Elisabeth Bayegan

Knowledge Representation for Relevance Ranking of Patient-Record Contents in Primary-Care Situations

Department of Computer and Information Science
Norwegian University of Science and Technology
NO-7491 Trondheim, Norway

NTNU Trondheim
Norges teknisk-naturvitenskapelige universitet
Doktor ingeniøravhandling 2002:144

Institutt for datateknikk og informasjonsvitenskap
IDI-rapport: 14/02

ISBN 82-471-5539-7
ISSN 0809-103X

Abstract

The computer-based patient-record system can play a significant role in physicians' decision-making process by, for instance, presenting them with information they need from the record, relevant to care situations. However, the patient-record system's contribution in decision making is often reduced to storing and presenting patient information as time-oriented logs of encounter events. As the record accumulates patient information over time, physicians loose overview over the contents, and the information becomes increasingly inaccessible for them.

Several realizations of problem-oriented patient records have attempted to structure the computer-based patient-record system. A persistent complaint against previous problem-oriented record systems refers to the way these systems compels the user to work. Despite less successful efforts, a problem-oriented view of data still offers a foundation for a well-structured patient record.

A knowledge-based patient-record system can support physicians during patient care in at least two ways; with a structure and a formalization that encodes information in the record system. A patient-record system with knowledge can reason about its content and use, and a process-aware system can recognize care situations and determine relevant information from the record to these situations. In this thesis, we present a way to represent knowledge for a Norwegian primary-care patient-record system. The knowledge representation enables encoding of record information—based on what primary-care physicians enter into it.

The knowledge-elicitation techniques that we have used in this thesis comprise: (1) an observation study of family physicians in clinical settings, (2) a study of patient-record contents , and (3) a qualitative and a quantitative survey among family physicians regarding the use and the usefulness of the patient-record systems in patient care. Results from these studies, in addition to their inherent empirical values, lay a foundation for further work and contribution of:

- An information model that supports a problem-oriented view of patient data
- A process model of primary-care activities , their goals and purposes, participants involved in the activities , information needs , and interactions with the patient-record system
- An ontology of clinical concepts that classify and represent the patient-record contents. The clinical concepts are associated with activities in the primary-care process

The problem-oriented structure organizes the patient-record contents into medical problems . The ontology classifies the patient-record contents according to clinical concepts. Moreover, the association between clinical concepts and activities enables a ranking of information contents to care activities.

For a practical evaluation of the knowledge model, we have developed a prototype implementation that simulates encounter-specific processes. The results demonstrate that the prototype can identify and extract information from the record, relevant to the various activities in the process model.

When compared to time-oriented patient-record systems, the problem-oriented information model, in combination with the clinical concepts and their associations to care activities, enables a method for providing a better overview over and easier access to patient-record information.

Contents

I Setting 1

List of Figures

List of Tables

Preface

This thesis is submitted to the Norwegian University of Science and Technology (NTNU) as part of the fulfillment for the degree *doktor ingeniør*. Most of the work leading to this dissertation is done while I participated in the HELPER project—Helpful ELectronic PatiEnt Records in Norwegian Primary Care—a collaboration between the Department of Computer and Information Science and the Department of Community Medicine and General Practice. Parts of the work have been done while I was on a research stay at Stanford Medical Informatics (SMI) group [177], Stanford University School of Medicine.

Acknowledgments

I thank my supervisor Associate Professor Øystein Nytrø for his collaboration during the four years of my doctorate. Øystein has given me answers and pointed out directions whenever I have asked him; it was his idea that I involve students from the course *Experts in Team* to perform parts of the surveys that are included in the fieldwork section (Section 7.2.1). I also thank Associate Professor Anders Grimsmo for advising me, as a computer scientist, on the challenges I had to face when working within the medical and primary-care field. Anders has provided me with the necessary clinical data to work with. Both Øystein and Anders have facilitated me, in the sense of giving me the freedom and responsibility of my own learning and understanding; they have let me pursue my research in the direction I myself have decided upon.

I would like to thank Professor and Head Mark A. Musen for hosting my visit to the SMI group, where I received valuable feed back on my thesis work. A special thank goes to Senior Research Scientist Samson Tu whom I collaborated and worked with during my stay. I would also like to thank the Protégé team—in particular, Ray Ferguson, Jenniffer Vendetti, Natasha Noy, and Mor Peleg—with whom I spent monthly meetings and weekly cookie times. I would also like to thank Louisa Crosby, and Carol Maxwell for her every-day smiles and *good mornings*.

I would like to thank the research group at the Department of Computer and Information Science (IDI), in particular people at the Division of Intelligent Systems (DIS), for the weekly meetings and discussions I could join in on. In particular, I thank Professor Agnar Aamodt, Professor Reidar Conradi—for the collaboration related to *Experts in Teams*—Professor Keith Downing, Professor Letizia Maria Jaccheri, and Associate Professor Pauline Haddow. I thank the many of my PhD colleagues and

friends, such as Piet van Remortel, Martin Thorsen Ranang, Zoran Constantinescu Fülöp, Amund Tveit, Xioameng Su, Terje Bratsethvik, and Inger Dybdahl Sørby. Special thanks for the great summer of 2002, which I spent with Christian Mönch, Marco Torchiano, Daniel Moody, and Diego Federici. A warm thank to my dear friend Roxana Diaconescu, whom I have shared four PhD years with—I have great memories!

From the Department of Community Medicine and General Practice, I thank Associate Professor Øystein Krüger for helping me with the data analysis in Chapter 7, Tom Christensen for letting me take part, as an observer, in his focus-group meetings. I thank Sigurd Førre for the collaboration we had between 1998 and 2000. I also thank Associate Professor Arild Faxvaag from the Department of Bone and Joint Disorders, for help and advice, and Hallvard Lærum from the Department of Neuroscience and Motion for interesting discussions on computer-based patient-record systems.

#blaa has been my *cyber-space community* for the last six years. From #blaa I give a special thank to Erik Naggum—for advise throughout both of my graduate studies—Asle Næss, Bjørn Remseth, and Jorge Barrios and Deanna Gibb-Wilkes. Great thanks to all of my friends.

Finally, with deepest gratitude to my dear family, my mother and father Astrid and Markus, my three sisters—Siv, Linn, and Charlotte—for their wonderful love and encouragement, and Christian Sasso for the great times we spend together!

Elisabeth Bayegan
December 18, 2002

Part I

Setting

Chapter 1

Toward a Helpful Patient-Record System

> I can try to provide a feel for the information processing load on a physician: a full-time general practitioner is currently expected to longitudinally follow a panel of 2000 to 2500 patients. Of course, the severity of illness varies, but it is clear that physicians need systems (computer or otherwise) to track the data pertaining to these patients, and turn it into working hypotheses for diagnosis, treatment and long term prognosis (Altman, 1998) [3].

1.1 The Computer-Based Patient-Record System

In 1991, the American Institute of Medicine (IOM) published a report on the computer-based patient-record system, for both American and European health care, and produced a revised edition in 1997 [51]. The revised edition reports on a shift from specialty care to primary care in the six years between 1991 to 1997. A corresponding shift is reported on the role of the computer-based patient-record system. In most western countries, in Europe and USA, health care increasingly move from paper-based records to computer-based patient-record systems. A computer-based record system is a constantly evolving concept, whose value and function is expected to grow with the constantly changing demands of the health-care environment and the improved technology [182]. The computer-based patient-record system has established itself as an *essential* technology for health care and especially within the area of primary care.

Primary care is a principal provider of patient care. Primary care, in countries such as Norway, Denmark, UK, and The Netherlands, functions as a gatekeeper for access to specialty care. Most family physicians run a practice by themselves, or with a small number of colleagues, where they invest in computer-based record systems that facilitate both clinical and administrative tasks.

1.1.1 Primary-Care Patient-Record Systems

In present day Norway, more than 90% of the family physicians use a computer-based patient-record system in their every-day work with patients. The record system primarily gives physicians efficiency gains on routine work such as printing of letters, referrals, documents, sick leaves, health certificates, and prescription of drugs [68].

The patient-record system serves many purposes; ranging from administrative and financial services, to legal documentation of patient care, to functioning as memory aid and decision support, to practical tasks. The record system contains many different types of information (e.g., administrative, financial, legal, government, clinical, demographics, etc.) and servers different users (e.g., laboratory assistants, administrative staff, nurses, physicians, etc.).

Family physicians have a particular interest in patient record-entries because of the impact entries have on their decision-making process. Record entries hold core information on patient conditions and events [41]. The extract in Table 1.1 shows a free-text entry taken from an existing patient-record system in Norway. Most entries are composed of descriptions of patients' symptoms, signs, complaints, reason for encounter, results from clinical examinations, drug prescriptions, etc. Patients often present more than one problem when they see a physician; so, a record entry may contain information on more than one problem. Physicians also tend to integrate family and social history in the encounter summaries, rather than sorting them under separate sections for family history and social circumstances. As a result, most information in the patient-record entries have value in different situations of patient care.

Table 1.1: Extract from a Norwegian patient-record entry.

131291JTA	Hypertension treatment for several years, tried without medication for a period this spring, but was suffering from headaches and started again with tenormin. BP 160/90 puls regular 62. Indicates a little dizziness, particularly when dark. Ophthalmoscope - no papilledema, bleeding or exudates, some minor blood vessel alterations. —HYPERTENSION ESSENTIAL NOS —HYPOTHYROIDISM RP:THYROXIN NA TBL 0.05 mg No:100 Pck.6 Reit:3 Pain left side abdominal-thorax evenings after going to bed, increasing last half year, particularly after meals. Pressure. Variable dyspepsia for years. Clin.ex.: Cor-pulm NA. Abd: BZ feces NA —STOMACH ACHE UNSPECIFIED Lab.test sent to hospital STOMACH ACHE UNSPECIFIED

1.1.2 Patient Records Accumulate over Time

Medical decision making and its impact on efficient and high-quality patient care depends on the linking of patient-specific information with medical knowledge. "…the educated physician should play the primary role in the two steps of decision making:

…retrieving relevant medical knowledge and linking it with patient data to identify information directly relevant to the patient's individual needs …" [195]. At the same time, extracting information from the record, that is relevant for decision making, is time consuming. Clinicians work under severe time constraints, which leave them little time left to search for information [193,196]. Even though family physicians benefit from using computer-based patient-record systems, efficiency gains do not make up for the perceived difficulties that disclose after 15 years of use:

- The patient record accumulates patient information (and record entries) over time. A typical feature of (Norwegian) primary-care practice is that patient care extends over many encounters. As time goes by, the patient record involves information on many medical problems that will, eventually and possibly, have some relationship to one another. Because of the many and short encounters that concern varying different problems, it becomes increasingly important for the physician to continue and follow up on previous encounters in a context of medical problems
- Family physicians lose overview over patient-record contents after a few years. Now that General-Practice List Systems[1] (GPLSs) have been introduced into Norwegian primary care [164], a long-term relationship will emphasize the need for continuing patient care and follow up from one encounter to another. A general-practice list system is a dedicated or fixed relationship between a family physician and a patient.

The fact that records accumulate patient information and the relationship between the family physician and the patient elongates over many years, is stressing physicians' needs for overview over and access to information in the patient records. Existing time-oriented patient records provide an inflexible and rigid structure, where information is, more or less, only accessible for the one who owns the information and knows where to search for it [70].

1.1.3 A Problem-Oriented View of Clinical Data

Unstructured patient records is not a novice problem, but to achieve *good computer-based solutions* of the patient record, that implements structure, still remains a challenge. A computer system can fully implement users' requirements, however, there is no guarantee that the users will accept and adopt to the new systems.

Several attempts have been made to implement structure in patient records. These attempts have had less success. A major reason for their lack of success is primarily that physicians (or clinicians) were forced to structure their data entry according to these systems. Dealing with data entry is still one of the grand challenges in health-care system development [45,88].

In 1968, Lawrence L. Weed [191] introduced the Problem-Oriented Medical Record (POMR) with aims to improve the structure and incompleteness of patient records. Weed proposed to separate a patient record into problems, to give number and name

---

[1] Or physicians' patient lists [50].

to each problem, and to record progress notes and care plans under the same problem number and name [27]. Problem orientation was introduced in an attempt to provide structure to the patient record. A problem-oriented patient record provides a context for constructive actions and for continuity of care, and it is a basis for a well-structured record [192].

The Subjective-Objective-Assessment-Plan Format

An essential characteristic of the POMR is the SOAP grid [20] or the SOAP format [190]. SOAP is an acronym for Subjective, Objective, Assessment, and Plan. Clinicians were advised to enter information according to these four categories as they were intended to reflect how physicians think and work.

PROMIS—a Gap Between Idea and Implementation

The idea of the problem-oriented patient record applied originally to the paper-based version of the patient record. The College of Medicine and the Medical Center Hospital in New England, USA, successfully implemented the *paper-based medical record* (i.e., the POMR). A majority of physicians, affiliated with the college or hospital, adopted the problem-oriented method because it proved more effective in clinical work, it became an official approach of the hospital, and it became part of the education of graduate physicians' record-keeping system. This same level of success did not apply for the computer-based version of POMR—the Problem-Oriented Medical Information System (PROMIS).

PROMIS was implemented by Weed and a team at a 750-bed ward of the Medical Center Hospital. An evaluation of the PROMIS system has been presented by Lundsgaarde et al. [99]. One of the principal reasons for why the PROMIS system did not gain widespread acceptance, suggested by Lundsgaarde et al., has to do with the implementation of the PROMIS system. The PROMIS system gave little room for maintaining the continuity of traditional practice or for clinicians to work individually with the system. To introduce PROMIS to the Medical Center Hospital required changes in all levels of their health-care delivery. PROMIS replaced consequently existing practices rather than functioning as an alternative or supplement to them. PROMIS was structured around four phases of clinical actions that were intended to make health care more complete, rigorous, and systematic in documenting clinical information. The four actions included a patient's past medical history, an enumerated problem list, diagnostic and treatment plans for each active problem, and progress notes outlining the diagnostic and therapeutic actions. Clinicians entered data into the system through video terminals (with touch-sensitive video screen) and attached typewriter keyboards. The system further guided clinicians' data entry in structuring the vocabulary, content, and organization of data. PROMIS added an explicit dimension of guidance on clinicians' clinical and therapeutic actions.

1.2 Support for a Problem-Oriented View

Despite less successful experiences with problem-oriented patient records, the idea of problem orientation still presents an intuitive and useful way to structure the patient record. A problem-oriented view of patient data serves at least two purposes:

- A problem-oriented view brings structure to the patient record by sorting and organizing information according to units of medical problems
- Physicians think and work in a problem-oriented way [6, 68, 191, 192].

1.2.1 Research Objectives

After more than a decade of using the computer-based patient record—except for the accumulation of information that contributes to loosing overview—family physicians claim that they are generally satisfied with their patient-record systems. Physicians need a patient-record structure that enables them a better overview over and access to patient-record information, than existing records do. With lack of overview in mind, our objectives are:

O1. To extend current patient-record systems with a problem-oriented view of patient data that facilitates a better structure of the patient record system, that again will give better overview over information in the record

O2. To add process knowledge to the record system as a way to overcome the challenge of forcing the user to work with and enter data into the system in a controlled way. A system that has knowledge of its content and use will be able to reason with its knowledge. If a patient-record system can recognize care situations it can present to physicians patient-specific information that is relevant to the clinical encounters. Thus, the user gets better access to the information in the record system; while at the same time, qualities, that family physicians already appreciate in their current record systems, are kept intact.

1.2.2 Research Method

This thesis will focus attention on developing parts of a knowledge model for a computer-based patient-record system. The knowledge model will contain process knowledge that enables support for a problem-oriented view of patient data. The methods used for the research work are characterized by steps in knowledge engineering:

- Fieldwork that will contribute with input to the knowledge model. We intend to use various knowledge-elicitation techniques in order to learn how family physicians think during patient care, and how they use and interact with the patient-record system. Relevant for us to use are both quantitative and qualitative methods [74, 165] in medical research, because we need to find out what family physicians themselves say about the record systems they use

- Knowledge modeling using well-known modeling methods. Besides well-established engineering methods, we will make use of tools to aid us in the knowledge-modeling process
- A prototype implementation of the knowledge model, which we will use to demonstrate process knowledge and a problem-oriented view of patient data
- An evaluation of components, demonstrated results, and knowledge engineering methods.

1.2.3 Research Hypotheses and Questions

Family physicians report that of the most prevalent and positive features with existing patient-record systems are two: (1) the record systems are easy to use and (2) the record systems do not bind them during patient care. We presume that it is possible to preserve and protect these features that already are sound and work well in existing patient-record systems. Our hypotheses and research questions are based on the following presumptions:

1. Previous experience with systems in health care has shown that clinicians are sensitive toward systems that lead and guide their actions; especially if they are forced to enter data in a strict manner into the system. We want to minimize this requirement, or potentially avoid it. The approach we have selected implies that patient-record system can itself formalize its data input. For this approach to succeed it is a prerequisite that the record system has means for recognizing its data, formalizing it, and reasoning with it. Adding knowledge to the system about primary-care activities, goals, purpose, and information needs from the patient record will enable the system to have knowledge about its information content and how this information is used in various care situations
2. The rigid data model that exists in many of the current patient-record systems, make it difficult, if not impossible, to reuse information from the patient record system, or to view patient data in relation to a specific medical problem. Today's information model typically supports only a time-oriented view of patient data; where patient data relate to a unique patient ID and date of entry. The record systems present data as a continuous log of events (i.e., a time-oriented or chronological record). We assume that we can extend and improve the systems' information model by adding support for a problem-oriented view and process knowledge. This extension of the information model will enable a model that is more flexible than already existing models, where several views of patient data are possible and information can be reused because the model adheres to a new knowledge formalism
3. Family physicians lack overview in current time-oriented patient records. A problem-oriented view of patient data will provide a well-structured patient record that further enables a better overview over and access to information in the record. We assume that structure, such as problem orientation, inherently gives a better overview than a chronological log of patient data

4. The SOAP format was a main feature in the problem-oriented patient record presented by Weed. An argument for introducing SOAP was that physicians apparently think and work in terms of subjective, objective, assessment, and plan headings. Is this assertion valid for Norwegian family physicians; do they work according to the SOAP format? Is it possible to see this SOAP structure in existing patient records? We assume that a problem-oriented model for a patient-record system is not necessarily dependent on a systematic documentation of medical information in terms of a SOAP format.

1.3 Thesis Results

The results and contribution of this thesis work is divided into three parts:

- Empirical studies involved observation of family physicians and the primary-care process, analysis of the information in patient records, and interviewing family physicians, about how they use their records system and how they see the record systems contribute in patient care
- A knowledge model for a problem-oriented patient record that structures the record into medical-problem units and associates information in the record with medical decision-making activities. In this way we can relate information in the patient record to relevant care situations, and thereby, both give an overview over and access to patient-record information
- A knowledge base that we have used to simulate specific medical decision activities and evaluate how the knowledge representation facilitates the extraction of relevant information to various clinical activities. The knowledge base can be used as a test bed for continuing clinical research on primary-care patient records.

1.4 Thesis Outline

This thesis can be divided into several parts, where one of the parts (Part III) contains four papers that present the methods, the knowledge model, the implementation, and the results of this thesis work. The four papers that are included in this thesis are: (1) *Ranking of Information in the Computerized Problem-Oriented Patient Record* [13], (2) *A Knowledge-Based, Problem-Oriented Patient Record System* [12], (3) *The Helpful Patient Record System: Problem Oriented and Knowledge Based* [15], and (4) *Ontologies for Knowledge Representation in the Computer-Based Patient Record System* [14]. We refer to these four papers when we speak about *the papers*, unless otherwise is specified. In addition to the four papers, there is a paper-introductory chapter (Chapter 8) and a closing chapter (Chapter 13) that discusses the results of the papers. The four papers are in Chapter 9, Chapter 10, Chapter 11, and Chapter 12, respectively.

