et al.*, 1996). Its major topics *pancreatitis* and *clarithromycin* are both present in the EMR of our patient. The major topic *etiology* (i.e., the causes of a disease) is not present in the EMR, but it is of the semantic type *Functional Concept*. Since our EMR does not store terms of this semantic type, the term *etiology* does not have to be present in the EMR to label the document relevant. Consequently, the second article is not discarded.

4.4   Experiments and Results

In the previous sections we presented two approaches to prevent a literature overload: (a) a reduction approach, comprising formulation and ranking (Section 4.2) and (b) a precision-improving approach (Section 4.3). Each approach involves a number of knowledge types, and the sequence of approaches eventually leads to the retrieval of a manageable number of documents (see Figure 4.4). However, it is not entirely clear which combination of knowledge types performs best for each approach.

To tune both approaches (i.e., to determine the optimal combination of knowledge types), we performed three subjective experiments. The purpose of the first and second experiment was to determine the best reduction approach. The purpose of the third experiment was to determine the best precision-improving approach.
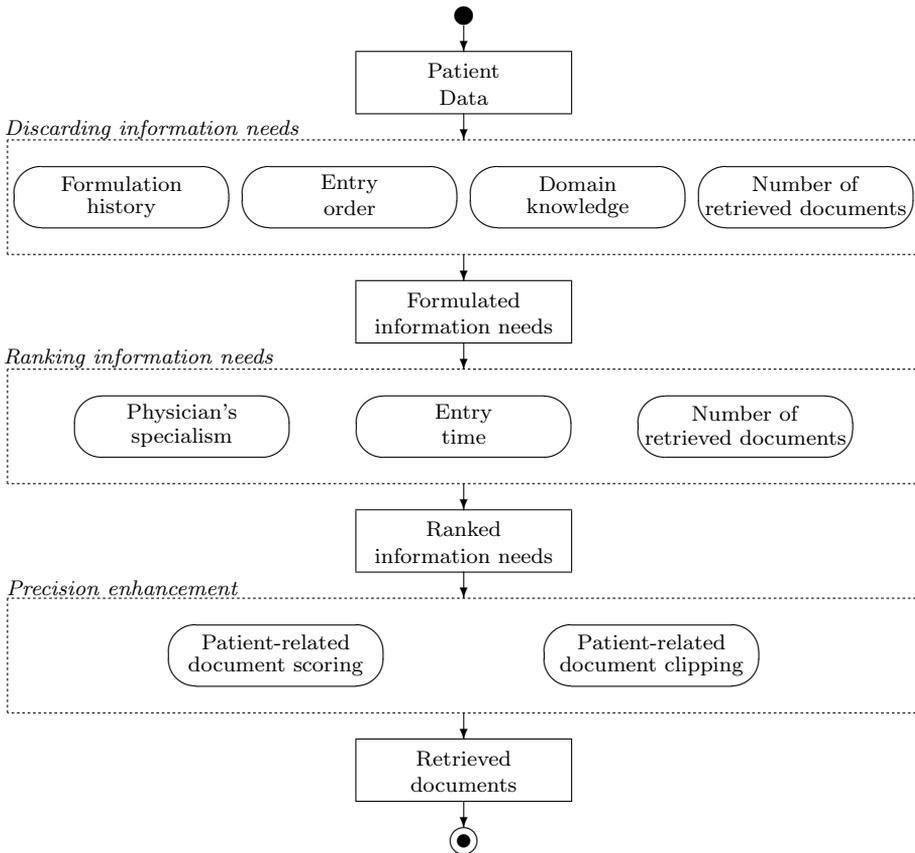
Figure 4.4: Schematic representation of the approaches to prevent a literature overload and their knowledge types.

In Subsection 3.7.2 we already described the test collection that was used in the abovementioned experiments. In Subsection 4.4.1 we present the first experiment. Subsection 4.4.2 discusses the second experiment.

4.4.1   Tuning the Reduction Approach

As mentioned in Subsection 4.2.1, the number of formulated information needs can be reduced in two ways: (a) discarding information needs and (b) ranking information needs. Each approach involves a number of knowledge types. However, it is not entirely clear which combination of knowledge types performs best for each approach. To tune both approaches (i.e., to determine the optimal combination of knowledge types), we performed two experiments. The purpose of the first experiment was to determine the best approach to discard information needs. The purpose of the second experiment was to determine the best approach to order information

Table 4.3: Settings in which the reduction approach was assessed.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Formulation history		•				•	•	•				•	•	•		•
Entry order			•			•			•	•		•	•		•	•
Domain knowledge				•			•		•		•	•		•	•	•
# Retrieved documents					•			•		•	•		•	•	•	•

needs and to select the best. In this subsection, both experiments are described separately.

Discarding Information Needs

In Subsection 4.2.1 we presented four knowledge types that may be used to determine whether an information need should be discarded.

- Formulation history

- Entry order

- Domain knowledge

- Number of retrieved documents

To determine the actual performance of these knowledge types (and combinations thereof) we conducted the first experiment. We let our agent formulate information needs using various combinations of the knowledge types. Since the knowledge types can be combined in any possible way, sixteen test runs were conducted (see Table 4.3).

We recall from Subsection 3.7.3 that our goal is to formulate a maximum average of ten information needs per interaction. Based on this goal we used one performance criterion when assessing the results of each of the test runs: the average number of information needs formulated per interaction.

The results of the experiment are shown in Figure 4.5. In this figure, the average number of information needs formulated per interaction is shown for each of the sixteen test runs. When assessing the results in Figure 4.5, we made five observations, the first four regarding the individual knowledge types.

- The use of the number of retrieved documents (test run 5) produces the largest reduction. This knowledge type reduces the average number of information needs from 315 to 5 (98.3% reduction).

- The use of the formulation history (test run 2) is the second-best knowledge type. The formulation history reduces the average number of information needs from 315 to 229 (27.3% reduction).
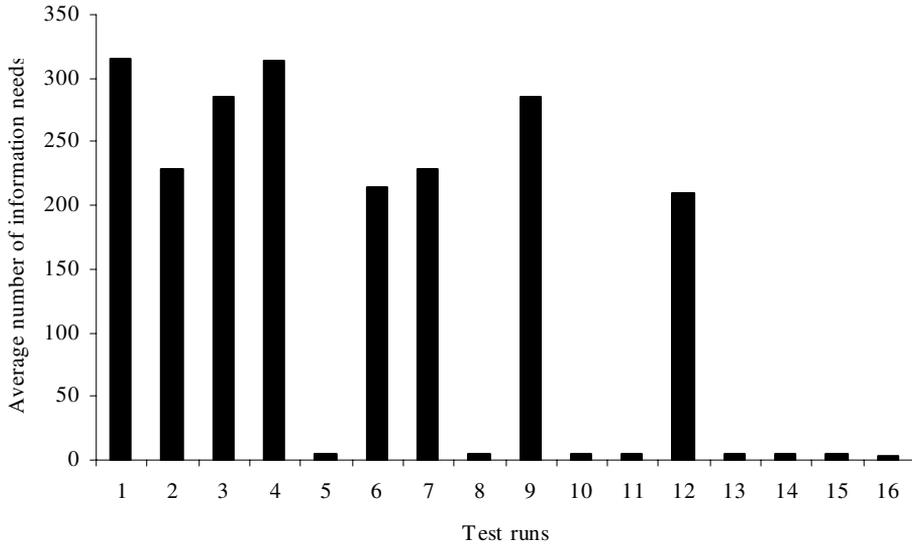
Figure 4.5: Average number of information needs formulated per interaction using sixteen combinations of knowledge types. The numbers of the test runs refer to Table 4.3.

- The use of the entry order (test run 3) is the third-best knowledge type. The entry order reduces the average number of information needs from 315 to 286 (9.2% reduction).

- The use of domain knowledge (test run 4) is the worst-performing knowledge type. Use of domain knowledge reduces the average number of information needs from 315 to 314 (0.3% reduction). This marginal performance might point out that the domain knowledge is too general with respect to the domain of anaesthesiology.

- Each combination of knowledge types results in a larger reduction than each of its subcombinations (e.g., compare test run 6 with test runs 2 and 3). Consequently, a combination of all four knowledge types (test run 16) results in the largest reduction from 315 to 3 (99.2% reduction).

From the five observations above we may draw two conclusions. First, we may conclude that the knowledge types can be ordered based on their performance. The order is (from best to worst): (1) number of retrieved documents, (2) formulation history, (3) entry order, and (4) domain knowledge. Second, we may conclude that a combination of all four knowledge types is the best approach to discard information needs. Therefore we choose this combination as our approach to discard information needs.

Table 4.4: Settings in which the ranking approach was assessed.

	R1	R2	R3	R4	R5	R6	R7	R8
Specialism		•			•	•		•
Entry time			•		•		•	•
# Retrieved documents				•		•	•	•

Ranking Information Needs

Above we observed that the average number of formulated information needs is reduced to 3 per interaction. Since this is a manageable number (see Subsection 3.7.3), ranking information needs and discarding all but the 10 most relevant ones might not be considered necessary. However, for some interactions, more than 10 information needs are formulated. This concerns interactions in which a large number of patient data are entered into the EMR (mostly the first interaction for a new patient). For these interactions, ranking and discarding information needs is highly beneficial.

In Subsection 4.2.2 we presented three types of knowledge that may be used to determine the interest value of each information need and rank the information needs accordingly.

- Specialism

- Entry time

- Number of retrieved documents

To determine the actual performance of these knowledge types (and combinations thereof) we conducted a second experiment. We let our agent formulate information needs using the discarding approach chosen in the previous subsection. To rank the formulated information needs, we used various combinations of the knowledge types presented in Subsection 4.2.2. Since these knowledge types can be combined in any possible way, eight test runs were conducted (see Table 4.4). Test run 1 uses none of the knowledge types and ranks the information needs at random. For test runs 2–8 (combining several knowledge types), the relevance of an information need is calculated as the average of the relevances according to the combined knowledge types.

When assessing the information needs selected by the ordering approach, we assumed that a physician prefers a selection of information needs (a) that contains as many distinct questions as possible and (b) that concerns as many distinct subjects as possible. Based on this assumption, we assessed the selected information needs using two criteria that we consider equally important. The first criterion is the *number of distinct questions* within the set of selected information needs. Two information needs are considered distinct questions if they are based on different templates. The second criterion is the *number of distinct subjects* within the set of selected information needs. Two information needs are considered to concern distinct subjects if their templates are instantiated with distinct active concepts. For instance, consider the information needs below.

Information Need 4.1 *What is the dose of Clarithromycin for this specific patient?*

Information Need 4.2 *What are the side effects of Clarithromycin for this specific patient?*

Information needs 4.1 and 4.2 are based on different templates (dosage vs. side effects). Consequently, they are distinct questions. However, they do concern the same subject (Clarithromycin).

Information Need 4.3 *What are the side effects of Amoxicillin for this specific patient?*

Information needs 4.2 and 4.3 concern different subjects (Clarithromycin vs. Amoxicillin). Consequently, they concern distinct subjects. However, they are based on the same template (side effects of Chemical).

Information Need 4.4 *What is the dose of Amoxicillin for this specific patient?*

Obviously, information needs may be distinct questions and distinct subjects at the same time. This is the case with Information needs 4.2 and 4.4 (side effects vs. dosage and Clarithromycin vs. Amoxicillin). However, information needs cannot be similar questions concerning similar subjects. In that case, the information needs would be equal. This is not possible since MIRA only formulates proper sets of information needs.

For each test run we calculated the average value per interaction for both criteria. The results of the experiment are shown in Figure 4.6. In this figure the average number of distinct questions and subjects are shown for each of the seven test runs. From the analysis of the results we made six observations.

- Test runs involving knowledge concerning the physician's specialism (test runs 2, 5, 6, and 8) result in a low variety of questions.

- Test runs involving knowledge concerning the entry time (test runs 3, 5, 7, and 8) result in a mixed variety of questions and subjects.

- Test runs involving knowledge concerning the number of retrieved documents (test runs 4, 6, 7, and 8) result in a high variety of subjects.

- Test run 1 (random ranking) results in a rather low variety of questions and an intermediate variety of subjects.

- Test run 2 results in the lowest variety of questions, whereas test run 3 achieved the highest variety of questions.

- Test run 3 results in the lowest variety of subjects, whereas test run 4 achieved the highest variety of subjects.
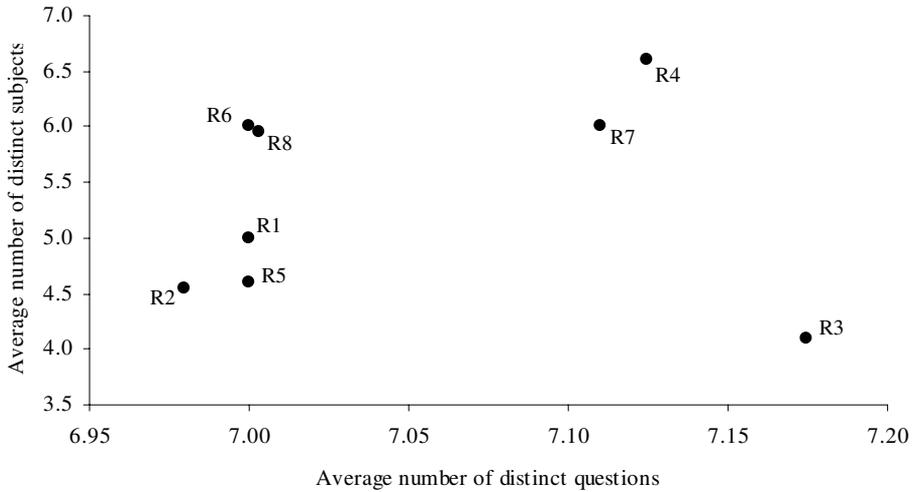
Figure 4.6: Average number of distinct questions and subjects for seven combinations of knowledge types. The numbers of the test runs refers to Table 4.4.

From these observations we may draw three conclusions. First, we may conclude that the use of the number of retrieved documents leads to a high variety of subjects. Second, we may conclude that the use of specialism leads to a low variety of questions. Third, we may conclude that the use of only the number of retrieved documents achieves an optimal combination of variety of questions and subjects. This is illustrated by the fact that the number of distinct questions plus the number of distinct subjects is optimal (i.e., highest) for use of the number of retrieved documents only (test run 3). Therefore, we expect the ordering approach to perform best when using only this knowledge type. However, we will perform an objective experiment based on assessments of physicians as well (see Subsections 5.3.3, 5.4.3, and 5.5.3).

Table 4.5: Four settings in which the precision enhancement approaches were assessed.

	A	B	C	D
Patient-related document scoring		●		●
Patient-related document clipping			●	●

4.4.2   Tuning the Precision-Enhancement Approach

According to Subsection 4.3 there are two approaches to enhance the precision of the set of retrieved documents: (a) patient-related document scoring and (b) patient-related document clipping. To find an optimal combination of both approaches we tested four combinations of these approaches (see Table 4.5). For each of the four combinations we determined the performance by manually assessing the relevance of each document with respect to the information need for which it was retrieved. Based on the relevance assessments, for each of the four combinations we calculated (a) the relative recall and (b) the precision of the document set. The results of the experiment are shown in Figure 4.7, in which each data point represents an interaction tested in a specific setting (see Table 4.5). Since, many interactions resulted in similar recall-precision combinations, there is considerable overlap among the data points. From the results we made four observations.

- Patient-related document scoring has hardly any influence on relative recall or precision.

- Patient-related document clipping increases relative recall with at most 33% for low numbers of retrieved documents; patient-related document clipping decreases relative recall with at most 4% at higher numbers of retrieved documents.

- Patient-related document clipping increases precision with at most 56% for low numbers of retrieved documents and at least 30% for higher numbers of retrieved documents.
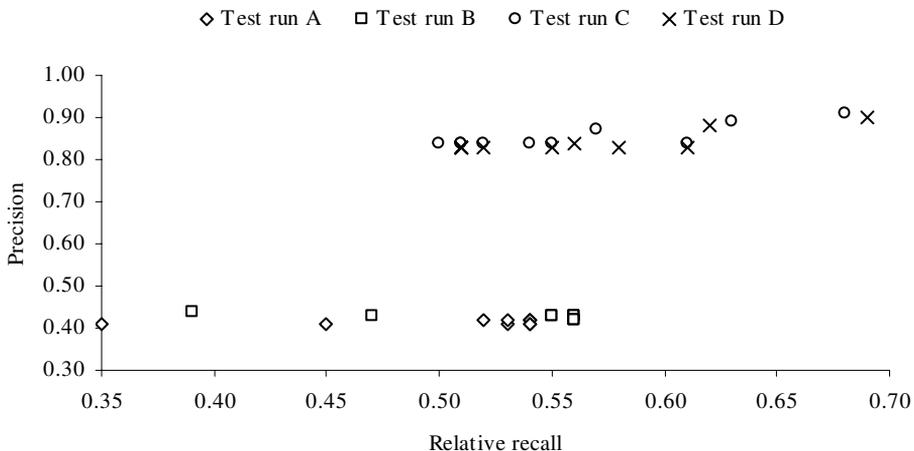
Figure 4.7: Relative recall and precision for four distinct precision-enhancement approaches. Each data point represents an interaction. The names of the test runs refer to Table 4.5.

- The number of retrieved documents has a higher influence on relative recall than on precision.

Based on these observations we may conclude that we may expect the precision to be highest when using only patient-related document clipping. However, we will perform a more objective experiment with assessments from physicians as well (see Subsection 5.3.1, 5.4.1, and 5.5.1).

4.5   Final Form of a Template

In this section, we describe the final form of a template, after incorporating all methods to prevent a literature overload. Each template is stated in XML and contains a number of predefined elements. Below, we provide an example by outlining the XML version of Template 4.5.

Template 4.5 *Is* <Chemical> *effective for* <Disease or Syndrome>*?*

```
<template>
<text>Is </text>
<slot>
   <st>Chemical</st>
   <mq>PD</mq>
   <order>1</order>
</slot>
<text> effective for </text>
<slot>
   <st>Disease or Syndrome</st>
   <mq>DT</mq>
   <order>0</order>
</slot>
<text>?</text>
<query>
   title:effect*
   meshTerms:effective
   abstract:effect*
</query>
</template>
```

The <text> element stores the text elements of the template. These elements are not explicitly used in the retrieval process. Nevertheless, they provide the possibility to list the template in natural language.

The <slot> element indicates the slots of a template. Within each <slot> element, there are one or more <st> elements, indicating the semantic types within each slot. Furthermore, each <slot> element contains a <mq> element, indicating the MeSH qualifier associated to the semantic type (see Subsection 3.5.2). Moreover,

each <slot> element that is temporally dependent contains an <order> element, indicating the proper chronological order of the slots in the template with respect to their entry time (see Subsection 4.2.1).

The <query> element contains the fixed query terms of the template. It specifies the query terms and the fields in which the terms should be found (see Subsection 3.5.2).

4.6   The Final Architecture of MIRA

Now we can describe the final architecture of MIRA as outlined in Figure 4.8.

The central module in MIRA is the *information-need formulator*. This module formulates the information needs. In this process, the information-need formulator receives information from five other modules, viz. (a) INMOD, (b) the data translator, (c) the information-need discarder, (d) the information-need ranker, and (e) the literature searcher.

The information-need formulator receives information on the available templates from INMOD. These templates are instantiated in order to formulate patient-related information needs.

From the *data translator*, MIRA receives translated patient data from the EMR. These data are used to instantiate the templates. As a consequence, the data translator receives data from the EMR directly, whereas the information-need formulator
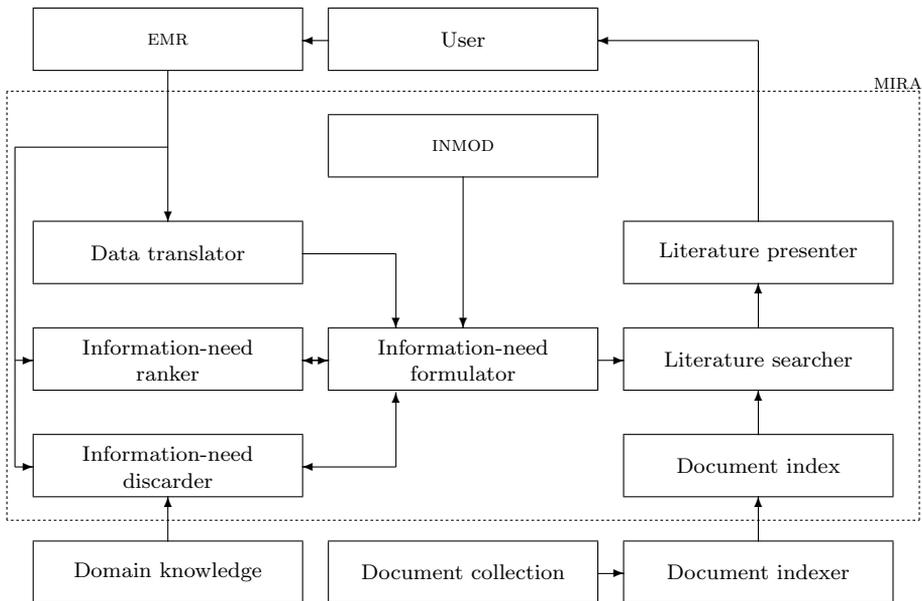
Figure 4.8: Schematic representation of the final version of MIRA. The arrows indicate flow of information.

only receives data from the EMR indirectly.

To discard information needs, the information-need formulator receives data from the *information-need discarder*. The data employed in this process are provided to the information-need discarder by the EMR, the information-need formulator, and a knowledge base containing domain knowledge.

From the *information-need ranker*, the information-need formulator receives information on the relevance ranking of the formulated information needs. To rank the information needs, the information-need ranker receives data from the EMR and the information-need formulator.

To start a literature search, the information-need translator presents the formulated information needs to the *literature searcher*, which retrieves appropriate literature from a *document collection* outside MIRA. However, the literature searcher does not receive data information from the document collection directly. Instead, it searches the *document index* (i.e., an index of the document collection), generated by the *document indexer*.

Once the appropriate documents have been retrieved, the literature searcher passes them to the *literature presenter*. From there, the literature is presented to the user.

From the outline in Figure 4.8, it is quite clear that the only direct interaction of the user with MIRA is via the *literature presenter*. The interaction consists of the user reading the literature presented by MIRA. Furthermore, the user interacts indirectly with MIRA via the EMR. Consequently, the only input from the user into MIRA is provided implicitly by entering data in the EMR. As mentioned earlier, the actual data entry by a physician into the EMR triggers MIRA to start IR.

4.7   Chapter Conclusions

In this chapter we discussed how the number of information needs can be reduced to a manageable size. With respect to these investigations we may draw five conclusions.

First, we may conclude that there are two ways to reduce the number of information needs: (a) discarding information needs directly, and (b) ranking information needs and selecting the ten most-relevant ones (Section 4.1).

Second, we may conclude that there are four knowledge types that may be used to determine whether an information need should be discarded, viz. (a) formulation history, (b) entry order, (c) domain knowledge, and (d) number of retrieved documents (Subsection 4.2.1).

Third, we may conclude that there are three knowledge types that may be used to determine the relevance of an information need: (a) specialism, (b) entry time, and (c) number of retrieved documents (Subsection 4.2.2).

Fourth, we may conclude that there are two ways to improve the precision of the set of retrieved documents: (a) patient-related document scoring and (b) patient-related document clipping (Section 4.3).

Fifth, we may conclude that the optimal formulation performance of MIRA is reached by using all four knowledge types as a discarding approach. According to

the results of our tuning experiment, this discarding approach reduces the average number of formulated information needs to 3 (99.2% reduction). Consequently, the discarding approach succeeds in reducing the formulated information needs to a manageable number. Furthermore, optimal ranking performance is expected by using knowledge concerning the number of retrieved documents (Subsection 4.4.1). Moreover, we expect optimal retrieval performance by using patient-related document clipping as our precision-enhancement approach (Subsection 4.4.2).

Chapter 5

Evaluation

In this chapter we provide an evaluation of MIRA. In Section 4.4 we have seen that
for tuning purposes it suffices to evaluate MIRA in a subjective way. However, for
obtaining an adequate insight into the actual performance, MIRA has to be evaluated
objectively, preferably by its future users: the physicians. An objective evaluation
is the topic of this chapter.

The outline of the chapter is as follows. In Section 5.1 we establish the four
criteria by which MIRA will be evaluated. Section 5.2 describes the physicians who
assessed the performance of MIRA. In Section 5.3 we describe the experimental set-
up. Section 5.4 discusses how the results obtained during the experiment are used
to determine the outcome of the application of each separate evaluation criterion. In
Section 5.5 the evaluation results and the final performance of MIRA are presented.
Section 5.6 discusses the evaluation results. Finally, in Section 5.7 we provide our
chapter conclusions.

5.1 Evaluation Criteria

As explained in the previous chapters, our retrieval system works in three steps: (1)
formulating information needs, (2) ranking the formulated information needs accord-
ing to relevance, and (3) retrieving documents based on the formulated information
needs. In the evaluation, we will assess the performance of MIRA on each of these
three steps separately. As an additional evaluation step, we will evaluate (4) the
unobtrusiveness of MIRA. Consequently, we established four evaluation criteria (see
Figure 5.1).

For evaluation purposes, it is important to note that the evaluation of the various
steps in the retrieval process and their corresponding evaluation criteria depend on
each other (see again Figure 5.1). Here we see that the evaluation of the *unobtru-
siveness* is dependent on the evaluation of the ranking, the formulation, and the
retrieval. For the other three evaluation criteria the dependency is as follows. The
evaluation of the *ranking* is dependent on the evaluation of the retrieval and the

```
┌─────────────────────────┐        ┌───────────────────────────┐
│  Retrieval performance  │───────▶│  Formulation performance  │
└─────────────────────────┘        └───────────────────────────┘
          │         ┌──────────────────────────┐         │
          │         │    Ranking performance    │◀────────│
          │         └──────────────────────────┘         │
          │                      │                        │
          │         ┌──────────────────────────┐         │
          └────────▶│      Unobtrusiveness      │◀────────┘
                    └──────────────────────────┘
```
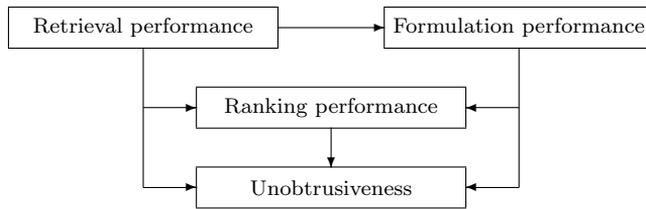
Figure 5.1: Dependencies among the four evaluation criteria of the retrieval process. An arrow from A to B denotes that B is dependent on A.

formulation (cf. Subsection 4.2.2). The evaluation of the *formulation* is dependent on the evaluation of the retrieval (cf. Subsection 4.2.1). The evaluation of the *retrieval* is independent of the evaluation of all other evaluation criteria. We note that the above-mentioned dependencies only hold for the evaluation criteria, not for the execution of the corresponding steps in the retrieval process.

To obtain reliable evaluation results despite the dependencies, we have to evaluate the steps in a specific order. Each evaluation criterion must be applied after the application of the criteria it depends upon. As a result the evaluation of the various steps does not match the order in which the steps are executed by MIRA. The evaluation criteria to be applied are given below in their proper order.

1. Retrieval performance

2. Formulation performance

3. Ranking performance

4. Unobtrusiveness

For the application of these four criteria, we conducted a single experiment. The main reason is to ensure that the time to be spent by the (busy) physicians is minimal. From the results of this single experiment, the performances of MIRA under all four criteria were calculated. We are grateful for the cooperation by the physicians.

5.2   Evaluation Persons

For the evaluation we were able to find nine qualified physicians to evaluate MIRA. The physicians are in different phases of their medical careers and have different levels of medical experience. These differences render the physicians particularly suitable as evaluation persons for MIRA, since we expect different results based on the various levels of medical experience (Fenichel, 1981; Borgman, 1987).

We distinguish three type of physicians, viz. *attending physicians*, *resident physicians*, and *recently graduated physicians*. For each type, three physicians were willing

Table 5.1: Physicians participating in the evaluation of MIRA.

Dutch translation	Type of physician	Experience level
Medisch specialist	Attending physician	High
Arts-assistent	Resident physician	Mediate
Basisarts	Recently graduated physician	Low

to participate in the evaluation. An attending physician (in Dutch: *medisch specialist*) has completed residency and practices medicine in a clinic or hospital, often focusing on the specialty learned during residency. Attending physicians have a high level of medical experience, due to many years of practice in the medical field. A resident physician (in Dutch: *arts-assistent*) is a physician in a stage of postgraduate medical training certification in a primary care or referral specialty. Due to their moderate experience in the medical field, resident physicians have an intermediate level of experience. A recently graduated physician (in Dutch: *basisarts*) is a physician who has just received his medical degree. The experience level of recently graduated physicians is low, because they have little practice in the medical field.

Table 5.1 summarizes the different types of physicians participating in our evaluation. All physicians were affiliated to the anaesthesiology department of the Catharina-ziekenhuis in Eindhoven, the Academic Medical Centre in Amsterdam, or the academisch ziekenhuis Maastricht in Maastricht.

5.3   Experimental Set-up

To evaluate MIRA's performance on the four criteria, four assessments had to be made, viz. (1) of the *relevance of the retrieved documents* (used in the evaluation of the retrieval performance), (2) of the *relevance of the formulated information needs* (used in the evaluation of the formulation performance), (3) of the *ranking of the formulated information needs* (used in the evaluation of the ranking performance), (4) of the *unobtrusiveness of* MIRA. Consequently, the experiment comprised four parts: *Assess*1, *Assess*2, *Assess*3, and *Assess*4.