Part I introduces a setting for the thesis, including a review of literature on methods and tools for design of ontologies and knowledge-based systems. The literature review is angled toward the medical domain and includes a review on clinical terminologies, vocabularies, and nomenclatures, as well

Chapter 2 introduces the impact of Artificial Intelligence (AI) theory and method applied in the medical domain. The chapter contains a presentation of important contributions from the field of AI in Medicine (AIM). The chapter introduces the CommonKADS approach and the Protégé-2000 System

Chapter 3 introduces ontologies, ontology design, typical features of ontologies, ontology-design aspects, etc. The chapter also contains a comparison of ontologies against other representations such as taxonomies, semantic nets, and object orientation (OO)

Chapter 4 introduces various clinical terminologies. Two important taxonomy development projects are presented; the UMLS and the GALEN Terminology Server

Part II presents three different empirical studies on Norwegian family physicians and the patient-record systems in primary care. The results emanating from these studies are two fold, the studies represent the selection of knowledge-elicitation techniques that we did, and they function as part of the input to the knowledge model

Chapter 5 presents results from an observation study of family physicians in clinical encounters and their use of the patient-record systems during patient care

Chapter 6 presents results from a study of the patient-record contents in primary care

Chapter 7 presents results from a qualitative and a quantitative survey on family physicians and their own opinions on the use and usefulness of patient-record systems in primary-care

Part III describes mainly the four papers. As an introduction to the papers is a chapter on the preliminaries to the knowledge model

Chapter 8 describes the approach and the method that is used to develop a knowledge representation for the patient-record system. The knowledge representation enables ranking of patient-record contents that are considered relevant to primary-care activities

Chapter 9 describes a two-layer framework for a problem-oriented, knowledge-based patient-record system. The paper introduces and discusses important notions in the framework such as *traces, decision frames, medical problems, remedies, process model* and *the encounter*

Chapter 10 describes the ideas behind a problem-oriented patient record. New ideas on the problem-oriented patient record are compared to the original ideas behind the Problem Orientated Medical Record that was introduced in the late 1960s. We argue in favor of a problem-oriented view of patient data

- **Chapter 11** describes the method that is used to create a knowledge model, which supports a problem-oriented view of patient data and ranks information in the patient records as relevant to care activities in the primary-care process
- **Chapter 12** describes in more detail the approach that is used to produce results from simulations on the knowledge base

Part IV contains a discussion, an evaluation, and concluding remarks on the contribution of the thesis. This part also contains a presentation of future work

- **Chapter 13** discusses the methods—as presented in the four papers—that are used
- **Chapter 14** evaluates the contribution in terms of the problem-oriented information model, the process model, the ontology, the simulation results, and the selection of knowledge-engineering methods that have been used. In connection with the evaluation follows a presentation of interesting work for the future
- **Chapter 15** summarizes the thesis work in concluding remarks

Part V contains appendices

- **Appendix A** contains a glossary of terms and expressions that are used in the thesis.

Chapter 2

Knowledge Engineering in Medicine

> In 1970 an article in the *Journal* [The New England Journal of Medicine] predicted that by the year 2000 computers would have an entirely new role in medicine, acting as a powerful extension of the physician's intellect. At the halfway point, how realistic does this projection seem? It is now clear that great progress has been made in understanding how physicians solve difficult clinical problems and in implementing experimental programs that capture at least a portion of human expertise. On the other hand, it has become increasingly apparent that major intellectual and technical problems must be solved before we can produce truly reliable consulting programs. Nevertheless, assuming continued research, it still seems possible that by the year 2000 a range of programs will be available that can greatly assist the physician (Swartz, Patil, and Szolovits, 1987) [169].

2.1 Introduction

The next three chapters present important areas of research within the field of Medical Informatics (MI): (1) Artificial Intelligence in Medicine (AIM) (this chapter), (2) ontology design and development (Chapter 3), and (3) representation of clinical information (Chapter 4).

This chapter is focused on applied AI theories and methods in medicine, from its early beginning in the late 1950s up to now. The AIM community has contributed significantly with tools and applications that also have benefited the AI community; for which some of them we will present in this chapter. The two communities merge in second-generation knowledge-based systems, where they share construction of ontologies and problem-solving methods as two important tasks in developing knowledge-based systems. The experiences that have been made through the multitude of systems that have been created, help explain current research status in the AIM field. We introduce some of the well-known systems that have been developed, such as MYCIN, de Dombal, INTERNIST-1, and PRODIGY. We conclude this chapter by going through some of the theory, methods, and tools that are mostly used in the medical informatics field.

Important in medical informatics are ontologies and the construction of knowledge bases through ontologies. Ontology development has become a popular approach in creation of knowledge-based systems, where they represent components that can be linked with problem-solving methods, another basic building block in knowledge-based systems creation. Chapter 3 deals with ontologies and vital properties related to ontologies (e.g., definition, typical features, existing ontologies, and tools for constructing ontologies).

Through the knowledge intensive tasks that are inherent in the medical domain, a substantial part of medical informatics research is concerned with how to formalize the domain, in terms of standardized terminologies and vocabularies. In Chapter 4 we focus on a few projects that have worked on representing and standardizing clinical information. Two of the most influential works on terminologies and vocabularies for clinical data are the GALEN Terminology Server and the UMLS.

2.2 Artificial Intelligence in Medicine

Medicine drew the attention of computer scientists in the late 1950s, early 1960s [2]. This happened principally because—as pointed out in Ledley and Lusted's *Reasoning Foundations of Medical Diagnosis* [93]—medicine involves so many stereotypical tasks (e.g., Boolean logic, symbolic inference, and Bayesian probability).

Medicine is a suitable domain for applying computer programs and the domain has functioned as a test bed for a series of application-development projects. These previous experiences have given valuable input to the design of expert and knowledge-based systems. "AI has embraced medical applications from its inception, and some of the earliest work in successful application of AI technology occurred in medical contexts " [3].

An expected result of building expert systems was for the system to appear as an expert in the (medical) domain. Even though expert systems have shown that they could make decisions that match experts' level, the expected results have not been met, where the system has taken over as an expert in the domain [115]. Research on computer-aided diagnostics had high hopes of representing clinical problems in mathematical formalisms, since medical reasoning tasks have a fair amount of structure—an attractive property for automation. Clinicians think in semi-structured ways about epidemiology, pathophysiology (the physiology of abnormal states), diagnosis, treatment, and prognosis. But, as the introductory quote says, there are still some remaining challenges—both intellectual and technical—before system developers manage to build AI applications, which helpfully assist clinicians in their clinical tasks.

2.2.1 Intellectual and Technical Challenges

Of the intellectual challenges that system developers have to overcome are for instance to find ways to represent the many, broad and complex domains in medicine (e.g., internal medicine, primary care, etc.). The medical domain is complicated and contains many interrelated parts, which present difficulties in capturing relevant knowledge in rules for a system. However, the task of representing knowledge is not a problem particular to AIM, but a general AI problem. The question of *what knowledge to represent and how to represent it*, in terms of a formalism, is a valid question and an abstruse one to answers. “The modeling and representation of human knowledge is still a key challenge in the knowledge-based systems community” [120].

Of the technical challenges that exist in the medical domain— that system developers have to take into consideration if a computer system is to operate successfully—are related to the role of the decision-support systems in clinical settings. Can clinicians fully trust a decision-support system? How should users deal with an expert systems that contains faults? We can tolerate an expert system that fails to identify a defect in a telephone network, more than we can tolerate a wrong diagnose on a seriously ill patient [2]. Many decision-support systems have been built and implemented with expert knowledge, but they have failed to incorporate into clinical practices because they were stand-alone applications, not integrated with other clinical systems [89]. Most expert systems required a strict data entry, more like a *question-and-answer* dialogue, which proved only suitable for the *super novice*. Enforcing strict data entry on clinicians have shown to be inappropriate for (experienced) clinicians.

2.2.2 From Diagnostic Systems to Advice Systems

Alan Rector suggested in [152] a paradigm shift, from making expert systems that fully automate the decision-making process to making systems that play a lesser role as an expert in a clinical setting. Alternatively the shift may be viewed as a shift from systems that advice to systems that inform [85].

Instead of making systems that play the role of an expert, system developers should focus on making systems that can participate in a decision-making process by informing clinicians with information that may be relevant for them. Now, in the medical domain, developers of second-generation knowledge-based systems have a focus toward designing active systems that can participate in a decision-making process, rather than taking over the expert role.

> Automation of expert tasks have grown a misleading term for two reasons. First, knowledge-intensive tasks are often so complex that full automation is simply an ill-directed ambition, bound to lead to wrong expectations. At the same time, knowledge systems can provide active rather than passive help, in contrast to most current automated systems, precisely because they store knowledge and are able to reason about it. On this basis, they can act and interact more actively with the user [167].

2.3 Expert and Decision-Support Systems in Medicine

Among the early AI systems in medicine, MYCIN [48] represented one of the most influential rule-based systems that served as a basis for many other experiments (e.g., EMYCIN and GUIDON [44], and PROSPECTOR [173]). Although MYCIN did not operate in routine clinical use, several formal experiments have validated its decision-making performance, which showed that MYCIN's decisions match the level of an expert (in selecting antibiotic therapy).

MYCIN also represented a famous example of a truth-functional system for uncertainty reasoning by introducing its *certainty factors* (CFs) model, widely used in expert systems in the late 1970s and early 1980s [161]. The domain-independent portions was extracted from MYCIN and was known as EMYCIN (Essential MYCIN). EMYCIN was used to develop other expert systems.

Since the introduction of MYCIN in the 1970s, rule-based systems became attractive as a mean for developing clinical decision-support systems [119].

2.3.1 MYCIN—An Expert System

MYCIN was intended to give physicians consultative advice on drug prescriptions to patients with potential serious bacterial infections [173]. Attending physicians had to make decisions in the absence of expert knowledge and certain information; from evidence of bacterial growth to a positive identification could be made was often between 12-48 hours. MYCIN, therefore, was able to assist physicians with discovering potential signs that could indicate or not an infection. MYCIN comprised five components in addition to the patient database:

1. The knowledge base of 450 rules. The knowledge base was the primary source of domain-specific knowledge. Each rule had a premise and an action. The premise contained a Boolean combination of predicates, where each clause of the premise had the following components:

   `<predicate><function><object><attribute><value>`

   MYCIN had a standard set of 24 predicate functions (e.g., SAME, KNOWN, DEFINITE), 80 attributes (e.g., IDENTIFY, SITE, SENSITIVITY), and 11 objects (e.g., ORGANISMS, CULTURE, DRUG), available for use as primitives in constructing rules. A premise could contain a conjunction of clauses, but it could not contain a disjunction. Writing separate rules for each clause solved the representation of disjunction of clauses in the premise. Rules had an inferential character because of their IF-THEN format:

   ```
   PREMISE: (AND (SAME CNTXT INFECT PRIMARY-BACTEREMIA)
                 (MEMBF CNTXT SITE STERILESITES)
                 (SAME CNTXT PORTAL GI))
   ACTION: (CONCLUDE CNTXT IDENT BACTEROIDES TALLY .7)
   ```

   IF: 1) The infection is primary-bacteremia, and
   2) The site of the culture is one of the sterile sites, and
   3) The suspected portal of entry of the organism is the gastro-intestinal tract.
   THEN: There is suggestive evidence .7 that the identity of the organism is bacteroides.

2. The consultation program gave advice on therapy
3. The question-answering program handled simple, natural language inquiries about the specific consultation, or the general knowledge base
4. The explanation program gave reasons why the performance program had requested an item of information or had drawn a particular conclusion. MYCIN could respond to inquiries of WHY and HOW; MYCIN explained reasons WHY asking a question or HOW to reach a conclusion by presenting a trace of its reasoning. Hence, MYCIN showed a recapitulation of the system's internal actions
5. The knowledge-acquisition program enabled the extension of the knowledge base through a dialogue with an expert. MYCIN handled an evolving knowledge base, where experts could add knowledge, as rules, to the knowledge base.

MYCIN incorporated a calculus of uncertainty called certainty factors. Certainty factors indicated the strength of a rule, related to probabilities [161]. The last element in a fact (i.e., represented as a quadruple) was the certainty factor:

```
(SITE CULTURE-1 BLOOD 1.0)
(IDENT ORGANISM-2 KLESBIELLA .25)
(IDENT ORGANISM-2 E.COLI .73)
(SENSITIVS ORGANISM-1 PENICILLIN -1.0)
```

2.3.2 Other Decision-Support Systems

Knowledge Couplers

Knowledge couplers are data capture and clinical guidance tools that provide decision support and management advice to clinicians [53]. Weed says about the knowledge couplers [195] that they are information tools; given a diagnostic or management problem, they will help to assure that all relevant diagnoses or management options are known, unique features are presented, and appropriate associations are made between the unique features and the diagnostic or management situation. A knowledge coupler collects related medical information in a computer program that the physician can use during patient care [27]. Knowledge couplers have a strong association with electronic clinical guidelines and protocols [46].

de Dombal

Also known as the Leed's system. de Dombal et al. showed in 1972 that it is possible to make a system that give clinically accurate diagnoses on abdominal pain and dyspepsia. The de Dombal system used probabilities to calculate the most likely diagnosis for acute abdominal pain, given a set of symptoms and signs. The system received input of 30 data items—these data were gathered and collected during physical examinations—which resulted in a diagnosis. de Dombal also showed that the use of a decision-support system helped to decrease unnecessary interventions, to reduce perforation rates among patients with appendicitis, and to decrease the use of surgical beds [190]. de Dombal's cute abdominal pain tool is one of the very few tools that has succeeded in being exported to another hospital (i.e., a Swedish hospital). The computer's accuracy then dropped from 80% to 50% [21].

ACORN

The ACORN system used initially probabilities for making decisions and required an input of 54 data items. The final version of the ACORN system builds on symbolic rules to decide whether emergency patients with chest pain need urgent admittance to the coronary care unit (CCU). The system is fed with clinical data gathered by nurses, it then makes a diagnostic decision relevant in selecting further steps of actions.

First the system performed poorly, nurses did not use the system because it was too complex, and it told nurses to do things they were not allowed to do (e.g., to admit a patient to the coronary care unit, which is formally a physician's job). After several iterations, the ACORN system ended up with less than 12 data input features and with a set of rules to complement the probabilistic inference; "a system based on probabilities alone was not sufficiently accurate and that additional symbolic rules were needed to complement the Bayesian analysis" [21].

Probabilistic reasoning was abandoned in the final version. In addition, system developers added an automated electrocadiogram (ECG) interpreting machine to get faster data input, and inference was solely based on rules. The ECG unit was added to the system after confirming that the machine would give more reliable data input to the system; "It seems, firstly, that future systems for computer-aided diagnosis should employ data from real life and not clinicians' estimate, and, secondly, that clinicians themselves cannot analyze cases in a probabilistic fashion, since often they have little idea of what the 'true' probabilities are" [52].

INTERNIST-1

INTERNIST-1 was an experimental clinical decision-support system, which was aimed specifically at medical diagnosis in internal medicine [173]. When presented with findings from a patient's initial history, physical examination, and laboratory findings, INTERNIST-1 used a rule-based reasoning method with a quasi-probabilistic scoring scheme (i.e., like MYCIN's certainty factors) to suggest likely disease candidates [101].

The group at University of Pittsburgh, PA, USA, attempted to cover the entire domain of internal medicine. INTERNIST-1 used symbolic rules to reason and had a knowledge base with 500 diagnoses, 3550 manifestations of disease, 2600 links among the 500 diagnoses, and 6500 interrelationships. For each diagnosis that was entered into the system, a disease profile was constructed, which consisted of findings (i.e., historical items, symptoms, physical signs, and laboratory abnormalities), which had been reported to occur in association with the disease [141]. INTERNIST-1 was later transferred to QMR.

QMR

Quick Medical Reference (QMR) is a diagnostic decision-support system. The program assists the user in the diagnostic process, where for each disease, the system presents a list of necessary clinical findings required for decision making [190]. QMR can suggest diagnoses, give advice on treatment strategies (i.e., an expert-consultant program), and explain relationships between symptoms and signs of a disease (i.e., an electronic textbook and diagnostic spreadsheets) [53]. The system has a knowledge base of diseases, diagnoses, findings, and laboratory information. The knowledge base contains knowledge on around 600 diseases and more than 4000 symptoms, signs, and laboratory data.

QMR has been reformulated to use probabilistic reasoning based on a decision theoretic version—the QMR-DT. With a decision theoretic system, Shwe et al. [101] mean "a system that conforms to the principles of decision theory. Decision theory uses the axioms of probability theory and utility theory to provide a framework for choosing among alternative courses of action ... Utility theory comprises a set of axioms for ascribing numerical values to outcomes of events".

PRODIGY

Prescribed RatiOnally with Decision Support in General Practice StudY (PRODIGY) is a system that provides decision support to family physicians (or general practitioners) in the UK by presenting clinical advice and therapeutic recommendations [181]. PRODIGY is embedded into existing clinical systems, but the execution model does not interact directly with the user, nor with the host patient-record system. Like other guideline systems (e.g., [121, 188, 189]), PRODIGY comprises a scenario-driven modeling of the guideline for medical decisions regarding diagnosis and current treatment.

2.4 Second-Generation Knowledge-Based Systems

The systems that were mentioned in the previous section are representatives from different eras of software production; from early mathematical and stand-alone applications to integrated knowledge systems. Early software production—the process from system specification to system maintenance—was merely focused on developing system functionality, rather than taking into consideration how these systems fit into the environment they were to operate within. The first expert systems in clinical settings required that clinicians change their habits and routines of working (e.g., de Dombal and ACORN). Later system developments in AI have been applications that integrate with existing systems, they are no longer stand-alone applications (e.g., PRODIGY and Knowledge Couplers).

Studer et al. [180] describe a paradigm shift in artificial intelligence software production from a transfer view to a modeling view. The transfer view views knowledge engineering as a process of extracting (or transferring) expert knowledge from humans to software programs. The transfer view is now substituted with the modeling view, where software and knowledge engineering is viewed as a process of modeling external, as well as internal aspects of software systems. External aspects are the organization and the environment of the software systems. Internal aspects are classical aspects of software engineering, such as requirements specification, data modeling, software design (e.g., architecture, platform, etc.), code implementation, code and system testing, etc. Knowledge and software engineering are now considered modeling activities [168].

Part of the shift in system engineering happened in response to the *software crisis* that occurred in the late 1960s. Software Engineering (SE) emerged as a discipline to assist software-system development with methods and languages for modeling. In the same way as software engineering was a response to the software crisis, Knowledge Engineering (KE) is a response to a same unsatisfactory situation of software production in the artificial intelligence field. Knowledge engineering has emerged as a discipline that contributes with methodological approaches, where the construction of knowledge-based systems has shifted from being an art into being an engineering discipline. The discipline contributes with methods, languages, and tools that are specialized for developing knowledge-based systems. This shift in methodology is also sometimes called a transfer from first-generation expert systems to second generation knowledge-based systems.

2.4.1 Epistemologic and Heuristic Adequacy

In the late 1950s and early 1960s part of the debate in artificial intelligence research communities was about the dichotomic issue of knowledge representation—what knowledge to represent and how to represent the knowledge for a machine to act intelligently. One discipline focused on how to make machines intelligent, in terms of a symbolic representation (e.g., first-order logic [124]), whereas another discipline focused on computational modeling of human intelligence in nonsymbolic representations (e.g., neural net [24]).

From a historical perspective on symbolic representation, in 1969 McCarty and Hayes [107] divided the nontrivial task of representing intelligence in computer systems into two: (1) epistemologic adequacy, and (2) heuristic adequacy. Epistemologic adequacy refers to the study of the kind of available facts in the world (i.e., how to represent the world in a computer system). The epistemologic level refers to what the systems knows [96]. Heuristic adequacy refers to the study of mechanisms that decide what to do and how to solve a problem. The heuristic level refers to how a computer system reasons efficiently with and about what it knows (ibid.).

2.4.2 The Knowledge Level

In 1982 Allen Newell proposed the idea of a *knowledge level*:

> The system at the knowledge level is the *agent*. The components at the knowledge level are *goals, actions,* and *bodies*. Thus, an agent is composed of a set of actions, a set of goals and a body. The medium at the knowledge level is *knowledge* (as might be suspected). Thus, the agent processes its knowledge to determine the actions to take. Finally, the behavior law is the *principle of rationality*: Actions are selected to attain the agent's goals [122].

Newell proposed that problem-solving agents can be characterized in terms of the knowledge and goals that can be attributed to them under the direction of the *Principle of Rationality*. An agent will always select rational actions; actions that lead to goal fulfillment. Knowledge-level modeling is concerned with what kind of knowledge is needed to accomplish a task or an activity. However, it is worth noting that knowledge-level modeling does not account for how the knowledge is represented or what kind of mechanisms are required to solve problems. Knowledge-level modeling is concerned with identifying the knowledge required to solve problems.

2.4.3 Content and Mechanism Theories

Chandrasekaran et al. [35] explain both the epistemological and heuristic adequacies and the knowledge level as attempts to characterize what it means to have a *content theory*. They further suggest that theories in AI fall into two broad categories: mechanism theories and content theories. Content theories provide potential terms for describing knowledge about a domain, whereas mechanism theories are concerned with reasoning strategies and problem-solving methods. According to Chandrasekaran et al. [34]:

> A knowledge-based system (KBS) has explicit representation of knowledge as well as an inference processes that operate on these representations to achieve a goal. An inference process consists of a number of inference steps, each step creating additional knowledge. The process of applying inference steps is repeated until the information needed to fulfill the requirements of the problem-solving goal or task is generated. Typically, both domain knowledge and possible inference steps have to be modeled and represented in some form [34].

From a pragmatic perspective—after more than 30 years of semi-successful experience in building knowledge-based systems, principles for a new discipline have emerged. The knowledge-based system communities now seek practical solutions to the challenging task of knowledge representation in computer systems. “Most workers in the artificial intelligence community now view intelligent computer systems as comprising the following four essential components: (1) a domain ontology . . . ; (2) a knowledge base . . . ,; (3) a problem-solving method . . . ; a set of mappings . . . ” [120] (the four components are presented and elaborated in more detail in Section 2.5.2).

2.5 Knowledge-Engineering Methods and Tools

From a philosophical perspective, the complexity of the world is overwhelming, and humans—as informaticians—need advice in deciding what in the world to attend to and what to ignore. Davis et al. [49] relate knowledge representation to ontologies, by stating that the selection of a representation means making a set of ontological commitments. The ontology determines what we can see of the world, bringing some parts of the world into focus while ignoring others [198].

The process of building knowledge-based systems can be seen as a modeling activity. Constructing knowledge-based systems requires a clear understanding of the relationship between tasks, knowledge, and inferences in the domain to be modeled [166, 167].

2.5.1 Ontologies and Problem-Solving Methods

Ontology development is a way to reduce the complexity of the world in models, because ontologies represent the domain structure of knowledge and enables knowledge sharing. Ontologies make up quintessentially content theories, where their main contribution is based on identifying specific classes of objects and relations that exist in some domain [35]. As Pérez and Benjamins [138] say:

> Ontologies and problem-solving methods are promising candidates for reuse in Knowledge Engineering. Ontologies define domain knowledge at a generic level, while problem-solving methods specify generic reasoning knowledge [138].

The advantage of splitting the production of knowledge-based systems into ontology development (content theories) and problem-solving methods is that it enables knowledge engineers to start out by developing ontologies and subsequently selecting a set of problem-solving methods (after getting acquainted with the concepts in the domain).

Once a computer tool supports the design and development of ontologies, the data structures created will adhere to machine interpretable formats, an attractive property that furthers the application development of knowledge-based systems [127]. Ontology design and development inhabit some attractive properties. For example, production of reusable ontologies can be combined with various problem-solving methods, and the use of ontology-development tools produces machine-interpretable data structures for use in applications [118].

2.5.2 A Stepwise Approach to Knowledge-Based Systems

Four components have been suggested to provide an explicit and flexible framework for constructing knowledge-based systems [120]:

1. A domain ontology that defines the primary concepts in the application area and relationships among those concepts
2. A set of problem-solving methods which encodes algorithms that can automate the task for which the intelligent system has been built
3. A knowledge base of detailed content knowledge about the domain of discourse in terms of a domain ontology
4. A set of mappings that define how the concepts represented in the domain ontology and corresponding knowledge base satisfy the input-output requirements of the particular problem-solving method.

The explicit framework split the task of constructing a knowledge-based system into small and doable tasks—reducing the cognitive distance between what knowledge to represent and how to represent it (or between the way in which a developer thinks about solving a problem and the language available for specifying a solution) [118].

The framework enables flexible and reusable components. "Ontologies may be mapped to multiple problem solvers ... Problem solvers may be mapped to multiple ontologies" [119]. Figure 2.1 presents the four steps.

Starting from the far left in Figure 2.1, the knowledge modeling process may start out with specifying the domain-specific ontology that the eventual application will need. The domain-specific ontology also provides the knowledge structure for the created knowledge base. From the far right in the figure, developers can select from a repository a set of generic problem-solving methods, that match the problem-solving tasks the application will require. These generic problem-solvers can be custom tailored to match specific needs of the application (i.e., the second step from the right, in Figure 2.1). The final step is to develop an *ontology mapper* between the domain ontology and the custom-tailored problem-solving methods that have been created for the specific knowledge-based application (i.e., the second step from the left, in Figure 2.1).

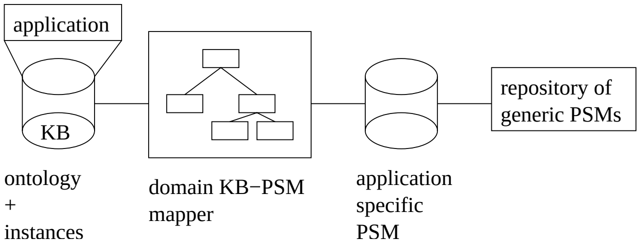


Figure 2.1: Four major steps in knowledge-base system (KBS) development: (1) the creation of ontologies and knowledge bases (KBs) for the application and domain; (2) the selection of problem-solving methods (PSMs) from a repository; (3) the adjustment of PSMs to the application and domain; and (4) the mapping among ontologies, knowledge bases, and the PSMs.

2.5.3 CommonKADS

The CommonKADS project introduces a set of methods for constructing knowledge-based systems. Dating back to 1983, these methods have evolved over the last two decades. A predecessor project to CommonKADS was KADS—a European Esprit project [168]. KADS appended their experience and knowledge from various engineering projects of building knowledge-based system. Three cornerstone principles underlie the KADS modeling approach [166]: (1) to create models is a means for coping with the complexity of the knowledge-engineering process; (2) a knowledge engineer can acquire the required modeling expertise through the KADS framework; and (3) the KADS framework supports reuse through generic model components and templates.

The CommonKADS project [167] demonstrates fully fledged theories, examples, and tables (schemes) necessary for knowledge modeling. The CommonKADS methodology help the knowledge engineer to analyze tasks, in such a way that the knowledge engineer can specify which problem solvers and methods are to be used and, moreover, to extract essential domain knowledge for knowledge modeling. The method includes six models [167, 180]:

1. An organization model that is based on the analysis of the organization structure and its features
2. A task model that describes hierarchically the tasks performed in the organization and further enables an analysis of task layout, inputs, outputs, preconditions, and performance criteria
3. An agent model that describes characteristics of agents (i.e., an agent can be a human, an information system, or any other entity that executes a task)
4. A knowledge model that makes explicit the types and structures of the knowledge used in performing tasks (in the task model)
5. A communication model that specifies the various interactions that occur among agents in the agent model
6. A design model that gives a technical system specification in terms of architecture, implementation platform, software modules, representation constructs, and computational mechanisms needed to implement features in the former models.

2.5.4 The Open Knowledge Base Connectivity Protocol

The Open Knowledge Base Connectivity (OKBC) protocol [36] specifies a knowledge model using a frame-based representation and a set of operations on these frames to access knowledge in knowledge bases (e.g., *find a frame matching a name*, *enumerate the slots of a frame*, *delete a frame*, etc.). A frame is a data structure that represents knowledge. According to Minsky [111], the inventor of frame-based systems, a frame can be described as:

> We can think of a frame as a network of nodes and relations. The top levels of a frame are fixed, and represent things that are always true about the supposed situation. The lower levels have many terminals—slots that must be filled by specific instances or data . . . A frame's terminals are normally already filled with "default" assignments [111].

Frames are often associated with an object-oriented (OO) representation because the language used to specify frames uses constructs commonly found in object-oriented languages (e.g., *classes*). The most common difference between a frame representation and an OO representation is that OO representations include definitions of methods on objects, whereas frames representations do not include definitions of methods on frames. Frame representations are merely concerned with knowledge representation of objects (or frames) in terms of properties, relations, and constraints.

The OKBC protocol specifies a set of operations for a generic interface to underlying knowledge representations, using the Knowledge Interchange Format (KIF) syntax [65]. The KIF language presents a declarative specification language, but it does not include commands for knowledge-base query or manipulation. The interface layer allows an application some independence from the idiosyncrasies of specific knowledge-representation systems, which further enables the development of generic tools. "The Open Knowledge-Base Connectivity protocol (OKBC) is a common query and construction interface for frame-based systems that facilitates . . . interoperability" [126]. Figure 2.2 illustrates how the OKBC APIs serve a common query and construction interface for frame-based systems.

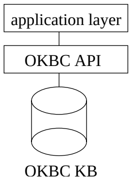


Figure 2.2: The Open Knowledge Base Connectivity (OKBC) protocol specifies a knowledge model for constructing knowledge bases (KBs) and a set of Application Programming Interfaces (APIs) for accessing objects in knowledge bases and perform a set of operations on them.

In addition to defining frames—that are primitive objects which represent entities in the domain of discourse—the OKBC protocol defines various other constructs [36]:

- A slot describes a property of a frame. Each slot has a value
- Facets describe properties of slots. Properties of slots are restrictions on slot values.
- Slots may have default values that are inherited by instances of frames with default slot values
- A class has a set of entities. Each of the entities in a class defines an instance of the class. Several classes can have the same instance. Meta classes have classes as instances
- Individuals that are not represented as classes are individuals. Thus the domain of discourse consists of individuals and classes
- A class frame is associated with a collection of *template slots* that describe slot values considered to hold for each instantiation of the class represented by the frame. Subclasses inherit the values of template slots
- Standards represent a collection of standard classes, facets, and slots with specified names and semantics. Knowledge bases do not have to include these standard classes, but if they do, they must satisfy the semantics specified by the OKBC protocol:

```
:THING //root (i.e., superclass) of class hierarchy
:CLASS //class of all classes
:INDIVIDUAL //class of non-class entity
:VALUE-TYPE //facet that specify type restriction
```

```
    :INVERSE //facet that specify inverse slot value
    :DOCUMENTATION //slot with text strings as value
    :SLOT-VALUE-TYPE //specify classes of slot values
    :SLOT-INVERSE //specify inverse relation of slot
```

The OKBC knowledge model comprises knowledge bases, classes, individuals, slots, and facets. The knowledge model serves as an implicit interlingua (i.e., a common knowledge representation for knowledge bases' constructs and operations) for communicating knowledge using OKBC and systems that use OKBC.

2.5.5 The Protégé-2000 System

The Protégé-2000 System is a general purpose tool for building knowledge-based systems [179]. The system offers a set of knowledge-acquisition tools to develop ontologies, knowledge bases, and other components in knowledge-based systems (i.e., *ontology mappers* that map domain-specific concepts to generic problem-solving methods and *schemas* for knowledge bases). Protégé-2000 complies to and implements the OKBC protocol [36] (Section 2.5.4). Furthermore, Protégé provides meta-class definitions. Meta-class instances are themselves classes. Meta-class definitions are useful for defining class templates in ontologies.

The Protégé *editor* allows the developer to easily specify concepts, attributes, constraints, and relationships in an ontology. Besides using the editor to structurally and formally define concepts and relationships, the editor allows for graphical modeling and knowledge instantiation (i.e., instances of concepts). The editor can be extended with three types of *plug-ins* [117] (2.3):

- User-interface widgets which are front-end plug-ins that extend the Protégé editor's user interface in terms of tabs and slot widgets
- Utility programs enhance knowledge-acquisition tasks (e.g., utility programs for WordNet (Section 3.3.1) and UMLS (Section 4.4))
- User-end applications which are back-end plug-ins that let users operate on Protégé's knowledge base. Protégé can store and retrieve ontologies and knowledge bases from ASCII files (with CLIPS syntax[1]), Resource Description Framework (RDF), and JDBC. JDBC provides mapping from the object-oriented frame schema to the relational database schema (RDBS).

2.6 Summary

This chapter has introduced some of the most influential expert and decision-support systems that have been made for the medical domain, such as the MYCIN system, de Dombal, INTERNIST-1, QMR, etc. Medicine has been an attractive domain for making AI applications because of its complexity, broad domain, multitude of stereotypical

---

[1] CLIPS is a production system [161].

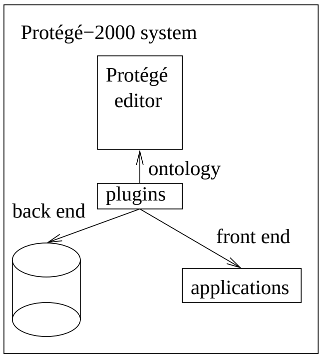


Figure 2.3: The Protégé-2000 system supports three kinds of plug-ins: (1) utility program plug-ins, (2) front-end plug-ins, and (3) back-end plug-ins.

tasks, etc. Yet, the medical domain presents some intellectual challenges with respect to capturing relevant knowledge and representing it for clinical utilization.

Knowledge-based systems differ from conventional software systems by having an explicit representation of the domain of interests and inference capabilities; most conventional software systems represent knowledge as procedural knowledge [154], where properties of behavior and inference are not explicitly defined. System acceptance has also proved to depend upon other factors such as integration with existing routines and health-information systems. For a system to gain acceptance among its users, system developers have to take into account organizational, as well as, environmental aspects, not just system specification, development, and maintenance.

System design has changed focus from first-generation expert systems to second-generation knowledge-based systems. First generation expert-system engineering focused on transferring human knowledge into computer system, while second-generation knowledge-system engineering focuses on organization modeling, in addition to data modeling. Second-generation knowledge-based system design involves knowledge engineering as a modeling activity, the separation of knowledge-system development into ontology and problem-solving methods, and a set of tools, protocols, and repositories of reusable components.

Chapter 3

Ontology Design

> What is an ontology? . . . What are the components of an ontology? What types of ontologies exist? . . . What methods should I use to build my own ontology? Which techniques are appropriate for each step? How do software tools support the process of building and using ontologies? What are the uses of ontologies? (Péreez and Bejamins, 1999) [138].

3.1 Introduction

The introductory quote presents many questions related to the concept of an *ontology*. These questions may arise because of the perceived confusion about what an ontology really is and how to distinguish an ontology from other existing notions such as taxonomies, OO-programming constructs, semantic networks, etc. Part of this confusion can perhaps be explained because the word *ontology*, in terms of its definition, originates from the field of philosophy rather than the field of computer science or medical informatics, where ontologies have become a commonly used term.

The AI community adopted the word *ontology* from philosophy in their quest to solve or try to progress in the knowledge representation problem—what knowledge to represent and how to represent the knowledge in computer systems. By adopting ontologies into the AI community followed the establishment of a purpose and an intention to the use of ontologies. Ontologies define a universe of discourse, they project a world model, and they represent a perspective or point of view. For computer scientists, the purpose of an ontology is to represent it in a formalism that can be further used in software programs. Computer scientists seek to make machine-interpretable data structures of the ontology definitions.

In this chapter we will present various definitions of the word ontology and discuss impacts of these definitions, which eventually will culminate in a definition that holds for a computer scientist, in order to create usable data structures that can be used in applications. We will present typical features and general properties of ontologies, how ontologies compare to other well-know constructs (e.g., taxonomies, oo constructs such as class and subclass concepts, inheritance, etc.). We also present some examples of existing ontologies.

3.2 Definition of Ontologies

"The word *ontology* has been taken from Philosophy, where it means a systematic explanation of Existence" [138]. Merriam-Webster's collegiate dictionary defines an ontology as a branch of metaphysics:

Definition 3.1 *by Merriam-Webster's Collegiate Dictionary*
Ontology is a particular theory about the nature of being or the kinds of existents [110].

The definition of ontology has been adopted by people working in the field of AI, in their search to understand and find means to represent intelligence, knowledge, and human reasoning in computer systems. If we understand the existence of things, we can define concepts. If we have concepts we can make classifications and classify our concepts into categories. If we have classifications we can define relationships. If we have concepts, classifications, and relationships, we can make rules and reason with our world model.

One of the most cited definitions of ontology in use among people working with AI (e.g., [23, 125, 138, 180]), comes from Gruber [71]:

Definition 3.2 *by Thomas Gruber*
An ontology is an explicit representation of a conceptualization [71].

Although the latter definition may be a *useless* definition—because a conceptualization may not obviously explain or clarify what an ontology is—it provides a starting point for a scientist with respect to understanding and using ontologies in a context of constructing computer programs and knowledge-based systems (from the Protégé mailing list, by Marc Goemans). To go from the latter definition and one step further, we need an explanation of conceptualization.

3.2.1 Conceptualization

An ontology represents a *point of view*, where certain features or attributes of objects become more relevant than others in a domain of interest. Thus the ontology determines what to perceive of the world, bringing some parts of the world into focus and removing others [197].

> The subject of *ontology* is the study of the *categories* of things that exist or may exist in some domain. The product of such a study, called *an ontology*, is a catalog of the types of things that are assumed to exist in a domain of interest D from the perspective of a person who uses a language L for the purpose of talking about D (John Sowa, pp.492, 2000).

Given a specific task, the necessary knowledge can be represented in a number of ways [95]. Designing a knowledge representation involves making a set of decisions about how and what to see of the world. A conceptualization represents a way to conceive the world and decide what to model in a knowledge representation. Studer et al. [180] defines conceptualization as: "Conceptualization refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon".

Gruber [71] defines *conceptualization* as: "an abstract, simplified view of the world that we wish to represent for some purpose". Gruber further suggests that the conceptualization should be specified at *The Knowledge Level*. In Section 2.4.2, we explained that the knowledge level deals with the knowledge that an agent holds, without taking into consideration how the knowledge is represented. Knowledge level modeling focuses on the knowledge that an agent needs to perform his or her goals, and does not deal with the particular symbol-level encoding, which is needed to make use of ontologies as data structures.

3.2.2 Ontologies as Data Structures

Ontologies are *a priori* not data structures. Once we wish to represent an ontology as a data structure, we have to take into consideration the limitations that are projected by the formalism we use to represent the ontology.

The formalism used to represent an ontology as a data structure imposes restrictions and limitations on the ontology, in terms of what is possible to express. Selecting a representation means making a set of ontological commitments [49]. By having to take computational aspects into consideration, the expressiveness of the ontology is reduced (e.g., expressiveness versus tractability) [28,49,97,111]. Because of the limitations of the formalism used to represent ontologies as data structures, it makes sense to distinguish between ontologies used for defining a domain and ontologies used as data structures in applications.

Ontologies that are not formally represented are *non-formal* ontologies (i.e., definitional). Ontologies that are formally represented as data structures are *formal* ontologies.

Table 3.1: Four different types of ontologies.

	Formal	Nonformal
Domain general	Domain ontology	Definitional
Application specific	Application ontology	Definitional

Some ontologies are intended to fully define a domain, whereas other ontologies are only intended for use in a specific application (i.e., application specific). For example, a *tractor* means different things to a farmer than to a mechanic, although the concept of tractor never actually changes. We may choose to classify *tractor* one way in an ontology geared toward farmers than in another geared toward mechanics (example taken from slides by David Koepsell). A domain ontology entails several, divergent application ontologies.

The ontology that is concerned with defining the general concept of tractor is a definitional ontology, or a *domain-general* ontology, while the ontology that uses the tractor concept for a specific purpose (e.g., in an application for farmers), or with a limited scope, is an *application-specific* ontology. An ontology can either be a general domain ontology or an application-specific ontology.

Table 3.1 presents four different types of ontologies that have different scope (e.g., domain general or application specific) and representation (e.g., formal or nonformal). In this thesis we mainly speak of ontologies that are formally represented. Hence, for our purpose, we use ontologies represented as data structures (the left-most column in Table 3.1). The types of formalisms that can be used to represent ontologies are for instance theorem proves (or logic programming languages), production systems, frame-based systems, description-logic formalisms [161], and deductive database systems [11, 55, 59, 60, 73, 163]. However, the formalism used to express an ontology often does not matter, since many knowledge-representation systems can import and export ontologies [127].

3.3 Properties of Ontologies

There exists no one correct way to model a domain—many viable alternatives often exist. The best solution almost always depends on the application that is to be developed, that is, how the anticipated application is going to extend the domain. Ontology development is an iterative approach, where concepts in the ontology refer to objects (physical or logical) and relationships in the domain of interest.

Designing good ontologies does not necessarily depend on creating a complete ontology of the domain, but an ontology should determine an appropriate conceptualization. By an appropriate conceptualization, we mean that an ontology should define general properties and relationships of concepts in the domain of interests. Ontologies should reflect *general properties* in the domain of discourse.