- *Assess*1: relevance of the retrieved documents (Subsection 5.3.1)

- *Assess*2: relevance of the formulated information needs (Subsection 5.3.2)

- *Assess*3: ranking of the information needs (Subsection 5.3.3)

- *Assess*4: unobtrusiveness of MIRA (Subsection 5.3.4)

Each part is discussed separately below.

5.3.1   Assessment of Relevance of the Retrieved Documents

The purpose of *Assess*1 is to obtain a relevance assessment for each retrieved document. The assessments are used in the evaluation of the retrieval performance (see

Table 5.2: *Assess*1: Assessment data presented to the physicians (comprising four data entries and three retrieved documents).

Patient X – Day 1

Data type	Data	Entry time
Disease or Syndrome	Coronary Artery Bypass	18:23:00
Disease or Syndrome	Kidney Failure, Acute	18:23:06
Chemical	Captopril	18:32:45
Chemical	Acetylcysteine	18:32:54

Indicate (on the evaluation form) for each article whether you would like to read the complete article (based on the title and abstract).

1. Predictors, prevention, and long-term prognosis of atrial fibrillation after coronary artery bypass graft operations.
Multiple trials have suggested the use of digoxin, digoxin and propranolol, or timolol to prevent atrial fibrillation after coronary artery bypass grafting. No trial has evaluated the efficacy of digoxin versus propranolol.

2. Pathogenesis and treatment of acute renal failure.
This article reviews the current understanding of the pathophysiologic sequence of events that culminate in an acute renal insult. The use of urinary indices to differentiate the physiologic causes of oliguria, namely, diminished intravascular volume or renal perfusion, from an established acute renal failure, is discussed for children and adolescents, as well as for neonates.

3. Cough due to captopril.
No abstract available.

Subsection 5.4.1). According to Saracevic (1996) there are five manifestations of relevance. For our purpose, we employ the notion of *situational relevance* or *utility*, which is defined as follows by Saracevic (1996).

Definition 5.1 (Situational Relevance) *Situational relevance is the relation between on the one hand, the situation, task, or problem at hand, and on the other hand, texts retrieved by a system or in the file of a system, or even in existence.*

Criteria by which situational relevance is inferred are: usefulness in decision making, appropriateness of information in resolution of a problem, and reduction of uncertainty. Other criteria may be introduced. To obtain the relevance assessments, we provided the physicians with patient data of one specific patient: *Patient X* (see upper part of Table 5.2) and the corresponding documents retrieved by MIRA (see lower part of Table 5.2). We recall from Subsection 3.7.2 that the document collection is a set of references from MEDLINE dating from 1988–1991. The patient data were divided into interactions (i.e., days). Table 5.2 represents the first interaction: *Day 1*. For this specific interaction we see four data entries (upper part of Table 5.2). Each data entry includes the semantic type (e.g., 'Disease or Syndrome'), data (e.g., Kidney Failure, Acute), and entry time (e.g., *18:23:00*) and three retrieved documents (lower part of Table 5.2). Due to space limitations, Table 5.2 shows only

Table 5.3: *Assess*1: Assessment form.

Patient X – Day 1

Indicate for each article whether you would like to read the complete article (based on the title and abstract)	Absolutely not	No	Yes	For sure	*If not, why not?*	The information in this article is known to me	The title and abstract suffice	The article is not relevant w.r.t. this patient
1. Predictors, prevention, and long-term prognosis of atrial fibrillation after coronary artery bypass graft operations.	☐	☐	☐	☐		☐	☐	☐
2. Pathogenesis and treatment of acute renal failure.	☐	☐	☐	☐		☐	☐	☐
3. Cough due to captopril.	☐	☐	☐	☐		☐	☐	☐

the titles of the documents and the first two sentences of their abstracts. However, in the actual experiment the physicians were provided with the complete abstracts to obtain a more reliable relevance assessment. In total, 11 tables (i.e., 11 days) similar to Table 5.2 were presented to the physicians (see Appendix D.1).[1]

Each physician was asked to indicate his assessments of the documents on an assessment form. Table 5.3 shows the assessment form of Patient X for Day 1. The left-hand side of the assessment form shows the titles of the documents (equal to the documents in Table 5.2). The right-hand side of the assessment form shows the possible assessments. The task of a physician during *Assess*1 was to indicate for each document whether he would like to read the entire article with respect to the specific patient. Assessments had to be given on a four-point scale (*absolutely not*, *no*, *yes*, *for sure*). We used a four-point scale because this type of scale avoids problems encountered in a five-point scale or a two-point scale. On the one hand, a five-point scale provides the physician with considerable freedom of choice, but it suffers from central tendency bias, i.e., the tendency of people to choose non-extreme assessments (Hully *et al.*, 2001). On the other hand, a two-point scale

---

[1]All data, except document titles and abstracts, were presented to the physicians in Dutch.

forces an explicit assessment, but leaves the physician with little freedom of choice. A four-point scale combines the advantages of both scales and thus constitutes an adequate compromise (Robertson, 1981).

When a document received a negative assessment (*absolutely not* or *no*), the physician had to indicate the reason for this assessment. The additional information was necessary to prevent documents from being labelled non-relevant for the wrong reasons. Possible reasons for negative assessment were as follows (see Table 5.3).

- The information in this article is known to me.

- The title and abstract suffice.

- The article is not relevant with respect to this patient.

From the assessments we derived the relevance of the documents with respect to the specific patient. Ultimately we were interested in whether a document was relevant with respect to the patient under investigation. Consequently, we had to map the answers from the four-point scale to a two-point scale. This was done as follows.

- If the assessment was *yes* or *for sure*, the document was labelled relevant.

- If the assessment was *no* or *absolutely not*, and the reason for this assessment was *The information in this article is known to me* or *The title and abstract suffice*, the document was labelled relevant.

- If the assessment was *no* or *absolutely not*, and the reason for this assessment was *The article is not relevant w.r.t. this patient*, the document was labelled non-relevant.

As a consequence, a document is labelled non-relevant only when the physician explicitly indicates that the document is not relevant with respect to the specific patient. This is exactly the criterion in which we are interested. Documents that are labelled non-relevant for other reasons could be eliminated by MIRA in other ways. For instance, documents that are known to the physician can be eliminated by employing a user model. However, this is outside the scope of our research. In accordance with the evaluation data, the physicians were provided with 11 tables (i.e., 11 days) similar to Table 5.3 for Patient X (see Appendix D.2).

5.3.2   Assessment of Relevance of the Information Needs

The purpose of *Assess*2 is to obtain a relevance assessment for each information need with respect to the specific patient. The assessments are used in the evaluation of the formulation performance (Subsection 5.4.2).

To obtain the relevance assessments we provided the physicians once more with the patient data which they had seen in *Assess*1 (see upper part of Table 5.4). In *Assess*2, we provided formulated information needs based on the patient data (see lower part of Table 5.4) instead of documents. In accordance to *Assess*1, the patient

Table 5.4: *Assess*2: Assessment data presented to the physicians (comprising four data entries and three formulated information needs).

Patient X – Day 1

Data type	Data	Entry time
Disease or Syndrome	Coronary Artery Bypass	18:23:00
Disease or Syndrome	Kidney Failure, Acute	18:23:06
Chemical	Captopril	18:32:45
Chemical	Acetylcysteine	18:32:54

Indicate (on the evaluation form) for each question whether you would like to read an article which provides an answer to this question for this specific patient.

1. What are the long-term outcomes for this specific patient who had Coronary Artery Bypass?

2. What are the treatment options for Kidney Failure, Acute for this specific patient?

3. What are the side effects of Captopril for this specific patient?

data were divided into 11 interactions. Table 5.4 represents Day 1 for Patient X. Each interaction was presented as outlined in Table 5.4.

The assessments had to be indicated on an assessment form. Table 5.5 shows the assessment form for Patient X for Day 1. The left-hand side of the assessment form shows the information needs (equal to the information needs in Table 5.4). The right-hand side of the assessment form shows the possible assessments. The task of the physicians in *Assess*2 was to indicate for each question (i.e., information need) whether he would like to read an article which provides an answer to the question with respect to this specific patient. Assessments had to be given on a four-point scale (*absolutely not*, *no*, *yes*, *for sure*). Additionally, when a question received a negative assessment (*absolutely not* or *no*), the physicians had to indicate the reason for this assessment. As with the document assessment, this additional information was necessary to prevent information needs from being labelled non-relevant for the wrong reasons. Possible reasons for negative assessment were as follows (see Table 5.5).

- The answer to this question is already known to me.

- The answer is irrelevant w.r.t. this patient.

- The answer is completely irrelevant.

From the assessments we derived the relevance of the information needs with respect to the specific patient. Disregarding the many possible approaches, in the end we were interested in whether an information need was relevant with respect to the patient under investigation. Consequently, we had to map the answers from the four-point scale to a two-point scale. This was done as follows.

Table 5.5: *Assess*2: Assessment form.

Indicate for each question whether you would like to read an article which provides an answer to this question for this specific patient.	**Absolutely not**	**No**	**Yes**	**For sure**	*If not, why not?* The answer to this question is already known to me	The answer is irrelevant w.r.t. this patient	The answer is completely irrelevant
Patient X – Day 1							
1. What are the long-term outcomes for this specific patient who had Coronary Artery Bypass?	☐	☐	☐	☐	☐	☐	☐
2. What are the treatment options for Kidney Failure, Acute for this specific patient?	☐	☐	☐	☐	☐	☐	☐
3. What are the side effects of Captopril for this specific patient?	☐	☐	☐	☐	☐	☐	☐

- If the assessment was *yes* or *for sure*, the information need was labelled relevant.

- If the assessment was *no* or *absolutely not*, and the reason for this assessment was *The answer to this question is already known to me*, the information need was labelled relevant.

- If the assessment was *no* or *absolutely not*, and the reason for this assessment was *The answer is irrelevant w.r.t. this patient* or *The answer is completely irrelevant*, the information need was labelled non-relevant.

As a consequence, an information need is labelled non-relevant only when the physician explicitly indicates that the information need is not relevant with respect to the specific patient or not relevant at all. This is exactly the criterion in which we are interested. Information needs that are labelled non-relevant for another reason (i.e., the answer to the question is already known to the physician) could be eliminated

by MIRA in other ways. For instance, information needs for which the answer is known to the physician may be eliminated by employing a user model. However, as mentioned before, this is outside the scope of our research. In accordance with the evaluation data, the physicians were provided with 11 tables (i.e., 11 days) similar to Table 5.5 for Patient X (see Appendix D.3).

5.3.3   Assessment of Ranking

The purpose of *Assess*3 is to obtain an assessment of the ranking of the formulated information needs. To obtain the assessment, we needed the physicians' relevance assessments of the information needs formulated by MIRA. Based on their relevance and their position in the ranking, we may derive the ranking performance.

Since the assessments of the formulated information needs were already obtained in *Assess*2, we may reuse these assessments. Consequently, *Assess*3 coincides fully with *Assess*2. The difference with respect to the evaluation is that the relevance assessments obtained in *Assess*3 will be used to determine MIRA's *ranking* performance (i.e., whether relevant information needs are ranked higher than non-relevant ones), whereas the assessments obtained in *Assess*2 will be used to determine MIRA's *formulation* performance (i.e., the proportion of relevant information needs among all formulated information needs).

5.3.4   Assessment of the Unobtrusiveness

The purpose of the assessment of the obtrusiveness was to measure the unobtrusiveness of MIRA. There are various forms of obtrusiveness for a retrieval system. We focused on three of such forms: interruption, literature overload, and retrieval of non-relevant documents.

The first form of obtrusiveness is the *interruption of the physicians' daily workflow*. This form of obtrusiveness may be quite severe (Abate, Shumway, and Jacknowitz, 1992). It necessitates multitasking by physicians which may lead to an increased number of medical errors (Collins *et al.*, 2007). For instance, in a study concerning a computerized physician order entry system, Collins *et al.* (2006) found that 43% of the medication errors are caused by clinical distractions. We attempted to minimize this form of obtrusiveness by coupling MIRA to the EMR. As a consequence, MIRA fits naturally into the physicians' work environment and the use of MIRA requires no additional effort from the physician. Furthermore, we ensured that MIRA presented itself only once, at the end of an interaction, via a blinking, minimized window in the task bar. In this way, we minimized the interruption of the physician's daily work. By evaluating the unobtrusiveness of MIRA, we determined whether we were successful in developing a system that does not interrupt the physicians' workflow.

The second form of obtrusiveness is a *literature overload*. The more severe the literature overload, the more obtrusive MIRA. In Chapter 4, we discussed how to reduce the number of articles retrieved by MIRA in order to prevent a literature

Table 5.6: *Assess*4: Assessment form.

	Absolutely not	No	Yes	For sure	*If not, why not?*	It interrupts my work	The number of articles presented is too large	The literature is non-relevant
Assume you would enter patient data in ICIS. Would you like to be provided (at the end of your system interaction) automatically with patient-related literature (as in this evaluation)?	□	□	□	□		□	□	□

overload. By evaluating the unobtrusiveness of MIRA, we determined whether we were successful in preventing literature overload.

The third form of obtrusiveness is the *retrieval of non-relevant documents*. The fewer non-relevant documents are retrieved by MIRA, the less obtrusive MIRA. In Section 4.3, we discussed how we may improve the precision of the literature retrieved by MIRA. By evaluating the unobtrusiveness of MIRA, we determined whether we were successful in reducing the number of non-relevant documents.

The assessments of the obtrusiveness of MIRA had to be established on an assessment form (see Table 5.6). The right-hand side of Table 5.6 shows the possible assessments. Each physician had to assess the unobtrusiveness only once during the experiment. Consequently, the physicians were provided with just one instance of Table 5.6. The task of the physicians in *Assess*4 was to indicate whether they would like to be provided with patient-related literature automatically, after entering the patient data into an EMR (see Table 5.6). Assessments had to be given on a four-point scale (*absolutely not*, *no*, *yes*, *for sure*). Additionally, when MIRA received a negative assessment, the physicians had to indicate the reason for this assessment. Possible reasons for negative assessment were derived from the three forms of obtrusiveness.

- It interrupts my work.

- The number of articles presented is too large.

- The literature is non-relevant.

From the assessments we derived the obtrusiveness of MIRA.

5.4 Evaluation Methodology

In the previous sections we established the evaluation criteria (Section 5.1) and the experimental set-up (Section 5.3) of our evaluation process. In this section we describe how the results obtained during the experiment are used to determine the outcome of the application of each evaluation criterion.

5.4.1 Evaluating the Retrieval Performance

The purpose of the evaluation of the retrieval performance is to compare the various retrieval methods and to decide which one performs best. The best performing method will be used in the final version of MIRA. We recall from Section 4.3 that we investigated two different retrieval enhancements to improve the set of retrieved documents.

- Patient-related document scoring (Subsection 4.3.1)

- Patient-related document clipping (Subsection 4.3.2)

Combining these enhancements leads to four possible retrieval methods (see Table 4.5). In our evaluation, we retrieved documents by means of each of the four retrieval methods. The documents so retrieved were collected into one proper set (not a bag) and assessed by the physicians (see Subsection 5.3.1). The assessments were used in the evaluation of each of the four retrieval methods.

With respect to the retrieval performance, two measures are important: *precision* and *recall*.

Precision

The precision is the proportion of retrieved documents that are relevant. The higher the precision (i.e, the more of the retrieved documents are relevant) the better. To express the value of this measure, we calculated, for each physician p, for each retrieval method m, and with respect to each interaction i within the experiment, the precision $P_{p,m,i}$ of the set of retrieved documents. $P_{p,m,i}$ is calculated by equation 5.1.

$$P_{p,m,i} = \frac{\# \text{ retrieved documents for } i \text{ by } m \text{ that are relevant according to } p}{\# \text{ retrieved documents for } i \text{ by } m} \quad (5.1)$$

The precision $P_{p,m}$ for retrieval method m according to physician p was calculated as the average of the precision values $P_{p,m,i}$ over all 11 interactions. $P_{p,m}$ is calculated by equation 5.2.

$$P_{p,m} = \frac{\sum_{i=1}^{11} P_{p,m,i}}{11} \quad (5.2)$$

In the best case the precision would be 1.0 (all retrieved documents are relevant); in the worst case, the precision would be 0.0 (none of the retrieved documents is relevant).

Recall

The recall is the proportion of relevant documents that are retrieved. The higher the recall (i.e., the more relevant documents are retrieved) the better. The recall R of the set of retrieved documents is calculated by equation 5.3.

$$R = \frac{\text{\# retrieved documents that are relevant}}{\text{\# relevant documents in the document collection}} \qquad (5.3)$$

To calculate the recall, relevance assessments are required for *all* documents in the collection. In our experiment, we only asked physicians to assess the retrieved documents; assessing all documents in the OHSUMED collection would be too time consuming for the physicians. Therefore, instead of the recall, we used the *relative recall*, which is derived from the recall and does not require relevance assessments for all documents in the collection. However, the relative recall may only be used to compare methods, not to make absolute statements concerning the methods separately. The relative recall $RR_{p,m,i}$ for retrieval method m was calculated with respect to each physician p and for each interaction i in the experiment. $RR_{p,m,i}$ is calculated by equation 5.4.

$$RR_{p,m,i} = \frac{\text{\# documents retrieved for } i \text{ by } m \text{ and relevant according to } p}{\text{\# documents retrieved for } i \text{ by any } m \text{ and relevant according to } p} \qquad (5.4)$$

The relative recall $RR_{p,m}$ for retrieval method m for a certain physician p was calculated as the average of all precision values for physician p over all 11 interactions within the experiment. $RR_{p,m}$ is calculated by equation 5.5.

$$RR_{p,m} = \frac{\sum_{i=1}^{11} RR_{p,m,i}}{11} \qquad (5.5)$$

In the best case the relative recall would be 1.0 (the method retrieves all relevant documents that are retrieved by all methods together); in the worst case, the precision would be 0.0 (the method retrieves no relevant documents at all).

Performance Criteria

In our tuning experiments (Subsection 4.4.2), we reached a precision of 0.90 when using test run 4 (using both enhancements). We consider the retrieval performance of MIRA successful if we reach the same value (or higher) during this objective experiment. The relative recall is solely employed to compare the performance of the methods, hence no success criterion can be set.

5.4.2 Evaluating the Formulation Performance

We now turn from the evaluation of document retrieval to the evaluation of information-need formulation. In Subsection 4.4.1 we established which of the formulation methods was most successful in reducing the number of formulated information

needs. We recollect that we used a formulation method incorporating four types of knowledge, viz. (a) formulation history, (b) entry order, (c) domain knowledge, and (d) number of retrieved documents. However, we have not yet evaluated MIRA's retrieval performance in practice. Therefore, the purpose of the evaluation of the formulation performance was solely to determine MIRA's formulation performance, not to determine which of the formulation methods performs best.

In the experiment, MIRA formulated information needs based on the patient data. The information needs were assessed by the physicians (see Subsection 5.3.2). The assessments were used in the evaluation of the formulation performance.

To express the quality of the set of formulated information needs with respect to their relevance, we calculated the precision $P_{p,i}$ of the set, with respect to each physician p and for each interaction i. $P_{p,i}$ is calculated by equation 5.6.

$$P_{p,i} = \frac{\# \text{ information needs formulated for } i \text{ and relevant according to } p}{\# \text{ information needs formulated for } i} \quad (5.6)$$

The precision P_p with respect to a certain physician p was calculated as the average of all precision values $P_{p,i}$ with respect to physician p over all n interactions that resulted in at least one information need. P_p is calculated by equation 5.7.

$$P_p = \frac{\sum_{i=1}^{n} P_{p,i}}{n} \quad (5.7)$$

In the best case, the precision would be 1.0 (all formulated information needs are relevant); in the worst case, the precision would be 0.0 (none of the formulated information needs is relevant).

Furthermore, the number of information needs formulated per interaction is also an important element of the formulation performance. Therefore, we calculate the average number of information needs formulated per interaction as well. Since we restrict the number of formulated information needs to 10 per interaction, the average number of information needs formulated per interaction is a number in the interval $[0, 10]$.

Performance Criteria

We consider the formulation performance of MIRA sufficient if the precision is above or equal to 0.90 and the average number of information needs formulated per interaction is between 0 and 10. Moreover, we note that we could neither calculate the recall nor the relative recall for the formulation performance as we did for the retrieval performance. There are two reasons. First, there is no such thing as a collection of possible information needs with relevance assessments, so we could not calculate the recall. Second, we did not compare various methods, so there was no relative recall to calculate.

5.4.3 Evaluating the Ranking Performance

The purpose of the evaluation of the ranking performance is (1) to calculate the performance of each ranking method for each specific physician and (2) to determine if either (a) there is one ranking method performing best or (b) the ranking performance is physician dependent. We recall from Subsection 4.2.2 that we used three different ranking criteria, viz. (a) specialism, (b) entry time, and (c) number of retrieved documents. Combining these criteria led to eight possible ranking methods (see Table 4.4). The ranking performance of each method can be derived from the relevance assessments of the information needs provided by the physicians (Subsection 5.3.3). As an example, Table 5.7 shows three possible rankings for a particular set of information needs.

The relevance assessment of an information need according to a physician is a Boolean: relevant or not relevant. In Table 5.7, such a relevance assessment is indicated by a plus sign or a minus sign, representing a relevant or a non-relevant information need, respectively. Consequently, the set of information needs presented in Table 5.7 comprises five relevant information needs (B, E, F, H, and L) and five non-relevant information needs (A, C, D, G, and K).

The relevance score of an information need according to MIRA is a number. After the relevance scores for all information needs are calculated, they are sorted with respect to their relevance. Subsequently, the relevance score of each information need is translated into a natural number corresponding to its rank in the sorted list. The higher the relevance score, the lower the rank. For instance, an information need with rank 2 has a higher relevance score than an information need with rank 7. Finally, the information needs with a rank higher than 10 are discarded. Consequently, the rank of a document is a number in the interval [1,10]. In Table 5.7 the rank of the information needs is indicated in the first column ('Rank').

The second column of Table 5.7 represents an optimal ranking of information needs. An optimal ranking reflects the relevance of the information needs according to the physician. As such, the information needs that are relevant according to the

Table 5.7: Three distinct rankings of an example set of information needs (+ represents a relevant information need; − represents a non-relevant information need).

Rank	Optimal ranking	Sub-optimal ranking	Worst ranking
1	needF (+)	needE (+)	needC (−)
2	needH (+)	needL (+)	needG (−)
3	needL (+)	needC (−)	needA (−)
4	needB (+)	needH (+)	needK (−)
5	needE (+)	needA (−)	needD (−)
6	needD (−)	needD (−)	needE (+)
7	needK (−)	needK (−)	needB (+)
8	needA (−)	needF (+)	needL (+)
9	needG (−)	needB (+)	needH (+)
10	needC (−)	needG (−)	needF (+)

physician are in the top of the ranking, whereas the information needs that are not relevant to the physicians are in the bottom.

The third column ('Sub-optimal ranking') shows a sub-optimal ranking, in which the information needs that are relevant according to the physician (indicated by a plus sign) are ranked first, second, fourth, eighth, and ninth. The fourth column ('Worst ranking') shows a worst-case ranking (i.e., all information needs that are relevant according to the physician are in the bottom of the ranking).

To express the ranking performance of a certain method we calculated the precision $P_{p,m,s}$ of the formulated information needs with respect to each physician p, each ranking method m, and each size s of the set of formulated information needs. A set of information needs of size s contains only the information needs of which the rank is smaller or equal to s. For instance, a set of size 5 contains only information needs with rank 1 to 5. To calculate $P_{p,m,s}$, we use the notion of relevance $relevant(p,m,i)$, which is the number of information needs for interaction i that are relevant according to physician p and ranked by method m. By means of $relevant(p,m,i)$ we may now calculate the precision $P_{p,m,s,i}$ with respect to each physician p, ranking method m, size s of the set of formulated information needs, and interaction i.

$$P_{p,m,s,i} = \frac{relevant(p,m,i) \text{ calculated over a set with size } s}{s} \qquad (5.8)$$

Since MIRA retrieves a maximum of ten information needs per interaction, $s \, \epsilon \, [1, 10]$. However, if s is larger than the number of information needs formulated for interaction i, $P_{p,m,s,i}$ is not defined. Consequently, we calculate $P_{p,m,s,i}$ only for those interaction for which s is smaller or equal to the number of formulated information needs.

From $P_{p,m,s,i}$ we may then calculate $P_{p,m,s}$ by averaging the values of $P_{p,m,s,i}$ over all interactions i. However, since $P_{p,m,s,i}$ is not defined for interactions in which no documents are retrieved, we calculate $P_{p,m,s}$ over the n interactions for which documents are retrieved. This is done by using equation 5.9.

$$P_{p,m,s} = \frac{\sum_{i=1}^{n} P_{p,m,s,i}}{n} \qquad (5.9)$$

To continue our example, equation 5.9 may be used to calculate $P_{p,m,s}$ for each ranking in Table 5.7 for various values of s. For instance, when $s = 4$, $P_{p,m,s}$ for the optimal ranking is 1.0, $P_{p,m,s}$ for the sub-optimal ranking is 0.75, and $P_{p,m,s}$ for the worst-case ranking is 0. From the obtained results we may derive which of the ranking methods performs best with respect to physician p. If one specific ranking method is best for the majority of physicians, then we may conclude that this is the best ranking method in general. However, if this is not the case, we may conclude that we cannot use the same ranking method for each physician. In that case, the ranking method to be used for a specific physician should be determined by user modelling. However, this is outside the scope of the thesis.

Here, we note again that we were unable to calculate the recall for the ranking performance. The reason was that we had no collection of possible information

needs with relevance assessments. Furthermore, to calculate the relative recall of the ranking performance would not help. Since all sets of formulated information needs were equal (only their order differs), the relative recall will not have any discriminating power.

5.4.4 Evaluating the Unobtrusiveness

To evaluate the unobtrusiveness of MIRA we used the assessments of the unobtrusiveness provided by the physicians (Subsection 5.3.4). The unobtrusiveness O_p, according to physician p, is either 1 (unobtrusive) or 0 (obtrusive). We consider MIRA successful if it is unobtrusive according to all nine physicians.

5.5 Evaluation Results

In the previous section we described how we envisaged to evaluate MIRA by applying each evaluation criterion. In this section, we present the results of the evaluation of the retrieval performance (see Subsection 5.5.1), the formulation performance (see Subsection 5.5.2), the ranking performance (see Subsection 5.5.3), and the unobtrusiveness (see Subsection 5.5.4). Table 5.8 shows the number of information needs formulated for each interaction and each precision-enhancement approach.

5.5.1 Retrieval Performance

Below we show two parts of the evaluation of the retrieval performance (in accordance with Subsection 5.4.1). Part 1 comprises the calculation of the precision ($P_{p,m,i}$) of the retrieved document sets and part 2 comprises the calculation of the relative recall ($RR_{p,m,i}$) of the retrieved document sets.

Part 1: Precision

The results of part 1 are shown in Table 5.9. The table shows $P_{p,m,i}$ of the set of retrieved documents for nine physicians (see Section 5.2), for 4 precision-enhancement approaches (cf. Table 4.5), and with respect to 11 interactions. Furthermore, the last column of the table shows the average precision values $P_{p,m}$. With respect to these results, we list the observations for each type of physician separately.

Table 5.8: Number of information needs formulated for eleven interactions and four distinct precision-enhancement approaches (A, B, C, and D).

	1	2	3	4	5	6	7	8	9	10	11	Average
A	10	9	2	0	3	0	3	0	0	4	3	3.1
B	10	9	2	0	3	0	3	0	0	4	3	3.1
C	6	2	0	0	1	1	0	1	0	0	1	1.1
D	6	2	0	0	1	1	0	1	0	0	1	1.1

Note: the number of documents per interaction is equal to the number of information needs.

Table 5.9: $P_{p,m,i}$ for nine distinct physicians, four distinct precision-enhancement approaches (A, B, C, and D), and eleven interactions.

	1	2	3	4	5	6	7	8	9	10	11	$P_{p,m}$
Attending physician 1												
A	0.22	0.22	NA	ND	0.33	ND	0.67	ND	ND	0.25	0.35	**0.34**
B	0.22	0.22	NA	ND	0.33	ND	0.67	ND	ND	0.25	0.35	**0.34**
C	0.20	0.00	ND	ND	0.33	ND	0.67	ND	ND	0.25	0.33	**0.37**
D	0.20	0.00	ND	ND	0.33	ND	0.67	ND	ND	0.25	0.33	**0.37**
Attending physician 2												
A	0.00	0.22	0.00	ND	0.67	ND	0.00	ND	ND	0.25	0.00	**0.16**
B	0.00	0.22	0.00	ND	0.67	ND	0.00	ND	ND	0.25	0.00	**0.16**
C	0.17	0.00	ND	ND	1.00	ND	0.00	ND	ND	0.00	0.00	**0.19**
D	0.17	0.00	ND	ND	1.00	ND	0.00	ND	ND	0.00	0.00	**0.19**
Attending physician 3												
A	0.10	0.56	0.00	ND	0.33	ND	0.00	ND	ND	0.00	0.00	**0.14**
B	0.10	0.56	0.00	ND	0.33	ND	0.00	ND	ND	0.00	0.00	**0.14**
C	0.17	0.00	ND	ND	0.00	ND	1.00	ND	ND	0.00	0.00	**0.19**
D	0.17	0.00	ND	ND	0.00	ND	1.00	ND	ND	0.00	0.00	**0.19**
Resident physician 1												
A	0.67	0.67	0.50	ND	0.67	ND	0.67	ND	ND	0.50	0.67	**0.59**
B	0.56	0.44	0.50	ND	0.67	ND	0.33	ND	ND	0.50	0.67	**0.52**
C	0.60	0.00	ND	ND	1.00	ND	1.00	ND	ND	0.00	1.00	**0.60**
D	0.60	0.00	ND	ND	1.00	ND	1.00	ND	ND	0.00	1.00	**0.60**
Resident physician 2												
A	0.40	0.44	0.50	ND	0.33	ND	0.33	ND	ND	0.00	0.33	**0.33**
B	0.40	0.44	0.50	ND	0.33	ND	0.33	ND	ND	0.00	0.33	**0.33**
C	0.50	0.00	ND	ND	0.00	ND	1.00	ND	ND	0.00	1.00	**0.42**
D	0.50	0.00	ND	ND	0.00	ND	1.00	ND	ND	0.00	1.00	**0.42**
Resident physician 3												
A	0.50	0.44	0.50	ND	0.33	ND	1.00	ND	ND	0.00	0.67	**0.49**
B	0.50	0.44	0.50	ND	0.33	ND	1.00	ND	ND	0.00	0.67	**0.49**
C	0.83	0.00	ND	ND	0.00	ND	1.00	ND	ND	0.00	1.00	**0.47**
D	0.83	0.00	ND	ND	0.00	ND	1.00	ND	ND	0.00	1.00	**0.47**
Recently graduated physician 1												
A	0.22	0.56	0.50	ND	0.67	ND	0.33	ND	ND	0.00	0.67	**0.42**
B	0.33	0.56	0.50	ND	0.67	ND	0.67	ND	ND	0.00	0.67	**0.48**
C	0.80	0.00	ND	ND	1.00	ND	1.00	ND	ND	0.00	1.00	**0.63**
D	0.80	0.00	ND	ND	1.00	ND	1.00	ND	ND	0.00	1.00	**0.63**
Recently graduated physician 2												
A	0.70	0.76	0.00	ND	0.67	ND	0.33	ND	ND	0.25	1.00	**0.52**
B	0.40	0.67	0.00	ND	0.67	ND	0.33	ND	ND	0.25	1.00	**0.47**
C	0.83	0.50	ND	ND	0.00	ND	0.00	ND	ND	0.00	1.00	**0.39**
D	0.83	0.50	ND	ND	0.00	ND	0.00	ND	ND	0.00	1.00	**0.39**
Recently graduated physician 3												
A	0.60	0.56	0.50	ND	0.67	ND	1.00	ND	ND	0.25	0.33	**0.56**
B	0.60	0.56	0.50	ND	0.67	ND	1.00	ND	ND	0.25	0.33	**0.56**
C	0.83	0.00	ND	ND	0.00	ND	1.00	ND	ND	0.00	1.00	**0.47**
D	0.83	0.00	ND	ND	0.00	ND	1.00	ND	ND	0.00	1.00	**0.47**

Note: Interactions resulting in no documents (indicated by 'ND') and interactions for which the information needs were not assessed by the physician (indicated by 'NA') are not used to calculate the average precision $P_{p,m}$.

- For the attending physicians, the average precision is quite low for each of the four retrieval methods. Furthermore, the average precision values of methods A and B are equal, as are the average precision values of methods C and D. Moreover, for attending physicians 1 and 2, there is not much difference between the average precision of the four methods in general.

- For the resident physicians, the average precision is in the medium range for each of the four retrieval methods. Furthermore, the average precision values of methods A and B are almost equal. The same holds for the average precision values of methods C and D. For resident physicians 1 and 3, there is not much difference between the average precision of the four methods in general.

- For the recently graduated physicians, the average precision is in the medium range for each of the four retrieval methods. Furthermore, the average precision values of methods C and D are equal for all three recently graduated physicians. For recently graduated physician 1, retrieval methods C and D produce better results than methods A and B, whereas the opposite is true for recently graduated physicians 2 and 3.

Based on these observations we may state that the positive influence of the precision-enhancement approaches on the precision of the retrieved document sets is only marginal. Moreover, in some cases the approaches even display a negative influence. None of the precision-enhancement approaches reaches our target value of 0.90 for any of the physicians.

Part 2: Relative Recall

The results of part 2 are shown in Table 5.10. It shows $RR_{p,m,i}$ of the set of retrieved documents for nine physicians (Subsection 5.2), for 4 precision-enhancements approaches (Table 4.5), and with respect to 11 interactions within the experiment. Furthermore, the last column of the table shows the average relative recall values $RR_{p,m}$.

With respect to these results, we list the observations for each type of physician separately.

- For the attending physicians, the average relative recall values of retrieval methods A and B are equal, as are the average relative recall values of methods C and D. Additionally, the average relative recall values of methods A and B are higher than the average relative recall values of methods C and D. For attending physician 1, the difference is particularly clear.

- For the resident physicians, the average relative recall values of methods A and B are roughly equal, as are the average relative recall values of methods C and D. Furthermore, the relative recall values of methods A and B are much higher than the average relative recall values of methods C and D. Moreover, for resident physician 1, method A performs better that method B.

Table 5.10: $RR_{p,m,i}$ for nine distinct physicians, four distinct precision-enhancement approaches (A, B, C, and D), and eleven interactions.

	1	2	3	4	5	6	7	8	9	10	11	$RR_{p,m}$
Attending physician 1												
A	0.67	1.00	0.00	ND	1.00	ND	0.67	ND	ND	1.00	1.00	**0.76**
B	0.67	1.00	0.00	ND	1.00	ND	0.67	ND	ND	1.00	1.00	**0.76**
C	0.33	0.00	0.00	ND	0.00	ND	0.33	ND	ND	0.00	1.00	**0.24**
D	0.33	0.00	0.00	ND	0.00	ND	0.33	ND	ND	0.00	1.00	**0.24**
Attending physician 2												
A	0.00	1.00	0.00	ND	0.67	ND	0.00	ND	ND	1.00	0.00	**0.38**
B	0.00	1.00	0.00	ND	0.67	ND	0.00	ND	ND	1.00	0.00	**0.38**
C	1.00	0.00	0.00	ND	0.33	ND	0.00	ND	ND	0.00	0.00	**0.19**
D	1.00	0.00	0.00	ND	0.33	ND	0.00	ND	ND	0.00	0.00	**0.19**
Attending physician 3												
A	0.50	1.00	0.00	ND	1.00	ND	0.00	ND	ND	0.00	0.00	**0.36**
B	0.50	1.00	0.00	ND	1.00	ND	0.00	ND	ND	0.00	0.00	**0.36**
C	0.50	0.00	0.00	ND	0.00	ND	1.00	ND	ND	0.00	0.00	**0.21**
D	0.50	0.00	0.00	ND	0.00	ND	1.00	ND	ND	0.00	0.00	**0.21**
Resident physician 1												
A	0.86	1.00	1.00	ND	0.67	ND	0.67	ND	ND	1.00	1.00	**0.88**
B	0.71	1.00	1.00	ND	0.67	ND	0.67	ND	ND	1.00	1.00	**0.86**
C	0.43	0.00	0.00	ND	0.33	ND	0.33	ND	ND	0.00	0.50	**0.23**
D	0.43	0.00	0.00	ND	0.33	ND	0.33	ND	ND	0.00	0.50	**0.23**
Resident physician 2												
A	0.67	1.00	1.00	ND	1.00	ND	0.33	ND	ND	0.00	1.00	**0.71**
B	0.67	1.00	1.00	ND	1.00	ND	0.33	ND	ND	0.00	1.00	**0.71**
C	0.50	0.00	0.00	ND	0.00	ND	0.33	ND	ND	0.00	1.00	**0.26**
D	0.50	0.00	0.00	ND	0.00	ND	0.33	ND	ND	0.00	1.00	**0.26**
Resident physician 3												
A	0.57	1.00	1.00	ND	1.00	ND	0.60	ND	ND	0.00	1.00	**0.74**
B	0.57	1.00	1.00	ND	1.00	ND	0.60	ND	ND	0.00	1.00	**0.74**
C	0.57	0.00	0.00	ND	0.00	ND	0.20	ND	ND	0.00	0.50	**0.18**
D	0.57	0.00	0.00	ND	0.00	ND	0.20	ND	ND	0.00	0.50	**0.18**
Recently graduated physician 1												
A	0.40	1.00	1.00	ND	0.67	ND	0.33	ND	ND	0.00	1.00	**0.68**
B	0.60	1.00	1.00	ND	0.67	ND	0.67	ND	ND	0.00	1.00	**0.70**
C	0.80	0.00	0.00	ND	0.33	ND	0.33	ND	ND	0.00	0.50	**0.28**
D	0.80	0.00	0.00	ND	0.33	ND	0.33	ND	ND	0.00	0.50	**0.28**
Recently graduated physician 2												
A	0.75	0.86	0.00	ND	1.00	ND	0.50	ND	ND	1.00	1.00	**0.73**
B	0.63	0.86	0.00	ND	1.00	ND	0.50	ND	ND	1.00	1.00	**0.71**
C	0.50	0.14	0.00	ND	0.00	ND	0.00	ND	ND	0.00	0.33	**0.14**
D	0.50	0.14	0.00	ND	0.00	ND	0.00	ND	ND	0.00	0.33	**0.14**
Recently graduated physician 3												
A	0.71	1.00	1.00	ND	1.00	ND	0.75	ND	ND	1.00	1.00	**0.92**
B	0.71	1.00	1.00	ND	1.00	ND	0.50	ND	ND	1.00	1.00	**0.89**
C	0.57	0.00	0.00	ND	0.00	ND	0.25	ND	ND	0.00	1.00	**0.26**
D	0.57	0.00	0.00	ND	0.00	ND	0.25	ND	ND	0.00	1.00	**0.26**

Note: Interactions resulting in no documents (indicated by 'ND') are not used to calculate the average relative recall $RR_{p,m}$.

- For the recently graduated physicians, the average relative recall values of methods A and B are quite similar to each other, whereas the average relative recall values of methods C and D are exactly equal. Furthermore, the average relative recall values of methods A and B are higher than the values of methods C and D. Moreover, for recently graduated physician 1, method B outperforms method A, whereas the opposite is true for recently graduated physicians 2 and 3.

Based on the above observations, we may state that the precision-enhancement approaches have considerable negative influence on the relative recall of the retrieved document sets.

Combining the results concerning precision and relative recall, we may state that retrieval methods C and D outperform methods A and B with respect to precision. At the same time, we may state that methods A and B outperform methods C and D with respect to relative recall. However, the superiority of methods A and B with respect to precision is rather small, whereas the superiority of methods C and D with respect to relative recall is quite high. Even when we analyzed the absolute numbers of retrieved documents, methods A and B outperform methods C and D. When methods C or D are used, the number of retrieved documents decreases drastically. As a result, the negative influence on the recall is significant, whereas the positive influence on the precision is only marginal. Therefore, we prefer methods A and B over methods C and D. For all physicians except recently graduated physician 1, retrieval method A performs at least as well as retrieval method B. For some physicians (resident physician 1 and recently graduated physicians 2 and 3) method A outperforms method B. Consequently, we choose method A as our final retrieval method.

5.5.2   Formulation Performance

As mentioned in Subsection 5.4.2, the purpose of the evaluation of the formulation performance is to calculate the precision $P_{p,i}$ of the set of formulated information needs and the average number of information needs formulated per interaction.

Table 5.11: $P_{p,i}$ for nine distinct physicians and eleven interactions.

	1	2	3	4	5	6	7	8	9	10	11	P_p
Attending physician 1	0.60	1.00	ND	ND	1.00	ND	1.00	ND	ND	1.00	1.00	**0.93**
Attending physician 2	1.00	1.00	1.00	ND	1.00	ND	1.00	ND	ND	0.50	0.33	**0.83**
Attending physician 3	1.00	1.00	1.00	ND	1.00	ND	1.00	ND	ND	1.00	1.00	**1.00**
Resident physician 1	0.80	1.00	1.00	ND	1.00	ND	1.00	ND	ND	1.00	1.00	**0.97**
Resident physician 2	0.80	1.00	1.00	ND	1.00	ND	1.00	ND	ND	1.00	1.00	**0.97**
Resident physician 3	1.00	1.00	1.00	ND	1.00	ND	1.00	ND	ND	1.00	1.00	**1.00**
Recently graduated physician 1	0.50	0.33	0.50	ND	0.67	ND	0.00	ND	ND	0.00	0.00	**0.29**
Recently graduated physician 2	0.80	0.78	1.00	ND	1.00	ND	0.33	ND	ND	0.25	1.00	**0.74**
Recently graduated physician 3	0.70	0.56	1.00	ND	1.00	ND	0.67	ND	ND	0.00	0.67	**0.66**

Note: Interactions resulting in no documents (indicated by 'ND') are not used to calculate the average precision P_p.

The results of the evaluation are shown in Table 5.11. The precision $P_{p,i}$ of the set of formulated information needs is shown for nine physicians (Section 5.2), and with respect to 11 interactions i within the experiment. As mentioned in the previous subsection, we used retrieval method A.[2]

From the results in Table 5.11, we may make three observations.

- For the attending physicians, the average P_p is 0.92, exceeding our target value of 0.90. When considering each attending physician separately, the precision values of attending physicians 1 and 3 exceed our target value as well. The precision value of attending physician 2 does not exceed our target value, but may still be considered quite high.

- For the resident physicians, the average P_p is 0.98, clearly exceeding our target value of 0.90 and the average precision value of the attending physicians. When considering each resident physician separately, the precision value of each of the physicians exceeds our target value of 0.90.

- For the recently graduated physicians, the average P_p is 0.56. This value is considerably lower than our target value of 0.90 and much lower than the average precision values for the attending physicians and the resident physicians. When considering each of the recently graduated physicians separately, none of the three precision values exceeds our target value of 0.90 either.

The average number of information needs formulated per interaction by method A is 3.1. This result concurs with the results of our tuning experiment in Subsection 4.4.1 and is within our target interval of $[0, 10]$.

5.5.3 Ranking Performance

As mentioned in Subsection 5.4.3, the purposes of the evaluation of the information-need ranking performance are (1) to determine the performance of each ranking method for each specific physician and (2) to determine if either (a) there is one ranking method performing best or (b) the ranking performance is physician dependent. The results of our experiments are discussed separately for the attending physicians, the resident physicians, and the recently graduated physicians.

The Attending Physicians

We calculated the precision $P_{p,m,s}$ according to attending physician 1 for document sets of different sizes obtained with eight different ranking methods $R1$ to $R8$ (see Table 4.4) and the precision for the optimal ranking method. From the results (see Table E.1) we may make four observations.

- $R3$ performs worst. Even for an information-need set of size 2, it is not able to obtain maximum precision. $R3$ is solely based on knowledge concerning the entry time of the patient data.

---

[2]Remember that we can make this choice in retrospect, since each physician assessed the documents retrieved by all four methods.

- $R5$, $R6$, and $R8$ produce the same precision values for all sizes of the document set. With these methods, maximum precision is maintained for document sets up to size 5. The only difference between the methods is the way in which the ranking criteria are combined. Additionally, the results are quite similar to $R1$ (random ranking).

- $R4$ and $R7$ produce the same precision values for all sizes of the document set. Maximum precision is maintained for document sets up to size 5; precision increases for document sets of size 7. The only difference between both methods is the use of knowledge concerning the entry time.

- $R2$ results in precision values that are closest to the precision values of the optimal ranking. Maximum precision is obtained for document sets up to size 6. Furthermore, at each document size, $R2$ obtains a precision value which is at least as high as the precision values of all other methods (except the optimal ranking) at the same size. $R2$ is solely based on knowledge concerning the physician's specialism.

Table E.2 shows the results for attending physician 2. From these results we may make three observations.

- $R3$ performs worst. Even for an information-need set of size 2, it obtains a precision of only 0.75.

- $R4$ and $R7$ produce the same precision values for all sizes of the document set. Maximum precision is obtained for document sets larger than size 5; precision increases for document sets of size 2. The only difference between the methods is the use of knowledge concerning the entry time.

- $R2$, $R5$, $R6$, and $R8$ result in precision values that are closest to the precision values of the optimal ranking. With these methods, maximum precision is obtained for document sets larger than size 5. The only difference between the methods is the way in which the ranking criteria are combined.

Table E.3 shows the results for attending physician 3. From these results we may observe that all rankings produce the same results, which are equal to the results of the optimal ranking. These results are caused by the fact that attending physician 3 labelled all information needs as relevant. Consequently, there is no difference between the various ranking methods.

Based on the results above, we may state that $R2$ produces the best ranking (or one of the best rankings) for each attending physician. Consequently, we use ranking method $R2$ to rank the information needs for the attending physicians.

The Resident Physicians

Table E.4 shows results for resident physician 1. From these results we may make four observations.

- $R1$ (random ranking) performs worst. Maximum precision is only reached for document sets of size 5.

- $R2$, $R5$, $R6$, and $R8$ produce the same precision values for all sizes of the document set. Maximum precision is only obtained for document sets of size 1. All of these methods use knowledge concerning the physician's specialism in combination with other ranking criteria.

- $R4$ and $R7$ produce the same precision values for all sizes of the document set. Maximum precision values are obtained for document sets up to size 2. In both methods knowledge concerning the number of results is used.

- $R3$ performs best; it produces the same precision values as the optimal ranking for all sizes of the document set. Furthermore, at each document size, $R3$ obtains a precision value which is at least as high as the precision values of all other methods (except the optimal ranking) at the same size.

Table E.5 shows results for resident physician 2. From these results we may make four observations.

- $R1$ (random ranking) performs worst. Maximum precision is only reached for document sets of size 5.

- $R4$, $R6$, $R7$, and $R8$ produce the same precision values for all sizes of the document set. Maximum precision is only obtained for document sets of size 4. All of these methods use knowledge concerning the number of results in combination with other ranking criteria.

- $R3$ performs better than $R1$, $R4$, $R6$, $R7$, and $R8$. Maximum precision values are obtained for document sets up to size 6. In this method, knowledge concerning the physician's specialism and the entry time is used.

- $R2$ and $R5$ perform best; these methods produce the same precision values as the optimal ranking for all sizes of the document set. Furthermore, at each document size, $R2$ and $R5$ obtain precision values which are at least as high as the precision values of all other methods at the same size.

Table E.6 shows results for resident physician 3. From these results we may observe that all rankings produce the same results, which are equal to the results of the optimal ranking. These results are due to the fact that resident physician 3 labelled all information needs as relevant. Consequently, there is no difference between the various ranking methods.

Based on the results above, we may state that $R3$ performs best for resident physician 1 and sufficiently well for resident physicians 2 and 3. Furthermore, we may state that $R2$ and $R5$ perform best for resident physician 2, sufficiently well for resident physician 3 and rather poorly for resident physician 1. Consequently, we use ranking method $R3$ to rank the information needs for the resident physicians.

The Recently Graduated Physicians

Table E.7 shows results for recently graduated physician 1. From these results we may make four observations.

- There is much similarity among the performances of the eight ranking methods.

- Maximum precision is never reached.

- $R4$ and $R7$ perform worst. Both methods employ knowledge concerning the number of retrieved documents.

- There is no ranking method that has a precision which is at least as high as the precision values of the other methods at each size s of the document set.

Table E.8 shows results for recently graduated physician 2. From these results we may make four observations.

- Maximum precision is never reached.

- $R4$, $R6$, $R7$, and $R8$ perform worst; their performance is quite similar. All four methods employ knowledge concerning the number of retrieved documents.

- $R2$ and $R5$ perform somewhat better. Both methods employ knowledge concerning the physician's specialism.

- $R1$ performs even better than the aforementioned methods.

- $R3$ performs best; this method employs knowledge concerning the entry time of the patient data. However, it does not produce precision values which are at least as high as the precision values of the other methods at each size s of the document set.

Table E.9 shows results for recently graduated physician 3. From these results we may make four observations.

- Maximum precision is never reached.

- $R4$ and $R7$ perform worst with rather similar performance. Both methods employ knowledge concerning the number of retrieved documents.

- $R3$, $R6$ and $R8$ perform somewhat better. These methods employ all three types of knowledge.

- $R2$ and $R5$ perform best; these method employs knowledge concerning the physician's specialism. However, it does not produce precision values which are at least as high as the precision values of the other methods at each size s of the document set.

Based on the results above, we may state that there is no single best ranking method for recently graduate physicians. However, we observed that methods $R4$ and $R7$ are not appropriate options according to all three physicians. Since both methods employ knowledge concerning the number of retrieved documents, apparently this knowledge type is not suitable. Furthermore, we may state that methods $R2$, $R3$, and $R5$ perform reasonably well for all three physicians. These three methods employ knowledge concerning the physician's specialism and the entry time of the patient data. Moreover, the methods do not employ knowledge concerning the number of retrieved documents, further supporting the inappropriateness of this knowledge type. Consequently, we may state that any ranking method may be used, as long as it does not employ knowledge concerning the number of retrieved documents.

5.5.4   Unobtrusiveness

As mentioned in Subsection 5.4.4, the purpose of the evaluation of the unobtrusiveness is to determine whether MIRA is unobtrusive.

Attending physician 3 indicated that he would like to be provided with patient-related literature at the end of his interaction with the EMR. However, only under the condition that the provided literature is 'to the point'. Indeed, attending physician 3 has the lowest values for retrieval performance (cf. Tables 5.9 and 5.10) and the highest value for formulation performance (cf. Table 5.11). The other two attending physicians indicated that they would not like to be provided with patient-related literature at the end of his interaction with the EMR. However, they did not specify the reason for their answer.

Two of the resident physicians (1 and 3) indicated that they would like to be provided with patient-related literature at the end of their interaction with the EMR. Resident physician 2 indicated that he would not like to be provided with patient-related literature, because the literature was not sufficiently relevant. Indeed, resident physician 2 had the lowest values for retrieval performance (cf. Tables 5.9 and 5.10).

Two of the recently graduated physicians (1 and 2) indicated that they would like to be provided with patient-related literature at the end of their interaction with the EMR. Recently graduated physician 3 indicated that he would not like to be provided with literature, because the number of retrieved documents is too high. Indeed, recently graduated physician 3 has high values for retrieval performance (cf. Tables 5.9 and 5.10), but despite the high retrieval performance, the high number of retrieved documents was perceived as obtrusive.

5.6   Discussion

As mentioned in Section 5.2, the results of the evaluation were obtained from the assessments of nine physicians. Since the number of physicians is quite small, it would be premature to draw any final conclusions from the obtained data. Therefore this section merely contains a discussion of the results and provides our *provisional*

conclusions regarding the retrieval performance (Subsection 5.6.1), the formulation performance (Subsection 5.6.2), the ranking performance (Subsection 5.6.3), and the unobtrusiveness (Subsection 5.6.4) of MIRA.

5.6.1   Retrieval

We discuss the retrieval performance with respect to precision and relative recall.

Precision

With respect to the precision $p_{p,m,i}$ of the retrieved documents (Subsection 5.5.1), we may draw three conclusions.

First, we may conclude that the average precision never reaches our target of 0.90. The lowest precision values (14 to 37%) are observed for the attending physicians. For the resident physicians and the recently graduated physicians, the precision values are higher: 33 to 60% and 39 to 63%, respectively.

Second, we may conclude that patient-related document scoring does not improve the precision of the retrieved document set. We observed that there is hardly any positive difference of retrieval method B over method A. Only for recently graduated physician 1, there is a positive difference of method B over method A. However, there is no difference between methods C and D. For some physicians (resident physician 1 and recently graduated physician 2), the precision of patient-related document scoring is worse than when using no precision enhancement. The minor influence of method B on the precision of the retrieved documents may be explained by the fact that the patient data added to the query do not occur frequently in the document collection. Furthermore, we observed that there is no difference between retrieval method C and method D. Since the only difference between methods A and C on the one hand, and methods B and D on the other hand, is the use of patient-related document scoring, we may conclude that this approach has no positive influence on the precision of the retrieved document set. Consequently, the addition of the data to the query does not make a substantial contribution to the retrieval process.

Third, we may conclude that patient-related document clipping improves the precision of the retrieved documents in some cases. We observed that for the attending physicians, two of the resident physicians, and one of the recently graduated physicians, retrieval methods C and D result in higher precision values than retrieval methods A and B. Since the only difference between methods A and B on the one hand, and methods C and D on the other hand, is the use of patient-related document clipping, we may conclude that for some of the physicians, this approach has a positive influence on the precision of the retrieved document set. However, for one of the resident physicians and two of the recently graduated physicians, the precision values of retrieval methods C and D are inferior to the results of retrieval methods A and B. Consequently, for these physicians, patient-related document clipping has a negative influence on the precision of the retrieved documents. The difference between the physicians might be explained by a difference in their proportion of relevance assessments. Attending physicians and resident physicians tend

to label only a few documents as relevant, whereas recently graduated physicians tend to label many documents as relevant. Due to this difference, discarding documents leads to a larger decrease in precision for recently graduated physicians than for attending physicians or resident physicians. Consequently, discarding documents makes a contribution to the retrieval process for attending physicians and resident physicians.

Relative recall

With respect to the recall $RR_{p,m,i}$ of the retrieved documents (Subsection 5.5.1), we may draw two conclusions.

First, we may conclude that patient-related document scoring has hardly any positive influence on the relative recall values. We observed that for all attending physicians and two of the resident physicians, there is no difference between the retrieval methods A and B, or between the retrieval methods C and D. For one of the resident physicians and two of the recently graduated physicians, patient-related document scoring has a marginally negative influence on the relative recall. For one of the recently graduated physicians, the influence is positive. Since the only difference between methods A and C on the one hand, and methods B and D on the other hand, is the use of patient-related document scoring, we may conclude that addition of patient data to the query has little positive influence on the recall of the retrieved document set.

Second, we may conclude that patient-related document clipping has a severely negative influence on the recall of the retrieved document set according to all physicians. This effect is most prominent. for the resident physicians and the recently graduated physicians. For the attending physicians, the effect is less severe. This difference in influence of patient-related document clipping may be explained by the fact that we use the *relative* recall. When not using document clipping, the resident physicians and the recently graduated physicians had high recall values, whereas the recall values for the attending physicians were quite low (cf. Table 5.10). When using document clipping, many documents were discarded. Since the attending physicians labelled many documents irrelevant, discarding of these documents had only little influence on the relative recall. However, since the resident physicians and the recently graduated physician labelled many documents relevant, the removal of these documents had a large influence on the relative recall. Overall, we may conclude that discarding documents has a negative influence on the relative recall of the retrieved documents.

Provisional Conclusions

Based on the conclusions with respect to precision and recall of the retrieved documents, we may draw three provisional conclusions with respect to the retrieval performance.

First, with respect to precision, patient-related document scoring has hardly any influence according to recently graduated physicians and only little positive influence

according to attending physicians and resident physicians. Patient-related document clipping has some positive influence. Consequently, with respect to precision, we would prefer either retrieval method A or C.

Second, with respect to relative recall, patient-related document scoring has hardly any positive influence and patient-related document clipping has a severely negative influence. Consequently, with respect to relative recall, we would prefer either method A or B.

Third, based on the abovementioned conclusions, we prefer retrieval method A for document retrieval. This method does not use patient-related document scoring and it avoids the negative influence of patient-related document clipping. Consequently, this method is optimal when taking both precision and relative recall into account.

5.6.2 Formulation

With respect to the precision $P_{p,i}$ of the formulated information needs (Subsection 5.5.2), we may provisionally conclude that for the attending physicians and the resident physicians, the formulation performance of MIRA is sufficient. For the recently graduated physicians, the formulation performance seems inadequate. However, many of the information needs that were labelled non-relevant by the recently graduated physicians resulted in documents that were labelled relevant. Consequently, the physicians are aware of the importance of the answer (i.e., the retrieved document), but unaware of the importance of the corresponding question (i.e., the formulated information need). Hence, the information need is implicit. Consequently, the results indicate that MIRA is capable of formulating implicit information needs. As a result, MIRA may be useful for educational purposes, i.e., in teaching recently graduated physicians which questions are important in the medical process.

Furthermore, we may conclude that the average number of information needs formulated per interaction (3.1) is well below the maximum of 10. Consequently, from this perspective, the formulation performance of MIRA may be considered adequate as well.

5.6.3 Ranking

We discuss the ranking performance with respect to the three types of physicians separately.

Attending Physicians

With respect to the precision $P_{p,m,s}$ of the ranked documents according to the attending physicians (Subsection 5.5.3), we may draw three conclusions.

First, we may conclude that the use of knowledge concerning the physician's specialism leads to a ranking that is quite similar to an optimal ranking. This may be explained by the fact the observed physicians are attending physicians. Since attending physicians are specialized in a narrow field of medicine, they probably prefer information needs tuned to their area of expertise.

Second, we may conclude that the entry time of the patient data is not useful as a ranking criterion. It has no effect or a negative effect on the ranking performance. This may be explained by the fact that the data were all entered in quite a small time period.

Third, we may conclude that the use of the number of retrieved documents is superior over the use of knowledge concerning the entry time, but still, the use of this criterion does not produce a ranking superior to a random ranking.

Resident Physicians

With respect to the precision $P_{p,m,s}$ of the ranked documents according to the three resident physicians (Subsection 5.5.3), we may draw three provisional conclusions.

First, we may conclude that the use of knowledge concerning the physician's specialism has a negative influence on the ranking performance.

Second, we may conclude that the use of knowledge concerning the entry time of the patient data results in the best ranking performance. This may be explained by the fact that the observed physicians are resident physicians. Since they are not yet completely specialized in one specific field, they might prefer information needs concerning various subjects. Therefore, tuning the information needs towards a specific specialism is not beneficial. In contrast, the physicians probably prefer the criteria of entry time, to ensure that the information needs concerning recent information are given priority over information needs concerning less recent information.