3.3.1 General Ontology Concepts

Most ontologies include a small number of general concepts. The generality of the concepts depend on the possible uses of the ontology, on the level of detail that the application requires, on personal preferences, and sometimes on the requirements to be compatible with other applications and ontologies:

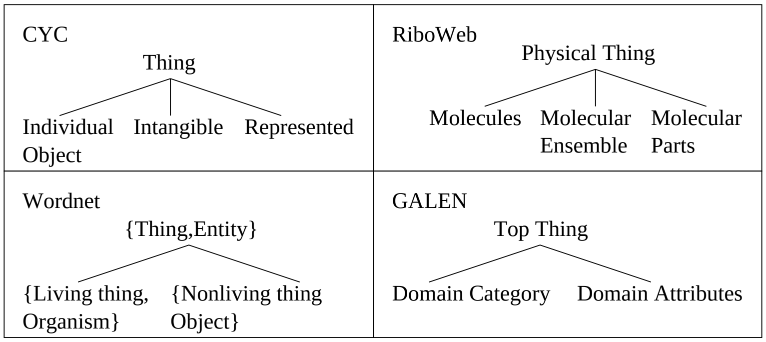
CYC: “The CYC program ... has as its task to capture the full spectrum of human knowledge and reasoning” [95]. The CYC project started in 1984 by a team led by Doug Lenat and later also Ramanathan V. Guha. CYC can be viewed as a universal schema consisting of around 1,000,000 general concepts spanning human reality [96]. This extremely large knowledge base deals with common-sense knowledge including time, space, substance, intention, causality, contradiction, uncertainty, belief, emotions, planning, etc. [23]. The knowledge base also includes complex concepts that cannot be described by a single word or *synsets* (i.e., sets of synonyms), such as *bachelor party*, JFK*’s assassination*, and *the first place one can remember calling home* [94]

WordNet: “WordNet has been defined humorously by its author, George A. Miller, as *a poor man’s* CYC” [23] . The WordNet project from Princeton University, New Jersey, USA, has produced a lexical database containing around 120,000 English word-sense pairs. As opposed to conventional dictionaries, that are built around individual words, words in WordNet—nouns, verbs, adjectives, and adverbs—are organized into sets of synonyms (synsets). WordNet deals with synonymy, hyponymy, entailment, antonymy, meronymy, and troponymy [94]. Wordnet is an online lexical reference system. Lexical objects have a semantic organization with reference to nouns, verbs, adjectives, and adverbs [125]

Riboweb: Riboweb is an example of an integrated *application* that constitutes a *knowledge base* that builds on an ontology. RiboWeb functions as an online data resource for the ribosome (a vital cellular apparatus). The concepts and the relations in the RiboWeb domain make up an ontology. The leaf instances of concepts, with values for attributes, create the knowledge in RiboWeb's knowledge base [4, 5]

Gene Ontology: In the context of genomic sequencing, the Gene Ontology Consortium has developed a Gene Ontology (GO) for biological processes, molecular functions, and cellular components. The Gene Ontology [47] functions as a general ontology for gene information (on biological process, molecular functions, and cellular components), like the UMLS (Section 4.4) is for biomedical information.

Figure 3.1: Four ontologies and their most general concepts; (1) CYC, (2) Wordnet, (3) RiboWeb, and (4) GALEN high-level ontology (the figure is partly based on [35]).



"Ontologies range in abstraction, from very general terms that form the foundation for knowledge representation in all domains, to terms that are restricted to specific knowledge domains". However, there is "no sharp division between domain-independent and domain-specific ontologies for representing knowledge" [35].

Figure 3.3.1 demonstrates four ontologies—CYC, Wordnet, RiboWeb, and GALEN—with their most general concepts. The four ontologies all have a common root concepts such as THING or ENTITY. Three of the most commonly used general concepts are THING, PROCESSES, and EVENTS [125]. However, as Figure 3.3.1 illustrates, the ontologies diverge immediately on the next level after the top-level root concept. For example, subconcepts to CYC's THING are INDIVIDUAL OBJECT, INTANGIBLE, and REPRESENTED, whereas subconcepts to Wordnet's THING are LIVING THING and NONLIVING THING.

3.3.2 Key Ontology Components

In addition to general concepts, ontologies are usually organized in taxonomies and typically contain modeling primitives such as classes, relations, functions, axioms and instances [138]. The components are described in Table 3.2.

Table 3.2: Key properties in an ontology.

Classes	A formal, explicit description of concepts, which are organized in a hierarchical classification of concepts from general to specific
Hierarchy	Represented with *is-a* relations (i.e., subsumption). The subsumption (i.e., inheritance from general to specific) property defines implicitly a taxonomy of the ontology [127]. A taxonomy is a restricted form of representation, where many properties and relationships cannot be expressed. A taxonomy of classes indicates inheritance through class and subclass memberships
Attributes	Describe properties of classes that can have a *value*. Attributes are associated with each concept
Constraints	Describe restrictions and properties on attribute and attribute values. Axioms specify additional restrictions on attributes
Relations	(and functions) Define dependencies between concepts, or classes in a taxonomy
Instances	Most specific concepts. Instances are created when they are instantiated from classes in the class hierarchy

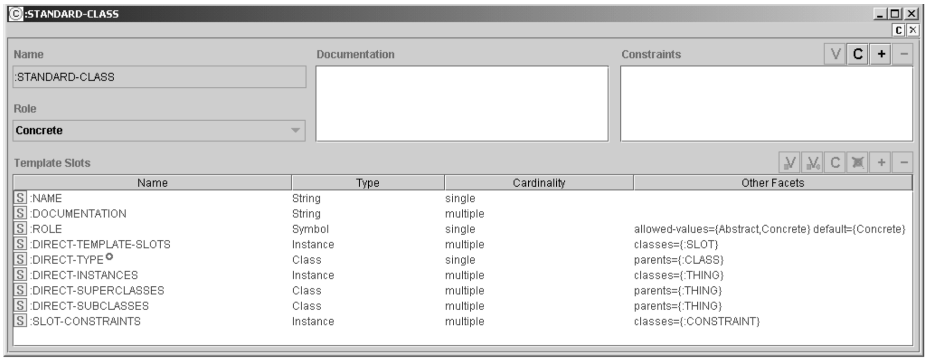
3.4 Ontologies in the Protégé-2000 System

The Protégé-2000 System contains a set of tools for the design and development of ontologies and knowledge bases. Ontologies created in Protégé are machine-readable data structures that can be used in further application development. The system complies with and implements the OKBC protocol (Section 2.5.4), which means that frames are the basic building blocks in Protégé.

Ontologies created in Protégé consists of *meta classes, classes, slots, facets,* and *axioms*. A meta class is a template for other classes, where instances of meta classes are themselves classes. Classes represent *concepts* in the world model that is being formalized. The root of a class hierarchy is the built-in class :THING. Slots correspond to *attributes* in Table 3.2 and facets correspond to *constraints*.

A class hierarchy in Protégé supports multiple inheritance [126] (i.e., a class C can have more than one superclass. Thus both class A and class B can be superclasses of class C). Classes, slots, and facets are all *first-class objects*. First-class objects are defined independently of any other object. Hence, a slot is a frame in itself and it is defined independently of any class. When a slot is attached to a frame in a user's ontology, the slot describes properties of that concrete frame and can hold a particular value.

Moreover, a slot can be attached to a frame in two ways, either as a *template slot* or as an *own slot*.



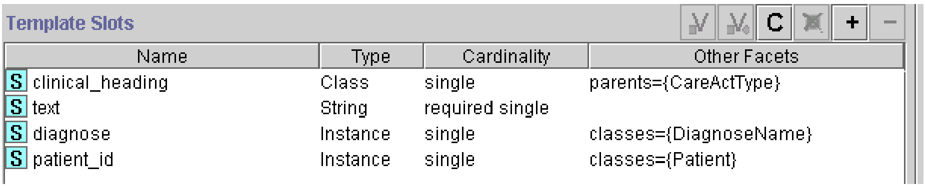
Name: :STANDARD-CLASS

Role: Concrete

Name	Type	Cardinality	Other Facets
S :NAME	String	single	
S :DOCUMENTATION	String	multiple	
S :ROLE	Symbol	single	allowed-values={Abstract,Concrete} default={Concrete}
S :DIRECT-TEMPLATE-SLOTS	Instance	multiple	classes={:SLOT}
S :DIRECT-TYPE	Class	single	parents={:CLASS}
S :DIRECT-INSTANCES	Instance	multiple	classes={:THING}
S :DIRECT-SUPERCLASSES	Class	multiple	parents={:THING}
S :DIRECT-SUBCLASSES	Class	multiple	parents={:THING}
S :SLOT-CONSTRAINTS	Instance	multiple	classes={:CONSTRAINT}

Figure 3.2: Illustration of how Protégé implements template slots for a class (:STANDARD-CLASS. Slots in the template are for instance :DOCUMENTATION and :NAME. Facets are specified on some of the slots, for instance on slots such as :DIRECT-TYPE, DIRECT-INSTANCE, and SLOT-CONSTRAINTS.

A template slot can only be associated with a class frame (as opposed to an instance frame). Template slots, therefore, describe properties of instances of a class, whereas an own slot describe properties of a class (or instance) itself (Figure 3.2). Own slots describe properties of a frame rather than properties of instances of that frame. Own slots do not propagate to subclasses or instances of the frame (class) to which they are attached. Own slots to a class are not inherited to subclasses, while template slots are.



Name	Type	Cardinality	Other Facets
S clinical_heading	Class	single	parents={CareActType}
S text	String	required single	
S diagnose	Instance	single	classes={DiagnoseName}
S patient_id	Instance	single	classes={Patient}

Figure 3.3: Four template slots for a class in a Protégé ontology. Forms are generated based on the class definitions which enable specification of individuals (i.e., class instances). The form will contain entry fields that control the input value on the slot fields, based on the restrictions given for each slot (i.e., facets).

Instances are created through *forms*. Forms are generated based on the class definition of slots and facets. Forms enable specification of class instances or individuals (i.e., the most specific instance of a class). These forms can then be custom tailored by the user. Figure 3.3 demonstrates four template slots for a class: (1) *clinical_heading* which has type class, the value of *clinical_heading* belongs to the class *CareActType*, (2) *text* which has type string (the value can be any string), (3) *diagnose* which has type instance, the value of *diagnose* is an instance of class *DiagnoseName*, and (4) *patient_id* which has type instance, the value of *patient_id* is an instance of class *Patient*. The form generated for the class in Figure 3.3 will include fields for these four slots, where the value of the slots are controlled by the form (restrictions are given by the slots' facets).

3.4.1 Ontologies and Semantic Networks

Semantic networks and frame-based systems use objects as nodes in a graph and where links between nodes represent binary relations. Nodes are organized in a taxonomic structure. More than one relationship can be represented, so, we can consider semantic networks as a generalization of taxonomies.

In frame systems, binary relations are represented by slots, whereas in semantic networks they are represented as arrows between nodes. "The choice between the frame metaphor and the semantic network metaphor determines whether you draw the resulting network as nested boxes or as graphs, but the meaning and implementation of the two types of systems can be identical" [161].

However, one problem with frame-based systems and semantic networks is that they do not have a formal inheritance semantics [23]. For instance, the transitive relationship between classes and instances are not very clear in frame-based systems and semantic networks.It is not very clear if a an `instance C` of a `subclass B` of `class A` also is an instance of `class A`. In Protégé every instance of `class B` is also an instance of `class A`.

3.4.2 Ontologies and Object-Oriented Programming

Ontology development differs from designing classes and relations in object-oriented programming. Object-oriented programming focuses primarily on methods on classes—a programmer makes design decisions based on the operational properties of a class, whereas an ontology designer makes these decisions based on the structural properties of a class. As a result, a class structure and relations among classes in an ontology are different from the structure of a similar domain in a program that has been designed based on the object-oriented paradigm.

Ontologies and Knowledge Bases

Different definitions and relationships exist between an ontology and a knowledge base. Some define a knowledge base as the inclusion of an ontology and its instantiation of individuals. Individual instances represent the most specific concepts in a knowledge base.

"An ontology together with a set of individual instances of classes constitutes a knowledge base" [127]. "We define a knowledge base as the combination of an ontology and an associated set of instances" [5]. "The concepts and the relations constitute our system's ontology. We create leaf instances of concepts, with values for attributes; the union of the ontology and the instances constitutes the knowledge base" [4].

> An ontology serves a different purpose than a knowledge base, and therefore a different notion of representation adequacy (McCarthy & Hayes, 1969) applies. A shared ontology need only describe a shared vocabulary

> for talking about a domain, whereas a knowledge base may include the knowledge needed to solve a problem or answer arbitrary queries about a domain [71].

Ontologies make up the knowledge structure in knowledge bases and we can further create advanced, structured knowledge bases. We can refer to static knowledge and dynamic knowledge, where dynamic knowledge is inferred or new knowledge. The quote by Gruber refers to an ontology that creates static knowledge, it does not include problem-solving methods or means to infer new knowledge. "…our system supports inheritance but does not use automatic deduction or constraints except for superclass and sub-class relationships" [4]. However, such a knowledge base (the latter one by [4]) can be used to solve problems and answer arbitrary questions.

3.4.3 Knowledge Bases and Relational Databases

An essential difference between a relational database (DB) and a knowledge base (KB) is the complex knowledge structure offered by an ontology in contrast to the flat relationships offered by tables in a relational database. "There used to be an easy answer to the KB versus DB question. Namely, entries in a DB were opaque tokens, whereas entries in a KB were interrelated …the DB community has adopted more and more of these AI-inspired and -developed ideas, so it is no longer easy to draw the line" [95].

Another difference between a knowledge base and a conventional database (relational) is that most knowledge bases contain explicit rules and components which can make inferences over the knowledge structure, and thereby provide some form of deductive retrieval facility. As a consequence the knowledge base contains not only the explicit facts and rules that are stored, but also what can be derived by applying the rules on the facts [23].

3.5 Summary

In this chapter we have presented the frame-based ontology, in which frames are the basic building block. Other constructs in a frame formalism are classes, slots, facets, and relations. Slots are properties of a class, facets are restrictions on slot values. Slots can specify relations between various classes in the ontology. A frame-based ontology is a comprehensive semantic network of nested relationships among classes in the domain. Classes are structured in a taxonomy, where most ontologies inherit from the root concept THING.

To construct an ontology involves representing a domain of interest that is independent of the operational behavior taken by the various applications that necessary employ the ontology. The explicit representation of a domain is independent of the various applications that operate within that domain. As ontologies define general domain knowledge, not operational knowledge, they are considered to inhabit potentially reusable components. An ontology as a data structure adheres to a well-known formalism, such as frame-based systems, description-logic systems, etc. However, ontologies that are machine readable do not tell how to do things, they merely represent data structures that exist to facilitate the creation of effective algorithms, which moreover enable the building of applications. So, the only thing an ontology can do is being *used*.

Chapter 4

Efforts on Representing Clinical Information

> No major terminology source can lay claim to being ideal for a computer-based patient record. However, based upon this analysis of releases for April 1995, SNOMED International is considerably more complete, has a compositional nature and a richer taxonomy. It suffers from less clarity, resulting from a lack of syntax and evolutionary changes in its coding schemes (Campbell et al., 1997) [33].

> The main problem with using SNOMED for coding patient information is that it is *too* expressive. Because there are few rules about how the post-coordination coding should be done, the same expression might end up being represented differently by different coders (Cimino, 1996) [42].

4.1 Introduction

In this chapter we introduce some existing vocabularies, terminologies, and various coding schemes (e.g., International Classification of Diseases and International Classification of Primary Care) that are used in health care to classify health conditions. We also introduce two major projects—UMLS and GALEN—that have worked on standardizing and unifying clinical vocabularies and terminologies.

The former project has focused on integration and standardization of clinical vocabularies and terminologies through the creation of a Metathesaurus and a Semantic Network. The Metathesaurus defines the concepts and the Semantic Network defines the relations. The UMLS functions as a reference for concepts and translations of concepts, in addition to the multitude of applications that can be implemented based on the Metathesaurus and Semantic Network.

The latter project, GALEN, has focused on developing a robust representation formalism to standardize representation of clinical information. GALEN developed their own formalism—GRAIL—to define semantic models of medical concepts that can be put together by specified rules to make composite expressions. GRAIL contains both the rules for composition and the formalism for representation. GALEN also constructed an engine—a Terminology Server—that delivers terminology services. A terminology server is a language independent and reusable shared system of concepts that supports the development and integration of clinical systems through a range of key terminological services [150]. Terminological services can for example be concept or reference models, code conversion between thesauri, nomenclatures, classifications, local controlled vocabularies, formal models , and multi-lingual support [149].

The strength in a terminology server lies in the uniform representation of medical concepts and the multitude of tasks that can be done based on the uniform representation [39]. The purpose of a terminology server is to be able to relate clinical information from various sources, combine them, and perform tasks on them. Terminology servers and metathesauri have been developed to mediate between different terminologies and vocabularies [40].

4.2 Standardizing Clinical Information

Standards for clinical information are important in medical informatics because they allow two or more people to cooperate in their work [72]. Computer-based patients records will have to represent patient data using standard vocabularies to facilitate clinical decision support, research, integration with other clinical systems and users, and patient care.

Data collected in health care for a primary purpose (i.e., primary care versus secondary care, tertiary care, etc.), most likely will need to be used in a multitude of ways, in for instance secondary and tertiary care (or for statistics or for governmental issues). However, once we seek to communicate between systems that operate in different fields with different purposes and needs, the problem arises, to deal with multiple and heterogeneous coding systems. A standard for coding patient data, for instance, is nontrivial when we consider the need for agreed-upon definitions, while at the same we need to support multiple users and multiple purposes. It is a challenge to agree on standards for medical terminologies [43,152]. We face at least three important challenges with regard to standardization of terminologies and coded vocabularies:

- Most terminologies and vocabularies are incompatible

- A standard terminology will have to serve conflicting purposes due to the multitude of tasks and users

- Standardizing terminologies will lead to the dichotomy dilemma of *expressiveness versus tractability* (Section 3.2.2) [151]. A language representation can not be too expressive if it is to be computational tractable [10]. However, a language that is not expressive enough risk the chance of not fulfilling its intentional purpose. The trade off between expressiveness and tractability is one of the major challenges to deal with when working on representations for clinical information.

4.3 Some Terminologies, Vocabularies and Nomenclatures

In the following we present some widely used classifications, vocabularies and nomenclatures. International Classification of Diseases (ICD) and International Classification of Primary Care (ICPC 1-2) [18, 91, 186] are one-axis coding schemes that are mostly used in primary and secondary care, respectively. ICPC is more used in Europe, wheres as ICD is widely used all over the world. The Systematized Nomenclature Of MEDicine (SNOMED) is a multi-axis coding scheme that is much more expressive than ICPC and ICD.

ICPC and ICD are enumerative classifications schemes, whereas SNOMED is a coded vocabulary that allows for richer coding of information in medical records. However, the main problem with SNOMED for coding patient information is that it is too expressive. Because there are few rules about how to use multiple codes, the same expression might be represented differently by different coders [42].

> No classification capture all concepts, although SNOMED did notably the most complete job. The system in major use in the United States, ICD-9-CM and CPT, fail to capture substantial clinical concepts. ICD-10 does not perform better than ICD-9-CM. The major clinical classification in use today incompletely cover the clinical content of patient records, thus analytic conclusion that depend on these systems may be suspect (Chute et al. [41]).

In SNOMED Reference Terminology (SNOMED-RT) and SNOMED Clinical Terms (SNOMED-CT), the problems in SNOMED of inconsistent use and multiple hierarchies has been tried fixed.

SNOMED-RT incorporates description logic to limit the expressiveness and to ensure computational tractability. In SNOMED-CT several coding schemes are tried merged, while keeping the semantic definitions in tact. For each concept in the terminologies, the original meaning of the concept is tried kept as sound as possible [176].

ICD is a classification, with up to four-digit codes, to encode purposes of morbidity and mortality. The World Health Organization (WHO) keeps mortality statistics based on ICD code reports

ICPC is a classification that encodes the *reason for encounter* as well as the diagnostic processes, interventions, preventions, administrative procedures, and the diagnosis. In ICPC the ability to represent concepts that are found in patient records is much greater through the multiple coding of terms. The classification has been extensively tested and found to be very practicable and reliable for use in primary care, with less than 3% recording error [20]. ICPC maps partially to ICD-9

SNOMED is a nomenclature that can be used to index, store, and retrieve information about a patient in a computer-based medical record [87]. The SNOMED nomenclature allows clinicians to record entities and observations related to a particular disease [20]. SNOMED is not a classification, it is a coded vocabulary of names and descriptions in health care that allows for multiple coding of terms along several axes such as topography (anatomic), morphology, etiology, function, procedures, and occupation. A primary strength of SNOMED is the ability to code complex clinical concepts [100]

SNOMED-RT is the international reference terminology that is designed to encourage consistent use of terms by explicitly defining relationships among terms

SNOMED-CT is a merge of SNOMED and Read Codes, to produce a single, joint clinical terminology. Read Codes were designed to support primary-care practitioners in the UK. Read codes are specifically designed for use in coding computer-based medical records.

Appendix A contains a description of terms, nomenclature, terminology, and vocabulary used in this chapter.

4.4 The Unified Medical Language System

The United States' National Library of Medicine (NLM) began a research and development project in 1986 to build a Unified Medical Language System (UMLS). The UMLS was designed primarily for use by system developers with the intentions to improve access to computer-based biomedical information [134].