Third, we may conclude that the use of knowledge concerning the number of retrieved documents has a positive influence on the ranking performance.

Recently Graduated Physicians

With respect to the precision $P_{p,m,s}$ of the ranked documents according to the recently graduated physicians (Subsection 5.5.3), we may draw two conclusions.

First, we may conclude that use of knowledge concerning the number of retrieved documents has a negative influence on the ranking performance. This may be explained by the fact that recently graduated physicians prefer a larger number of retrieved documents than resident physicians or attending physicians. Consequently, if information needs are labelled less relevant if they result in a larger number of retrieved documents, the information needs with the smallest number of retrieved documents are ranked as most relevant, whereas the opposite should be the case for recently graduated physicians.

Second, we may conclude that for the recently graduated physicians , it is not quite clear which ranking method performs best. This may be explained by the fact that the observed physician is a recently graduated physician. In general, physicians who just graduated have relatively little medical experience. Consequently, they have little preference of one information need over another. Therefore, the information needs do not have to be ranked according to specific criteria.

Provisional Conclusions

Based on the conclusions with respect to the precision of the ranked information needs, we may draw two provisional conclusions with respect to the ranking performance.

First, we may conclude that there is not one ranking method performing best for all physicians. Consequently, the ranking method to be used has to be based on the physician's level of experience.

Second, we may conclude that for attending physicians, ranking works best when using knowledge concerning the physician's specialism, for resident physicians, ranking works best when using knowledge concerning the entry time of the patient data, and recently graduated physicians have no explicit preference for any of the ranking criteria.

5.6.4   Unobtrusiveness

With respect to the unobtrusiveness of MIRA (Subsection 5.4.4) we may draw two provisional conclusions.

First, according to one of the attending physicians, MIRA is unobtrusive, provided that the presented literature is 'to the point'. According to the other two attending physicians, MIRA is obtrusive, although we have no information on their reason to think so. The reason might be that most of the literature is known to them.

Second, according to the resident physicians and the recently graduated physician, MIRA is not obtrusive. This may be explained by the fact that, when compared to attending physicians, the medical experience of resident physicians and recently graduated physicians is relatively low. Therefore, they have more need for patient-related information when interacting with the EMR.

5.7   Chapter Conclusions

In this chapter we evaluated MIRA by applying four evaluation criteria (Subsection 5.1) on the assessments of nine physicians (Subsection 5.2). With respect to the evaluation we may draw four provisional conclusions.

First, with respect to the retrieval performance, we may conclude that the investigated precision-enhancements do not enhance the precision of the retrieved document set. Consequently, the final version of MIRA will not incorporate any of the precision-enhancement approaches.

Second, with respect to the formulation performance, we may conclude that the performance of MIRA is adequate according to the attending physicians and the resident physicians. For the recently graduated physicians, the results indicate that MIRA is capable of generating implicit information needs and retrieving documents answering the implicit information needs.

Third, with respect to the ranking performance, we may conclude that the ranking method to be used depends on the type of physician using MIRA. Attending physicians prefer ranking according to their specialism, resident physicians prefer

ranking according to entry time of the patient data, and recently graduated physicians have no explicit preference for any of the ranking criteria.

Fourth, with respect to the unobtrusiveness of MIRA, we may conclude that according to the attending physicians, MIRA is obtrusive. However, according to the resident physicians and the recently graduated physicians, MIRA is unobtrusive. The difference in assessments may be attributed to the difference in experience among the physicians.

Chapter 6

Conclusions and Future Research

In this chapter we complete our research, give recommendations, and provide directions for future research. In Section 6.1 we answer the research questions presented in Chapter 1. Then, Section 6.2 discusses our answer to our twofold problem statement. Subsequently, in Section 6.3 we provide a discussion of our research. Finally, Section 6.4 presents two directions for future research.

6.1 Answers to Research Questions

In Section 1.3, we formulated five research questions. This section provides an answer to each of these questions, based on the conclusions from the previous chapters.

Research question 1 (RQ1): *How can the information needs of a physician be modelled?*

In Chapter 2 we described our efforts in modelling a physician's information needs. We established that it is possible to model a physician's information needs based on templates. We identified 181 information needs (see Appendix A). These information needs were identified by a literature study and interviews with physicians (Section 2.3).

Based on the identified information needs we developed INMOD, a model of a physician's information needs, consisting of 102 templates. The templates were constructed by analysis, abstraction, and refinement of the identified information needs (Section 2.4).

The templates constitute a model in the sense that they are abstractions of reality, i.e., a physician's information needs. INMOD can be used to formulate a physician's information needs in a specific patient situation. Moreover, we showed that INMOD is viable and may be employed in other medical domains. Consequently,

our answer to RQ1 is that a physician's information needs may be modelled based on templates.

Furthermore, from our modelling effort we may conclude that modelling of information needs is an important step in the process of automatically formulating information needs. Without a model, it is impossible to handle all possible information needs that physicians may have.

Research question 2 (RQ2): *How can patient-related information needs be formulated?*

In Chapter 3 we presented our approach for transforming the templates in INMOD into patient-related information needs. We arrived at the conclusion that an adequate approach comprises four steps. First, useful templates are selected based on their hierarchical position and the available data in the EMR (Section 3.1). Second, appropriate patient data are selected from the patient's EMR based on their semantic types (Section 3.2). Third, the patient data are translated into English (Section 3.3). Fourth, the templates are instantiated with the translated patient data (Section 3.4). Consequently, our answer to RQ2 is that we can formulate patient-related information needs, based on the above-mentioned procedure.

Furthermore, we may conclude that the formulated information needs may be used as a starting point for literature retrieval. In order to do this, the information needs are transformed into queries. These queries are executed by a regular IR system on MEDLINE (in our research a subset of MEDLINE), which results in the retrieval of patient-related literature (Section 3.5).

Furthermore, we may conclude that our approach for formulating information needs is adequate and can be generalized to other EMRs, as long as they use a clear information structure. However, the evaluation indicated that the number of formulated information needs per interaction and the number of retrieved documents per interaction are still high (Section 3.7). In this way, physicians would be overloaded with information. Consequently, the basic retrieval approach is not adequate.

Research question 3 (RQ3): *How can different medical terminologies be mapped?*

In Chapter 3, we mentioned that Dutch non-standardized terms in the EMR present a real obstacle to our approach of retrieving literature. To overcome this obstacle, we developed two distinct translation mechanisms: translation based on a manually constructed mapping: MANUTRANS (Subsection 3.3.1) and automatic translation: AUTOTRANS (Subsection 3.3.2). Both mechanisms were evaluated based on their effectiveness and scalability (Section 3.6). A general answer to RQ3 is that both mechanisms are useful methods for translation of Dutch medical terms into UMLS concepts.

However, comparison of both methods provides us with a more specific answer to RQ3 (Subsection 3.6.2). On the one hand, for our purpose MANUTRANS was much more effective than AUTOTRANS. It translated 73.6% of the terms correctly, against 47.1% for AUTOTRANS. On the other hand, AUTOTRANS outperformed MANUTRANS

in terms of scalability. However, since we value effectiveness above scalability for the purpose of our research, we chose MANUTRANS to be employed in our retrieval system.

We note that this answer is only valid when the EMR uses a non-standard medical terminology. If the EMR uses a standard medical terminology, mapping may be performed directly via UMLS. Consequently, our final answer to RQ3 is that non-standard medical terminologies are best mapped by employing a manually constructed mapping.

Research question 4 (RQ4): *How can a literature overload be prevented?*

In Chapter 4 we discussed two ways to prevent a literature overload: (a) reducing the number of information needs, and (b) improving the precision of the set of retrieved documents.

The number of information needs may be reduced by using additional knowledge (Subsection 4.1.3). In this respect we investigated four types of knowledge: (a) formulation history, (b) entry order, (c) domain knowledge, and (d) number of retrieved documents (Subsection 4.2.1). We evaluated all possible combinations of these knowledge types. From the results of the evaluation we may conclude that the combination of all four knowledge types leads to a reduction of 99.2% in the number of information needs (from 315 to 3 per interaction). Furthermore, when using this combination of knowledge types, MIRA formulated information needs with a precision of 0.92 for the attending physicians, 0.98 for the resident physicians, and 0.56 for the recently graduated physicians.

Subsequently, the formulated information needs are ranked according to their relevance (Subsection 4.1.3). For some interactions, the number of formulated information needs exceeds 10. In that case, only the 10 most relevant information needs will be presented to the user. With respect to the ranking we investigated three additional types of knowledge: (a) specialism, (b) entry time, and (c) number of retrieved documents. We evaluated all possible combinations of these knowledge types (Subsection 5.4.3). According to our evaluation results, we may conclude that the specific combination to be used depends on the experience level of the physician using MIRA (Subsection 5.5.3). For the attending physicians, the best ranking is obtained when using only knowledge concerning the physician's specialism. For the resident physicians, the best ranking is obtained when using only knowledge concerning the entry time of the patient data. For the recently graduated physicians, there is not one best ranking method.

With respect to the precision enhancement of the retrieved literature we investigated two distinct methods: (a) patient-related document scoring and (b) patient-related document clipping (Section 4.3). We evaluated all four combinations of methods (Subsection 5.4.1). From our evaluation we observed that the use of patient-related document scoring hardly any positive influence on the precision and hardly any positive influence on the relative recall. However, the use of patient-related document clipping had a significantly negative influence on the recall, opposed to a marginally positive influence on the precision. Consequently, since patient-

related document scoring is hardly useful and patient-related document clipping has a negative influence, we may conclude that it is best to use none of the precision-enhancement approaches.

Consequently, our answer to RQ4 is that a literature overload may be prevented by (a) reducing the number of information needs by using four types of knowledge and (b) ranking the formulated information needs and presenting the 10 most relevant ones (if the number of information needs exceeds 10). The ranking method to be used depends on the experience level of the physician.

Research question 5 (RQ5): *How can the* IR *system be designed to be unobtrusive?*

We designed MIRA to be unobtrusive in three respects. First, MIRA is coupled directly to the EMR and works in the background. Thus, MIRA requires no input from the physician. The only interaction of the physician with MIRA is the presentation of the results once per interaction. Since none of the physicians indicated that their daily workflow would be interrupted by MIRA, we may conclude that MIRA is indeed unobtrusive in this respect.

Second, MIRA is designed not to overload the physician with literature. The number of information needs is reduced and only a single article is presented per information need. Nevertheless, one of the recently graduated physicians still considered the number of documents retrieved per interaction too high. As a result, we may conclude that prevention of a literature overload is an appropriate method of ensuring unobtrusiveness. However, this method is not yet exploited to its full extent and has to be improved.

Third, MIRA is designed to provide physicians with relevant literature only. Still, one of the resident physicians considered the retrieved documents not sufficiently relevant. Additionally, one of the attending physicians indicated that he would like to be provided with literature by MIRA, provided that the literature is 'to the point'. Consequently, we may conclude that the retrieval of relevant literature is an important method of ensuring unobtrusiveness. However, in the current version of MIRA, this method is not yet adequately implemented.

Our answer to RQ5 is that MIRA can be designed to be unobtrusive by (a) preventing workflow interruptions, (b) preventing literature overload, and (c) preventing retrieval of non-relevant literature. In its current version, MIRA is only unobtrusive with respect to the prevention of workflow interruptions. With respect to the prevention of literature overload and the prevention of retrieval of non-relevant literature, MIRA is still obtrusive for some physicians. Consequently, the latter two methods are not performing adequately and have to be improved. Furthermore, two of the attending physicians labelled MIRA as obtrusive without specifying a specific reason. As a result, we have to consider the possibility that there are other reasons for MIRA to be obtrusive. Future research may focus on the detection of these reasons and the development of methods to eliminate them.

6.2   Answer to the Problem Statement

In this section we provide an answer to our twofold problem statement as posed in Chapter 1. The answer is based on the answers to the five research questions discussed in the previous section.

Problem statement:

PS1: *To what extent can a physician's information needs – implicit and suppressed – be formulated automatically?*

PS2: *To what extent can the automatically formulated information needs be used as a starting point for the retrieval of relevant and patient-related literature?*

For PS1, we may conclude that it is possible to formulate a physician's information needs automatically, even if the information needs are implicit. This is achieved by creating a model, INMOD, of a physician's information needs and instantiating INMOD with translated patient data (cf. the answers to RQ1, RQ2, and RQ3). From the evaluation of the formulated information needs, we may conclude that their average precision ranges from 0.56 to 0.98 depending on the type of physician using MIRA. For two of the three types of physicians, the average precision exceeds our target value of 0.90.

For PS2, we may conclude that the automatically formulated information needs may be used as a starting point for literature retrieval. However, this is only feasible under four conditions, viz. (1) the number of information needs should be reduced to a manageable size, (2) the information needs should be ranked according to their relevance, (3) the precision of the retrieved documents should be sufficiently high, and (4) MIRA should be unobtrusive (cf. the answers to RQ4 and RQ5). This research resulted in a fully functional version of MIRA that was evaluated by physicians in an intensive care environment. From the evaluation of the documents retrieved by MIRA, we may conclude that their precision does not reach our target value of 0.90.

6.3   Discussion

In this section, we discuss four topics with respect to our research, viz. (1) the ranking approach, (2) the evaluation persons, (3) the document collection, and (4) the generalizability of our conclusions.

Our first point of discussion concerns the ranking of information needs. As mentioned in Subsection 4.2.2, we used three knowledge types for the ranking of information needs, viz. (a) specialism, (b) entry time, and (c) number of retrieved documents. For each of these knowledge types, we assumed a linear relationship between the knowledge type and the rank of an information need. The question is whether this assumption is correct. Obviously, there are many other functions that could describe the relation between the knowledge type and the rank of an

information need. However, to preserve model parsimony, we chose to use a linear function. If a specific function ought to be designed for each template separately, more complex and exact functions would be more appropriate. Since we did not design functions for each template separately, we employed a linear relationship that fits all templates and consequently, all possible information needs. Furthermore, ranking is employed to discard the least relevant information needs when more than 10 information needs are formulated. However, in Subsection 5.5.2, we observed that the average number of formulated information needs is 3.1. Consequently, in most cases no information needs have to be discarded. Nevertheless, in these cases the ranking is also useful, because it provides the physician with additional information concerning the relevance of the information needs.

Our second point of discussion concerns the evaluation persons. During the evaluation, MIRA was assessed by nine physicians (Section 5.2). We acknowledge that it would have been beneficial if MIRA had been assessed by a larger group of physicians. In this way a more reliable evaluation would have been obtained. However, nine physicians was the maximum number available to us during our research period.

The third point of discussion is the employed document collection. In our research we used the Ohsumed document collection, which is a subset of MEDLINE articles published from 1987–1991 (Subsection 3.7.2). This set is limited in two respects. First, the number of documents is rather small. Second, the documents in the set are quite old. As a consequence, the information in the articles is not up to date, which may influence the precision of the retrieved documents negatively. Therefore, it would have been more beneficial to use a larger and more recent set of documents, preferably, the current complete MEDLINE collection itself. However, searching MEDLINE directly presents us with two problems. First, searching MEDLINE via the ENTREZ programming utilities is subject to strict user requirements.[1] For instance, running a query only once every three seconds. This causes MIRA to run too slow for actual use. The only possible way to search the complete MEDLINE collection is by searching an offline copy of the collection. However, such a collection was not available to us for use during our research. Second, ENTREZ does not rank documents according to relevance like LUCENE (Subsection 3.5.3). Consequently, it is not possible to determine which document is most relevant and present only that document to the user. Because of these two restrictions, we employed the Ohsumed document collection instead of MEDLINE itself.

The fourth point of discussion is the generalizability of our conclusions. With respect to RQ1, we concluded that a physician's information needs may be modelled by templates and that such a model is necessary to handle all of a physician's information needs. Since the templates in INMOD were not developed for any medical domain in particular, INMOD may be used in any medical domain. However, we tested INMOD solely in the particular field of anaesthesiology. Consequently, our conclusions concerning INMOD are not necessarily generalizable to any other medical domain. With respect to RQ2, we concluded that information needs may be formulated automatically by means of a four-step process based on patient data in

---

[1]http://eutils.ncbi.nlm.nih.gov/entrez/query/static/eutils_help.html#UserSystemRequirements

the EMR. This conclusion is generalizable in two respects. First, the process may be applied to any possible EMR, as long as it uses a clear information structure. Second, the process may be applied to any medical domain in which structured EMR data are available, along with domain-specific templates. However, the conclusion of automatic formulation of information needs is not necessarily generalizable with respect to the diagnostic process. The reason is that information needs related to diagnosis are hard to anticipate, since they rely on physicians' thought processes instead of EMR data. With respect to RQ3, we concluded that non-standard medical terminologies are best mapped by means of a translation approach employing a manually constructed mapping (MANUTRANS). This conclusion is not necessarily generalizable to other medical domains. In our manually constructed mapping, we used specific terms from the field of anaesthesiology. However, when using terms from other medical domains, it is possible that an automatic translation approach (AUTOTRANS) outperforms MANUTRANS. Consequently, when employing MIRA in another medical domain, a new domain-specific version of MANUTRANS has to be developed and compared to AUTOTRANS. In general, the use of UMLS for the translation of patient data is generalizable to other medical domains. With respect to RQ4, we concluded that a literature overload may be prevented by employing additional knowledge. This conclusion is generalizable to other medical domains. The knowledge types employed by MIRA are based on the specific templates in INMOD. Since INMOD is solely tested in the domain of anaesthesiology, this holds for the knowledge types as well. Consequently, we may not conclude that the specific knowledge types employed by MIRA perform equally well in other medical domains. Furthermore, we concluded that the use of document scoring and document clipping do not have a positive influence on the precision of the retrieved documents. We expect this conclusion to be generalizable to other medical domains, since these methods are not based on domain-specific features. With respect to RQ5 we concluded that our current version of MIRA is only unobtrusive with respect to the prevention of workflow interruptions may be obtrusive. With respect to the prevention of the prevention of a literature overload and the prevention of the retrieval of non-relevant literature, MIRA is still obtrusive. Despite the fact that the obtrusiveness was assessed specifically in the field of anaesthesiology, the evaluated forms of obtrusiveness apply to other medical domains as well. Consequently, it is highly probable that our conclusion concerning unobtrusiveness is generalizable to other medical domains.

6.4   Future Research

With respect to our research we suggest three directions of future research.

Our first suggestion is to investigate to what extent a user model improves the performance of MIRA, since user models have been successfully employed in the medical domain before (see e.g., Pratt and Sim, 1995; Quintana, 1998). In the absence of a user model, information needs are only formulated with respect to the physician's specialism. However, even among physicians within the same specialism, the preference for specific information needs will vary. A user model could be used to tune the

formulated information needs to the preferences of a specific physician. We expect that extending MIRA with a user model would be beneficial to the performance of MIRA, but the correctness of this expectation has to be confirmed by future research. The exact set of preferences that has to be taken into account by the user model is a subject of investigation as well. A straightforward preference would be a physician's personal interest, but it is highly probable that other preferences may be applicable too.

Our second suggestion for future research is to investigate how the incorporation of user feedback influences the performance and the unobtrusiveness of MIRA. User feedback is an important topic in IR research (cf. Sufyan Beg, 2005; Kelly and Fu, 2006). On the one hand, user feedback is a powerful tool for improving retrieval performance. On the other hand, user feedback requires an extra effort from the user, increasing the obtrusiveness of MIRA. Consequently, it would be interesting to investigate how user feedback may be obtained without disturbing the physician in his daily work and to what extent the obtained user feedback may improve MIRA's performance.

Our third suggestion of discussion is to extend the evaluation of MIRA. As mentioned in Chapter 5, we evaluated MIRA based on the physicians' assessments of output by MIRA. After implementation of the above mentioned directions for future research, another interesting approach would be to perform an online evaluation. In such a setting, in which physicians actually use MIRA in their daily work, a more reliable assessment of MIRA's performance might be obtained.

In addition to the suggestions for future research stated above, we suggest three design issues that may be improved by future development.

First, MIRA's performance might be improved by basing INMOD on a more extensive investigation of a physician's information needs. In our research we conducted a literature study, interviews, and a registration procedure. The most results were obtained by means of the literature study, but it would be interesting to obtain more information needs from physicians themselves.

Second, it would be beneficial to extend MIRA to handle multiple languages and terminologies. Terminology mapping and language mapping are major research topics within medical informatics and bioinformatics (cf. Lee, Supekar, and Geller, 2006b; Lu *et al.*, 2006). However, since terminology mapping and language mapping were not the aim of our current research, we restricted MIRA to handling Dutch terms from the specific terminology used by the ICIS. This was achieved by terminology mapping via UMLS. Yet, it would be an improvement if MIRA were extended to other languages and terminologies by means of a more general of more extensive terminology mapping approach. If successful, such an approach would significantly extend the employability of MIRA.

Third, it would be interesting to link MIRA to other EMRs. For the purpose of our current research, we linked MIRA to only one particular EMR: the ICIS. In Chapter 3 we concluded that our approach for formulating information needs is extendable to other EMRs, as long as they use a clear data structure. Consequently, it would be beneficial to use this advantage and link MIRA to multiple EMRs.

References

Abate, M. A., Shumway, J. M., and Jacknowitz, A. I. (1992). Use of two online services as drug information sources for health professionals. *Methods of information in medicine*, Vol. 31, No. 2, pp. 153–158.

Anderson, J. G., Jay, S. J., Anderson, M., and Hunt, T. J. (2003). Evaluating the impact of information technology on medication errors: a simulation. *Journal of the American Medical Informatics Association*, Vol. 10, No. 3, pp. 292–293.

Anderson, J. G. (2004). Information technology for detecting medication errors and adverse drug events. *Expert Opinion of Drug Safety*, Vol. 3, No. 5, pp. 449–455.

Balas, E. A. and Boren, S. A. (2000). Managing clinical knowledge for health care improvement. *Yearbook of Medical Informatics* (eds. J. Bemmel and A. T. McCray), pp. 65–70, Schattauer Publishing Company, Stuttgart, Germany.

Barnett, G. O. (2006). Report to the National Institutes of Health Division of Research Grants Computer Research Study Section on Computer Applications in Medical Communication and Information Retrieval Systems as Related to the Improvement of Patient Care and the Medical Record – September 26, 1966. *Journal of the American Medical Informatics Association*, Vol. 13, No. 2, pp. 127–135.

Belmont, E. and Waller, A. A. (2003). The role of information technology in reducing medical errors. *Journal of Health Law*, Vol. 36, No. 4, pp. 615–625.

Berkenstadt, H., Yusim, Y., Katznelson, R., Ziv, A., Livingstone, D., and Perel, A. (2006). A novel point-of-care information system reduces anaesthesiologists' errors while managing case scenarios. *European Journal of Aneasthesiology*, Vol. 23, No. 3, pp. 239–250.

Borgman, C. L. (1987). The study of user behavior on information retrieval systems. *ACM SIGCUE Outlook*, Vol. 19, Nos. 3–4, pp. 35–48.

Braun, L. M. M., Wiesman, F., and Sprinkhuizen-Kuyper, I. G. (2002). Information retrieval from historical corpora. *Proceedings of the 3rd Dutch Information Retrieval Workshop (DIR2002)* (eds. D. Hiemstra, W. Kraaij, and M. F. Moens), pp. 106–112, Katholieke Universiteit Leuven, Leuven, Belgium.

Braun, L. M. M., Wiesman, F., van den Herik, H. J., Hasman, A., and Korsten, H. H. M. (2005). Towards Automatic Formulation of a Physician's Information Needs. *Journal of Digital Information Management*, Vol. 3, No. 1, pp. 40–47. Also published in *Proceedings of the 5th Dutch-Belgian Information Retrieval Workshop (DIR 2005)*, (ed. R. van Zwol), pp. 25-32. Centre for Content and Knowledge Engineering, Utrecht, The Netherlands.

Braun, L. M. M., Wiesman, F., van den Herik, H. J., and Hasman, A. (2006). Avoiding Literature Overload in the Medical Domain. *Proceedings of the 20th International Congress of the European Federation for Medical Informatics (MIE 2006)* (eds. A. Hasman, R. Haux, J. van der Lei, E. de Clercq, and F. H. Roger France), Vol. 124, pp. 497–502, Maastricht, The Netherlands.

Braun, L. M. M., Wiesman, F., Hasman, A., van den Herik, H. J., and Korsten, H. H. M. (2007). Towards Patient-Related Information Needs. *International Journal of Medical Informatics*, Vol. 76, Nos. 2–3, pp. 246–251.

Catchpole, K., Bell, M. D., and Johnson, S. (2008). Safety in anaesthesia: a study of 12,606 reported incidents from the UK National Reporting and Learning System. *Aneasthesia*, Vol. 63, No. 4, pp. 340–346.

Cimino, J. J., Johnson, S. B., Aguirre, A., Roderer, N., and Clayton, P. D. (1992). The MEDLINE Button. *Proceedings of the Annual Symposium on Computer Applications in Medical Care*, pp. 81–85.

Cimino, J. J., Elhanan, G., and Zheng, Q. (1997). Supporting infobuttons with terminological knowledge. *Proceedings of the AMIA Annual Fall Symposium*, pp. 528–532.

Cimino, J. J., Li, J., Bakken, S., and Patel, V. L. (2002). Theoretical, Empirical, and Practical Approaches to Resolving the Unmet Information Needs of Clinical Information System Users. *Proceedings of the AMIA Annual Symposium*, pp. 170–174.

Cimino, J. J. (2006). Use, Usability, Usefulness, and Impact of an Infobutton Manager. *Proceedings of the AMIA Annual Symposium*, pp. 151–155.

Collins, S., Currie, L., Bakken, S., and Cimino, J. J. (2006). Interruptions during the use of a CPOE system for MICU rounds. *Proceedings of the AMIA Annual Symposium*, p. 895.

Collins, S., Currie, L., Patel, V., and Cimino, J. J. (2007). Multitasking by clinicians in the context of CPOE and CIS use. *Proceedings of* MEDINFO *2007*, Vol. 12, pp. 958–962.

Cucina, R. J., Shah, M. K., Berrios, D. C., and Fagan, L. M. (2001). Empirical Formulation of a Generic Query Set for Clinical Information Retrieval Systems.

MedIinfo2001. Proceedings of the 10th World Congress on Medical Informatics, London, United Kingdom (eds. V. Patel, R. Rogers, and R. Haux), pp. 181–185, IOS Press, Amsterdam, The Netherlands.

De Vries Robbé, P. F., Beckers, W. P. A., and Zanstra, P. E. (1988). *MEDES. Het prototype.* Onderzoeksgroep Medische Informatie- en Besliskunde, Academische Ziekenhuis Groningen, Groningen, The Netherlands. (In Dutch).

Degnan, D., Merryfield, D., and Hultgren, S. (2004). Reaching out to clinicians: implementation of a computerized alert system. *Journal for Healthcare Quality: official publication of the National Association for Healthcare Quality*, Vol. 26, No. 6, pp. 26–30.

Deville, G. (2001). Corpus-based sublanguage modelling for NLP applications: a tentative methodology. *Proceedings of the International Colloquium on Trends in Special Language & Language Technology* (eds. R. Temmerman and M. Lutjeharms), Standaard Editions, Antwerp, Belgium.

Ely, J. W., Osheroff, J. A., and Ebell, M. H. (1999). Analysis of Questions Asked by Family Doctors Regarding Patient Care. *British Medical Journal*, Vol. 319, No. 7206, pp. 358–361.

Ely, J. W., Osheroff, J. A., Ebell, M. H., Chambliss, M. L., Vinson, D. C., Stevermer, J. J., and Pifer, E. A. (2002). Obstacles to answering doctors' questions about patient care with evidence: qualitative study. *British Medical Journal*, Vol. 324, No. 7339, pp. 710–716.

Fenichel, C. H. (1981). Online Searching: Measures That Discriminate among Users with Different Types of Experiences. *Journal of the American Society for Information Science*, Vol. 32, No. 1, pp. 23–32.

Furman, C. (2005). Implementing a patient safety alert system. *Nursing Economics*, Vol. 23, No. 1, pp. 42–45.

Gale, J. (1982). Some Cognitive Components of the Diagnostic Thinking Process. *British Journal of Educational Psychology*, Vol. 52, No. 1, pp. 64–76.

Gamble, S. (1996). Hospital Libraries Enhance Patient Care and Save Money. *Journal of the Alberta Association of Library Technicians*, Vol. 23, No. 2, pp. 10–12.

Gorman, P. N. (1995). Information Needs of Physicians. *Journal of the American Society for Information Science*, Vol. 46, No. 10, pp. 729–736.

Grossman, D. A. and Frieder, O. (2004). Retrieval Strategies. *Information Retrieval*, Vol. 15 of *The Information Retrieval Series*, Chapter 2. Springer.

Grundmeijer, H. G. L. M., Reenders, K., and Rutten, G. E. H. M. (1999). *Het Geneeskundig Proces. Van Klacht naar Therapie.* Elsevier Gezondheidszorg, Maarssen, The Netherlands. (In Dutch).

Haas Binder, P., Stoever, J., Dickinson Dent, S., Hill, T., and Krause, C. (1997). Expert Describes Phases of Medical Record Computerization. *FP Report*. Vol. 2, No. 7, AAFP News Department, Kansas City, MO.

Hersh, W. and Hickam, D. (1994). Use of a multi-application computer workstation in a clinical setting. *Bulletin of the Medical Library Association*, Vol. 82, pp. 382–389.

Hersh, W., Buckley, C., Leone, T., and Hickam, D. (1994). OHSUMED: An interactive retrieval evaluation and new large test collection for research. *Proceedings of the 17th Annual ACM SIGIR Conference*, pp. 192–201.

Hully, S. B., Cummings, S. R., Browner, W. S., Grady, D., Hearst, N., and Newman, T. B. (2001). *Designing Clinical Research*. Lippincott Williams & Wilkins, Philadelphia, 2nd edition.

Jennings, N. R. (2000). On agent-based software engineering. *Artificial Intelligence*, Vol. 117, No. 2, pp. 277–296.

Jensen, M. (2004). Medical terminology. http://www.gen.umn.edu/faculty sta/jensen/1135/med term activites/default big list.html.

Jerome, R. N., Giuse, N. B., Gish, K. W., Sathe, N. A., and Dietrich, M. S. (2001). Information needs of clinical teams: analysis of questions received by the Clinical Informatics Consult Service. *Bulletin of the Medical Library Association*, Vol. 89, No. 2, pp. 177–185.

Kelly, D. and Fu, X. (2006). Eliciting better information need descriptions from users of information search systems. *Information Processing & Management*, Vol. 43, No. 1, pp. 30–46.

Klein, M. S., Ross, F. V., Adams, D. L., and Gilbert, C. M. (1994). Effect of online literature searching on length of stay and patient care costs. *Academic Medicine: Journal of the Association of American Medical Colleges*, Vol. 69, No. 6, pp. 489–495.

Kohn, L., Corrigan, J., and Donaldson, M. (eds.) (2000). *To Err is Human: Building a Safer Health System*. National Academies Press, Washington, DC.

Koopmans, R. and Offringa, M. (2003). De Juiste Vragen Stellen. *Inleiding in Evidence-Based Medicine* (eds. M. Offringa, W. Assendelft, and R. Scholten), Chapter 2, pp. 11–17. Bohn Stafleu Van Loghum, Houten, The Netherlands. (In Dutch).

Kopec, D., H, K. M., Reinharth, D., Rothschild, O., and Castiglione, J. A. (2003). Human errors in medical practice: systematic classification and reduction with automated information systems. *Journal of Medical Systems*, Vol. 27, No. 4, pp. 297–313.

Koshy, R. (2005). Navigating the information technology highway: computer solutions to reduce errors and enhance patient safety. *Transfusion*, Vol. 45, No. 4, pp. 189S–205S.

Latoszek-Berendsen, A., Talmon, J., and Hasman, A. (2006). With Good Intentiions. *Proceedings of the 20th International Congress of the European Federation for Medical Informatics (MIE 2006)* (eds. A. Hasman, R. Haux, J. van der Lei, E. de Clercq, and F. H. Roger France), Vol. 124 of *Studies in Health Technology and Informatics*, pp. 509–514, Maastricht, The Netherlands.

Latoszek-Berendsen, A., Talmon, J., Clercq, P. de, and Hasman, A. (2007). With good intentions. *International Journal of Medical Informatics*, Vol. 76, No. 3, pp. 440–446.

Ledley, R. S. and Lusted, L. B. (1960). The use of electronic computers in medical data processing: aids in diagnosis, current information retrieval, and medical record keeping. *IRE Transactions on Medical Electronics*, Vol. 7, pp. 31–47.

Lee, M., Cimino, J., Zhu, H. R., Sable, C., Shanker, V., Ely, J., and Yu, H. (2006a). Beyond information retrieval – medical question answering. *Proceedings of the AMIA Annual Symposium*, pp. 469–473.

Lee, Y., Supekar, K., and Geller, J. (2006b). Ontology integration: experience with medical terminologies. *Computers in Biology and Medicine*, Vol. 36, Nos. 7–8, pp. 893–919.

Leibovitch, L., Levy, Y., and Shoenfeld, Y. (1996). Pancreatitis Induced by Clarithromycin. *Annals of Internal Medicine*, Vol. 125, No. 8, p. 701.

Lemieux, M. and Bordage, G. (1992). Propositional versus Structural Semantic Analyses of Medical Diagnostic Thinking. *Cognitive Science*, Vol. 16, pp. 185–204.

Lu, W. H., Lin, R. S., Chan, Y. C., and Chen, K. H. (2006). Overcoming terminology barrier using Web resources for cross-language medical information retrieval. *Proceedings of the Annual Symposium of the American Medical Association 2006 (AMIA)*, pp. 519–523.

Luhn, H. P. (1958). The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*, Vol. 2, No. 2, pp. 159–165.

Manning, C. D., Raghavan, P., and Schütze, H. (2007). Scoring, term weighting & the vector space model. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY.

Maviglia, S. M., Yoon, C. S., Bates, D. W., and Kuperman, G. (2006). KnowledgeLink: Impact of Context-Sensitive Information Retrieval on Clinicians' Information Needs. *Journal of the American Medical Informatics Association*, Vol. 13, No. 1, pp. 67–73.

McCarn, D. B. and Leiter, J. (1973). On-line services in medicine and beyond. *Science*, Vol. 181, No. 4097, pp. 318–324.

Merry, P. (1997). Slow to Learn. *The Health Service Journal*, Vol. 107, pp. 28–29.

Miller, R. A., Jamnback, L., Giuse, N. B., and Masarie, F. E. (1991). Extending the Capabilities of Diagnostic Decision Support Programs through Links to Bibliographic Searching: Addition of 'Canned MeSH Logic' to the Quick Medical Reference (QMR) Program for Use with Grateful Med. *Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care (SCAMC), Washinton, DC* (ed. P. Clayton), pp. 150–155, McGraw-Hill, Inc., New York, NY.

Nanus, B. (1960). The use of electronic computers for information retrieval. *Bulletin of the Medical Library Association*, Vol. 48, No. 3, pp. 278–91.

NeVeMeDis (2004). Wat is een Medische Fout?
http://www.nevemedis.nl/fouten/ft_watismedfout.htm. (In Dutch).

NICTIZ (2007). Programma van Eisen voor een Goed Beheerd Zorgsysteem (GBZ).
https://www.nictiz.nl/uploaded/FILES/AORTA%20release%20mei%202007/
Programma%20van%20Eisen%20GBZ%20v2.0.pdf. (In Dutch).

NLM (2006a). UMLS Metathesaurus. Fact Sheet.
http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html.

NLM (2006b). UMLS Semantic Network. Fact Sheet.
http://www.nlm.nih.gov/pubs/factsheets/umlssemn.html.

NLM (2006c). UMLS SPECIALIST Lexicon. Fact Sheet.
http://www.nlm.nih.gov/pubs/factsheets/umlslex.html.

NLM (2006d). Unified Medical Language System. Fact Sheet.
http://www.nlm.nih.gov/pubs/factsheets/umls.html.

NLM (2007). Medical Subject Headings. Fact Sheet.
http://www.nlm.nih.gov/pubs/factsheets/mesh.html.

NLM (2008). UMLS Knowledge Sources. April Release, 2008AA Documentation.
http://www.nlm.nih.gov/research/umls/umlsdoc.html.

Norman, G., Young, M., and Brooks, L. (2007). Non-analytical models of clinical reasoning: the role of experience. *Medical Education*, Vol. 41, No. 12, pp. 1140–1145.

Offringa, M., Assendelft, W., and Scholten, R. (2003). Inleiding. *Inleiding in Evidence-Based Medicine* (eds. M. Offringa, W. Assendelft, and R. Scholten), Chapter 1, pp. 1–10. Bohn Stafleu Van Loghum, Houten, The Netherlands. (In Dutch).

O'Leary, D. (2000). Focus on Reducing Medical Errors. Statement of the Joint Commission on Acreditation of Healthcare Organizations. February 22, 2000, Washington, DC.

Oppenheim, M. I., Mintz, R. J., Boyer, A. G., and Frayer, W. W. (2000). Design of a clinical alert system to facilitate development, testing, maintenance, and user-specific notification. *Proceedings of the AMIA Symposium*, pp. 630–634.

Patel, V. L., Arocha, J. F., and Zhang, J. (2004). Thinking and Reasoning in Medicine. *Cambridge Handbook of Thinking and Reasoning*. Cambridge University Press, New York, NY.

Powsner, S. M. and Miller, P. L. (1992). Automated online transition from the medical record to the psychiatric literature. *Methods of Information in Medicine*, Vol. 31, No. 3, pp. 169–174.

Powsner, S. M., Riely, C. A., Barwick, K. W., Morrow, J. S., and Miller, P. L. (1989). Automated bibliographic retrieval based on current topics in hepatology: hepatopix. *Computers and Biomedical Research*, Vol. 22, No. 6, pp. 552–564.

Pratt, W. and Sim, I. (1995). Physician's information customizer (PIC): using a shareable user model to filter the medical literature. *Proceedings of* MEDINFO *1995*, Vol. 8, Pt. 2, pp. 1447–1451, Vancouver, Canada.

Quintana, Y. (1998). Intellingent Medical Information Filtering. *International Journal of Medical Informatics*, Vol. 51, Nos. 2–3, pp. 197–204.

Rada, R., Barlow, J., Bijstra, D., Potharst, J., Vries Robbé, P. de, and Zanstra, P. (1992). OAR: Open Architecture for Reasoning Applied to Connection Patient Records to Medical Literature. *Advances in Medical Informatics* (eds. J. Noothoven van Goor and J. P. Christensen), pp. 287–294, IOS Press, Amsterdam, The Netherlands.

Ramos, K., Linscheid, R., and Schafer, S. (2003). Real-time information-seeking behavior of residency physicians. *Family Medicine*, Vol. 35, No. 4, pp. 257–260.

Raschke, R. A., Gollihare, B., Wunderlich, T. A., Guidry, J. R., Leibowitz, A. I., Peirce, J. C., Lemelson, L., Heisler, M. A., and Susing, C. (1998). A computer alert system to prevent injury from adverse drug events: development and evaluation in a community teaching hospital. *Journal of the American Medical Association*, Vol. 280, No. 15, pp. 13170–1320.

Rassiat, E., Michiels, E., Jouve, C. L., Sgro, C., Faivre, J., and Hillon, P. (2003). Acute pancreatitis after clarithromycin and beta-methasone. *Gastroentérologie clinique et biologique*, Vol. 27, No. 1, p. 123. (In French).

Reddy, M. C., Pratt, W., Dourish, P., and Shabot, M. M. (2002). Asking Questions: Information Needs in a Surgical Intensive Care Unit. *Proceedings of the Annual AMIA Symposium (AMIA'02)* (ed. I. S. Kohane), pp. 647–651, Americal Medical Informatics Association, Bethesda, MD.

Robertson, S. E. (1981). The Methodology of Information Retrieval Experiment. *Information Retrieval Experiment* (ed. K. Sparck Jones), Chapter 2, pp. 9–31. Butterworths, London, UK.

Saracevic, S. (1996). Relevance Reconsidered. *Information Science: Integration in Perspectives. Proceedings of the Second Conference on Conceptions of library and Information Science (CoLIS 2)*, pp. 201–218, Copenhagen, Denmark.

Schouwenberg, B. J. J. W. and Deinum, J. (2003). Acute Pancreatitis after a Course of Clarithromycin. *The Netherlands Journal of Medicine*, Vol. 61, No. 7, pp. 266–267.

Singh, H., Naik, A. D., Rao, R., and A, P. L. (2008). Reducing diagnostic errors through effective communication: harnessing the power of information technology. *Journal of General Internal Medicine: Official Journal of the Society for Research and Education in Primary Care Internal Medicine*, Vol. 23, No. 4, pp. 489–494.

Smith, R. (1996). What Clinical Information Do Doctors Need? *British Medical Journal*, Vol. 313, No. 7064, pp. 1062–1068.

Strasberg, H. R., Hubbs, P. R., Rindfleisch, T. C., and Melmon, K. L. (1999). Analysis of Information Needs of Users of the Stanford Health Information Network for Education. *Proceedings of the Annual AMIA Symposium (AMIA'99)* (ed. N. M. Lorenzi), pp. 965–969, American Medical Informatics Association, Bethesda, MD.

Sufyan Beg, M. M. (2005). User feedback based enhancement in web search quality. *Information Sciences*, Vol. 170, Nos. 2–4, pp. 153–172.

Swanson, D. R. (1960). Searching Natural Language Text by Computer. *Science*, Vol. 132, No. 3434, pp. 1099–1104.

Swets, J. A. (1963). Information Retrieval Systems. *Science*, Vol. 141, No. 3577, pp. 245–250.

Ter Hofstede, A. H. M., Proper, H. A., and Weide, T. P. van der (1996). Query Formulation as an Information Retrieval Problem. *The Computer Journal*, Vol. 39, No. 4, pp. 255–274. British Computer Society.

Tran, T. D., Gracelon, N., Burgun, A., and Le Beux, P. (2004). Experiments in cross-language medical information retrieval using a mixing translation module. *MedInfo2004. Proceedings of the 11th World Congress on Medical Informatics, San Fransisco, USA* (eds. M. Fieschi, E. Coiera, and Y. C. Li), pp. 946–949, IOS Press, Amsterdam, The Netherlands.

Van Mulligen, E. M. (1999). UMLS-Based Access to CPR Data. *International Journal of Medical Informatics*, Vol. 53, Nos. 2–3, pp. 125–131.

Van Rijsbergen, C. J. (1979). *Information Retrieval*. Department of Computer Science, University of Glasgow. Butterworths, London, UK, 2nd edition.

Vermeulen, I. B., Bohte, S. M., Somefun, D. J. A., and La Poutré, J. A. (2006). Improving Patient Schedules by Multi-agent Pareto Appointment Exchanging. *Proceedings of 2006 IEEE International Conference on E-Commerce Technology (CEC/EEE 2006)*. San Fransisco, CA.

Vermeulen, I. B., Bohte, S., and La Poutré, J. A. (2007a). Adaptive patient scheduling with dynamic resource usage. *ERCIM News*, Vol. 69, pp. 53–54.

Vermeulen, I. B., Bohte, S. M., Elkhuizen, S. G., Lameris, J. S., Bakker, P. J. M., and La Poutré, J. A. (2007b). Adaptive Optimization of Hospital Resource Calendars. *Proceedings of the 11th Conference on Artificial Intelligence in Medicine (AIME 07)* (eds. R. Bellazzi, A. Abu-Hanna, and J. Hunter), number 4594 in Lecture Notes in Computer Science, pp. 305–315.

Vermeulen, I. B., Bohte, S. M., Somefun, K., and La Poutré, J. A. (2007c). Multi-agent Pareto appointment exchanging in hospital patient scheduling. *Service Oriented Computing and Applications*, Vol. 1, No. 3, pp. 185–196.

Weigand, H., Wieringa, R., Meyer, J.-J., and Starmans, R. J. C. M. (2004). SIKS Basic Course: 'Research Methods and Methodology for IKS'. Lunteren, The Netherlands.

Westbrook, J. I., Coiera, E. W., and Gosling, A. S. (2005). Do online information retrieval systems help experienced clinicians answer clinical questions? *Journal of the American Medical Informatics Association*, Vol. 12, No. 3, pp. 315–321.

Wieringa, R. and Heerkens, H. (2003). Requirements Engineering as Problem Analysis: Methodology and Guidelines. Reading material for SIKS Basic Course 'Research Methods and Methodology for IKS'. University of Twente, Enschede, The Netherlands.

Wiesman, F., van den Herik, H. J., and Hasman, A. (2004). Information Retrieval by Metabrowsing. *Journal of the American Society for Information Science and Technology*, Vol. 55, No. 7, pp. 565–578.

Zeng, Q. and Cimino, J. J. (2000). Providing Multiple Views to Meet Physician Information Needs. *Proceedings of the 33rd Hawaii International Conference on System Sciences (CD-rom)* (ed. R. H. Sprague), Computer Society Press, IEEE, Los Alamitos, CA.

Zhang, J., Patel, V. L., and Johnson, T. R. (2002). Medical Error: Is the Solution Medical or Cognitive? *Journal of the American Medical Informatics Association*, Vol. 9 (supplement), pp. 75–77.

Appendices

In Appendix A we present the information needs that we identified by means of a literature survey and interviews with physicians. Appendix B provides the templates that constitute our model of physicians' information needs. In Appendix C we list the heuristics that are used to translate patient data. Appendix D provides the materials used during the evaluation of MIRA. In Appendix E we present the evaluation results of MIRA.

Appendix A

Identified Information Needs

In this appendix we present the information needs identified by means of a literature survey (Section A.1) and by means of interviews with physicians (Section A.2). The format of some of the information needs is rather similar to the format of our templates. They can be distinguished by the fact that in our templates, the slots are indicated by angle brackets ('<' and '>'), whereas in the identified information needs, square brackets ('[' and ']') are used.

A.1   Literature Survey

Reddy et al. *(2002)*

- What is the protocol for doing an apnea test?
- When is myocardial infarction most likely to strike post-operatively?

Jerome et al. *(2001)*

- Is pyridoxine deficiency implicated in the etiology of seizures?
- What is the role of helical copumted tomography in assessing aortic disruption?
- What are the differential diagnoses in a patient with lymphadenopathy and constitutional symptoms when standard diagnoses like lymphoma and autoimmune disorders have been ruled out?
- What is the Antley-Bixler syndrome?
- What percentage of patients with deep vein thrombosis develop silent pulmonary embolism?
- What are the long-term outcomes of a neonate who had a middle cerebral infarction?
- Can you provide some information for a patient's family about vancomycin-resistant enterococci?
- What are the treatment options for hypereosinophilic syndrome?
- What are the side effects of giving magnesium for high blood pressure?

- Do lumbar punctures help prevent the development of hydrocephalus following intraventricular hemorrhage?

Cucina et al. *(2001)*

- What is the definition of [MANIFESTATION] / [INVESTIGATION] / [PATHOLOGY] / [THERAPY]?
- What are the risk factors for [MANIFESTATION] / [PATHOLOGY]?
- What is the etiology of [MANIFESTATION] / [PATHOLOGY]?
- Can [MANIFESTATION] / [INVESTIGATION] / [PATHOLOGY] / [THERAPY] cause [MANIFESTATION] / [PATHOLOGY]?
- What is the differential diagnosis of [MANIFESTATION] / [PATHOLOGY]?
- What distinguishes [MANIFESTATION] / [PATHOLOGY] from [MANIFESTATION] / [PATHOLOGY]?
- How can [MANIFESTATION] / [INVESTIGATION] be used in the evaluation of [MANIFESTATION] / [INVESTIGATION] / [PATHOLOGY] / [THERAPY]?
- How can [MANIFESTATION] / [INVESTIGATION] / [PATHOLOGY] / [THERAPY] be evaluated?
- How can [THERAPY] be used in the treatment of [MANIFESTATION] / [INVESTIGATION] / [PATHOLOGY] / [THERAPY]?
- What are the treatments for [MANIFESTATION] / [INVESTIGATION] / [PATHOLOGY] / [THERAPY]?
- How can [THERAPY] be used in the prevention of [MANIFESTATION] / [PATHOLOGY]?
- What are the performance characteristics of [MANIFESTATION] / [INVESTIGATION] / [THERAPY] in the setting of [MANIFESTATION] / [INVESTIGATION] / [PATHOLOGY] / [THERAPY]?
- How does [INVESTIGATION] / [THERAPY] compare with [INVESTIGATION] / [THERAPY] in the setting of [MANIFESTATION] / [INVESTIGATION] / [PATHOLOGY] / [THERAPY]?
- Is [INVESTIGATION] / [THERAPY] contraindicated by [MANIFESTATION] / [INVESTIGATION] / [PATHOLOGY] / [THERAPY]?
- What are the sequelae and prognosis of [MANIFESTATION] / [PATHOLOGY]?
- What are the physical properties of [PATHOLOGY] / [THERAPY]?

Grundmeijer et al. *(1999)*

- Wat zijn mogelijke diagnoses voor [KLACHT] / [SYMPTOOM]?
- Wat is de incidentie van [DIAGNOSE] / [AANDOENING]?
- Wat is de kans op [DIAGNOSE] / [AANDOENING] bij een patiënt met [KLACHTEN]?
- Wat is de kans op [DIAGNOSE] / [AANDOENING] bij een [LEEFTIJD]-jarige [GESLACHT] met [KLACHTEN]?
- Wat is de kans op [DIAGNOSE] / [AANDOENING] bij een patiënt met [KLACHTEN] en [CONTEXT]?

- Wat is de 'gouden standaard' test voor [DIAGNOSE] / [AANDOENING]?
- Wat is de sensitiviteit van [TEST] bij een patiënt met [KLACHTEN]?
- Wat is de specificiteit van [TEST] bij een patiënt met [KLACHTEN]?
- Wat is de positief voorspellende waarde van [TEST] bij een patiënt met [KLACHTEN]?
- Wat is de negatief voorspellende waarde van [TEST] bij een patiënt met [KLACHTEN]?
- Wat zijn de voordelen van [TEST]?
- Wat zijn de nadelen van [TEST]?
- Welke test is geïndiceerd bij een patiënt met [KLACHT]?
- Is de indicatie voor [BEHANDELING] juist voor een patiënt met [DIAGNOSE] / [AANDOENING]?
- Wat is de efficacy van [BEHANDELING] voor een patiënt met [DIAGNOSE] / [AANDOENING]?
- Hoe is de effectiveness van [BEHANDELING] voor een patiënt met [DIAGNOSE] / [AANDOENING]?
- Hoe is de efficiency van [BEHANDELING] voor een patiënt met [DIAGNOSE] / [AANDOENING]?
- Is [BEHANDELING] zinvol?
- Wat zijn de voordelen van [BEHANDELING]?
- Wat zijn de nadelen van [BEHANDELING]?
- Is [MEDICAMENT] geïndiceerd bij een patiënt met [DIAGNOSE] / [AANDOENING]?
- Wat is de doeltreffendheid van [MEDICAMENT]?
- Wat zijn de bijwerkingen van [MEDICAMENT]?
- Hoe groot is het risico op bijwerkingen bij gebruik van [MEDICAMENT]?
- Wat is de interactie van [MEDICAMENT1] met [MEDICAMENT2]
- Is [MEDICAMENT] contrageïndiceerd?
- Wat zijn de kosten van [MEDICAMENT]?
- Wie moet [BEHANDELING] uitvoeren?
- Waar moet [BEHANDELING] uitgevoerd worden?
- Hoe moet [BEHANDELING] uitgevoerd worden?
- Hoe moet / kan een patiënt met [DIAGNOSE] / [AANDOENING] verder begeleid worden na [BEHANDELING]?
- Welke professionele zorg is nodig bij een patiënt met [DIAGNOSE] / [AANDOENING] na [BEHANDELING]?
- Hoe kan [DIAGNOSE] / [AANDOENING] voorkomen worden?
- Is preventie van [DIAGNOSE] / [AANDOENING] doeltreffend?

Ely et al. *(1999)*

- What is the cause of [SYMPTOM]?

- What is the dose of drug [DRUG]?
- How should I manage [DISEASE] / [FINDING]?
- How should I treat [DISEASE] / [FINDING]?
- What is the cause of physical finding [FINDING]?
- What is the cause of test finding [FINDING]?
- Could this patient have [DISEASE] / [CONDITION]?
- Is test [TEST] indicated in situation [SITUATION]?
- What is the drug of choice for [CONDITION]?
- Is drug [DRUG] indicated in situation [SITUATION]?

Smith (1996)

- How best to sort out diabetic neuropathy, vascular pain, discogenic pain, and musculoskeletal causes for back and leg pain in a 70 year old woman (without testing all of them)?
- For diagnosis of deep vein thrombosis, how good is ultrasound?
- Does ultrasound obviate the need for a venogram in case of deep vein thrombosis?
- Can ultrasound exclude the diagnosis in case of deep vein thrombosis?
- Is there a connection between asthma and migraines?
- Can isosorbide dinitrate and/or Lopresor be responsible for itching and rash?
- In an 88 year old woman with dysphagia due to past laryngeal cancer, now in respiratory failure due to aspiration, what is the physician's role in aggressiveness of care decisions when the patient's family has unrealistic expectations?
- In a 60 year old woman with an elevated set rate, guiaic-positive stool attributed to haemorrhoids, and 35 pound weight loss, what work up is appropriate, taking cost and potential benefit into account?

Gorman (1995)

- In a patient with refractory headaches, now benefiting from a calcium channel blocker, is there a specific drug or dose that has been shown to work?
- After 2 courses of antibiotics in a physician's daughter with bronchitis, what treatment is appropriate for persistent symptoms?
- In an octogenarian with anemia, angina, and a history of transients ischemic attacks, with a normal creatinine, iron, and mean corpuscular volume, who refuses a bone marrow exam, what diagnostic and therapeutic options are there?
- Is it safe to use ibuprofen in a 50-year-old man with a history of colon cancer, now reporting dysuria, who has cellular casts in his urine?
- Does Norpace cause fatigue?
- What are the costs, risk and usefulness of dipyridamole thallium scanning in a patient with chronic obstructive lung disease, claudication, and angina pectoris?

- In a woman with sclerosing adenosis on breast biopsy and family history of breast cancer, who requires estrogen therapy to control symptoms, how can the risk of breast cancer be lowered?

- In an 88-year-old woman with dysphagia due to past laryngeal cancer, now in respiratory failure due to aspiration, what is the physician's role in aggressiveness of care decisions when the patient's family has unrealistic expectations?

- For a child with exacerbation of steroid dependent asthma and varicella exposure, how do you give varicella immune globulin and where do you get it?

- Is meclizine effective for labyrinthitis?

- In a man with vague intermittent abdominal and back pain, what additional information will be most useful and what is the complete differential diagnosis?

- Can aspirin or an antiplatelet agent be used as prophylaxis against pulmonary embolism in an elderly woman with unexplained oxygen disaturation and no clinical risk factors for pulmonary embolism?

- In a woman with history of delivering at 33 weeks, now having Braxton-Hicks contractions at 32 weeks, on terbutaline and bedrest, in breech position, is c-section indicated if labor cannot be stopped?

- How can I distinguish and manage chest pain in an older woman with known coronary disease, status post angioplasty of the left anterior descending coronary artery, arthritis which precludes treadmill testing, esophagitis, inadequate personality which complicates history, given that dipyridamole testing is 180 miles away?

- In a patient with steroid dependent chronic obstructive lung disease, does the risk of renal or gastrointestinal complications outweigh the benefit of non-steroidal anti-inflammatory therapy for degenerative joint disease?

- Can an insule-dependent diabetic be certified as a commercial driver?

De Vries Robbé (1988)

- Geef een beschrijving van de farmacodynamiek van [MEDICAMENT] bij [ZIEKTE].

- Geef een beschrijving van de pathofysiologie van [BEVINDING] in geval van [ZIEKTE]?

- Geef een beschrijving van de pathogenese en de verdere gevolgen van [ZIEKTE1] in geval van [ZIEKTE2].

- Geef een beschrijving van de pathogenese en de verdere gevolgen van [BEVINDINGEN] / [ZIEKTE]?

- Geef een beschrijving van de pathologische anatomie in geval van [BEVINDINGEN] / [ZIEKTE]?

- Geef een beschrijving van de pathologische anatomie van [BEVINDINGEN] / [ZIEKTE]?

- Geef een beschrijving van de voor- en nadelen van [ONDERZOEK1] ten opzichte van [ONDERZOEK2] in geval van [BEVINDINGEN] / [ZIEKTE].

- Geef een beschrijving van het pathofysiologisch mechanisme dat ten grondslag ligt aan [BEVINDINGEN] / [ZIEKTE]?

- Hoe duur is [THERAPIE]? Hoe hoog is de dosering van [THERAPIE] in geval van [BEVINDINGEN] / [ZIEKTE]?

- Hoe hoog is de dosering van [THERAPIE] voor [BEVINDINGEN] / [ZIEKTE]?
- Hoe hoog is de dosering van [THERAPIE] voor [BEVINDINGEN] / [ZIEKTE1] en [ZIEKTE2]?
- Hoe hoog is de frequentie van [BEVINDING] in geval van [THERAPIE] voor [ZIEKTE]?
- Hoe hoog is de frequentie van [BEVINDING] in geval van [ZIEKTE]?
- Hoe hoog is de frequentie van [BEVINDINGEN] in geval van [ZIEKTE]?
- Hoe hoog is de frequentie van [BEVINDINGEN]?
- Hoe hoog is de frequentie van [ZIEKTE] bij [BEVINDINGEN]?
- Hoe hoog is de frequentie van [ZIEKTE]?
- Hoe lang duurt [ONDERZOEK] in geval van [BEVINDINGEN] / [ZIEKTE]?
- Hoe lang moet [THERAPIE] in geval van [BEVINDINGEN] / [ZIEKTE] worden gegeven?
- Hoe luidt de prognose gegeven [BEVINDINGEN] / [ZIEKTE]?
- Hoe luidt de prognose gegeven [BEVINDINGEN]?
- Hoe luidt de prognose gegeven [THERAPIE] gegeven [BEVINDINGEN] / [ZIEKTE]?
- Hoe luidt de prognose gegeven [ZIEKTE]?
- Hoe moet [ONDERZOEK] worden uitgevoerd bij [BEVINDING]?
- Hoe moet [THERAPIE] in geval van [BEVINDINGEN] / [ZIEKTE] worden uitgevoerd?
- Hoe moet [THERAPIE] worden uitgevoerd?
- Hoe zijn de behandelresultaten van [THERAPIE] met betrekking tot [BEVINDINGEN] / [ZIEKTE]?
- Hoe zijn de behandelresultaten van [THERAPIE] voor [BEVINDINGEN] / [ZIEKTE1] in geval van [ZIEKTE2]?
- Hoe zijn de behandelresultaten van [THERAPIE] voor [ZIEKTE]?
- Hoeveel tijd verloopt er tussen het optreden van [BEVINDINGEN] en het ontstaan van [ZIEKTE]?
- Kan met behulp van onderzoek meer zekerheid worden verkregen over de prognose van [BEVINDINGEN] / [ZIEKTE]?
- Met welk onderzoek kan meer zekerheid worden verkregen over de prognose van [BEVINDINGEN] / [ZIEKTE]?
- Moet [BEVINDING] in geval van [ZIEKTE] worden gemeld?
- Waardoor wordt [BEVINDING] bevorderd?
- Wat is de duur van [THERAPIE] voor [BEVINDINGEN] / [ZIEKTE]?
- Wat is de duur van [THERAPIE] voor [BEVINDINGEN] / [ZIEKTE1] en [ZIEKTE2]?
- Wat is de farmacodynamiek van [THERAPIE]?
- Wat is de frequentie van [BEVINDINGEN] bij [ZIEKTE] in [POPULATIE]?
- Wat is de frequentie van [ZIEKTE] bij [POPULATIE]?
- Wat is de invloed op de prognose van [BEVINDINGEN] / [ZIEKTE] als nog niet voor een therapie wordt gekozen?
- Wat is de kosten/batenverhouding van [ONDERZOEK] gegeven [BEVINDINGEN] / [ZIEKTE]?