Application programs may interpret and refine user queries, map user's terms to appropriate controlled vocabularies and classification schemas, interpret natural language, and assist in structured data creation [108]. The UMLS contains three *knowledge sources* [185]: (1) the Metathesaurus, (2) the SPECIALIST, and (3) the Semantic Network.

4.4.1 The UMLS Metathesaurus

The Metathesaurus contains information about biomedical concepts and terms from many controlled vocabularies and classifications used in, for instance, patient-record systems, health-care systems, bibliographic and full-text databases, and expert and knowledge systems. The Metathesaurus comprises biomedical terminology from over 60 biomedical vocabularies and classifications, and which links many different names for the same concepts (i.e., synonyms). At present the thesaurus includes more than 800,000 concepts and more than 2 million concept names (some in multiple languages) [131].

Computer programs can use the Metathesaurus to interpret and interact with users and other biomedical applications by converting biomedical terminology from one system to another system with reference to the thesauri. The Metathesaurus is organized by concept or meaning.

4.4.2 The UMLS SPECIALIST Lexicon

The SPECIALIST lexicon is a general English lexicon that includes many biomedical terms that provides lexical information needed for the SPECIALIST Natural Language Processing (NLM) System. The lexicon entry for each word or term records syntactic, morphological (i.e., word formation), and orthographic (i.e., proper writing) information.

Eleven syntactic categories, or parts of speech, are recognized by the lexicon recognizes: verbs, nouns, adjectives, adverbs, auxiliaries, modal, prepositions, conjunctions, complementizers (i.e., a function word that combines with a clause to form a subordinate clause), and determiners [133].

4.4.3 The UMLS Semantic Network

The Semantic Network represents more than 130 basic semantic concept found in the Metathesaurus and may be considered as a basic ontology for that domain [108]. It also includes more than 50 relationships that may hold between the semantic types.

Semantic types represent nodes in the Network, relationships represent links between the nodes. Seven major groupings of semantic types exist: (1) organisms, (2) anatomical structures, (3) biological functions, (4) chemical properties, (5) events, (6) physical objects, and (7) concepts or ideas. The *is-a* relationship is the primary relationship. The *is-a* link provides an inheritance property and allows nodes in a hierarchy to inherit properties from higher-level nodes.

In addition to the *is-a* relationship, is a set of non-hierarchical relations between the types that has been identified and grouped into five relational categories: (1) physically related to (e.g., *part-of* and *contains*), (2) especially related to (e.g., *location-of* and *surrounds*), (3) temporally related to (e.g., *precedes* and *co-occurs with*), (4) functionally related to (e.g., *performs* and *indicates*, and (5) conceptually related to (e.g., *measures*, *property-of* and *diagnoses*) [132].

4.4.4 Compatibility with UMLS

Because UMLS integrates so many biomedical terminologies, consecutive application development projects can use the UMLS Metathesaurus and Semantic Network to be in compliance with both the UMLS and other terminologies. For example, the Digital Anatomist Symbolic Knowledge Base [159] is an example of how an anatomical ontology can be in compliance with UMLS Semantic Network and Metathesaurus. The UMLS Metathesaurus and Semantic Network has been extended (Figure 4.1) by the anatomical ontology by assigning new concepts to one of the top-level classes in the UMLS. Classes and subclasses in the anatomical ontology correspond to the high-level semantic types in the UMLS Semantic Network.

4.5 GALEN

The General Architecture for Languages, Encyclopædias, and Nomenclatures in Medicine (GALEN) project was a consortium of universities and vendors. The project was a three-phase European Union (EU) funded project, where GALEN was committed to making reusable and application-independent representations for medical concepts. They developed standards for representing coded patient information to be used in applications for medical records, clinical user interfaces and clinical information systems [72]. But the work in GALEN also included systems for natural-language un-

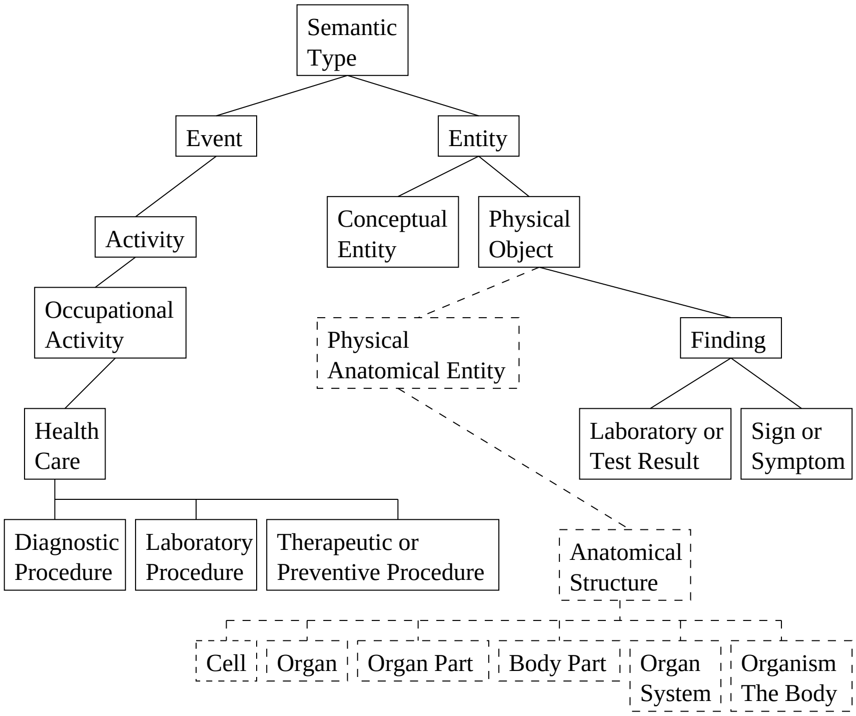


Figure 4.1: Part of the UMLS Semantic Network. The Anatomical Ontology (with dashed lines) complies with a UMLS higher-level concept.

derstanding, clinical decision support, and management of coding and classification schemes.

The second phase of GALEN—GALEN-IN-USE—focused on the practical use of a standardized terminology system, thus demonstrating GALEN technology in use in medical procedures (i.e., surgical procedures). GALEN-IN-USE developed tools and methods to assist in the collaborative construction and maintenance of surgical-procedure classifications [158]. Open GALEN followed after GALEN and GALEN-IN-USE, a not-for-profit organization, with a dedication to bring GALEN to the public as *open source* [1].

4.5.1 GALEN Terminology Server

To meet the needs for a unified controlled medical vocabulary, GALEN separated concepts into their component parts, organized parts into nearly pure hierarchies, and subsequently recombined them into composite representations. This approach was demonstrated in the GALEN *Terminology Server*. The terminology-server architecture discriminated strictly among *concepts*, *languages*, *coding systems*, and *general inference modules* [142]. The terminology server produced a unified controlled medical vocabulary through four components: (1) a CORE model, (2) a GRAIL formalism, (3) a code-conversion module, and (4) a multi-lingual module. The four components are illustrated in Figure 4.2.

The two latter modules, *the code-conversion module* and *the multi-lingual module* will not be discussed in this text. These two modules have received little attention in publications. Most of the work that we have read (and PEN&PAD) [66, 76, 142–150, 156–158] concerns the two former modules, *the* CORE *model* and *the* GRAIL *language*.

The interaction between the terminology server and the back-end databases have been modeled in Figure 4.2 in two ways, to illustrate two different and possible architectures for using a terminology server; these will be discussed in Section 4.5.4.

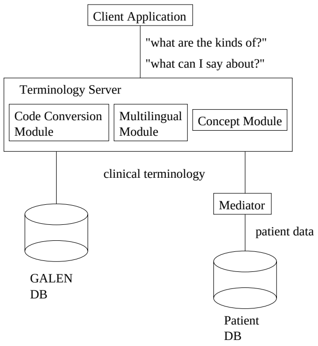


Figure 4.2: Components in a GALEN framework using the GALEN Terminology Server and its three modules: (1) the code-conversion module that converts different coding schemas, (2) the multi-lingual module that represents language constructs from various languages, and (3) the concept module. The concept module integrates the CORE model and the GRAIL representation language.

4.5.2 The Common Reference Model

The COmmon REference model (CORE model) represents a set of medical concepts that are used by various applications in the medical domain. CORE is based on a theory about the generation and subsumption of composite concepts. Composite concepts are created by *all and only* sensible medical concepts (e.g., men can not be pregnant) through four design principles [146]: (1) to separate the concepts into their apparently elementary parts, (2) to form clean taxonomies of the parts, (3) to recombine the parts by defining the original concepts in terms of the elementary parts, and (4) to classify the composite concepts according to formal rules.

The medical concept model is a general taxonomy of generic concepts:

- The top-level schema of the CORE model is GALEN's high-level ontology [149] (see Figure 3.3.1 in previous chapter for GALEN's high-level ontology), that involves a semantic network of nodes and links. The ontology represents language-independent concepts with a basic structure of substances, processes, and structures, coupled with a system of modifiers that can be used to intuitively model most medical concepts. Two important types of relationships is used to organize the taxonomy: (1) *kind-of* (i.e., *is-a* or subsumption), and (2) *part-of*—sometimes called *generic* or *partitive* because the subsumption relationship is also a partial-order relation [176]. Table 4.1 presents some constructs from the GRAIL language

- CORE serves as an *interlingua* [146]. An interlingua is a meaning representation in a language-independent canonical form. The idea is for the interlingua to represent all sentences that mean the *same* thing in the same way [83] (i.e., a common knowledge representation of medical concepts).

4.5.3 GALEN Representation and Interpretation Language

In response to standard description logics' (e.g., KL-ONE [155], CLASSIC, LOOM, and BACK) poor ability to cope with medical terminologies, the GALEN Representation And Interpretation Language (GRAIL) was developed specifically for this task [76]. GRAIL is GALEN's knowledge representation language, which makes up the formal language of medical concept representation [144].

Description logics have special features to support the requirements of coordinating taxonomies, particularly for coordinating the *kind-of* relation (in addition to being *reflexive*, *transitive*, and *antisymmetric*. GRAIL adds novel features to support the part-whole (concept inclusion) and other transitive relations, but it lacks a number of other constructs, like disjunction [142].

4.5.4 Terminology Servers and Mediators

The GRAIL kernel provides the rules for combining existing entities into new composite entities (Table 4.1). The GRAIL language contains a set of keywords and tokens (e.g., *which*, *name*, -) and a set of entities and attributes from the CORE model.

GALEN's approach of separating semantic concepts in the medical domain from the rules that form composite expressions is based on the assumption that there exists

Table 4.1: Constructs from the GRAIL language (based on [146]).

GRAIL language constructs	Example
Entity	Lung, Tumor, JohnSmith, IntrinsicallyPathologicalStructure, Opacity, Effusion, Excising, RadiographicObservation22, Lobe
'ACompositeEntity'	RightLung, LeftPleuralSpace, LeftPleuralEffusion, LeftLowerLobeAtelectasis, PostOperativeChange, SurgicalRole
attribute	hasSeverity, hasLocation, hasForm, hasProgress, hasDiagnose, shows, isStateOf, presence, performance, isMainlyCharacterizedBy, isEnactmentOf, playClinicalRole, actsSpecificallyOn, isSpecific
value	severe, plateLike, new
attribute-Value	hasSeverity-severe
Entity *which* < attribute-Value1 ... attribute-ValueN>	Opacity *which* <hasForm-plateLike, hasProgress-new>
Expression *name* 'Name'	(Effusion *which* hasLocation PleuralSpace) *name* 'PleuralEffusion'
Entity-Attribute-Entity	RadiographicObservation22- shows- (presence *which* isStateOf-'SurgicalClip')

a distinctive *conceptual level* or *terminology level* and an *assertion level*. GALEN separated the *linguistic level* from the *general inferences* that can be done based on the linguistic level [143]. The same inferences can then be used for multiple languages, where the code conversion module and the multi-lingual module (Figure 4.2) would translate expressions with different codes and languages (on the terminology and linguistic level).

Figure 4.2 demonstrates two different architectures using the GALEN Terminology Server. One architecture solution (the left side of the figure) is based on a database using concepts recognized in the GALEN CORE model. This solution requires that the database is represented in the GRAIL format, while the other solution (the right side of the figure) is based on a mediator architecture [9]. The mediator is an intermediate layer between the terminology server and the database, where the mediator functions as a mapper between the underlying database (which has no relation to concepts in the CORE model) and the concepts in the CORE model. The Tzolkin System [172] is an example of a system that uses such a mediator architecture. The latter solution requires the development of a *mapper* or translator (e.g., an *ontology mapper* as described in Section 2.5.2) that can transmit between expressions referred to in the database to expressions referred to in the concept module.

The GALEN Terminology Server contains a very detailed ontology for some parts of medicine, but has very poor coverage of the medical domain as a whole (statement taken from Werner Ceuster's presentation at MIE 2002 in Budapest, Hungary). To use the GALEN Terminology Server for other areas of medicine than surgical procedures would imply to extend the CORE module with additional medical concepts that are commonly in use in, for instance, primary care. Adding concepts may not be a trivial task, since the construction of composite expressions can get a bit arcane. For example, to express *excision of lobe lung*, a GRAIL composition would look like the following (example taken from Mark A. Musen's presentation in computer science course CS 270A / BMI 210A at Stanford University, fall 2001):

```
(SurgicalDeed which
   isMainlyCharacterisedBy (performance which
      isEnactmentOf (Excising which
```

```
        playClinicalRole Surgical Role) which
           actsSpecificallyOn (Lobe which
              isSpecific SolidDivisionOfLung)))
```

4.5.5 PEN&PAD

We present the Practitioners Entering Notes and Practitioners Accessing Data (PEN&PAD) project because of its relationship to the GALEN projects. The GRAIL formalism presented in Section 4.5.3 builds on the Structured Meta Knowledge (SMK) formalism that was developed in the PEN&PAD project. The PEN&PAD project produced a vocabulary and a formalism—SMK—that enabled the expression of direct observations through a structured vocabulary [42].

PEN&PAD attempted to produce a *sound* patient-record system [152]; that is, the computer-based patient record model satisfied the requirements for a *faithful* and structured patient record. Rector et al. explain [147] their devotion to faithfulness:

> The clinicians' observations must be recorded in the form in which they are made, if the record is to be faithful to the process of care. Clinicians freely mix statements about symptoms, signs and diagnoses at various levels of abstraction and detail. Our experience is that clinicians find it natural to find terms of different levels of diagnostic precision occurring within the same list of choices, for example "cough" and "acute bronchitis". In different situations, clinicians appear to perceive the presenting problem at either level of abstraction. For this reason, we draw no sharp distinction between "diagnosis", "symptoms", and "signs" and treat them all as "observations". There is strong evidence that expert clinicians perform much of their diagnosis by pattern matching. Forcing clinicians to generate a *post hoc* rationalisation distorts the record of the clinical process and fails to communicate an authentic model of the clinicians' observations to other clinicians [147].

The faithfulness principle involved the *direct observations* [148]. The patient-record system should be able to remain faithful to the direct observation of what physicians have heard, seen, thought, and done. To manage the aspect of faithfulness—the true state of the patient (i.e., what actually occurred)—a two-level architecture for the patient record has been presented [147]:

- The facts level described the direct observations of "what doctors have heard, seen, thought and done"
- The meta level described "what can be said about what doctors have heard, seen, thought and done". Intentions with the meta level was to infer knowledge on the *the clinical dialogue* (i.e., what physicians have heard, seen, and thought) and on *the decision-making actions* (i.e., what physicians have done).

An immediate benefit that comes from the two-level architecture proposed in PEN&PAD has to do with the possible distinction that can be made between the direct observations—actual facts—and the various interpretations (i.e., the meta level) that may exist based on the facts. Section 6.4.2 and 13.2 will go more into detail on how the two-level architecture for the computer-based patient record is advantageous for this thesis work.

One intention with the GALEN Terminology Server is to be able to represent the complete content of medical records. With such a complete formalism, it is also possible to use and reuse the clinical information for many different users, purposes, and tasks. With such a comprehensive formalism comes the challenging task of defining and representing medical concepts and relationships, in addition to formalizing the rules for how concepts can be combined to form composite and sensible expressions. To implement such a formalisms has proved to be tedious and time consuming, both the UMLS and the GALEN project started in the mid 1980s and are still evolving.

Because we are not seeking to completely represent every clinical concept in patient records, but to provide an information model for the record system, we have been looking into works by UK's National Health Services' Clinical Headings project and the Health Level 7's Reference Information Model for clinical communication. To communicate clinical information implies that both sender and receiver have a common understanding of the contents that is communicated. In this respect, a common understanding of clinical concepts is exactly what we are looking for; some means to define and reference the contents in patient records.

4.6 Clinical Communication

Clinicians generally use clinical terms different from each other, or they use them with different meanings. In an attempt to standardize the communication of clinical concepts, both the National Health Services (NHS)—with their Clinical Headings Project [136, 139, 171]—and the Health Level 7 (HL7)—with their Reference Information Model (RIM) [184]—have produced concepts that define and reference clinical information.

4.6.1 Clinical Headings

One of the NHS Clinical Headings Project's main objective has been to identify a limited number of headings that clinicians needed for communicating clinical information, and that could be agreed upon by every participant involved in the communication. A common understanding of what to communicate is a prerequisite, so, the headings function as a tool for common reference and navigation of clinical information.

> Thousands of clinical messages are sent and received everyday by healthcare professionals. The communications are documents, documents contain sections, the names of the sections are Headings [130].

The project produced a set [84, 92, 170] of *clinical headings* (Table 4.2). Clinical headings refer to expressions and phrases that clinicians use to name sections of their communications (e.g., symptoms, examination findings, test results, diagnosis, etc.).

Table 4.2: Clinical Headings (based on [92]).

Health Characteristics Headings	
History	Terms elicited by normal listening and questioning the patient and / or carer, or through existing information in the clinical record, which describe the patient's state of health
Family history	Terms which describe health and illness of specified family members
Social circumstance	Terms which describe the patient's environment, social network, and their care takers
Examination findings	Terms which describe the features of a patient typically elicited by a clinician during the acts of observation, inspection, percussion, palpation, and auscultation
Test results	Terms which describe the features of a patient or part thereof elicited using additional technologies
Diagnosis	Terms for all relevant diseases, disorders and syndromes
Outcome	The change in the state of a person due to one or more interventions or their absence, when evaluated by the clinician, patient or multi-professional team
Actions Headings	
Assessment	Any activity which is undertaken to find out more information about a patient's state of health and social well being
Treatment	Any activity which is undertaken to maintain or improve the patient's state of health and well being, relieve distress or reduce risk
Clinical administration	Any administrative act undertaken by a clinician, that supports the assessment or treatment of a patient, but which in itself is not investigatory
Participation	Any activity of sharing information and decision making about a specified patient with the patient or their representative
Role View Headings	
Problems	Clinical information items that a specific clinician has highlighted for further action
Alerts	Critical information items that a clinician must be aware of before making decisions about a patient
Reason of encounter	Information items that identify why a specific health care or social professional is involved with a patient at this particular time
Time Modified Headings	
Past	Information from the past, which occurred in the past and are discontinuous with the present and need to be communicated
Present	Either health characteristics, which are being exhibited by a patient or actions which are being constantly or repeatedly enacted on them, at the time of recording
Outlook	The health characteristics that a clinician or team envisaged for a patient in the future
Goal	A future view containing one or more health characteristics which describe the state that a patient should achieve in a specified time frame with the help of their care professionals
Expected outcome	A future view containing one or more health characteristics which describes the projected steady-state condition of the patient after treatment (or its lack)
At Risk	A future view containing one or more health characteristics which describe conditions which the patient has not yet got but may do so in the future without some action
Plan	Actions which a clinician or team intend to do at a future time

4.6.2 HL7 Reference Information Model

Health Level 7[1] (HL7) provides a standard for interchange of clinical data. An original goal was to provide a standard for the exchange of data among hospital-computer applications that eliminated, or substantially reduced, the hospital-specific interface programming and program maintenance that was required at that time.

A major purpose of the HL7 Version 3 Reference Information Model (RIM) is to secure a shared and consistent meaning among the parties that communicate. The model defines a standard for communication of health information between loosely coupled information systems, but is not a data model. RIM is a model for the specification and creation of HL7 messages, but is not a model for database storage. The RIM model is a collection of subject areas, scenarios, classes, attributes, use cases, actors, trigger events, and interactions that depict the information needed to specify HL7 messages. In this sense, it is more than a data-interchange standard for specifying the format for data exchange.

RIM uses a language similar to the Universal Modeling Language (UML). As HL7 is primarily concerned with information exchange and not information storage, everything in the model is a subject of communication. RIM consists of five subjects, or subject areas [184]:

1. RIM_Acts: An instance of *Act* represents a record of an intentional action in health

---

[1] HL7 refers to the applications level (level 7) of the International Standards Organization / Open Systems Interconnection (ISO / OSI) model.

care

2. RIM_Entities: Physical things or organizations, or groupings of physical things
3. RIM_Roles: Devices used to administer therapeutic agents
4. RIM_Communication_Infrastructures: A set of classes related to the definition and control of message-based communication between loosely coupled information systems
5. RIM_Unassigned: Yet to be assigned.

Our interest in looking at standards for communication of clinical information has to do with two properties: (1) the common understanding of what is communicated in messages between participants involved in the communication and (2) the format in which information is communicated. Neither clinical headings nor the RIM are strictly formal representations. Clinical headings are merely language descriptions, whereas HL7 messages are specified by way of the RIM.

As our intentions are *not* to completely represent the contents of patient records, communication of clinical information through clinical headings and HL7 messages are ways to refer to clinical information without the requirement of full-fledged formalisms for representing content. A disadvantage of not having a strict formalisms for representing content is the chance of ambiguous and uncertain inclusions and interpretations of communicated information.

Clinical headings need to be formally defined in terms of what information belongs to each heading, but the contents of the headings can be referred to as, for instance, string units. These string units would have a label as opposed to pure free text. Clinical headings and HL7 messages intermediate representations between full-fledged formalisms and pure free text.

4.7 Summary

In this chapter we have briefly gone through some important classifications, terminologies, and vocabularies that are used to represent clinical information in health care. To standardize clinical nomenclatures is a challenging tasks due to the multiple users, users' needs, and heterogeneous health systems. Both the UMLS project and the GALEN projects started in the mid 1980s. The fact that these projects have been going on for 10-15 years is an indication of how difficult it is to represent medical information. In contrast to their approach in formalizing clinical concepts are other projects such as the NHS Clinical Heading and the HL7 RIM which have suggested semi-formal representations for clinical information.

The terminologies, standards, and vocabularies presented in this chapter have varying degrees of formal representations. We classify these nomenclatures and their degree of formality on a conceptual scale. The left side of the scale defines no formal representations and the right side defines full-fledged formalisms. On the very left side of this scale we place pure free text (e.g., textual descriptions), on the very right side of the scale we place formal logics (e.g., first-order logic). Classifications such as

ICD and ICPC are placed somewhere on the left-half of the scale; SNOMED, SNOMED-RT, and SNOMED-CT are placed somewhere in the middle; the UMLS is placed in the right-half of the scale; and the GALEN Terminology Server is placed somewhere right after. In contrast to the Terminology Server are clinical headings and their nonformalism that puts them close to pure free text on the very left side of the scale.

Part II

Empirical Studies

Chapter 5

Activities in the Primary-Care Process

5.1 Introduction

This chapter reports on a study of the primary-care process. The purpose of the study has been to observe family physicians during patient care.

Of the many services the computer-based patient-record system provides family physicians with, one of the main services has to do with the provision of information to their medical decision-making process. In efforts to build computer-based diagnostic systems, for instance, systems that support clinical decision-making processes, several surveys have been conducted to characterize physicians' information needs in patient care [128, 183].

Berner et al. [22] address the importance of usefulness when evaluating computer-based diagnostic systems: "two major issues need to be addressed: accuracy and usefulness". To specify what accuracy and usefulness implies, Berner et al. suggest that it is necessary to understand the clinical context in which computer-based patient-record systems appear to be most useful, or how they actually perform. No systematic assessment exists yet of how computer-based patient-record systems perform in different clinical situations.

The information in patient records are mostly produced during the clinical encounter, where physicians perform a set of care activities that produce and require patient information. Information necessary for physicians decision-making process can be divided into two categories: (1) information that can be retrieved from patient records and (2) information that can be retrieved from other information sources than patient records. A major aim with the study is to identify and characterize physicians' interaction with the patient-record systems. An interaction with the patient record system is part of a care activity. We want to identify the various care activities for the primary-care process.

To our awareness, no surveys have been conducted in Norway so far, with the purpose of investigating physicians' information needs during patient care, and how much of this information can be supported by information from the patient records. The surveys that have been performed have had a focus on physicians' use of information technology (e.g., the exploitation of the Internet) [29], rather than on how physicians work with and use the patient-record system, and in what terms does the system contribute with useful and needed information to their decision-making activities.

Smith [174] concluded that around 40% of physicians' questions can be answered with information from the patient records, whereas a quarter of the questions require a synthesis of patient data and medical knowledge. Gardner [64] suggested that physicians' request for information be classified into two categories: (1) information that can be retrieved from the patients' medical records, and (2) medical knowledge about the present problem. Kuilboer et al. [89] found that 49% of the questions that physicians ask during patient care can be answered by retrieving information from the patient records. Medical Records Institute's annual survey on electronic medical records reported that in 1999 [78] and in 2002 [79], 72% and 75.7%, respectively, of the physicians claim they need *patient data* from the patient records for various different purposes in health care.

The study will be accomplished through an engineer's approach to observation. Software engineering concerns itself with all aspects of software production, from system specification to system maintenance [175]. Knowledge engineering concerns itself with the same aspects as software engineering, but with a particular focus on transforming knowledge (i.e., problem-solving expertise) from a knowledge source into a computer program. Important in knowledge engineering is the modeling of additional aspects of the computer system—not just the conventional software application—such as the organization and the environment in which the prospective system will operate within.

5.2 Materials and Methods

This study includes a two-month fieldwork comprising 100 hours of observation of five family physicians and a specialist and chief physician in rheumatology.

5.2.1 Selection of Clinics

We spent 90 hours of observation during 11 days, from May throughout June 2000, observing five family physicians in four different primary-care clinics in Norway. We also spent two days, of five hours each day, observing a specialist in rheumatology.

The clinics we selected for the study were related to university environments in Oslo or Trondheim (University of Oslo and Norwegian University of Science and Technology (NTNU), respectively). All physicians that agreed to host an observer had a relationship to one of the universities, either through teaching or through research, within public health and primary care.

Two of the clinics in Oslo were used to hosting medical students from the Department of Community Medicine at the University of Oslo; they therefore, accepted to have an observer during patient encounters. Likewise, in the two clinics in Mid-Norway, physicians in the clinics were particularly interested in research related to the computer-based patient-record systems and were, for that reason, willing to let us observe them during patient care.

Each clinic employed 3-4 physicians and 3-4 clinicians (Table 5.1). The clinics had a local laboratory and modem connection to an external laboratory.

Table 5.1: Demography of primary-care clinics.

	Clinic 1	Clinic 2	Clinic 3	Clinic 4	Specialist
Patient-record system	Profdoc	Profdoc	Infodoc	Profdoc	Doculive
Record-system platform	DOS	DOS	Win	DOS	Win
# Physicians[a]	4	4	4	3	8
# Clinicians[b]	4	4	4	3	-
Local laboratory	yes	yes	yes	yes	yes
External laboratory	modem	modem	modem	modem	paper
Front office	yes	yes	yes	yes	-
Physician office	yes	yes	yes	yes	yes

---

[a] Family physicians and specialist physician in Rheumatology.

[b] Nurses, laboratory assistants, and ancillary staff (e.g., receptionist).

5.2.2 Setting of Observation

The method we used in this study combines in situ observations, unstructured interviews, and think-aloud protocols [183, 200], where we could freely ask questions in between patient encounters and where physicians communicated their thoughts to us with respect to the patient-record system. The physicians took the time in between patient encounters to willingly answer our questions and talk about their own opinions on the usefulness of the computer-based record system. A conversation could continue over a whole day, because the physicians would make sure they kept the appointment times as prompt as possible.

The physicians had experience in having medical students that participated in their encounters, and they informed every patient about us and our participation as observers in the encounter. Physicians would introduce us to the patients and say something like:

> *Today we are having some observers here. They are doing research on the computer-based patient-record system. Is it OK for you that they are present in the office? They are not going to watch you, they are going to watch me and how I work with my computer system.*

On one occasion we had to leave the office. For the other encounters we watched the whole encounter, including those that involved clinical examinations and clinical surgery. We sat behind the physician such that we could see the patient-record on the screen, the physician, and how he or she worked with the record system. As a consequence, we sat close to both the patient and the physician and felt very present in their conversations. We were even invited to partake in a conversation with a patient, his wife, and the physician. If the physician left the office to speak to a laboratory assistant or a clinician in the front office, several patients initiated a conversation with us asking us about our research and work. There was only one observer present in the clinical encounter.

Field Notes

Because we sat so close to both the patient and the physician, we did not want to disturb or draw the attention of either the patient or the physician by writing notes; we rarely took any notes during the observations. Writing notes would impose on the conversation between them (we could theoretically have written while we talked to the physicians in between encounters. However, the conversations were so compact that we hardly had any time to write while we talked). After the end of the day, we wrote summary reports.

Appendix A contains an explanation of terminology for this chapter. We will make a note that we see the clinical encounter as the main process in which physicians treat patients. Pre-encounter and post-encounter assessment are activities that happen before and after the clinical encounter, respectively. Activities have been separated into two categories; administrative and clinical. We focus on clinical activities that have an impact on family physicians medical decision-making process.

5.2.3 Models for Analysis

Components in a Process Model

The Workflow Management Coalition [187] has defined a workflow model, originally for optimizing business processes. Workflow modeling has also proved successful in a number of different settings within the medical domain, for instance, modeling of organizational knowledge in a hospital setting [140, 178] and modeling of biological processes [137]. Because work processes involve a number of different people and organizational units, they incorporate boundary spanning both within and outside the organization. Workflow models emphasize the integration of procedures and application tools between participants performing activities in a workflow process [82]. The workflow paradigm defines generic concepts such as processes, activities, participating actors (participants), and transition requirements from one activity to another [106, 113].

ActionWorkflow

Winograd and Flores [199] developed the ActionWorkflow loop (Figure 5.1). Each loop has a customer and a performer and consists of a sequence of four phases: (1) preparation, (2) negotiation, (3) performance, and (4) assessment [1]. A model based on ActionWorkflow loops comprises a set of loops connected by lines that show dependencies in terms of workflow between activities. Activities modeled by the ActionWorkflow loop identify a customer and a performer of an activity who further agree upon a set of conditions of satisfaction, which determines the completion of an activity (fulfillment of conditions of satisfaction for both performer and customer) [7].

---

[1] Assessment, in the context of workflow modeling, is not to be confused with clinical assessment that leads to a medical diagnose on a patient (i.e., diagnostics).

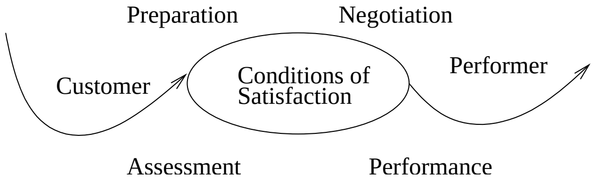


Figure 5.1: The ActionWorkflow loop developed by Winograd and Flores [199].

Organization and Activity Models

According to Schreiber et al. [167] there are six components in an organization model, these are presented in Table 5.2. “The organization model supports the analysis of the major features of an organization to discover problems and opportunities for KBS development, as well as possible effects a KBS could have when fielded” [168]. In addition to the organization model, there is the activity model which “describes, at a general level, the tasks that are performed or will be performed in the organization where the expert system will be installed” (ibid.). Table 5.3 illustrates the components in the activity model.

Table 5.2: Components in the organizational model.

Organization component	Description
Structure	Structure of the organization
Process	Various processes in the organization
Participants	Participants in the organization
Resources	Application tools used in the processes
Knowledge	Knowledge needed to perform activities and processes. Activities are modeled using the components listed in Table 5.3
Culture	Unwritten rules, habits of working and communication, etc. in the organization

Table 5.3: Components in the activity model.

Activity component	Description
Process	The name of the process that includes the activity
Activity	Name of activity
Purpose	Purpose of the activity
Data	Data (or information) used by the activity
Participant	Participants in the activity

The ActionWorkflow loop in Figure 5.1 introduces a very useful feature—the ability to distinguish between the customer and the performer of an activity. Participants in an activity play different *roles*, in this case, performer, customer, or both. The ability to differentiate roles becomes very useful when we want to model family physicians’ interaction with the patient record system.

We want to determine physicians' information use and needs for each activity in the primary-care process and we distinguish between *information user* and *information owner*.

Definition 5.1 *The participant who enters information into the patient-record system plays the role of* Information owner. *An information owner* provides *information to the patient record.*

Definition 5.2 *The participant who uses information from the patient-record system plays the role of* Information user. *An information user* depends *on information in the patient record.*

Although we do not aim to use workflow systems to *automate* clinical processes, we use workflow models to analyze activities and their participants in terms of information owner and information user (Figure 5.2). In addition, we pick out relevant components from the organization model and the activity model to enable us to model the primary-care process; these components are listed in Table 5.4.

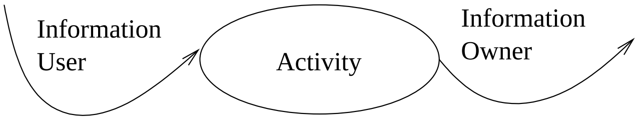


Figure 5.2: The ActionWorkflowloop adjusted to model primary-care physicians' interaction with the computer-based patient-record system for each activity in the primary-care process.

Table 5.4: Selected components used to model the primary-care organization and activities.

Category	Description
Organization	Clinical offices
Process	The primary-care process
Subprocess	Subprocesses within the primary-care process
Activity	Activities within primary-care subprocesses (activities are related to the medical decision-making process and physicians' interaction with the patient-record system)
Participant	Participants in the clinical setting modeled with the two following roles: (1) *Information owner* and (2) *Information user*
Type of activity	Administrative or clinical. Our focus is on clinical activities that participate in physicians decision-making process
Application tool	The computer-based patient-record system
Data	Clinical data in the computer-based patient record (relevant to clinical decision making)

5.3 Results

5.3.1 Pre- and Post-Encounter Assessment

Physicians made use of the pre and post phase to the clinical encounter; two phases in which no patient was present in the physician's office. During the post-encounter phase, physicians usually terminated actions that they did not complete during the encounter, such as finishing the encounter summaries, printing prescriptions and referrals, writing messages to themselves for the next encounters with the patients, etc.

Get into Context

In the pre-encounter assessment phase, before the patient entered the office, physicians used the patient-record system to *get into context*. In this pre phase, physicians spent everything from a few seconds to a couple of minutes preparing themselves by looking at the patient record, switching from one encounter context to another.

In pre encounter, physicians looked for *memory triggers* which helped them remember something about the patient: *I have made a bet with this patient about him losing weight. He needs to lose 5 kg. He will bring his wife with him today, and most likely, we will have a conversation about why he does not feel motivated to lose weight* or *This patient comes to me because she just needs to talk, she feels lonely*.

Vital information that helped them to switch context from one patient encounter to another were information such as patients' name, ID, and reason for encounter, which they accessed through their electronic daily schedule (part of the appointment system). Clinicians in the front office entered this information into the appointment system when they received patients' request for an appointment. Figure 5.3 illustrates how information flows among the clinical offices, via the computer-based record-system.

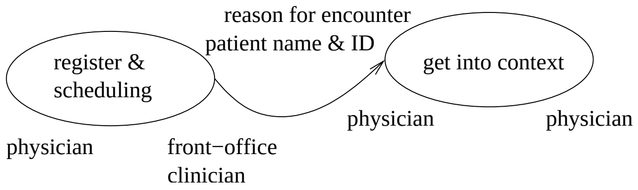


Figure 5.3: The pre-encounter assessment phase, where the activity *get into context* depends on information from other activities such as *register & scheduling*. The information flow between these activities consists of clinical concept categories such as *patient name*, *patient* ID, and *reason for encounter*.

5.3.2 Primary-Care Clinical Encounters

Most of the patient encounter happened in the physician office. The interaction among the physician office and the two other clinical offices, laboratory and front offices, were limited to performance of laboratory tests (Figure 5.4) and practical tasks of administering patients (Figure 5.3).

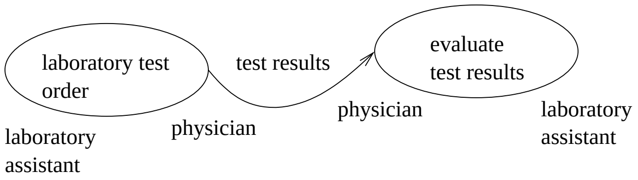


Figure 5.4: The laboratory test-order and -result procedure demonstrate how information flows among the different clinical offices, and how physicians depended on information from both the front office (as in the pre-encounter assessment phase, Section 5.3.1) and the (local) laboratory. In the first activity, where the physician specifies the *laboratory test order*, the laboratory assistant has the role of the customer (information user). While in the activity of *evaluate test results* the roles have changed and the physician takes the role of the customer and the laboratory assistant has the role of the performer (information owner).

Physicians scheduled 15 minutes for each encounter (in two specific situations they scheduled a double encounter due to the reason for encounter). During these minutes they had to get into context, listen to a patient's problem statement, have a dialogue with the patient, perform clinical examinations, order laboratory tests, write, fill out, and print forms, letters, health certificates, prescriptions, and write a summary of the clinical encounter. We can characterize physicians' interaction with the patient-record system in terms of the following observations:

- Physicians spent a significant part of the encounter introductory on patients' problem statement (Figure 5.5). A patient's problem statement usually included: (1) a reason for encounter, (2) a set of symptoms, (3) complaints, (4) a more specific description of the problem, and some times (5) a repetition of the patient's relevant problem history. Patients often presented more than one problem

- Physicians read from and wrote in the record during all phases of the encounter, while the patient explained his or her problem, while the patient waited, while the patient prepared him- or herself for a clinical examination, or after the patient had left the office

- During telephone encounters, physicians demonstrated positive effects of using the patient-record system. Physicians had at least two–three telephone encounters (or house calls (Appendix A for glossary)) in between regular encounters (or office encounters (Appendix A for glossary)) each day

- Most physicians performed clinical examinations of the patients. Figure 5.6 demonstrates the information flow from the actual clinical examination to the recording of corresponding findings in the patient record.

- Assessment and treatment activities involved active use of the patient-record system. Table 5.5 lists the clinical activities we identified for the primary-care process. Even though each clinical encounter were unique, they all followed a regular pattern that included a limited set of care activities. An encounter started with a patient's problem statement and then followed with a set of care activities. We revealed 24 different clinical activities and 13 subprocesses as part of the primary-care process.

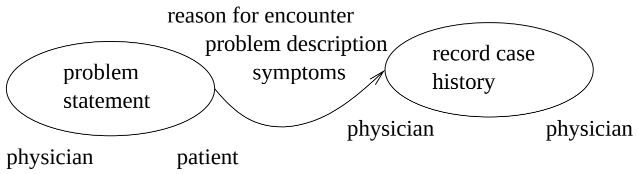


Figure 5.5: When a patient gives his or her problem statement, the physicians is requesting information (i.e., the physician is the information user or customer) from the patient (i.e., the patient is the information owner or performer). In the activity *record case history* the physician writes case history information into the patient record and uses it later in his or her own decision-making process (i.e., the physician is then both information owner and user).

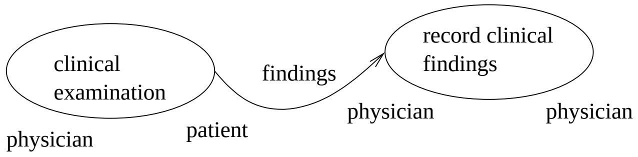


Figure 5.6: The information flow between the actual activity perform *clinical examination* and the activity *record clinical findings* into the patient record.

Reflections from Family Physicians

Physicians expressed their opinions regarding the use of the patient-record system and its usefulness in patient care. The following lists some of physicians' most prevalent expressions:

- Flexibility between modules: A large part of the clinical encounters required moving between modules such as the scheduling and billing, the patient record, the laboratory, and the prescription module. Physicians considered moving between different modules as one of the advantages with their patient-record systems
- To have overview: Physicians mentioned that having an overview played an important part in their decision-making process and that current patient records did not provide good means for getting an overview. An important aspect of having an overview had to do with *knowing what physicians have asked patients before, and what physicians have previously done*

Table 5.5: Clinical processes and care activities.