- Wat is de kosten/batenverhouding van [THERAPIE] voor [BEVINDINGEN] / [ZIEKTE]?

- Wat is de oorzaak van [BEVINDING] bij [THERAPIE] voor [ZIEKTE]?

- Wat is de patiëntbelasting van [ONDERZOEK] gegeven [BEVINDINGEN] / [ZIEKTE]?

- Wat is de patiëntbelasting van [THERAPIE] gegeven [BEVINDINGEN] / [ZIEKTE]?

- Wat is de voorspellende waarde van een positieve / negatieve uitslag van [ONDERZOEK] voor aan- / afwezigheid van [BEVINDINGEN] / [ZIEKTE]?

- Wat is het tijdsbeslag van [ONDERZOEK]?

- Wat zijn de kosten van [ONDERZOEK]?

- Wat zijn de risico's van [ONDERZOEK]?

- Welke ziekten komen in aanmerking bij [BEVINDING]?

- Welke bevindingen komen in aanmerking in geval van [BEVINDINGEN] / [ZIEKTE]?

- Welke bijwerkingen heeft [THERAPIE]?

- Welke bijwerkingen van [THERAPIE] voor [BEVINDINGEN] / [ZIEKTE] zijn niet meer acceptabel?

- Welke gevolgen worden voorkomen als [THERAPIE] wordt uitgesteld?

- Welke invloed heeft [THERAPIE] voor [BEVINDINGEN1] / [ZIEKTE] op [BEVINDINGEN2]?

- Welke lokalisaties komen in aanmerking bij [BEVINDINGEN] / [ZIEKTE]?

- Welke micro-organismen kunnen [BEVINDINGEN] / [ZIEKTE] veroorzaken?

- Welke onderzoeken komen gegeven [BEVINDINGEN] / [ZIEKTE] in principe in aanmerking?

- Welke onderzoeken komen gegeven [THERAPIE] voor [BEVINDINGEN] / [ZIEKTE] in principe in aanmerking?

- Welke onderzoeken komen in aanmerking gegeven [BEVINDINGEN] / [ZIEKTE]?

- Welke onderzoeken komen in aanmerking voor [BEVINDING] gegeven [THERAPIE]?

- Welke therapieën hebben een direct aangrijpingspunt in [PATHOFYSIOLOGISCH MECHANISME]?

- Welke therapieën komen gegeven [BEVINDINGEN] / [ZIEKTE] in principe in aanmerking?

- Welke therapieën komen in aanmerking gegeven [BEVINDINGEN] / [ZIEKTE]?

- Welke therapieën komen in aanmerking gegeven [BEVINDINGEN] / [ZIEKTE1] en [ZIEKTE2]?

- Welke therapieën komen in aanmerking gegeven [BEVINDINGEN] / [ZIEKTE] en [THERAPIE]?

- Welke therapieën kunnen [BEVINDINGEN] / [ZIEKTE] voorkomen?

- Welke therapieën voor [BEVINDINGEN] / [ZIEKTE] kunnen [BEVINDINGEN] veroorzaken?

- Welke van de in aanmerking komende onderzoeken heeft de minst ongunstige invloed op [BEVINDINGEN] / [ZIEKTE]?

- Welke voorkeurslokalisaties komen in aanmerking bij [BEVINDINGEN] / [ZIEKTE]?

- Welke ziekten komen in aanmerking in geval van [BEVINDINGEN]?

- Welke ziekten komen in aanmerking in geval van [BEVINDINGEN] bij [THERAPIE] voor [ZIEKTE]?

- Welke zijn de bijwerkingen van [THERAPIE]?

- Welke zijn de interacties van [THERAPIE] met andere therapieën?

- Welke zijn de risico's van [ONDERZOEK]?

A.2   Interviews

- Wat is anno 2004 het optimale onderzoeks- en behandelingstraject voor [AANDOENING]?

- Kan [OPERATIE] uitgevoerd worden bij een patiënt met [CONDITIE]?

- Wat is EBM voor een patiënt met [AANDOENING]?

- Wat zijn de complicaties van [BEHANDELING]?

- Wat zijn de resultaten van [BEHANDELING]?

- Wat is een mogelijke oorzaak van [SYMPTOOM / BEVINDING]?

- Hoe moet [LABORATORIUMONDERZOEK] geinterpreteerd worden?

- Hoe moet [LICHAMELIJK ONDERZOEK] geinterpreteerd worden?

- Wat is [ZIEKTE]?

Appendix B

Information-Need Templates

In this appendix, we list the complete set of templates (in text format) in INMOD. The XML format of a template as used by MIRA can be found in Section 4.5. The templates are derived from the information needs in Appendix A by means of analysis (Subsection 2.4.1), abstraction (Subsection 2.4.2), and refinements (Subsection 2.4.3).

- Can a patient with <DISEASE OR SYNDROME> be certified as a commercial driver?
- What are the long-term outcomes of a patient who had <DISEASE OR SYNDROME>?
- Give a description of the pathogenesis and the further consequences of <DISEASE OR SYNDROME>.
- Is prevention of <DISEASE OR SYNDROME> effective?
- How can <DISEASE OR SYNDROME> be prevented?
- What are the treatment options for <DISEASE OR SYNDROME>?
- Does <CHEMICAL> cause <DISEASE OR SYNDROME>?
- Is <CHEMICAL> effective for <DISEASE OR SYNDROME>?
- Give a description of the pharmacodynamics of <CHEMICAL> in <DISEASE OR SYNDROME>.
- After a course of <CHEMICAL> in a patient with <DISEASE OR SYNDROME>, what treatment is appropriate for persistent symptoms?
- Is <CHEMICAL> indicated in a patient with <DISEASE OR SYNDROME>?
- For a patient with exacerbation of <DISEASE OR SYNDROME> and <FINDING>, how do you give <CHEMICAL> and where do you get it?
- Is it safe to use <CHEMICAL> in a patient with a history of <DISEASE OR SYNDROME>, now reporting <SIGN OR SYMPTOM> who has <FINDING>?
- In a patient with a history of <DISEASE OR SYNDROME>, now having <SIGN OR SYMPTOM>, on <CHEMICAL> and <THERAPEUTIC OR PREVENTIVE PROCEDURE>, is <THERAPEUTIC OR PREVENTIVE PROCEDURE> indicated?

- How is the effectiveness of <SMALL>HEALTH CARE ACTIVITY</SMALL> for a patient with <SMALL>DISEASE OR SYNDROME</SMALL>?

- How is the efficiency of <SMALL>HEALTH CARE ACTIVITY</SMALL> for a patient with <SMALL>DISEASE OR SYNDROME</SMALL>?

- How should / can a patient with <SMALL>DISEASE OR SYNDROME</SMALL> be supported after <SMALL>HEALTH CARE ACTIVITY</SMALL>?

- Is <SMALL>HEALTH CARE ACTIVITY</SMALL> indicated in a patient with <SMALL>DISEASE OR SYNDROME</SMALL>?

- In a patient with <SMALL>DISEASE OR SYNDROME</SMALL> and <SMALL>SIGN OR SYMPTOM</SMALL> with <SMALL>FINDING</SMALL> , who refuses <SMALL>HEALTH CARE ACTIVITY</SMALL> , what diagnostic and therapeutic options are there?

- How can I distinguish and manage <SMALL>SIGN OR SYMPTOM</SMALL> in a patient with known <SMALL>DISEASE OR SYNDROME</SMALL> , status <SMALL>THERAPEUTIC OR PREVENTIVE PROCEDURE</SMALL>, <SMALL>PATHOLOGIC FUNCTION</SMALL>, which complicates history, given that <SMALL>HEALTH CARE ACTIVITY</SMALL> is 180 miles away?

- Should <SMALL>FINDING</SMALL> be reported in case of <SMALL>DISEASE OR SYNDROME</SMALL>?

- Give a description of the pathophysiology of <SMALL>FINDING</SMALL> in case of <SMALL>DISEASE OR SYNDROME</SMALL>.

- In a patient with <SMALL>FINDING</SMALL> due to <SMALL>DISEASE OR SYNDROME</SMALL> , what work up is appropriate, taking cost and potential benefit into account?

- What is the cause of <SMALL>FINDING</SMALL> in case of <SMALL>THERAPEUTIC OR PREVENTIVE PROCEDURE</SMALL> for <SMALL>DISEASE OR SYNDROME</SMALL>?

- How high is the frequency of <SMALL>FINDING</SMALL> in case of <SMALL>THERAPEUTIC OR PREVENTIVE PROCEDURE</SMALL> for <SMALL>DISEASE OR SYNDROME</SMALL>?

- Which diseases are eligible in case of <SMALL>FINDING</SMALL> in case of <SMALL>THERAPEUTIC OR PREVENTIVE PROCEDURE</SMALL> for <SMALL>DISEASE OR SYNDROME</SMALL>?

- Is there a connection between <SMALL>DISEASE OR SYNDROME</SMALL> and <SMALL>DISEASE OR SYNDROME</SMALL>?

- Give a description of the pathogenesis and further consequences of <SMALL>DISEASE OR SYNDROME</SMALL> in case of <SMALL>DISEASE OR SYNDROME</SMALL>.

- In a patient with <SMALL>DISEASE OR SYNDROME</SMALL>, does the risk of <SMALL>PATHOLOGIC FUNCTION</SMALL> outweigh the benefit of <SMALL>THERAPEUTIC OR PREVENTIVE PROCEDURE</SMALL> for <SMALL>DISEASE OR SYNDROME</SMALL>?

- Can <SMALL>CHEMICAL</SMALL> be used as prophylaxis against <SMALL>DISEASE OR SYNDROME</SMALL> in a patient with <SMALL>DISEASE OR SYNDROME</SMALL> and no clinical risk factors for <SMALL>DISEASE OR SYNDROME</SMALL>?

- In a patient with <SMALL>DISEASE OR SYNDROME</SMALL> due to <SMALL>DISEASE OR SYNDROME</SMALL>, now in <SMALL>PATHOLOGIC FUNCTION</SMALL> due to <SMALL>PATHOLOGIC FUNCTION</SMALL>, what is the physician's rol in aggressiveness of care decisions when the patient's family has unrealistic expectations?

- Which therapies are eligible given <SMALL>FINDING | DISEASE OR SYNDROME</SMALL>, and <SMALL>DISEASE OR SYNDROME</SMALL>?

- Which therapies are eligible given <SMALL>FINDING | DISEASE OR SYNDROME</SMALL> and <SMALL>DISEASE OR SYNDROME</SMALL>?

- How are the treatment results of <small>THERAPEUTIC OR PREVENTIVE PROCEDURE</small> for <small>FINDING | DISEASE OR SYNDROME</small> in case of <small>DISEASE OR SYNDROME</small>?

- How high is the dose of <small>CHEMICAL</small> for <small>FINDING | DISEASE OR SYNDROME</small> and <small>DISEASE OR SYNDROME</small>?

- What is the efficacy of <small>LABORATORY PROCEDURE</small> for a patient with <small>FINDING | DISEASE OR SYNDROME</small>?

- Which examinations are eligible given <small>FINDING | DISEASE OR SYNDROME</small>?

- Which findings are eligible in case of <small>FINDING | DISEASE OR SYNDROME</small>?

- Which of the eligible examinations has least unfavorable influence on <small>FINDING | DISEASE OR SYNDROME</small>?

- Can more certainty be obtained about the prognosis of <small>FINDING | DISEASE OR SYNDROME</small> by means of examination?

- Give a description of the pathologic anatomy of <small>FINDING | DISEASE OR SYNDROME</small>.

- Which therapies can prevent <small>FINDING | DISEASE OR SYNDROME</small>?

- What is the prognosis in case of <small>FINDING | DISEASE OR SYNDROME</small>?

- What is the influence on the prognosis of <small>FINDING | DISEASE OR SYNDROME</small> if no therapy is chosen yet?

- With which examination can more certainty be obtained about the prognosis of <small>FINDING | DISEASE OR SYNDROME</small>?

- How should I manage <small>FINDING | DISEASE OR SYNDROME</small>?

- How should I treat <small>FINDING | DISEASE OR SYNDROME</small>?

- Which therapies are eligible given <small>FINDING | DISEASE OR SYNDROME</small>?

- What is the prognosis in case of <small>THERAPEUTIC OR PREVENTIVE PROCEDURE</small> in case of <small>FINDING | DISEASE OR SYNDROME</small>?

- How high is the dose of <small>THERAPEUTIC OR PREVENTIVE PROCEDURE</small> for <small>FINDING | DISEASE OR SYNDROME</small>?

- Which side effects of <small>THERAPEUTIC OR PREVENTIVE PROCEDURE</small> for <small>FINDING | DISEASE OR SYNDROME</small> are not acceptable anymore?

- Which examinations are eligible given <small>THERAPEUTIC OR PREVENTIVE PROCEDURE</small> given <small>FINDING | DISEASE OR SYNDROME</small>?

- Which therapies are eligible given <small>FINDING | DISEASE OR SYNDROME</small> and <small>THERAPEUTIC OR PREVENTIVE PROCEDURE</small>?

- How are the treatment results of <small>THERAPEUTIC OR PREVENTIVE PROCEDURE</small> for <small>FINDING | DISEASE OR SYNDROME</small>?

- What is the influence of <small>THERAPEUTIC OR PREVENTIVE PROCEDURE</small> for <small>FINDING | DISEASE OR SYNDROME</small> on <small>FINDING</small>?

- How long does <small>HEALTH CARE ACTIVITY</small> take in case of <small>FINDING | DISEASE OR SYNDROME</small>?

- What is the cost-benefit ratio of <small>HEALTH CARE ACTIVITY</small> given <small>FINDING | DISEASE OR SYNDROME</small>?

- What is the burden for the patient of <HEALTH CARE ACTIVITY> given <FINDING | DISEASE OR SYNDROME>?

- What is the predictive value of a positive / negative result of <HEALTH CARE ACTIVITY> for presence / absence of <FINDING | DISEASE OR SYNDROME>?

- Give a description of the advantages and disadvantages of <HEALTH CARE ACTIVITY> in relation to <HEALTH CARE ACTIVITY> in case of <FINDING | DISEASE OR SYNDROME>.

- Which therapies for <FINDING | DISEASE OR SYNDROME> can cause <FINDING>?

- What are the costs of <HEALTH CARE ACTIVITY>?

- Is <HEALTH CARE ACTIVITY> useful?

- What are the risks of <HEALTH CARE ACTIVITY>?

- Who has to perform <HEALTH CARE ACTIVITY>?

- How can <HEALTH CARE ACTIVITY> be evaluated?

- Where should <HEALTH CARE ACTIVITY> be performed?

- How should <HEALTH CARE ACTIVITY> be performed?

- How long does <HEALTH CARE ACTIVITY> take?

- Give a description of the advantages and disadvantages of <HEALTH CARE ACTIVITY>.

- How can <HEALTH CARE ACTIVITY> be used in the evaluation of <HEALTH CARE ACTIVITY>?

- How does <HEALTH CARE ACTIVITY> compare with <HEALTH CARE ACTIVITY> in the setting of <SIGN OR SYMPTOM | PATHOLOGIC FUNCTION>?

- What are the performance characteristics of <HEALTH CARE ACTIVITY> in the setting of <SIGN OR SYMPTOM | PATHOLOGIC FUNCTION>?

- Is <HEALTH CARE ACTIVITY> contraindicated by <SIGN OR SYMPTOM | PATHOLOGIC FUNCTION>?

- How should <HEALTH CARE ACTIVITY> be performed in case of <FINDING>?

- What is the effectiveness of <CHEMICAL>?

- What are the side effects of <CHEMICAL>?

- How great is the risk of side in the use of <CHEMICAL>?

- What are the costs of <CHEMICAL>?

- What is the dose of <CHEMICAL>?

- Is <CHEMICAL> contraindicated?

- What is the interaction of <CHEMICAL> with <CHEMICAL>?

- Does <CHEMICAL> cause <SIGN OR SYMPTOM>?

- What are the side effects of <CHEMICAL> for <SIGN OR SYMPTOM>?

- In a patient with <SIGN OR SYMPTOM>, now benefiting from <CHEMICAL>, is there a specific drug or dose that has been shown to work?

- What are the pharmacodynamics of <THERAPEUTIC OR PREVENTIVE PROCEDURE>?

- Which consequences are prevented if <THERAPEUTIC OR PREVENTIVE PROCEDURE> is postponed?

- What are the interactions of <THERAPEUTIC OR PREVENTIVE PROCEDURE> with other therapies?

- What are the side effects of <THERAPEUTIC OR PREVENTIVE PROCEDURE>?

- How can <THERAPEUTIC OR PREVENTIVE PROCEDURE> be used in the prevention of <PATHOLOGIC FUNCTION>?

- How can <THERAPEUTIC OR PREVENTIVE PROCEDURE> be used in the treatment of <PATHOLOGIC FUNCTION>?

- What is the negative predictive value of <LABORATORY PROCEDURE> in a patient with <FINDING>?

- What is the positive predictive value of <LABORATORY PROCEDURE> in a patient with <FINDING>?

- What is the sensitivity of <LABORATORY PROCEDURE> in a patient with <FINDING>?

- What is the specificity of <LABORATORY PROCEDURE> in a patient with <FINDING>?

- What are the risk factors of <PATHOLOGIC FUNCTION>?

- What are the sequelae and prognosis of <PATHOLOGIC FUNCTION>?

- What distinguishes <PATHOLOGIC FUNCTION> from <PATHOLOGIC FUNCTION>?

- What are the treatments for <SIGN OR SYMPTOM | PATHOLOGIC FUNCTION>?

- What is the drug of choice for <SIGN OR SYMPTOM | PATHOLOGIC FUNCTION>?

- What are the physical properties of <PATHOLOGIC FUNCTION | HEALTH CARE ACTIVITY>?

- Which therapies have a direct point of application in <PHENOMENON OR PROCESS>?

- Can <PHENOMENON OR PROCESS | HEALTH CARE ACTIVITY> cause <PHENOMENON OR PROCESS>?

Appendix C

Heuristics

This appendix lists the heuristics used in the automatic translation of patient data (Subsection 3.3.2).

<div align="center">Table C.1: Heuristics.</div>

Lexical heuristics			
(→	ok →	gi→ gastrointestinaal	intermit→ intermittent
) →	ic →	gyn→ gynaecologisch	extr→ extremiteit
op →	zgn →	iv→ intraveneus	
icu →	po→ per os	impl→ implantatie	

Morphological heuristics			
aal → al	eus → ous	megalie → megaly	plastie → plasty
aat → ate	filie → philia	metrie → metry	plegie → plegia
air → ary	filie → phily	morf → morph	pneu → pnea
ale → al	fobie → phobia	odynie → odynia	poiese → poiesis
algie → algia	fonie → phonia	oide → oid	porose → porosis
alis → al	geen → gen	ologie → ology	prandiaal → prandial
atresie → atresia	genese → genesis	oloog → ologist	praxie → praxia
capnie → capnia	genetisch → genic	ome → oma	ptose → ptosis
centese → centesis	graaf → graph	oom → oma	ptyse → ptysis
cidaal → cidal	grafie → graphy	opie → opia	sarcoom → sarcoma
clase → clasis	iaal → ial	opsie → opsy	sclerose → sclerosis
clasie → clasia	iatrie → iatry	orrhagie → orrhagia	scoop → scope
clyse → clysis	ictaal → ictal	orrhee → orrhea	scopie → scopy
coc → coccus	icus → ician	ose → osis	scopisch → scopic
crien → crine	ie → ia	ostomie → ostomy	se → sis
criet → crit	ine → in	otomie → otomy	spasme → spasm
cyt → cyte	isch → ic	oxie → oxia	stase → stasis

Table C.1: (continued)

dese → desis	isme → ism	parese → paresis	stenose → stenosis
droom → drome	ite → itis	pathie → pathy	tie → tion
eaal → eal	lepsie → lepsy	penie → penia	tocie → tocia
ectase → ectasis	logie → logy	pepsie → pepsia	toir → tory
ectomie → ectomy	lyse → lysis	phagie → phagia	tomie → tomy
ectopie → ectopia	lyt → lyte	phorie → phoria	tripsie → tripsy
emese → emesis	lytisch → lytic	physe → physis	trophie → trophy
emie → emia	malacie → malacia	plasie → plasia	
ese → esis	manie → mania	plasma → plasm	

Appendix D

Evaluation Materials

In this appendix we present the patient data (Section D.1) and the evaluation forms (Sections D.2 and D.3) used in the evaluation of our system. Since MIRA was evaluated by Dutch physicians, the evaluation forms are in Dutch (except the information needs and the retrieved documents).

D.1   Evaluation Data

Table D.1: Evaluation materials.

	Day 1	
Opname-indicatie	MVR	19:08:00
Opname-indicatie	CABG Enkelvoudig	19:08:00
Opname-indicatie	Acuut Myocard Infarct	19:08:00
Opname-indicatie	Myocardinfarct < 8 weken	19:08:00
Behandeling	Arterie lijn op OK	19:18:43
Medicatie	Acenocoumarol	19:18:43
Behandeling	Beademing	19:18:43
Behandeling	Maagsonde	19:18:44
Onderzoek	ECG dagelijks	19:18:44
Behandeling	Perifeer infuus	19:18:44
Medicatie	Glucose/Zout	19:18:44
Behandeling	Halsinf./subclavia op OK	19:18:44
Behandeling	Catheter a Demeure	19:18:44
Behandeling	Fysiotherapie	19:18:45
Medicatie	Piritramide	19:18:45
Behandeling	Basiszorg	19:18:45
Behandeling	Thoraxdrain	19:18:45
Onderzoek	X-thorax dagelijks	19:18:45

Table D.1: (continued)

Medicatie	Propofol	19:18:46
Medicatie	Cefazoline	19:18:46
Medicatie	Fragmin	19:18:46
Medicatie	Alfentanyl	19:18:46
Medicatie	Thiamazol	19:19:04
Medicatie	Levothyroxine	19:19:16

Day 2

Behandeling	Swan Ganz op OK	00:47:50
Behandeling	Pacemaker AAN	00:48:10
Behandeling	Pacemaker standby	00:48:18
Behandeling	Low flow bed	13:23:09
Onderzoek	TEE	13:23:39
Medicatie	Midazolam	15:18:22
Medicatie	Sucralfaat	15:18:37
Medicatie	Dobutamine	15:19:11
Medicatie	Adrenaline	15:19:19
Medicatie	Nitroglycerine-iv	15:19:25
Medicatie	Oculetem simplex	15:19:47
Medicatie	Nutrison Protein Plus MF	15:20:10
Medicatie	Glucose/Zout	15:20:55

Day 3

Onderzoek	Wegen dagelijks	00:26:31
Onderzoek	Synacthen	08:28:52
Medicatie	Acetylcysteine	12:50:02
Medicatie	Ipratropiumbr. Salbutamol	12:50:08
Behandeling	Wisselligging	13:33:34

Day 4

Medicatie	Nutrison Protein Plus MF	11:29:03
Medicatie	Glucose/Zout	11:29:49
Behandeling	Weanen	13:10:44
Behandeling	Pacemaker standby	21:44:57
Behandeling	Sonde-Voeding	21:45:30
Behandeling	Ballonneren	21:46:15
Behandeling	Bronchiaal toilet	21:46:24
Behandeling	Wondzorg overig	21:46:39

Day 5

Onderzoek	BEE	09:21:05
Medicatie	Furosemide-iv	10:18:08
Medicatie	Amiodaron	14:06:47

Table D.1: (continued)

Day 6		
Onderzoek	Sputum kweek	08:43:18
Medicatie	Spironolacton	13:53:11

Day 7		
Onderzoek	Sputum kweek	05:43:22
Medicatie	Captopril	11:32:06
Medicatie	Nutrison Protein Plus MF	11:33:34
Medicatie	Glucose/Zout	11:34:21
Onderzoek	Bloedkweek 1	16:16:12
Onderzoek	Bloedkweek 2	16:16:22
Onderzoek	Wond kweek	16:16:31
Onderzoek	Kweek sheath	16:16:45
Onderzoek	Urine kweek	16:17:19

Day 8		
Medicatie	Captopril	09:41:15
Medicatie	Fosfaatdrank	09:43:11
Onderzoek	Bloedkweek 1	21:45:09
Onderzoek	Cito GRAM + sputumkweek	21:45:15
Onderzoek	Urine kweek	21:45:21

Day 9		
Onderzoek	Bloedkweek 1	05:16:34
Onderzoek	Bloedkweek 2	05:16:41
Onderzoek	I.V Catheter kweek overig	05:16:49
Behandeling	Perifeer infuus 2	07:12:11
Medicatie	Amoxi/Clavulaan	12:00:33
Onderzoek	Urine kweek	14:12:32
Onderzoek	Cito GRAM + sputumkweek	14:12:47
Onderzoek	Wond kweek	14:12:56

Day 10		
Onderzoek	Bloedkweek 1	05:16:34
Onderzoek	Bloedkweek 2	05:16:41
Onderzoek	I.V Catheter kweek overig	05:16:49
Behandeling	Perifeer infuus 2	07:12:11
Medicatie	Amoxi/Clavulaan	12:00:33
Onderzoek	Urine kweek	14:12:32
Onderzoek	Cito GRAM + sputumkweek	14:12:47
Onderzoek	Wond kweek	14:12:56

Table D.1: (continued)

	Day 11	
Medicatie	Hydrochloorthiazide	10:26:03
Medicatie	Glucose 5%	14:06:10

D.2 Document Assessment Form

Table D.2: Assessment form for documents.

Indicate for each article whether you would like to read the complete article (based on the title and abstract).	Absolutely not	No	Yes	Absolutely	If not, why not?	I already know the information in this article	The title and abstract suffice	The article is not relevant w.r.t. this patient
1. Clinical aspects of silent myocardial ischemia. Effects of treatment.	☐	☐	☐	☐		☐	☐	☐
2. Early percutaneous transluminal coronary angioplasty or coronary bypass surgery following thrombolytic treatment of acute myocardial infarction.	☐	☐	☐	☐		☐	☐	☐
3. Improvement in long-term survival among patients hospitalized with acute myocardial infarction, 1970 to 1980. The Minnesota Heart Survey.	☐	☐	☐	☐		☐	☐	☐
4. Methimazole, carbimazole, and congenital skin defects.	☐	☐	☐	☐		☐	☐	☐
5. Outcome of coronary artery bypass grafting in black persons.	☐	☐	☐	☐		☐	☐	☐
6. Perioperative acute myocardial infarction after valve replacement.	☐	☐	☐	☐		☐	☐	☐
7. Predictors, prevention, and long-term prognosis of atrial fibrillation after coronary artery bypass graft operations.	☐	☐	☐	☐		☐	☐	☐
8. Reduced high density lipoproteins as a risk factor after acute myocardial infarction.	☐	☐	☐	☐		☐	☐	☐
9. Secondary prevention trials after acute myocardial infarction.	☐	☐	☐	☐		☐	☐	☐

Table D.2: (continued)

Indicate for each article whether you would like to read the complete article (based on the title and abstract).	Absolutely not	No	Yes	Absolutely	If not, why not?	I already know the information in this article	The title and abstract suffice	The article is not relevant w.r.t. this patient
10. Single-dose antibiotic prophylaxis for biliary surgery. Cefazolin vs moxalactam.	☐	☐	☐	☐		☐	☐	☐
11. Subclinical hyperthyroidism and reduced bone density as a possible result of prolonged suppression of the pituitary-thyroid axis with L-thyroxine.	☐	☐	☐	☐		☐	☐	☐
12. Variability of L-thyroxine replacement dose in elderly patients with primary hypothyroidism.	☐	☐	☐	☐		☐	☐	☐
13. Why does antihypertensive treatment prevent stroke but not myocardial infarction?	☐	☐	☐	☐		☐	☐	☐
14. Comparative effects on hemodynamics of enoximone (MDL 17,043), dobutamine and nitroprusside in severe congestive heart failure.	☐	☐	☐	☐		☐	☐	☐
15. Comparison of subcutaneous injection and high-dose inhalation of epinephrine–implications for self-treatment to prevent anaphylaxis.	☐	☐	☐	☐		☐	☐	☐
16. Early tolerance to hemodynamic effects of high dose transdermal nitroglycerin in responders with severe chronic heart failure.	☐	☐	☐	☐		☐	☐	☐
17. Effect of nitroglycerin during hemodynamic estimation of valve orifice in patients with mitral stenosis.	☐	☐	☐	☐		☐	☐	☐
18. Epinephrine staining of a soft contact lens. Case report.	☐	☐	☐	☐		☐	☐	☐

Table D.2: (continued)

Indicate for each article whether you would like to read the complete article (based on the title and abstract).	Absolutely not	No	Yes	Absolutely	If not, why not?	I already know the information in this article	The title and abstract suffice	The article is not relevant w.r.t. this patient
19. Hemodynamic effects of diazepam, flunitrazepam, and midazolam in patients with ischemic heart disease: assessment with a radionuclide approach.	☐	☐	☐	☐		☐	☐	☐
20. Lack of effect of sucralfate on prednisone bioavailability.	☐	☐	☐	☐		☐	☐	☐
21. Lactic acidosis and insulin resistance associated with epinephrine administration in a patient with non-insulin-dependent diabetes mellitus.	☐	☐	☐	☐		☐	☐	☐
22. Methemoglobinemia from intravenous nitroglycerin: a word of caution.	☐	☐	☐	☐		☐	☐	☐
23. The effects of age, epinephrine, and operative site on duration of caudal analgesia in pediatric patients.	☐	☐	☐	☐		☐	☐	☐
24. Liver injury after oral and rectal administration of N-acetylcysteine for meconium ileus equivalent in a patient with cystic fibrosis.	☐	☐	☐	☐		☐	☐	☐
25. Nitrate tolerance: the lack of effect of N-acetylcysteine.	☐	☐	☐	☐		☐	☐	☐
26. Amiodarone-induced thrombocytopenia.	☐	☐	☐	☐		☐	☐	☐
27. Papillopathy caused by amiodarone.	☐	☐	☐	☐		☐	☐	☐

Table D.2: (continued)

Indicate for each article whether you would like to read the complete article (based on the title and abstract).	Absolutely not	No	Yes	Absolutely	If not, why not?	I already know the information in this article	The title and abstract suffice	The article is not relevant w.r.t. this patient
28. Short- and long-term effects of furosemide on lung function in infants with bronchopulmonary dysplasia.	☐	☐	☐	☐		☐	☐	☐
29. Thyroid dysfunction during chronic amiodarone therapy.	☐	☐	☐	☐		☐	☐	☐
30. A comparison of hemodynamic effects of one-month oral captopril and enoximone treatment for severe congestive heart failure.	☐	☐	☐	☐		☐	☐	☐
31. Angiotensin-converting enzyme inhibitors: considerations regarding proteinuria.	☐	☐	☐	☐		☐	☐	☐
32. Captorpil-induced liver dysfunction.	☐	☐	☐	☐		☐	☐	☐
33. Cough due to captopril.	☐	☐	☐	☐		☐	☐	☐
34. Na+, K+, and BP homeostasis in man during furosemide: effects of prazosin and captopril.	☐	☐	☐	☐		☐	☐	☐
35. Cochlear neural degeneration without hair cell loss in two patients with aminoglycoside ototoxicity.	☐	☐	☐	☐		☐	☐	☐
36. Cromolyn sodium inhibits the increased responsiveness to methacholine that follows ultrasonically nebulized water challenge in patients with asthma.	☐	☐	☐	☐		☐	☐	☐
37. Eye injuries caused by directed jets of water from a fire hose.	☐	☐	☐	☐		☐	☐	☐

Table D.2: (continued)

Indicate for each article whether you would like to read the complete article (based on the title and abstract).	Absolutely not	No	Yes	Absolutely	*If not, why not?*	I already know the information in this article	The title and abstract suffice	The article is not relevant w.r.t. this patient
38. Single-dose gentamicin therapy of recurrent urinary tract infection in patients with normal urinary tracts.	☐	☐	☐	☐		☐	☐	☐
39. Hypercapnia during total parenteral nutrition with hypertonic dextrose.	☐	☐	☐	☐		☐	☐	☐
40. Low-dose glucose infusion in patients who have undergone surgery. Possible cause of a muscular energy deficit.	☐	☐	☐	☐		☐	☐	☐
41. The effect of protein ingestion on the metabolic response to oral glucose in normal individuals.	☐	☐	☐	☐		☐	☐	☐

D.3 Information-Needs Assessment Form

Table D.3: Assessment form for information needs.

Indicate for each article whether you would like to read the complete article (based on the title and abstract).	Absolutely not	No	Yes	Absolutely	If not, why not?	The answer to this question is already known to me	The answer is irrelevant w.r.t. this patient	The answer is completely irrelevant
1. How can Myocardial Infarction be prevented for this specific patient?	☐	☐	☐	☐		☐	☐	☐
2. Give a description of the pathogenesis and the further consequences of Myocardial Infarction for this specific patient.	☐	☐	☐	☐		☐	☐	☐
3. Is prevention of Myocardial Infarction effective for this specific patient?	☐	☐	☐	☐		☐	☐	☐
4. What are the long-term outcomes of a patient who had Coronary Artery Bypass for this specific patient?	☐	☐	☐	☐		☐	☐	☐
5. What are the long-term outcomes of a patient who had Myocardial Infarction for this specific patient?	☐	☐	☐	☐		☐	☐	☐
6. What are the risk factors of Myocardial Infarction for this specific patient?	☐	☐	☐	☐		☐	☐	☐
7. What are the side effects of Methimazole for this specific patient?	☐	☐	☐	☐		☐	☐	☐
8. What are the side effects of Thyroxine for this specific patient?	☐	☐	☐	☐		☐	☐	☐

Table D.3: (continued)

Indicate for each article whether you would like to read the complete article (based on the title and abstract).	Absolutely not	No	Yes	Absolutely	If not, why not?	The answer to this question is already known to me	The answer is irrelevant w.r.t. this patient	The answer is completely irrelevant
9. What are the treatment options for Coronary Artery Bypass for this specific patient?	☐	☐	☐	☐		☐	☐	☐
10. What are the treatment options for Myocardial Infarction for this specific patient?	☐	☐	☐	☐		☐	☐	☐
11. What is the dose of Cefazolin for this specific patient?	☐	☐	☐	☐		☐	☐	☐
12. What is the dose of Thyroxine for this specific patient?	☐	☐	☐	☐		☐	☐	☐
13. What are the side effects of Epinephrine for this specific patient?	☐	☐	☐	☐		☐	☐	☐
14. What are the side effects of Nitroglycerin for this specific patient?	☐	☐	☐	☐		☐	☐	☐
15. What is the dose of Epinephrine for this specific patient?	☐	☐	☐	☐		☐	☐	☐
16. What is the dose of Nitroglycerin for this specific patient?	☐	☐	☐	☐		☐	☐	☐
17. What is the effectiveness of Dobutamine for this specific patient?	☐	☐	☐	☐		☐	☐	☐
18. What is the effectiveness of Epinephrine for this specific patient?	☐	☐	☐	☐		☐	☐	☐

Table D.3: (continued)

Indicate for each article whether you would like to read the complete article (based on the title and abstract).	Absolutely not	No	Yes	Absolutely	If not, why not?	The answer to this question is already known to me	The answer is irrelevant w.r.t. this patient	The answer is completely irrelevant
19. What is the effectiveness of Midazolam for this specific patient?	☐	☐	☐	☐		☐	☐	☐
20. What is the effectiveness of Nitroglycerin for this specific patient?	☐	☐	☐	☐		☐	☐	☐
21. What is the effectiveness of Sucralfate for this specific patient?	☐	☐	☐	☐		☐	☐	☐
22. What are the side effects of Acetylcysteine for this specific patient?	☐	☐	☐	☐		☐	☐	☐
23. What is the effectiveness of Acetylcysteine for this specific patient?	☐	☐	☐	☐		☐	☐	☐
24. How great is the risk of side effects in the use of Amiodarone for this specific patient?	☐	☐	☐	☐		☐	☐	☐
25. What are the side effects of Amiodarone for this specific patient?	☐	☐	☐	☐		☐	☐	☐
26. What is the effectiveness of Furosemide for this specific patient?	☐	☐	☐	☐		☐	☐	☐
27. How great is the risk of side effects in the use of Captopril for this specific patient?	☐	☐	☐	☐		☐	☐	☐
28. What are the side effects of Captopril for this specific patient?	☐	☐	☐	☐		☐	☐	☐

Table D.3: (continued)

Indicate for each article whether you would like to read the complete article (based on the title and abstract).	Absolutely not	No	Yes	Absolutely	If not, why not?	The answer to this question is already known to me	The answer is irrelevant w.r.t. this patient	The answer is completely irrelevant
29. What is the effectiveness of Captopril for this specific patient?	☐	☐	☐	☐		☐	☐	☐
30. What are the side effects of Gentamicins for this specific patient?	☐	☐	☐	☐		☐	☐	☐
31. What are the side effects of Water for this specific patient?	☐	☐	☐	☐		☐	☐	☐
32. What is the dose of Gentamicins for this specific patient?	☐	☐	☐	☐		☐	☐	☐
33. What is the dose of Water for this specific patient?	☐	☐	☐	☐		☐	☐	☐
34. What are the side effects of Glucose for this specific patient?	☐	☐	☐	☐		☐	☐	☐
35. What is the dose of Glucose for this specific patient?	☐	☐	☐	☐		☐	☐	☐
36. What is the effectiveness of Glucose for this specific patient?	☐	☐	☐	☐		☐	☐	☐

Appendix E

Evaluation Results

In this appendix we present the ranking performance of our system with respect to attending physicians (Tables E.1 to E.3), resident physicians (Tables E.4 to E.6), and recently graduated physicians (Tables E.7 to E.9). The results are discussed in Subsection 5.5.3.

Table E.1: $P_{p,m,s}$ according to attending physician 1, obtained with eight different ranking methods ($R1$ to $R8$) for information-need sets of different sizes s, and $P_{p,m,s}$ for the optimal ranking method (Opt).

Size	R1	R2	R3	R4	R5	R6	R7	R8	Opt
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	1.00	1.00	0.92	1.00	1.00	1.00	1.00	1.00	1.00
3	1.00	1.00	0.94	1.00	1.00	1.00	1.00	1.00	1.00
4	1.00	1.00	0.83	1.00	1.00	1.00	1.00	1.00	1.00
5	1.00	1.00	0.70	1.00	1.00	1.00	1.00	1.00	1.00
6	0.92	1.00	0.67	0.92	0.92	0.92	0.92	0.92	1.00
7	0.86	0.93	0.64	0.93	0.86	0.86	0.93	0.86	1.00
8	0.88	0.88	0.69	0.88	0.81	0.81	0.88	0.81	0.94
9	0.83	0.83	0.72	0.83	0.78	0.78	0.83	0.78	0.89
10	0.70	0.70	0.50	0.60	0.60	0.60	0.60	0.60	0.70

Table E.2: $P_{p,m,s}$ according to attending physician 2, obtained with eight different ranking methods ($R1$ to $R8$) for information-need sets of different sizes s, and $P_{p,m,s}$ for the optimal ranking method (Opt).

Size	R1	R2	R3	R4	R5	R6	R7	R8	Opt
1	0.88	0.88	0.75	0.75	0.88	0.88	0.75	0.88	1.00
2	0.86	0.93	0.79	0.86	0.93	0.93	0.86	0.93	0.93
3	0.78	0.83	0.78	0.83	0.83	0.83	0.83	0.83	0.83
4	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table E.3: $P_{p,m,s}$ according to attending physician 3, obtained with eight different ranking methods ($R1$ to $R8$) for information-need sets of different sizes s, and $P_{p,m,s}$ for the optimal ranking method (Opt).

Size	R1	R2	R3	R4	R5	R6	R7	R8	Opt
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table E.4: $P_{p,m,s}$ according to resident physician 1, obtained with eight different ranking methods ($R1$ to $R8$) for information-need sets of different sizes s, and $P_{p,m,s}$ for the optimal ranking method (Opt).

Size	R1	R2	R3	R4	R5	R6	R7	R8	Opt
1	0.86	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0.86	0.93	1.00	1.00	0.93	0.93	1.00	0.93	1.00
3	0.89	0.94	1.00	0.94	0.94	0.94	0.94	0.94	1.00
4	0.83	0.83	1.00	0.83	0.83	0.83	0.83	0.83	1.00
5	1.00	0.80	1.00	0.80	0.80	0.80	0.80	0.80	1.00
6	0.80	0.83	1.00	0.83	0.83	0.83	0.83	0.83	1.00
7	0.83	0.86	1.00	0.86	0.86	0.86	0.86	0.86	1.00
8	0.86	0.88	1.00	0.88	0.88	0.88	0.88	0.88	1.00
9	0.89	0.89	0.94	0.89	0.89	0.89	0.89	0.89	0.94
10	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80

Table E.5: $P_{p,m,s}$ according to resident physician 2, obtained with eight different ranking methods ($R1$ to $R8$) for information-need sets of different sizes s, and $P_{p,m,s}$ for the optimal ranking method (Opt).

Size	R1	R2	R3	R4	R5	R6	R7	R8	Opt
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	0.92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	0.90	1.00	1.00	0.90	1.00	0.90	0.90	0.90	1.00
6	0.92	1.00	1.00	0.92	1.00	0.92	0.92	0.92	1.00
7	0.93	1.00	0.93	0.93	1.00	0.93	0.93	0.93	1.00
8	0.94	1.00	0.88	0.94	1.00	0.88	0.94	0.88	1.00
9	0.94	0.94	0.89	0.94	0.94	0.89	0.94	0.89	0.94
10	0.90	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80

Table E.6: $P_{p,m,s}$ according to resident physician 3, obtained with eight different ranking methods ($R1$ to $R8$) for information-need sets of different sizes s, and $P_{p,m,s}$ for the optimal ranking method (Opt).

Size	R1	R2	R3	R4	R5	R6	R7	R8	Opt
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table E.7: $P_{p,m,s}$ according to recently graduated physician 1, obtained with eight different ranking methods ($R1$ to $R8$) for information-need sets of different sizes s, and $P_{p,m,s}$ for the optimal ranking method (Opt).

Size	R1	R2	R3	R4	R5	R6	R7	R8	Opt
1	0.43	0.43	0.43	0.14	0.43	0.43	0.14	0.43	0.57
2	0.36	0.36	0.29	0.21	0.36	0.36	0.21	0.36	0.50
3	0.28	0.28	0.22	0.28	0.28	0.28	0.28	0.22	0.39
4	0.33	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.50
5	0.40	0.50	0.40	0.30	0.50	0.30	0.30	0.30	0.60
6	0.33	0.50	0.42	0.25	0.50	0.25	0.25	0.25	0.50
7	0.36	0.43	0.36	0.29	0.43	0.29	0.29	0.29	0.43
8	0.31	0.44	0.38	0.31	0.44	0.31	0.31	0.31	0.38
9	0.42	0.39	0.33	0.33	0.39	0.28	0.33	0.28	0.33
10	0.60	0.50	0.40	0.40	0.50	0.40	0.40	0.40	0.40

Table E.8: $P_{p,m,s}$ according to recently graduated physician 2, obtained with eight different ranking methods ($R1$ to $R8$) for information-need sets of different sizes s, and $P_{p,m,s}$ for the optimal ranking method (Opt).

Size	R1	R2	R3	R4	R5	R6	R7	R8	Opt
1	0.71	0.71	0.86	0.57	0.71	0.57	0.57	0.57	1.00
2	0.79	0.79	0.71	0.57	0.79	0.79	0.57	0.79	0.86
3	0.72	0.67	0.61	0.56	0.67	0.61	0.61	0.61	0.78
4	0.58	0.58	0.58	0.42	0.58	0.45	0.50	0.50	0.75
5	0.80	0.80	0.80	0.60	0.80	0.63	0.70	0.70	1.00
6	0.83	0.75	0.75	0.58	0.75	0.67	0.58	0.67	1.00
7	0.79	0.71	0.79	0.64	0.71	0.71	0.64	0.71	1.00
8	0.75	0.75	0.81	0.69	0.75	0.75	0.69	0.75	0.94
9	0.78	0.78	0.83	0.72	0.78	0.72	0.72	0.72	0.83
10	0.80	0.80	0.80	0.70	0.80	0.70	0.70	0.70	0.80

Table E.9: $P_{p,m,s}$ according to recently graduated physician 3, obtained with eight different ranking methods ($R1$ to $R8$) for information-need sets of different sizes s, and $P_{p,m,s}$ for the optimal ranking method (Opt).

Size	R1	R2	R3	R4	R5	R6	R7	R8	Opt
1	0.43	0.86	0.71	0.57	0.86	0.71	0.57	0.71	0.86
2	0.50	0.79	0.71	0.71	0.79	0.64	0.69	0.64	0.86
3	0.44	0.61	0.61	0.61	0.61	0.50	0.61	0.61	0.72
4	0.44	0.42	0.42	0.33	0.42	0.33	0.33	0.33	0.67
5	0.50	0.60	0.60	0.50	0.60	0.60	0.60	0.60	1.00
6	0.50	0.50	0.58	0.50	0.50	0.67	0.50	0.58	0.92
7	0.54	0.50	0.64	0.57	0.57	0.64	0.57	0.64	0.86
8	0.60	0.50	0.69	0.56	0.63	0.63	0.56	0.63	0.75
9	0.65	0.56	0.61	0.61	0.61	0.61	0.61	0.61	0.67
10	0.70	0.63	0.70	0.70	0.70	0.70	0.70	0.70	0.70

Summary

Over the past years, various studies have proven that the retrieval of relevant, patient-related literature is vital to the quality of care. The number of medical errors has reached alarming heights and there is evidence that the use of information technology may reduce the number of medical errors. However, several obstacles in the medical domain obstruct successful medical information retrieval (IR).

In this thesis we investigate how to support physicians in the IR process. We believe that tailored IR support provides physicians with patient-related information and improves the quality of care. We investigate how to overcome two specific obstacles to medical IR: (a) the inadequate expression of information needs and (b) the time-consuming nature of the IR task. Consequently, we formulate a two-fold problem statement.

Problem statement (PS)

PS1: *To what extent can a physician's information needs – implicit and suppressed – be formulated automatically?*

PS2: *To what extent can the automatically formulated information needs be used as a starting point for the retrieval of relevant and patient-related literature?*

The problem statement results in five research questions. PS1 gives rise to three research questions (RQ1, RQ2, and RQ3), PS2 gives rise to two research questions (RQ4 and RQ5).

RQ1: *How can the information needs of a physician be modelled?*

RQ2: *How can patient-related information needs be formulated?*

RQ3: *How can different medical terminologies be mapped?*

RQ4: *How can a literature overload be prevented?*

RQ5: *How can the IR system be designed to be unobtrusive?*

The answers to the research questions are incorporated into a medical IR system: the Medical Information Retrieval Agent (MIRA).

In Chapter 2 we focus on RQ1: *How can the information needs of a physician be modelled?* We explain that modelling of information needs is an important step in the process of automatically formulating information needs. Without a model, it is impossible to handle all possible information needs that physicians may have. We use two distinct methods to identify the physicians' information needs, viz. a literature study and interviews with physicians. These methods result in a set of 181 identified information needs. Subsequently, the information needs are transformed into information-need templates: abstract representations of information needs. For the transformation, three methods are used: analysis, abstraction, and refinement. The application of these methods results in a set of 102 templates. These templates constitute INMOD, a model of physicians' information needs. INMOD is used to formulate a physician's information needs in a specific care situation.

Chapter 3 focuses on RQ2: *How can patient-related information needs be formulated?* and RQ3: *How can different medical terminologies be mapped?* In response to RQ2, we design an approach to bridge the gap from patient data to retrieved literature. The approach comprises five steps: (1) selecting templates, (2) extracting appropriate patient data, (3) translating patient data, (4) formulating information needs, and (5) retrieving literature. From the evaluation of MIRA we may conclude that our approach is feasible and can be generalized to other electronic medical records (EMRs), as long as they use a clear information structure that is programmatically accessible. However, the evaluation indicates that the number of formulated information needs per interaction is still high, as is the number of retrieved documents per interaction. In this way, physicians would be overloaded by information. So, we may conclude that our approach (up to this point) is not sufficiently adequate. In response to RQ3, we explain that Dutch non-standardized terms in the EMR present an obstacle to our approach. To address this obstacle, we investigate two different translation mechanisms, an automatic translation mechanism (AUTOTRANS) and a translation mechanism employing a manually constructed mapping (MANUTRANS). Experiments show that MANUTRANS provides us with the best translation results.

Chapter 4 focuses on RQ4: *How can a literature overload be prevented?* We investigate two approaches meant to prevent a literature overload. The first approach is to reduce the number of information needs. One way to achieve this is to discard information needs directly. There are four knowledge types that may be used to determine whether an information need should be discarded, viz. (a) formulation history, (b) entry order, (c) domain knowledge, and (d) number of retrieved documents. According to our experiments, the optimal formulation performance of MIRA is reached by using all four knowledge types as a discarding approach. The results of our tuning experiment indicate that this discarding approach reduces the average number of formulated information needs to a manageable number. The second way to reduce the number of information needs is to rank information needs according to their relevance and select the ten most-relevant ones (if more than ten information needs are formulated). There are three knowledge types that may be used to deter-

mine the relevance of an information need: (a) specialism, (b) entry time, and (c) number of retrieved documents. Optimal ranking performance is expected by using knowledge concerning the number of retrieved documents. The second approach to prevent a literature overload is to improve the precision of the retrieved documents. We investigate two precision enhancements: (a) patient-related document scoring and (b) patient-related document clipping. Optimal retrieval performance is expected when using patient-related document clipping as our precision-enhancement approach.

Chapter 5 provides an evaluation of MIRA and answers RQ5: *How can the* IR *system be designed to be unobtrusive?* The assessment of MIRA's performance is provided by nine physicians with three distinct levels of medical experience, viz. attending physicians, resident physicians, and recently graduated physicians. MIRA is evaluated with respect to four evaluation criteria: (a) retrieval performance, (b) formulation performance, (c) ranking performance, and (d) unobtrusiveness. With respect to the retrieval performance, we observe that the investigated precision enhancements (patient-related document scoring and patient-related document clipping) do not enhance the precision of the retrieved document set. Consequently, the final version of MIRA does not incorporate any of the precision-enhancement approaches. The evaluation of the formulation performance indicates that MIRA's performance is adequate according to the attending physicians and the resident physicians. For the recently graduated physicians, the results indicate that MIRA is capable of generating implicit information needs and retrieving documents answering the implicit information needs. With respect to the ranking performance, we observe that the ranking method to be used depends on the type of physician using MIRA. Attending physicians prefer ranking according to their specialism, resident physicians prefer ranking according to entry time of the patient data, and recently graduated physicians have no explicit preference for any of the ranking criteria. With respect to the unobtrusiveness of MIRA (in response to RQ5), we may state that MIRA can be designed to be unobtrusive by (a) preventing workflow interruptions, (b) preventing literature overload, and (c) preventing retrieval of non-relevant literature. The current version of MIRA is unobtrusive with respect to prevention of workflow interruptions. However, MIRA is obtrusive with respect to literature overload and the retrieval of non-relevant literature. Consequently, MIRA has to be improved with respect to the latter two aspects.

In Chapter 6 we use the answers to the five research questions to give an answer to our problem statement. With respect to PS1, we may conclude that it is possible to formulate a physician's information needs automatically. This statement holds for explicit, implicit, and suppressed information needs. With respect to PS2, we may conclude that the information needs formulated by MIRA can be used as a starting point for information retrieval under certain conditions. Finally, Chapter 6 discusses four topics with respect to our research and provides three directions for future research.

Samenvatting

De laatste jaren hebben verschillende studies laten zien dat de retrieval van relevante, patiënt-gerelateerde literatuur essentieel is voor de kwaliteit van de medische zorg. Het aantal medische fouten heeft een alarmerende hoogte bereikt en er zijn duidelijke aanwijzingen dat het gebruik van informatietechnologie dit aantal kan verminderen. Succesvolle medische *information retrieval* (IR) wordt echter gehinderd door verschillende obstakels in het medische domein.

In dit proefschrift beschrijven we hoe artsen ondersteund kunnen worden in het IR-proces. Wij denken dat gespecialiseerde IR-ondersteuning artsen kan voorzien van patiënt-gerelateerde informatie die kan leiden tot een verbetering van de kwaliteit van de zorg. In dit kader onderzoeken wij hoe twee specifieke obstakels van medische IR overwonnen kunnen worden: (a) de inadequate expressie van informatiebehoeften en (b) de tijdrovende aard van het IR-proces. Derhalve formuleren we een tweeledige probleemstelling.

Probleemstelling (PS)

PS1: *In welke mate kunnen de informatiebehoeften – impliciet en onderdrukt – van een arts automatisch geformuleerd worden?*

PS2: *In welke mate kunnen de automatische geformuleerde informatiebehoeften gebruikt worden als een uitgangspunt voor de* retrieval *van relevante en patiënt-gerelateerde literatuur?*

De probleemstelling leidt tot vijf onderzoeksvragen. PS1 leidt tot drie onderzoeksvragen (RQ1[1], RQ2, en RQ3), PS2 tot twee onderzoeksvragen (RQ4 en RQ5).

RQ1: *Hoe kunnen de informatiebehoeften van een arts gemodelleerd worden?*

RQ2: *Hoe kunnen patiënt-gerelateerde informatiebehoeften geformuleerd worden?*

RQ3: *Hoe kunnen verschillende medische terminologieën gematcht worden?*

---

[1]De afkorting *RQ* is afgeleid van het engelstalige *Research Question.*

RQ4: *Hoe kan een overdaad aan literatuur voorkomen worden?*

RQ5: *Hoe kan het* IR-*systeem zo ontworpen worden dat het niet hinderlijk is?*

De antwoorden op de onderzoeksvragen worden opgenomen in een medisch IR-systeem: de Medische Information Retrieval Agent (MIRA).

In Hoofdstuk 2 richten we ons op RQ1: *Hoe kunnen de informatiebehoeften van een arts gemodelleerd worden?* We leggen uit dat het modelleren van informatiebehoeften een belangrijke stap is in het proces van de automatische formulering van informatiebehoeften. Zonder een model is het onmogelijk om alle potentiële informatiebehoeften van artsen te formuleren. We gebruiken twee verschillende methoden om de informatiebehoeften van artsen te identificeren, namelijk een literatuurstudie en interviews met artsen. Deze methoden resulteren in een totaal van 181 geïdentificeerde informatiebehoeften. Vervolgens worden de informatiebehoeften getransformeerd in informatiebehoefte-*templates*: abstracte representaties van informatiebehoeften. Voor de transformatie gebruiken we drie methoden: analyse, abstractie, en verfijning. De toepassing van deze methoden resulteert in een verzameling van 102 templates. Deze templates vormen INMOD, een model voor de informatiebehoeften van artsen, dat gebruikt wordt om genoemde informatiebehoeften te formuleren in een specifieke zorgsituatie.

Hoofdstuk 3 richt zich op RQ2: *Hoe kunnen patiënt-gerelateerde informatiebehoeften geformuleerd worden?* en RQ3: *Hoe kunnen verschillende medische terminologieën gematcht worden?* Met betrekking tot RQ2 ontwerpen we een benadering die een brug slaat tussen patiëntdata en literatuur. Deze benadering omvat vijf stappen: (1) selectie van *templates*, (2) extractie van de juiste patiëntdata, (3) vertaling van patiëntdata, (4) formulering van informatiebehoeften, en (5) retrieval van literatuur. Uit de evaluatie van de benadering mogen we concluderen dat deze haalbaar is en gegeneraliseerd kan worden naar andere electronische medische dossiers (EMD, in het Engels: EMR), zolang deze een duidelijke informatiestructuur gebruiken die toegankelijk is voor MIRA. Daarnaast maakt de evaluatie duidelijk dat het aantal informatiebehoeften dat per interactie geformuleerd wordt nog erg hoog is, evenals het aantal documenten dat gepresenteerd wordt. Hierdoor zouden artsen een overdaad aan literatuur te verwerken krijgen. We mogen dus concluderen dat de benadering (tot zo ver) nog niet adequaat is. Met betrekking tot RQ3 laten we zien dat de nederlandstalige, ongestandaardiseerde termen in het EMR een obstakel vormen voor onze benadering. Om dit obstakel te elimineren onderzoeken we twee vertalingsmechanismen: een automatisch vertalingsmechanisme (AUTOTRANS) en een vertalingsmechanisme uitgaande van een handmatig geconstrueerde projectie (MANUTRANS). Experimenten wijzen uit dat MANUTRANS tot de beste resultaten leidt.

Hoofdstuk 4 richt zich op RQ4: *Hoe kan een overdaad aan literatuur voorkomen worden?* We onderzoeken twee benaderingen die moeten voorkomen dat een arts een overdaad aan literatuur te verwerken krijgt. De eerste benadering is het verminderen van het aantal geformuleerde informatiebehoeften. Eén manier om dit te bereiken is het direct verwerpen van de informatiebehoeften. Er zijn vier kennistypes die

gebruikt kunnen worden om te bepalen of een informatiebehoefte verworpen moet worden: (a) formuleringsgeschiedenis, (b) invoervolgorde, (c) domeinkennis, en (d) aantal gevonden documenten. Uit onze experimenten blijkt dat MIRA een optimale formulering bereikt wanneer alle vier de kennistypes in de verwerpingsmethode aanwezig zijn. De resultaten van deze experimenten geven aan dat het aantal geformuleerde informatiebehoeften op die manier wordt verminderd tot een aanvaardbaar aantal. De tweede manier om het aantal informatiebehoeften te verminderen is door ze te ordenen op grond van hun relevantie en de beste tien te selecteren (aangenomen dat er meer dan tien informatiebehoeften geformuleerd zijn). Drie kennistypes kunnen gebruikt worden om de relevantie van een informatiebehoefte te bepalen: (a) specialisme, (b) invoertijdstip, en (c) aantal gevonden documenten. Op grond van onze experimenten mogen we een optimale ordening verwachten bij gebruik van alleen het laatste kennistype. De tweede manier om te voorkomen dat artsen een overdaad aan literatuur te verwerken krijgen is het verbeteren van de precisie van de gevonden documenten. We onderzoeken twee precisie-verbeteringen: (a) de patiënt-gerelateerde document-score en (b) de patiënt-gerelateerde document-vermindering. We verwachten een optimale precisie bij gebruik van patiënt-gerelateerde document-vermindering.