No	Subprocess	Activity
1	Register & scheduling	Identify & register patient
		Record reason for encounter
		Schedule appointment & type
2	Pre-encounter assessment	Prepare patient
		Get into context
3	Assessment	Record patient's problem statement
		Identify possible hypotheses
		Identify new or existing problem
		Set diagnose
4	New or existing problem?	Determine context
		Continuation of problem?
		Sequelae of problem?
		Subproblem?
5	Clinical examination	Perform clinical examination
		Evaluate examination findings
6	Laboratory test	Laboratory test order
		Evaluate test results
7	Referral	Specify referral
8	Discharge	Evaluate discharge
9	Treatment	Determine treatment therapy
10	Surgical therapy	Give surgical therapy
11	Drug therapy	Give drug therapy
12	Physical therapy	Give physical therapy
13	Post-encounter assessment	Finalize encounter

- A feeling of ownership toward the patient record: All clinics had developed their own templates for laboratory tests. A couple of physicians had developed templates for treatment (e.g., pregnancy, maternity check up, and diabetes). Two physicians had installed and made available electronic knowledge sources. One physician developed his personal way of remembering and retrieving patient information by making his own customized overview. Another physician held his personal overview of patients' history of illness

- Interaction between paper and electronic media: Physicians considered the interaction between paper and the computer-based patient record as inconvenient. This inconvenience was especially mentioned in relation to receiving hospitals discharges. The hospitals usually sent discharges on paper, which then had to be entered manually into the patient-record system. First, the physicians had to read and select what information to enter, then he or she passed the discharge to the front office, where another clinician had to write the marked information into the computer-based patient record.

5.3.3 Specialist-Care Encounters

The most distinct difference between the rheumatology clinic and the primary-care clinics had to do with the use of the patient-record system. Although the specialist had a patient-record system in his office, he did not use it. He used the paper-based patient record instead, claiming that it was more complete than the computer-based patient record. Another visible difference between the two types of clinics had to do with who wrote the encounter narratives; while family physicians entered themselves information into the patient record, the specialist recorded his summary on tape and sent the tape to a typing pool.

5.4 Discussion

The organization of the primary-care clinics seemed to work efficiently and constructively among the various clinical offices, with respect to participants, activities, and information flow between the clinical offices.

Physicians had access to the record system at all times during the care process, where most of the encounter happened in the physician office. As a consequence, physicians used the record system quite actively. Physicians even expressed a feeling of ownership toward their record system. That physicians themselves both entered and retrieved the record information appeared to motivate them in writing long and detailed encounter summaries. Physicians were also willing to spend extra time in maintaining and keeping information updated in the record.

On the one hand, physicians perceived their record system as a flexible system, where moving between modules such as scheduling, billing, and patient record, was convenient and easy. The clinics had one record system that performed all tasks ranging from administration, to governmental obligations, and clinical encounter tasks. Physicians also expressed enthusiasm toward the integrated messaging system. Due to electronic messages paper notes had been completely removed, more or less, from their desks. On the other hand, physicians perceived their record system as inconvenient with regard to the interaction between paper and the computer-based system.

5.4.1 Limitations of Methods

We have conducted an observation study of family physicians in Norway. Observations of family- and specialist-care physicians have served two important purposes:

1. To learn about the primary-care process, in terms of care activities and family physicians' interaction with the record system
2. To elicit knowledge from the primary-care process, which we will use as input to a set of knowledge models.

5.4.2 The Engineering Discipline of Constructing Models

Knowledge elicitation deals with the acquisition of data and material needed for knowledge modeling. Of the various knowledge-elicitation techniques that exists, interviews—both structured and unstructured—represent the most commonly used ones. Observation and unstructured interview techniques suit the early stages of the modeling process (i.e., during organizational analysis or at the start of the knowledge identification phase), because they involve observing the expert actually solving problems in the domain of interest. "It has been shown repeatedly that the context in which one encodes information is the best one for recall" [167].

Observations of the primary-care process, therefore, give input to the modeling of care activities, with respect to the clinical organization structure, process activities, participants, and information use and needs for each activity. A primary purpose of the study has been to learn how family physicians work with the computer-based patient-record system, where we have divided observations into three categories: (1) observations related to the organizational structure of primary-care clinics, (2) observations related to activities, participants, and information needs in the clinical process, and (3) observations related to how physicians work with and use the patient-record system.

The authors of the CommonKADS methodology [167] emphasize that a success criterion for users' acceptance of a computer system includes modeling of organizational aspects, such as processes, tasks, actions, goals, intentions, participants or actors, resources, data, data flow, etc. A system's critical success factor depends on how well organizational issues have been dealt with. Computer systems must integrate well within the overall (business) process to be accepted by their users in their tasks as knowledge providers. The process of building a knowledge-based system has been approached as a modeling activity. Building a knowledge-based system means building models with the aim of integrating domain expertise into the computer system models [180].

Transcription of Field Notes

To transcribe unstructured field notes into text that we could utilize as input to our knowledge models, we used a *template analysis style* [104, 105]. The template analysis style employs a set of existing templates for structured analysis. The templates are used to produce text that enable us to identify *units and categories* that we can then place in models and tables. The templates we used for our models and tables have been given in the previous Section 5.2.3.

The results of this study is based on the field notes we wrote after the observations. We did neither do any recording of voice nor any recording of video during the observations; which would have enabled us, if we did, to study the observations in more detail afterward. Reasons why we did not do any recording has to do with us not wanting to disturb the clinical encounters. An important purpose for our study was to learn about the clinical encounter, in terms of how physicians relate to the patient, the patient-record system, and other clinicians during patient care. As the clinical encounter was very intimate, we would have spoiled this intimacy (that is one of the reasons why we did not take any field notes during the encounter). We believe that it was equally important to maintain this intimacy, which we have confirmed that we obtained, since we were invited to partake in the clinical encounter between the physician and the patient.

An Ideal Primary-Care Process?

Results from this study is based entirely on observations that we have done from the selected set of primary-care clinics and so, it is important that these observations are based on a representative selection of Norwegian primary-care clinics. Furthermore, the importance of having observations based on a representative set of Norwegian primary-care clinics is emphasized since we do not have any prior experience with the primary-care process, also taking into consideration that we are engineers and not physicians. We divide our results into two groups of results, results that are directly dependent on the observations that we have made—which implies that they will be directly affected by a potential bias that may exist if the clinics are not representative—and results that are not directly dependent on the observations and, as a consequence, they are not so sensitive to potential biases that may exist.

The selected set of clinics and physicians all have a relationship to universities and research environments that have a particular interest in primary care and computer-based patient-record systems. This explicit relationship may represent a bias, in terms of a non-representative selection of primary-care clinics in Norway. Some of the assumptions that we make in a later section (Section 5.4.3) may be based on observations that are non-representative.

For example, we do believe that we have captured a set of relevant generic activities and subprocesses that represent the primary-care process, because these activities are independent of physicians' interests in computer-based patient-record systems (i.e., the set of generic primary-care activities belong to the group of results that are not directly dependent on a representative selection of primary-care clinics). The validity of the existence of care activities are more dependent on physicians' medical knowledge (i.e., physicians select a set of care activities dependent on their medical expertise) than they are dependent on physicians' particular interest in computer-based patient-record systems' impact on patient care. However, it may be likely that the physicians we have observed were particularly motivated to use the patient-record system for reading and writing into it and for this reason they are biased and our findings are non-representative. In one of our results we assume that the patient record is a *content provider*, based on the fact that physicians actually write the amount of information that we say they do. If this assumption is not representative (i.e., physicians do not use the patient-record system as much as we say that they do), the validity of this indication should be questioned.

5.4.3 Uniqueness for Every Patient, Yet a Generic Process

To model the primary-care process, with the action-workflow loop, proved to be easier than anticipated because of the generic nature of the primary-care process. Workflow modeling that originally was intended for modeling business processes has also proved suitable for modeling processes in the medical domain. Several projects have successfully modeled various clinical settings, ranging from using workflow systems in health-care organizations [37, 114] to using workflow systems in biological processes [137].

Despite a generic pattern in the primary-care process activities, every patient encounter was unique. The uniqueness of each encounter lay in the patient and the patient's medical condition, and how these two features induced a unique combination of activities in order to handle the patient's problem(s). The unique set of activities is projected in the patient record as a summary of the patient's problem statement and documentation of relevant information to that statement. The content in the patient record can reproduce the uniqueness of the clinical process, given the generic set of activities in the primary-care process. This way of looking at the primary-care process comprising generic activities, where the content of the activity projects in the patient record, emphasizes the importance of the patient record as a container for useful and needed information relevant for various decisions that appear during patient care. Our findings may indicate that the patient-record system helps and enforces physicians in being able to treat every patient uniquely.

5.4.4 The Patient Record as a Context Provider

Our impression was that the patient-record system played a significant role in helping physicians with the task of getting into context. Each physician had an electronic daily schedule that presented patients' name, ID, and reason for contact. Often this information proved to be sufficient for physicians to trigger their memory. If not, they could open the patient record and read the previous encounter summaries. If this did not suffice for them to remember anything about the patient, physicians would, in some circumstances, search for and read list of diagnoses and list of medication to get a picture of the patient's medical condition.

5.4.5 To Have an Overview

Physicians claimed that part of having an overview had to do with *knowing what physicians have asked the patients before, and what actions have been done to them*. This statement suggests that if a patient record should contribute to giving physicians overview—to reveal what physicians have asked and what they have done before—physicians should have a common way to reference and communicate patient-record information.

Much in the same way as clinical headings have been suggested as a way to communicate clinical information (Section 4.2), we suggest that a set of clinical concepts, which classify and represent the patient-record content, is a way to obtain a common understanding of what has been said and what has been done in preceding patient-care activities.

To communicate clinical information implies a common understanding of what is communicated. A set of clinical concepts can provide a tool for recognizing record information and for understanding what has been done. For physicians to determine what has been previously asked and done, they need to know, first of all, what the patient record contains. Our suggestion implies that a set of clinical concepts can function as a tool for physicians' common understanding of record information, which again render possible a better overview of patient-record content.

5.4.6 Primary-Care Processes and Activities

Table 5.5 presents the subprocesses and corresponding care activities that we identified for the primary-care process. We have added one activity to the pre-encounter assessment phase that we did not actually observe—*Prepare patient*. This activity refers to preparative tasks intended to do before the clinical encounter starts, such as announcing the patient's arrival, preparing the patient's record, taking a patient's urinary test or a blood-sample test, etc.

Set diagnose is counted as a separate assessment activity because most clinical encounters include the activity of determining a patient's diagnose (be it symptom or condition based). We consider the explicit activity of setting a diagnose as a significant activity, especially in connection with a problem-oriented patient-record system, which will most likely require that some kind of association be made between patient-record content and the name of a medical problem. The activity of associating patient-record content with a medical problem name can be compared with the activity of determining a diagnose on a patient's condition.

The activity *Identify possible hypotheses* refers to a mental activity that we have made explicit based on the hypotetico-deductive method [19, 75, 98]. The hypotetico-deductive method is a process where medical data acquisition (through case history, clinical examinations, and laboratory tests) and treatment are intimately tied to an ongoing process of hypothesis generation [77]. Physicians generally have to chose among three decision paths when they are presented with a patient's problem statement: (1) is it a new problem?, (2) is it related to an existing problem?, or (3) is it a sequelae of an existing problem (i.e,. a complication or a consequence of a previous condition)? In addition to deciding a problem statement's relation with a patient's existing medical condition, physicians have to make sure that the problem statement is not related to something severe or serious. The three decision paths generate three activities in our set of generic primary-care activities: (1) continuation of problem?; (2) sequelae of problem?; and (3) subproblem?

The treatment activities that we have listed in Table 5.5 correspond to treatment activities in Grimsmo's et al. [69] process model for primary care in Norway.

5.5 Conclusion

Our main focus in this observation study has been to identify physicians' interaction with the patient-record system. From the result of this study we can describe physicians' interaction with the record system in terms of the *roles* the *participants* can have, the *activities* the participants perform, and the use of the record system determined by *information flow* among activities.

We have identified a limited set of generic subprocesses and care activities for the primary-care process that enable us to model the primary-care clinic as an organization that performs processes, which are composed of a set of activities, where each activity defines a purpose and is performed by one or more participants.

Norwegian primary-care clinics have organized themselves in a way that enable them to work efficiently and constructively with the patient-record system. The encounter is the main clinical process, in which a majority of physicians' medical decisions are made. Most of the encounter takes place in the physician office. Thus, family physicians have access to the record system at all times during patient care and they can retrieve patient information whenever they need or want.

Physicians have proved to spend a significant part of the 15 minutes they devote to patient care in writing detailed and long encounter summaries into the patient record. Part of the motivation for spending so much time, may be that physicians play both roles of information owner and user; they write themselves the summaries and use this information later in their decision-making process. Physicians use actively the record system during all activities of patient care, for both reading and writing.

The patient-record system plays two important roles: (1) it functions as a *context provider* when physicians switch contexts from one patient encounter to another and (2) it functions as a *content provider* for the generic primary-care process. Because physicians only spend a short time between consecutive encounters they use the patient-record system to quickly switch patient contexts. Despite the fact that the primary-care process is repetitive and generic, each encounter is unique. The uniqueness in each patient encounter lies in the combination of the patient and the patient's medical problems. The patient's medical problems are projected in the patient's record, which means that part of the content that makes every encounter unique is in the patient record. The two roles that are supported by the record system—a content provider and a context provider—indicate that the patient-record system is an essential tool for family physicians in Norwegian primary care.

Chapter 6

Categories of Patient-Record Information

6.1 Introduction

This chapter reports on a study of patient-record contents from a Norwegian primary-care record system. The purpose of the study is to identify a set of clinical concepts that classify and represent the patient-record contents.

Several studies have been performed with intentions to identify physicians' information needs during patient care [62, 116, 135]. Information needs can be associated with questions that arise in physicians encounters with patients. For example, questions regarding a patient's diagnostics, test orders, drug treatment, etc. To identify the many information sources that can answer these questions, one way is to classify the questions into categories of clinical concepts. Once physicians' information needs are categorized into clinical concepts, they can be related to various information sources.

Most of the studies reported in literature targets physicians' information needs in hospital settings. Many of the identified information sources were sources that extend beyond the patient-record system (i.e., medical literature, specialized sites on the world wide web, external computer-based knowledge sources, etc.).

From a previously conducted observation study of the primary-care process, we discovered that family physicians use the patient-record system extensively during all phases of the clinical encounter, for both reading and writing (Chapter 5). One of the reasons for the extensive use may be explained by the fact that physicians have access to the record system at any time in patient care. The study was particularly focused on physicians' interaction with the record system in care activities. Physicians spend a significant amount of time writing into the patient record, detailed and long encounter summaries; they also retrieve the information they write. Family physicians were motivated to enter information they eventually will use in their decision-making process, into the record system; for this reason, we suppose that a significant part of family physicians' information needs can be retrieved from the primary-care patient records.

To support continuity of care and to manage medical problems over successive encounters, problem orientation was suggested as a way to structure information in the patient records (Section 1.1.3). Long and nonorganized patient records hinder easy access to information. Physicians work in time-pressured environments and usually do not have time to search for information. Hence, a problem-oriented view of patient data is intended to provide easily access to information through units of medical problems. Information that are related to the same medical problem are organized into a problem unit.

As part of the problem-oriented medical record, the SOAP format was suggested to guide physicians way of entering information into the records. SOAP refers to four categories of information; Subjective, Objective, Assessment, and Plan, respectively. Physicians were encouraged to enter information according to the SOAP categories because they were intended to reflect how physicians think and work.

Problem-oriented records have yet to prove their success in clinical settings. However, if the SOAP format is meant to reflect how physicians work and think, we might be able to detect this structure in the contents that are in patient records. And more interestingly to ask, do these categories have any relationship to physicians information needs? The information that physicians enter into the patient records are a reflection of their information needs. A study of patient-record contents from a Norwegian primary-care system will serve several purposes: (1) to identify family physicians information needs that can be related to what physicians write in their encounter summaries; (2) to categorize the contents into a set of clinical concepts that classify and represent the patient-record contents; and (3) to evaluate the relevance between the identified set of clinical concepts and the SOAP categories.

6.2 Materials and Methods

The material used for this study is taken from a record system that is widely used in Norwegian primary care—Profdoc[1]. Profdoc is one of three major primary-care patient-record systems (the other two are WinMed and Infodoc). The study covers 12 patient records that have evolved over a period of nine years; from 1991 to 2000. Between 15-20 primary-care physicians account for the contents in the patient records.

We had access to the database files[2], but not the database schemas (the description of the database) [55]. To learn about the relationships between categories of information in the patient records, we reverse engineered the database files to reconstruct some of the most important database schemas. The reverse-engineering process has contributed to a design of a patient-record system in terms of modules. The modules are depicted in Figure 6.1.

Criteria for Selecting a Set of Patient Records

A set of criteria influenced the selection of patient records from the record system. The four criteria are related to characteristics that we wished to study with respect to

---

[1] Profdoc DOS version 2 (Rel 19970401).

[2] The major file formats (e.g., DBF, DBT, and NTX) are produced from Clipper—a programming language and superset of dBASE III+.

a future problem-oriented patient-record system. The selected patient records should contain medical problems that would enable us to study the following features:

1. Problems with slow onset and few or unspecific symptoms that could enforce a long time of progression, which emphasize the importance of being able to sort out specific and related information. During ongoing episodes of care, other problems may evolve and possibly interact and, thereby, create a confusing medical picture. A patient's medical conditions should be reflected through a list of problems and subproblems
2. Problems that are diagnostically challenging, which often lead to dead ends and several assessment iterations. With these kind of problems it is important to maintain and be able to look back on the case history apart from other assessment information in the record
3. Problems with chronic diseases that require continuous shared care, and where communication between primary and secondary care needs to function well. Communication between primary and secondary care is dependent on information from the patient records
4. Problems that represent severe diseases, frequent complications, and sequelae[3] as a result of medications and conditions.

When we selected patient records we looked for records that satisfied the listed criteria. We picked four problems: (1) rheumatic conditions that involve a variety of different symptoms and signs, and equally many conditions and types of disease [54], and where the conditions have a challenging diagnostics [80]; (2) hypothyroidism that often has diffuse symptoms such as alterations in behavior, moods, and cognition, which represent challenges in diagnostics and treatments [26]; (3) osteoporosis is a risk factor; and (4) hypertension, which also is a risk factor for cardiovascular diseases [109]. Hypertension is often used as an example condition in medical-informatics literature [189] and thereby, provides the possibilities for comparison and integration with other efforts in modeling and system-design, in the future.

6.3 Results

We studied in total more than 1000 record entries in order to pick out a set of 12 patient records that fulfilled the criteria listed in Section 6.2. The 12 selected patient records together constitute 922 entries, where each record contained a chronological log of time-stamped entries.

Six patient records belonged to female patients and the remaining six records belonged to male patients. The patients ranged in age from 27 to 75 years old, and their records spanned from seven to nine years. Most of the patients had five to twelve different treating physicians. Table 6.1 lists demographic information on the selected patient records. Table 6.2 lists the frequency of diagnosis for the four conditions we

---

[3] Appendix A contains a glossary of terms used in this chapter.

Table 6.1: Demography of 12 selected patient records.

Age	Gender	# Entries	# Diagnoses	Period	# Physicians
75	F	85	42	9	-
49	M	107	31	7	12
27	M	51	12	7	-
41	F	48	15	9	-
30	M	68	11	8	8
53	M	51	9	7	5
38	F	94	23	8	12
34	F	107	22	7	-
32	M	35	18	7	-
25	F	67	37	8	11
85	F	161	39	9	11
73	F	48	22	9	6

Table 6.2: The frequency of the four conditions that satisfied the four criteria for selection of patient records.

ICPC	Main title	#	%
T86	HYPOTHYROIDISM / MYXEDEMA	34	7
K86	UNCOMPLICATED HYPERTENSION	33	6
L88	RHEUMATOID ARTHRITIS / ALLIED COND.	29	6
L95	OSTEOPOROSIS	4	1

picked as selection criteria. Table 6.3 lists the 14 most frequent ICPC diagnoses in the 12 selected patient records.

6.3.1 A Time-Oriented Record System

The patient record system contained different modules that performed a multitude of tasks required in primary-care practice (e.g., administrative, governmental obligations, clinical decision making, etc.). The presentation of modules in Figure 6.1 is not complete with respect to the patient-record system, but we show the ones that provided information relevant to physicians' medical decision-making process; for instance the laboratory module, the drug module, the diagnose module, the administrative module, and the patient-record module.

Categories of Patient-Record Content

We identified a set of categories of clinical information which are listed in Table 6.4; the result is a three-level hierarchy of classes, subclasses, and subsubclasses. We use these categories to refer to clinical information that characterize the information in the patient records.

Table 6.3: The 14 most frequent ICPC diagnoses in the 12 selected patient records.