Hoofdstuk 5 geeft een evaluatie van MIRA en beantwoordt RQ5: *Hoe kan het IR-systeem zo ontworpen worden dat het niet hinderlijk is?* De beoordeling van MIRA's prestatie wordt gegeven door negen artsen met verschillende niveaus van medische ervaring, namelijk specialisten, arts-assistenten, en basisartsen. MIRA wordt geëvalueerd op grond van vier evaluatiecriteria: (a) *retrieval*, (b) formulering, (c) ordening, en (d) hinderlijkheid. Met betrekking tot de *retrieval* merken we op dat de precisie-verbeteringen (patiënt-gerelateerde document score en patiënt-gerelateerde document vermindering) de precisie van de gevonden documenten niet daadwerkelijk verbeteren. Om die reden is geen van beide benaderingen opgenomen in de uiteindelijke versie van MIRA. De evaluatie van de formuleringsprestatie maakt duidelijk dat MIRA's prestatie adequaat is volgens de specialisten en de arts-assistenten. Met betrekking tot de basisartsen maken de evaluatieresultaten duidelijk dat MIRA in staat is om impliciete informatiebehoeften te formuleren en documenten te vinden die aan de informatiebehoeften beantwoorden. Met betrekking tot de ordening merken we op dat de te gebruiken ordeningsmethode afhangt van de arts (type arts) door wie MIRA gebruikt wordt. Specialisten geven de voorkeur aan een ordening op grond van hun specialisme, arts-assistenten geven de voorkeur aan een ordening op grond van het invoertijdstip van de patiëntdata, en basisartsen hebben geen expliciete voorkeur. Met betrekking tot de hinderlijkheid van MIRA (als antwoord op RQ5) kunnen we vermelden dat MIRA als niet-hinderlijk kan worden ontworpen door het voorkomen van (a) werkonderbrekingen, (b) een overdaad aan literatuur, en (c) het presenteren van niet-relevante literatuur. De huidige versie van MIRA is niet-hinderlijk met betrekking tot de werkonderbrekingen. Met betrekking tot een overdaad aan literatuur en de prestentatie van niet-relevante literatuur is MIRA volgens sommige artsen echter wel hinderlijk. Hieruit blijkt dat MIRA nog verbeterd moet worden op deze laatste twee gebieden.

In Hoofdstuk 6 gebruiken we de antwoorden op de vijf onderzoeksvragen om een

antwoord te geven op onze probleemstelling. Met betrekking tot PS1 mogen we concluderen dat informatiebehoeften van een arts automatisch geformuleerd kunnen worden. Dit is het geval voor expliciete, impliciete, en onderdrukte informatiebehoeften. Met betrekking tot PS2 mogen we concluderen dat de informatiebehoeften die MIRA formuleert, gebruikt kunnen worden als een uitgangspunt voor IR onder bepaalde voorwaarden. Tenslotte bespreekt Hoofdstuk 6 vier onderwerpen met betrekking tot ons onderzoek en wijst het op drie richtingen voor toekomstig onderzoek.

Curriculum Vitae

Loes Braun was born in Kerkrade, the Netherlands, on May 13, 1980. She studied Knowledge Engineering at Maastricht University, graduating in 2002 in the field of Artificial Intelligence (cum laude). During her education, she joined the Institute for Knowledge and Agent Technology (IKAT) as a research assistant in the Agent Technology Group.

After her graduation, she stayed with IKAT (later renamed into the Maastricht ICT Competence Centre) as a Ph.D. researcher under the supervision of prof. dr. H.J. van den Herik and prof. dr. ir. A. Hasman. As a result of her joined interest in agent technology and medicine, she wrote her doctoral dissertation as part of the Medical Information Agents project (funded by the Netherlands Organisation for Scientific Research (NWO), in the framework of the TOKEN programme, grant number 634.000.021). Her research was published in several international refereed journals and she presented her work at numerous conferences and workshops, resulting in various awards and tokens of recognition (see list of publications).

During her research, she also participated in other research projects. She worked on opponent modelling as part of the ROLEC project (NWO grant number 512.066.406) and on educational information retrieval in cooperation with the European Group on Tumour Markers. Furthermore, she lectured undergraduate courses on Computer Science, graduate courses on Agent Technology, and was coordinator of part of the master curriculum on Artificial Intelligence.

As of 2007 she is studying Medicine at the Erasmus MC, Erasmus University Rotterdam, the Netherlands. Her current research interests are various topics residing in the overlapping area of medicine and artificial intelligence.

Publications

The investigations performed during my Ph.D. research resulted in the following publications.

Braun, L.M.M., Wiesman, F., Herik, H.J. van den, and Hasman, A. (2003). MIRA: a Medical Information Retrieval Agent. Proceedings of the 14th Dutch-Belgian Artificial Inftelligence Conference (BNAIC2003), (eds. T. Heskes, P. Lucas, L. Vuurpijl, and W. Wiegerinck), pp.51–58. Nijmegen, The Netherlands.

Braun, L.M.M., Wiesman, F., Herik, H.J. van den, and Hasman, A. (2003). MIRA: een Medische Information Retrieval Agent. Proceedings van het 21ste Medische Informatica Congres (MIC 2003). Veldhoven, The Netherlands. (In Dutch).

Braun, L.M.M., Wiesman, F., Herik, H.J. van den, Hasman, A., and Korsten, H.H.M. (2004). From Patient Data to Information Needs. Proceedings of the 22nd Medical Informatics Congress (MIC 2004) & 5th Belgian e-Health Conference, (eds. F.H. Roger France, E. de Clercq, G. de Moor, and J. van der Lei), Studies in Health Technology and Informatics, volume 110, pp. 27–34. Brussel, Belgium.

Braun, L.M.M., Wiesman, F., and van den Herik, H.J. (2005). Towards Automatic Formulation of a Physician's Information Needs. Proceedings of the 5th Dutch-Belgian Information Retrieval Workshop (DIR 2005), (ed. R. van Zwol), pp. 25–32. Utrecht, The Netherlands.

Braun, L.M.M., Wiesman, F., van den Herik, H.J., and Hasman, A. (2005). Agent Support in Medical Information Retrieval. Working Notes of the IJCAI-05 Workshop on Agents Applied in Health Care, (ed. A. Moreno), pp. 16–25. Edinburgh, United Kingdom.

Braun, L.M.M., Wiesman, F., Herik, H.J. van den, Hasman, A., and Korsten, H.H.M. (2005). Towards Patient-Related Information Needs. Proceedings of the 19th International Congress of the European Federation for Medical Informatics (MIE 2005), (eds. R. Engelbrecht, A. Geissbuhler, C. Lovis, and G. Mihalas), Studies in Health Technology and Informatics, volume 116, pp. 911–916. Geneva, Switzerland.

Braun, L.M.M., Wiesman, F., Herik, H.J. van den, and Hasman, A. (2005). Tailoring Information Needs. Proceedings of the 23rd Medical Informatics Congress (MIC2005). Veldhoven, The Netherlands.

Braun, L.M.M., Wiesman, F., Herik, H.J. van den, Hasman, A., and Korsten, H.H.M. (2005). Towards Automatic Formulation of a Physician's Information Needs. (Extended abstract). Proceedings of the 17th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2005), (eds. K. Verbeeck, K. Tuyls, A. Nowé, B. Manderick, and B. Kuijpers), pp.323–324. Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten. Brussel, Belgium.

Braun, L.M.M., Wiesman, F., van den Herik, H.J., Hasman, A., and Korsten, H.H.M. (2005). Towards Automatic Formulation of a Physician's Information Needs. Journal of Digital Information Management, 3(1): pp. 40–47. Also published in Proceedings of the 5th Dutch-Belgian Information Retrieval Workshop (DIR 2005).

Braun, L.M.M., Wiesman, F., Herik, H.J. van den, and Hasman, A. (2006). Avoiding Literature Overload in the Medical Domain. Proceedings of the 20th International Congress of the European Federation for Medical Informatics (MIE 2006), (eds. A. Hasman, R. Haux, J. van der Lei, E. de Clercq, and F.H. Roger France), Studies in Health Technology and Informatics, volume 124, pp. 497–502. Maastricht, The Netherlands.

Braun, L.M.M., Wiesman, F., Herik, H.J. van den, and Hasman, A. (2006). Proactive Identification of a Physician's Information Needs (Best Poster Award). Proceedings of the 17th European Conference on Artificial Intelligence (ECAI 2006), (eds. G. Brewka, S. Coradeschi, A. Perini, and P. Traverso), pp. 723–724. Riva del Garda, Italy.

Wiesman, F., Braun, L.M.M., Hasman, A., and van den Herik, H.J. (2006). Information Retrieval in Medicine: The Visual and the Invisible. Information Technology, 48(1): pp. 24–32.

Braun, L.M.M., Wiesman, F., Hasman, A., van den Herik, H.J., and Korsten, H.H.M. (2007). Towards Patient-Related Information Needs. International Journal of Medical Informatics, 76(23): pp. 246–251. Also published in Proceedings of the 19th International Congress of the European Federation for Medical Informatics (MIE 2005).

SIKS Dissertation Series

1998

1. Johan van den Akker (CWI[1]) *DEGAS - An Active, Temporal Database of Autonomous Objects*
2. Floris Wiesman (UM) *Information Retrieval by Graphically Browsing Meta-Information*
3. Ans Steuten (TUD) *A Contribution to the Linguistic Analysis of Business Conversations within the Language/Action Perspective*
4. Dennis Breuker (UM) *Memory versus Search in Games*
5. Eduard W. Oskamp (RUL) *Computerondersteuning bij Straftoemeting*

1999

1. Mark Sloof (VU) *Physiology of Quality Change Modelling; Automated Modelling of Quality Change of Agricultural Products*
2. Rob Potharst (EUR) *Classification using Decision Trees and Neural Nets*
3. Don Beal (UM) *The Nature of Minimax Search*
4. Jacques Penders (UM) *The Practical Art of Moving Physical Objects*
5. Aldo de Moor (KUB) *Empowering Communities: A Method for the Legitimate User-Driven Specification of Network Information Systems*
6. Niek J.E. Wijngaards (VU) *Re-Design of Compositional Systems*
7. David Spelt (UT) *Verification Support for Object Database Design*
8. Jacques H.J. Lenting (UM) *Informed Gambling: Conception and Analysis of a Multi-Agent Mechanism for Discrete Reallocation*

2000

1. Frank Niessink (VU) *Perspectives on Improving Software Maintenance*
2. Koen Holtman (TU/e) *Prototyping of CMS Storage Management*

---

[1]Abbreviations: SIKS – Dutch Research School for Information and Knowledge Systems; CWI – Centrum voor Wiskunde en Informatica, Amsterdam; EUR – Erasmus Universiteit, Rotterdam; KUB – Katholieke Universiteit Brabant, Tilburg; KUN – Katholieke Universiteit Nijmegen; RUL – Rijksuniversiteit Leiden; RUN – Radboud Universiteit Nijmegen; TUD – Technische Universiteit Delft; TU/e – Technische Universiteit Eindhoven; UL – Universiteit Leiden; UM – Universiteit Maastricht; UT – Universiteit Twente, Enschede; UU – Universiteit Utrecht; UvA – Universiteit van Amsterdam; UvT – Universiteit van Tilburg; VU – Vrije Universiteit, Amsterdam; RUN – Radboud Universiteit Nijmegen.

3 Carolien M.T. Metselaar (UvA) *Sociaal-organisatorische Gevolgen van Kennistech-nologie; een Procesbenadering en Actorperspectief*

4 Geert de Haan (VU) *ETAG, A Formal Model of Competence Knowledge for User Interface Design*

5 Ruud van der Pol (UM) *Knowledge-based Query Formulation in Information Retrieval*

6 Rogier van Eijk (UU) *Programming Languages for Agent Communication*

7 Niels Peek (UU) *Decision-Theoretic Planning of Clinical Patient Management*

8 Veerle Coupé (EUR) *Sensitivity Analyis of Decision-Theoretic Networks*

9 Florian Waas (CWI) *Principles of Probabilistic Query Optimization*

10 Niels Nes (CWI) *Image Database Management System Design Considerations, Algo-rithms and Architecture*

11 Jonas Karlsson (CWI) *Scalable Distributed Data Structures for Database Manage-ment*

2001

1 Silja Renooij (UU) *Qualitative Approaches to Quantifying Probabilistic Networks*

2 Koen Hindriks (UU) *Agent Programming Languages: Programming with Mental Mod-els*

3 Maarten van Someren (UvA) *Learning as Problem Solving*

4 Evgueni Smirnov (UM) *Conjunctive and Disjunctive Version Spaces with Instance-based Boundary Sets*

5 Jacco van Ossenbruggen (VU) *Processing Structured Hypermedia: A Matter of Style*

6 Martijn van Welie (VU) *Task-based User Interface Design*

7 Bastiaan Schonhage (VU) *Diva: Architectural Perspectives on Information Visualiza-tion*

8 Pascal van Eck (VU) *A Compositional Semantic Structure for Multi-Agent Systems Dynamics*

9 Pieter Jan 't Hoen (RUL) *Towards Distributed Development of Large Object-Oriented Models, Views of Packages as Classes*

10 Maarten Sierhuis (UvA) *Modeling and Simulating Work Practice BRAHMS: a Mul-tiagent Modeling and Simulation Language for Work Practice Analysis and Design*

11 Tom M. van Engers (VU) *Knowledge Management: The Role of Mental Models in Business Systems Design*

2002

1 Nico Lassing (VU) *Architecture-Level Modifiability Analysis*

2 Roelof van Zwol (UT) *Modelling and Searching Web-based Document Collections*

3 Henk Ernst Blok (UT) *Database Optimization Aspects for Information Retrieval*

4 Juan Roberto Castelo Valdueza (UU) *The Discrete Acyclic Digraph Markov Model in Data Mining*

5 Radu Serban (VU) *The Private Cyberspace Modeling Electronic Environments In-habited by Privacy-Concerned Agents*

6 Laurens Mommers (UL) *Applied Legal Epistemology; Building a Knowledge-based Ontology of the Legal Domain*

7 Peter Boncz (CWI) *Monet: A Next-Generation DBMS Kernel For Query-Intensive Applications*

8 Jaap Gordijn (VU) *Value Based Requirements Engineering: Exploring Innovative E-Commerce Ideas*

9 Willem-Jan van den Heuvel (KUB) *Integrating Modern Business Applications with Objectified Legacy Systems*

10 Brian Sheppard (UM) *Towards Perfect Play of Scrabble*

11 Wouter C.A. Wijngaards (VU) *Agent Based Modelling of Dynamics: Biological and Organisational Applications*

12 Albrecht Schmidt (UvA) *Processing XML in Database Systems*

13 Hongjing Wu (TU/e) *A Reference Architecture for Adaptive Hypermedia Applications*

14 Wieke de Vries (UU) *Agent Interaction: Abstract Approaches to Modelling, Programming and Verifying Multi-Agent Systems*

15 Rik Eshuis (UT) *Semantics and Verification of UML Activity Diagrams for Workflow Modelling*

16 Pieter van Langen (VU) *The Anatomy of Design: Foundations, Models and Applications*

17 Stefan Manegold (UvA) *Understanding, Modeling, and Improving Main-Memory Database Performance*

2003

1 Heiner Stuckenschmidt (VU) *Ontology-based Information Sharing in Weakly Structured Environments*

2 Jan Broersen (VU) *Modal Action Logics for Reasoning About Reactive Systems*

3 Martijn Schuemie (TUD) *Human-Computer Interaction and Presence in Virtual Reality Exposure Therapy*

4 Milan Petković (UT) *Content-based Video Retrieval Supported by Database Technology*

5 Jos Lehmann (UvA) *Causation in Artificial Intelligence and Law – A Modelling Approach*

6 Boris van Schooten (UT) *Development and Specification of Virtual Environments*

7 Machiel Jansen (UvA) *Formal Explorations of Knowledge Intensive Tasks*

8 Yong-Ping Ran (UM) *Repair-based Scheduling*

9 Rens Kortmann (UM) *The Resolution of Visually Guided Behaviour*

10 Andreas Lincke (UT) *Electronic Business Negotiation: Some Experimental Studies on the Interaction between Medium, Innovation Context and Cult*

11 Simon Keizer (UT) *Reasoning under Uncertainty in Natural Language Dialogue using Bayesian Networks*

12 Roeland Ordelman (UT) *Dutch Speech Recognition in Multimedia Information Retrieval*

13 Jeroen Donkers (UM) *Nosce Hostem – Searching with Opponent Models*

14 Stijn Hoppenbrouwers (KUN) *Freezing Language: Conceptualisation Processes across ICT-Supported Organisations*

15 Mathijs de Weerdt (TUD) *Plan Merging in Multi-Agent Systems*

16 Menzo Windhouwer (CWI) *Feature Grammar Systems - Incremental Maintenance of Indexes to Digital Media Warehouse*

17 David Jansen (UT) *Extensions of Statecharts with Probability, Time, and Stochastic Timing*

18 Levente Kocsis (UM) *Learning Search Decisions*

2004

1 Virginia Dignum (UU) *A Model for Organizational Interaction: Based on Agents, Founded in Logic*

2 Lai Xu (UvT) *Monitoring Multi-party Contracts for E-business*

3 Perry Groot (VU) *A Theoretical and Empirical Analysis of Approximation in Symbolic Problem Solving*

4 Chris van Aart (UvA) *Organizational Principles for Multi-Agent Architectures*

5 Viara Popova (EUR) *Knowledge Discovery and Monotonicity*

6 Bart-Jan Hommes (TUD) *The Evaluation of Business Process Modeling Techniques*

7 Elise Boltjes (UM) *Voorbeeld$_{IG}$ Onderwijs; Voorbeeldgestuurd Onderwijs, een Opstap naar Abstract Denken, vooral voor Meisjes*

8 Joop Verbeek (UM) *Politie en de Nieuwe Internationale Informatiemarkt, Grensregionale Politiële Gegevensuitwisseling en Digitale Expertise*

9 Martin Caminada (VU) *For the Sake of the Argument; Explorations into Argument-based Reasoning*

10 Suzanne Kabel (UvA) *Knowledge-rich Indexing of Learning-objects*

11 Michel Klein (VU) *Change Management for Distributed Ontologies*

12 The Duy Bui (UT) *Creating Emotions and Facial Expressions for Embodied Agents*

13 Wojciech Jamroga (UT) *Using Multiple Models of Reality: On Agents who Know how to Play*

14 Paul Harrenstein (UU) *Logic in Conflict. Logical Explorations in Strategic Equilibrium*

15 Arno Knobbe (UU) *Multi-Relational Data Mining*

16 Federico Divina (VU) *Hybrid Genetic Relational Search for Inductive Learning*

17 Mark Winands (UM) *Informed Search in Complex Games*

18 Vania Bessa Machado (UvA) *Supporting the Construction of Qualitative Knowledge Models*

19 Thijs Westerveld (UT) *Using Generative Probabilistic Models for Multimedia Retrieval*

20 Madelon Evers (Nyenrode) *Learning from Design: Facilitating Multidisciplinary Design Teams*

2005

1 Floor Verdenius (UvA) *Methodological Aspects of Designing Induction-based Applications*

2 Erik van der Werf (UM) *AI Techniques for the Game of Go*

3 Franc Grootjen (RUN) *A Pragmatic Approach to the Conceptualisation of Language*

4 Nirvana Meratnia (UT) *Towards Database Support for Moving Object data*

5 Gabriel Infante-Lopez (UvA) *Two-Level Probabilistic Grammars for Natural Language Parsing*

6 Pieter Spronck (UM) *Adaptive Game AI*

7 Flavius Frasincar (TU/e) *Hypermedia Presentation Generation for Semantic Web Information Systems*

8 Richard Vdovjak (TU/e) *A Model-driven Approach for Building Distributed Ontology-based Web Applications*

9 Jeen Broekstra (VU) *Storage, Querying and Inferencing for Semantic Web Languages*

10 Anders Bouwer (UvA) *Explaining Behaviour: Using Qualitative Simulation in Interactive Learning Environments*

11 Elth Ogston (VU) *Agent Based Matchmaking and Clustering - A Decentralized Approach to Search*

12 Csaba Boer (EUR) *Distributed Simulation in Industry*

13 Fred Hamburg (UL) *Een Computermodel voor het Ondersteunen van Euthanasiebeslissingen*

14 Borys Omelayenko (VU) *Web-Service Configuration on the Semantic Web; Exploring how Semantics Meets Pragmatics*

15 Tibor Bosse (VU) *Analysis of the Dynamics of Cognitive Processes*

16 Joris Graaumans (UU) *Usability of XML Query Languages*

17 Boris Shishkov (TUD) *Software Specification Based on Re-usable Business Components*

18 Danielle Sent (UU) *Test-selection Strategies for Probabilistic Networks*

19 Michel van Dartel (UM) *Situated Representation*

20 Cristina Coteanu (UL) *Cyber Consumer Law, State of the Art and Perspectives*

21 Wijnand Derks (UT) *Improving Concurrency and Recovery in Database Systems by Exploiting Application Semantics*

2006

1 Samuil Angelov (TU/e) *Foundations of B2B Electronic Contracting*

2 Cristina Chisalita (VU) *Contextual Issues in the Design and Use of Information Technology in Organizations*

3 Noor Christoph (UvA) *The Role of Metacognitive Skills in Learning to Solve Problems*

4 Marta Sabou (VU) *Building Web Service Ontologies*

5 Cees Pierik (UU) *Validation Techniques for Object-Oriented Proof Outlines*

6 Ziv Baida (VU) *Software-aided Service Bundling - Intelligent Methods & Tools for Graphical Service Modeling*

7 Marko Smiljanic (UT) *XML Schema Matching – Balancing Efficiency and Effectiveness by means of Clustering*

8 Eelco Herder (UT) *Forward, Back and Home Again - Analyzing User Behavior on the Web*

9 Mohamed Wahdan (UM) *Automatic Formulation of the Auditor's Opinion*

10 Ronny Siebes (VU) *Semantic Routing in Peer-to-Peer Systems*

11 Joeri van Ruth (UT) *Flattening Queries over Nested Data Types*

12 Bert Bongers (VU) *Interactivation - Towards an E-cology of People, our Technological Environment, and the Arts*

13 Henk-Jan Lebbink (UU) *Dialogue and Decision Games for Information Exchanging Agents*

14 Johan Hoorn (VU) *Software Requirements: Update, Upgrade, Redesign - towards a Theory of Requirements Change*

15 Rainer Malik (UU) *CONAN: Text Mining in the Biomedical Domain*

16 Carsten Riggelsen (UU) *Approximation Methods for Efficient Learning of Bayesian Networks*

17 Stacey Nagata (UU) *User Assistance for Multitasking with Interruptions on a Mobile Device*

18 Valentin Zhizhkun (UvA) *Graph Transformation for Natural Language Processing*

19 Birna van Riemsdijk (UU) *Cognitive Agent Programming: A Semantic Approach*

20 Marina Velikova (UvT) *Monotone Models for Prediction in Data Mining*

21 Bas van Gils (RUN) *Aptness on the Web*

22 Paul de Vrieze (RUN) *Fundaments of Adaptive Personalisation*

23 Ion Juvina (UU) *Development of Cognitive Model for Navigating on the Web*

24 Laura Hollink (VU) *Semantic Annotation for Retrieval of Visual Resources*

25 Madalina Drugan (UU) *Conditional log-likelihood MDL and Evolutionary MCMC*

26 Vojkan Mihajlović (UT) *Score Region Algebra: A Flexible Framework for Structured Information Retrieval*

27 Stefano Bocconi (CWI) *Vox Populi: Generating Video Documentaries from Semantically Annotated Media Repositories*

28 Borkur Sigurbjornsson (UvA) *Focused Information Access using XML Element Retrieval*

2007

1 Kees Leune (UvT) *Access Control and Service-Oriented Architectures*

2 Wouter Teepe (RUG) *Reconciling Information Exchange and Confidentiality: A Formal Approach*

3 Peter Mika (VU) *Social Networks and the Semantic Web*

4 Jurriaan van Diggelen (UU) *Achieving Semantic Interoperability in Multi-agent Systems: a Dialogue-based Approach*

5 Bart Schermer (UL) *Software Agents, Surveillance, and the Right to Privacy: a Legislative Framework for Agent-enabled Surveillance*

6 Gilad Mishne (UvA) *Applied Text Analytics for Blogs*

7 Natasa Jovanovic' (UT) *To Whom It May Concern - Addressee Identification in Face-to-Face Meetings*

8 Mark Hoogendoorn (VU) *Modeling of Change in Multi-Agent Organizations*

9 David Mobach (VU) *Agent-based Mediated Service Negotiation*

10 Huib Aldewereld (UU) *Autonomy vs. Conformity: an Institutional Perspective on Norms and Protocols*

11 Natalia Stash (TU/e) *Incorporating Cognitive/Learning Styles in a General-Purpose Adaptive Hypermedia System*

12 Marcel van Gerven (RUN) *Bayesian Networks for Clinical Decision Support: A Rational Approach to Dynamic Decision-Making under Uncertainty*

13 Rutger Rienks (UT) *Meetings in Smart Environments; Implications of Progressing Technology*

14 Niek Bergboer (UM) *Context-based Image Analysis*

15 Joyca Lacroix (UM) *NIM: a Situated Computational Memory Model*

16 Davide Grossi (UU) *Designing Invisible Handcuffs. Formal investigations in Institutions and Organizations for Multi-agent Systems*

17 Theodore Charitos (UU) *Reasoning with Dynamic Networks in Practice*

18 Bart Orriens (UvT) *On the Development and Management of Adaptive Business Collaborations*

19 David Levy (UM) *Intimate Relationships with Artificial Partners*

20 Slinger Jansen (UU) *Customer Configuration Updating in a Software Supply Network*

21 Karianne Vermaas (UU) *Fast Diffusion and Broadening Use: a Research on Residential Adoption and Usage of Broadband Internet in the Netherlands between 2001 and 2005*

22 Zlatko Zlatev (UT) *Goal-oriented Design of Value and Process Models from Patterns*

23 Peter Barna (TU/e) *Specification of Application Logic in Web Information Systems*

24 Georgina Ramírez Camps (CWI) *Structural Features in XML Retrieval*

25 Joost Schalken (VU) *Empirical Investigations in Software Process Improvement*

2008

1 Katalin Boer-Sorbán (EUR) *Agent-based Simulation of Financial Markets: a Modular, Continuous-time Approach*

2 Alexei Sharpanskykh (VU) *On Computer-Aided Methods for Modeling and Analysis of Organizations*

3 Vera Hollink (UvA) *Optimizing Hierarchical Menus: a Usage-based Approach*

4 Ander de Keijzer (UT) *Management of Uncertain Data - towards Unattended Integration*

5 Bela Mutschler (UT) *Modeling and Simulating Causal Dependencies on Process-aware Information Systems from a Cost Perspective*

6 Arjen Hommersom (RUN) *On the Application of Formal Methods to Clinical Guidelines, an Artificial Intelligence Perspective*

7 Peter van Rosmalen (OU) *Supporting the Tutor in the Design and Support of Adaptive E-learning*

8 Janneke Bolt (UU) *Bayesian Networks: Aspects of Approximate Inference*

9 Christof van Nimwegen (UU) *The Paradox of the Guided User: Assistance can be Counter-effective*

10 Wauter Bosma (UT) *Discourse Oriented Summarization*

11 Vera Kartseva (VU) *Designing Controls for Network Organizations: a Value-Based Approach*

12 Jozsef Farkas (RUN) *A Semiotically Oriented Cognitive Model of Knowledge Representation*

13 Caterina Carraciolo (UvA) *Topic Driven Access to Scientific Handbooks*

14 Arthur van Bunningen (UT) *Context-Aware Querying; Better Answers with Less Effort*

15 Martijn van Otterlo (UT) *The Logic of Adaptive Behavior: Knowledge Representation and Algorithms for the Markov Decision Process Framework in First-Order Domains*

16 Henriette van Vugt (VU) *Embodied Agents from a User's Perspective*

17 Martin Op 't Land (TUD) *Applying Architecture and Ontology to the Splitting and Allying of Enterprises*

18 Guido de Croon (UM) *Adaptive Active Vision*

19 Henning Rode (UT) *From Document to Entity Retrieval: Improving Precision and Performance of Focused Text Search*

20 Rex Arendsen (UvA) *Geen bericht, goed bericht. Een onderzoek naar de effecten van de introductie van elektronisch berichtenverkeer met de overheid op de administratieve lasten van bedrijven*

21 Krisztian Balog (UvA) *People Search in the Enterprise*

22 Henk Koning (UU) *Communication of IT-Architecture*

23 Stefan Visscher (UU) *Bayesian network models for the management of ventilator-associated pneumonia*

24 Zharko Aleksovski (VU) *Using background knowledge in ontology matching*

25 Geert Jonker (UU) *Efficient and Equitable Exchange in Air Traffic Management Plan Repair using Spender-signed Currency*

26 Marijn Huijbregts (UT) *Segmentation, Diarization and Speech Transcription: Surprise Data Unraveled*

27 Hubert Vogten (OU) *Design and Implementation Strategies for IMS Learning Design*

28 Ildiko Flesch (RUN) *On the Use of Independence Relations in Bayesian Networks*

29 Dennis Reidsma (UT) *Annotations and Subjective Machines - Of Annotators, Embodied Agents, Users, and Other Humans*

30 Wouter van Atteveldt (VU) *Semantic Network Analysis: Techniques for Extracting, Representing and Querying Media Content*