ICPC	Main title	#	%
P76	DEPRESSIVE DISORDER	37	7
P98	OTHER / UNSPEC.PSYCHOSIS	29	6
L83	SYNDROMES OF CERVICAL SPINE	12	2
L84	OSTEOARTHRITIS OF SPINE	11	2
L92	SHOULDER SYNDROME	10	2
R74	URI (HEAD COLD)	10	2
R83	OTHER INFEC.RESPIR.SYSTEM	10	2
L79	SPRAINS & STRAINS OF OTH JOINTS	8	2
T93	LIPID METABOLISM DISORDER	7	2
N17	VERTIGO / DIZZINESS (*excl.* H82)	6	1
R78	ACUTE BRONCHITIS / BRONCHIOLITIS	6	1
U71	CYSTITIS / OTHER URIN.INFECT.NEC	6	1
W30	POST BIRTH CTRL	6	1
X311	CYTOLOGICAL CLIN.EX. CERVIX	6	1

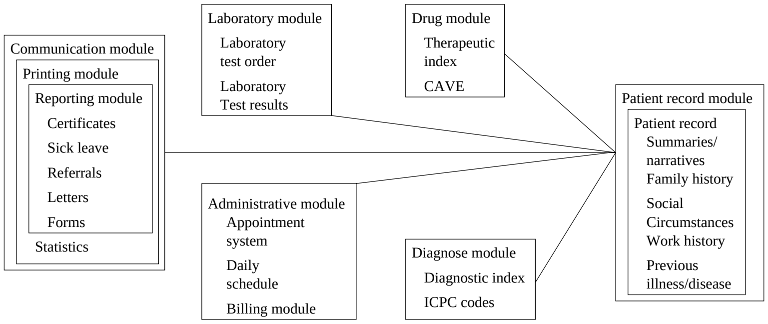


Figure 6.1: Some modules that make up the primary-care patient-record system (e.g., the patient record module, the diagnose module (which uses the ICPC coding scheme), the laboratory module, the drug module, the administrative module, etc.). Each module contains a set of sections, such as *Summaries/narratives*, *Family history*, *Social Circumstances*, *Daily schedule*, *Laboratory test results*, etc.

Table 6.4: Clinical concepts structured as a three-level hierarchy.

Concept	Subconcept	Subsubconcept
CASE HISTORY	PROBLEM STATEMENT	REASON FOR ENCOUNTER
		SYMPTOM
		DESCRIPTION
		COMPLAINT
	FAMILY HISTORY	
	SOCIAL CIRCUMSTANCES	
	PREVIOUS ILLNESS / DISEASE	
	CAVE	
	OBSERVATION	SIGN
		COMMENT
		HYPOTHESIS
		OUTLOOK
ASSESSMENT	CLINICAL EXAMINATION	
	CLINICAL FINDING	
	LABORATORY TEST ORDER	
	TEST RESULT	
	REFERRAL	
	DISCHARGE	
DIAGNOSE	ICPC	
TREATMENT	SURGICAL THERAPY	
	DRUG THERAPY	
	PHYSICAL THERAPY	

A Chronological Patient-Data Log

We chose to focus on the patient-record module because it contained the main patient record. With the main patient record we mean the chronological log of encounter summaries. Besides the main record, the patient-record module contained other *sections* (e.g., list of diagnoses, list of drugs (past and present), list of previous illness and disease, family and work history (i.e., patient demographics)). To access these sections, the user had to exit the main patient record. From now on, when we speak of the patient record we refer to both the main record containing the encounter summaries and the sections of the patient-record module that involve family history, social circumstances, laboratory test results, previous illness and diseases, etc.

Table 6.5 illustrates an extract of the chronological log comprising encounter summaries. All of the patient records that we studied spanned several pages:

- The patient record contained both semi-encoded and free text. Most of the free text described information on patients' SYMPTOMS, COMPLAINTS, PROBLEM DESCRIPTION, and physician's HYPOTHESIS or general COMMENTS, while semi-encoded text described DIAGNOSE, DRUG THERAPY, LABORATORY TEST ORDER, SICK LEAVE, and REFERRAL. In addition:
    - Some parts of the free text had a *de facto* notation. A *de facto* notation was used to separate text belonging to different categories. For example, to

Table 6.5: An extract from the time-oriented patient record.

120396MGH	Sent test to micro lab: Urine cult.
140396MGH	RP: PREDNISOLON Tabl 20mg No: 25 Pck. 11 Patient had an SR of 105. Feels better today, still stiff in shoulder muscles, can barely be touched. Assumes this is a poly myalgia. CE: Abdomen soft and not sensitive, no palpable selections. Hepas bearly palpable under the ribbon, lien not palpable. Mamma: NA, no glands swollen. Start with treatment of Prednisolon 40mg daily for a week. Will then have ctrl, then further evaluation of dosage and pot. further assessment. She uses Albyl-E, but this will seponate until later. She should reduce Prednisolon dosage before she starts with this again. —MYALGIA SHOULDER
150396MGH	Attention high ferritin
190396MGH	Tel from daughter. Patient immediately better after starting with steroids, have both diarrhea and vomiting. Reduce dosage, give also Cimal. She uses Imodium from before. —POLYMYALGIA RHEUMATICA RP: CIMAL Tabl 400mg No:30 RP: PREDNISOLON Tabl 5mg No:200 Pck.11

separate text belonging to CLINICAL EXAMINATION and DISCHARGE summaries from the rest of the text (e.g., a patient's SYMPTOMS, REASON FOR ENCOUNTER, and PROBLEM DESCRIPTION). A *de facto* notation of *Clin.ex and CE* and *EEE, PPP, X-RAY, and MIC* was used for information on categories of CLINICAL EXAMINATION and DISCHARGE summaries, respectively

  - Some of the free text was semi-encoded text. We have called a string of semi-encoded text for a *trace*. A trace represents an action pattern, that is, a trace represents a string of information that reflects an action that has been taken (e.g., a laboratory test order or a drug prescription). Semi-encoded text traces had prefixes such as —, ***, *RP:*, *Lab. test order to*, or *Referral*. The traces that we identified are listed in Table 6.6.

- Information categories such as PATIENT ID, PATIENT NAME and REASON FOR ENCOUNTER was produced in the administrative module

Characteristics of *The Entry*

The patient record was a chronology of encounter summaries, where each entry was signed with a date stamp and initials of a responsible clinicians:

- An entry often summarized multiple problems (Table 6.7). In entries that summarized multiple problems, it was not always clear which part of the free text belonged to which problem. There is no way in the record system to relate free text in the encounter summaries to the various problem that it describes

Table 6.6: Categories of clinical information and corresponding example traces.

Category	Example trace
LABORATORY TEST	*Lab.test order to* hospital *HYPOTHYROIDISM* (the laboratory test results appeared in another section of the patient-record module (Table 6.7))
DRUG PRESCRIPTION	*RP: THYROXIN NA TBL 0.05 mg No:100 Pk. 6 Reit: 3*
DIAGNOSE	*—HEADACHE TENSION*
SICK LEAVE	****SICK LEAVE NOTE for RHEUMATOID ARTHRITIS L88 100*
REFERRAL	*Referral*hospital *for CARCINOMA BASOCELLULARE CUTIS*

Table 6.7: An entry with an encounter summary on multiple problems.

050392AJT	Has used Thyr.Na 0.1mgx1. Feels less heavy in body and freezes more easily. Hypothyr from -89. Clin.ex: Not palp. struma. BP 170/95. puls reg 66, cor-, pulm-. NF- dyspepsia of onion, light fried food, stools early, up to 7 times daily for ca. 10 yrs, last yr. ca. 3 times daily. For several yrs. periodically stomach aches, soft high amylase. prvs chron. pancreatitis. —HYPOTHYROIDISM—DIS. DIGEST. SYSTEM. NOC EXCL. K96 RP: PANCREON COMP. FORTE Tabl. No:200 1-2x3 Lab.test to *hospital* HYPOTHYROID SUBST. Pain and stiff neck and shoulder for sev. yrs periodically. Now spoken myalgia neck muscl. —MYALGIA NECK Refer. phys. ther. MYALGIA NECK 8 x

- Some entries had a particular relation to each other, like the three entries illustrated in Table 6.8. The first entry in Table 6.8 contains a *complete* encounter summary. The two following entries neither contain a PROBLEM DESCRIPTION nor a DIAGNOSE, but their contents are related to the first entry. We refer to entries that relate to each other, like the ones in Table 6.8, as a *session*

- Many entries contained redundant information. Redundant information were especially related to categories such as PREVIOUS ILLNESS / DISEASE and DIAGNOSE (Table 6.9 and Table 6.7). With diagnostic redundancy we mean that the diagnoses varied slightly with respect to the given codes (Table 6.10)

- Physicians seemed to enter text freely into the record system with no adherence to the SOAP sequence. Despite no structure on text entries, we classified the summary information into a set of *core* concepts such as a patient's PROBLEM STATEMENT, CLINICAL FINDING, REFERRAL, LABORATORY TEST ORDER, DIAGNOSE, DRUG THERAPY, etc. Most entries included a diagnose

- Physicians tended to integrate FAMILY HISTORY (e.g., *A good deal of heart disease in family, father dead of heart 62 yrs old* and *A lot of heart disease on father's side, father, brother, sister, uncle.*), WORK HISTORY (e.g., *Will try increase activity, exercise, siml, and diet, ctrl chol*), and SOCIAL CIRCUMSTANCES (e.g., *Conversation about cholesterol*) in the encounter summary, even though, there were dedicated sections for this kind of information

- Physicians sometimes made explicit their thoughts or OBSERVATION, as opposed to actions or a patient's problem statement. Some examples of information categories that do not refer directly to care activities (or actions) are presented in Table 6.11.

Table 6.8: A session with three entries.

040393KGT	Noticed that he was bleeding from a mole on his back for appr. a month ago Clin.ex: Two moles are observed above left scapula. The most medial have a little crust. Unsure if this is just a senil wart. The two moles are removed and the wound is closed with three mattress sutures. The most lateral mole is marked with a thread and sent to histologic examination. The sutures are to be removed in two weeks. —MOLE Histologic examination of: Two moles from the back. NEUVUS
100393KGT	HIST PREP *hospital* 1183 / 93 taken 19930304 (moles): 1: Basal cell papilloma 2: Dermal neuvus. Both benign.
180393KGT	maa: Remove sutures. Looks good.

6.4 Discussion

The patient records that we studied were time oriented. Hence the main record is a chronological log of successive encounter summaries. The chronological log contains

Table 6.9: An entry demonstrating information redundancy with respect to PREVIOUS ILLNESS / DISEASE (compare with Table 6.7).

131291ATJ	Hypertension treatment for several years, tried without medication for a period this spring, but was suffering from headaches and started again with TERNORMIN. BP 160/90 puls regular 62. Indicates a little dizziness, particularly when dark. Ophthalmoscopy - no papilledema, bleeding or exudates, some minor blood vessel alterations. —HYPERTENSION ESSENTIAL NOS —HYPOTHYROIDISM RP: THYROXIN NA TBL 0.05 mg No:100 Pck.6 Reit:3 Pain left side abdominal-thorax evenings after going to bed, increasing last half year, particularly after meals. Pressure.Variable dyspepsia for years. Clin.ex.: Cor-pulm NA. Abd: BZ feces NA —STOMACH ACHE UNSPECIFIED Lab.test sent to hospital STOMACH ACHE UNSPECIFIED

Table 6.10: Variations in diagnoses given to a patient.

Diagnose description	Variations in given diagnoses and descriptions
Myalgia neck / shoulder	Stiff neck
	Tension headache
	Headache
	Shoulder syndrome
	Tendinitis shoulder
	Myalgia neck
	Myalgia shoulder

Table 6.11: Categories of clinical information that are not directly related to activities or a patient's problem statement.

Category	Example
HYPOTHESIS	*Assume this descend from muscle-sinew apparatus, but also evaluation from physiotherapist. Will in addition try NSAIDs for two weeks*, *Assume this is about a poly myalgia rheumatica*, and *Will now try suspect ischemia*
COMMENT	*Looks good* and *Pas. had an SR of 105*
OUTLOOK	*Will start treatment with Prednisolon*
HYPOTHESIS	*Unsure if this is just a senil wart. The two moles are removed and the wound is closed with three mattress sutures*
OUTLOOK	*Start with treatment of Prednisolon 40mg daily for a week. Will then have ctrl, then further evaluation of dosage and pot. further assessment. She uses Albyl-E, but this will seponate until later. She should reduce Prednisolon dosage before she starts with this again.*
PREVIOUS ILLNESS / DISEASE	*Has RA since 1988*, *Hypothyr from -89*, and *Hypertension treatment for several years*

episodic-oriented entries. Entries appear as separate units in the record and there are seldom signs to interrelationships among entries, even though, they are related.

An entry is the basic unit in the record and summarizes the clinical encounter, where each entry is signed by a time stamp and initials of a responsible clinician. The patient records usually comprise several pages of text; they are verbose, rich, and extensive in words.

We have identified a set of concepts that categorize and represent the patient-record content (Table 6.4). These concepts build and extend on four main concepts which are CASE HISTORY, ASSESSMENT, DIAGNOSE, and TREATMENT. In Table 6.12 we have demonstrated the use of these concepts on an encounter summary. The encounter-summary text is broken down into information units that are labeled with a clinical concepts from Table 6.4.

Table 6.12: Clinical concepts demonstrated in use on an encounter summary.

Information unit	Clinical concept
Hypertension treatment for several years,	PREVIOUS ILLNESS / DISEASE
tried without medication for a period this spring,	DESCRIPTION
but was suffering from headaches	SYMPTOM
and started again with TERNORMIN.	COMMENT / DRUG THERAPY
BP 160 / 90 puls regular 62.	CLINICAL EXAMINATION
	CLINICAL FINDING
Indicates a little dizziness,	SYMPTOM
particularly when dark.	DESCRIPTION
Ophthalmoscopy - no papilledema, bleeding or exudates, some minor blood vessel alterations.	CLINICAL EXAMINATION
	CLINICAL FINDING
—HYPERTENSION ESSENTIAL NOS	DIAGNOSE (ICPC)
—HYPOTHYROIDISM	DIAGNOSE
RP: THYROXIN NA TBL 0.05 mg No:100 Pck.6 Reit:3	MEDICAL THERAPY
Pain left side abdominal-thorax	SYMPTOM
evenings after going to bed, increasing last half year, particularly after meals.	DESCRIPTION
Pressure.	SYMPTOM
Variable dyspepsia for years.	PREVIOUS ILLNESS / DISEASE
Clin.ex.: Cor-pulm NA. Abd: BZ feces NA	CLINICAL EXAMINATION
	CLINICAL FINDING
—STOMACH ACHE UNSPECIFIED	DIAGNOSE
Lab.test sent to hospital STOMACH ACHE UNSPECIFIED	LABORATORY TEST ORDER
HB 14.5; SR 18; AVF 6 neg; USTIX abs-; K 4.0; KREAT 90; BILTO 12; ALP 179; ALAT 24; ASAT 24; LD 358; G-GT 26; AMYLIA *306; FT4 *6; TSH *34.1	LABORATORY TEST RESULTS

6.4.1 Limitations of Methods

We have studied patient records from one patient-record system. As a consequence our findings in this study may be biased toward idiosyncrasies related to the specific primary-care clinic that hosts the patient-record system, including the patient records that we studied. We assume that each clinic carry with it a culture, a set of habits, and routines that influence the recording of patient information. Maybe this clinic has a tradition of writing longer narratives than other clinics? Maybe this clinic have a habit of writing orderly and detailed summaries that are not representative for primary-care in Norway? Our findings are only based on data from one primary-care clinic in Norway. How do we know if our findings are based on wrong assumptions?

However, from our observation study we have already confirmed that physicians spend a significant time of the encounter in writing long narratives. This was observed in five different primary-care clinics and one specialty-care clinic. In addition, we intend to perform a quantitative survey among family physicians about their own opinions regarding their use and the usefulness of the patient-record system. Based on these results we may be able to generalize from a quantitative amount of data that represent family physicians in Norwegian primary care.

6.4.2 A Set of Clinical Concepts

Our concept hierarchy starts with four concepts; CASE HISTORY, ASSESSMENT, DIAGNOSTICS, and TREATMENT. The four concepts divide the primary-care process into four phases, where CASE HISTORY functions as input to a set of activities in the care process (e.g., ASSESSMENT, DIAGNOSTICS, and TREATMENT). This is not a novel way to divide the primary-care process—each problem is either a diagnostic problem (i.e., the physician tries to determine what caused the problem) or a management problem (i.e., the physician knows the cause, but he or she tries to choose the best option for treatment) [195]. Instead of having two major phases, we separate diagnostics from assessment (or vice versa), causing the primary-care process to be split into three phases.

Separation between Observation and Interpretation

A record system that can separate between CASE HISTORY and the various activities based on the CASE HISTORY, is able to distinguish *interpretations* from *observations*. An interpretation leads to a set of actions. As there may exist many interpretations based on the observations, we would like to keep the interpretations apart.

The value of keeping interpretations separate (and also the corresponding set of actions) become evident when we want to say something about what has been done (i.e., the set of actions) and what the nature of the problem is (i.e., the set of symptoms and signs in the case history). “In many dynamic, uncertain, and fast-paced environments, there is no single right way to make decisions” [195].

Physicians generate several hypotheses based on a patient’s case history [68, 98], but they select only one to pursue at a time. Kluge [86] describes that a diagnosis is based on a physicians particular interpretation of the data in the patient record: “a diagnosis may be interpreted as a path taken by a health care professional with a certain ... patient profile ... of stored data about a specific patient”. A set of actions defines a path, a path is guided by an interpretation, and an interpretation depends on each individual physician.

If a hypothesis (and the corresponding interpretation) leads to a dead end, the physician needs go back to the initial CASE HISTORY. In these occasions, to go back implies to go back to the initial CASE HISTORY (and possible newer history added to it) and start with another hypothesis and corresponding interpretation. For this reason, we want to be able to distinguish between input to physicians' care process and the actions done based the input.

Comparing Different Sets of Clinical Concepts

The set of clinical concepts we have listed in Table 6.4 bears resemblance to various other projects that also aim to represent patient-record content.

The *Clinical Headings* project from the National Health Services (NHS) defined a set of headings for communicating clinical information (Section 4.2). To communicate implies that the sender and the receiver have a common understanding of what to communicate. In the same way as with clinical headings, the clinical concepts in Table 6.4 are categories that reference and classify patient-record information.

Health Level 7 (HL7) Reference Information Model (RIM) defines a set of *Subject Areas* (Section 4.6.2). A subject area is defined through a set of properties such as participating roles and actions (i.e., acts). In the same way as with HL7 RIM, we can specify participants, roles, activity, and purpose for each clinical concept.

Other projects such as [62, 116, 135], have focused on developing concepts that represent the content of patient records. These projects have merely focused on the hospital records and bear little resemblance with the concepts which we have identified for the primary-care patient record. An exception to these projects is the GALEN project, that has built a Terminology Server for the purpose of encoding content in hospital patient records (Section 4.2). GALEN has developed a model of medical concepts and a set of rules to combine these concepts into composite medical concepts. In this way, GALEN can represent the content of medical records by using a description-logic language—GRAIL—which the GALEN project implemented for their own purpose. In comparison with GALEN, our approach is more top-down, in the sense that we can represent the content in terms of a sentence, part of a sentence, or several sentences. However, we cannot represent the content in each sentence, for that we need a more detailed formalism. Our approach represents a middle way between a complete formalism that represents fully the content in medical records and the clinical-headings approach that has no formalism at all. With our approach we can represent the content of patient records in terms of clinical concepts.

6.4.3 Lack of Overview

Two of our findings indicate that physicians lack overview over content in patient records:

- Redundant information in the patient record, especially redundant information on categories related to PREVIOUS ILLNESS / DISEASE and (2) DIAGNOSE. The redundant information tell us what type of information physicians need because they have been made explicit. The presence of redundant information proves that physicians lack overview of what has been asked and done in the past. The information they need is there, but they do not have the time or the means to retrieve the information they want and need

- Physicians developed their own *de facto* notation to distinguish among various types of information (e.g., clinical findings and discharges summaries). The *de facto* notation is an attempt to bring structure to the patient record, a structure that the record system did not provide.

Physicians rely on a record structure to have overview. The episodic-oriented entries are in contrast to what physicians have said before (Chapter 6.4.3)—that part of having overview is *knowing what physicians have asked the patients and what has been done to them*. The redundant information indicates that physicians do not have the time to search and find the information when they need it. The *de facto* notation demonstrates that physicians need a structure, for them to make use of the information that is in the patient records.

6.4.4 No Strict SOAP Structure

After classifying information units in the record entries with our clinical concepts, we did not find any common order when compared with the SOAP structure. Some clinical-concept categories were found in a majority of the encounter summaries—*core* information such as PROBLEM STATEMENT, CLINICAL FINDINGS, REFERRALS, LABORATORY TEST ORDER, and DRUG PRESCRIPTIONS.

That we did not find any resemblance to the SOAP format may suggest that physicians did not have a common way of working (i.e., a common sequential way of thinking), in terms of writing information into the patient record. But since we found a set of core information in a majority of the encounter summaries, this core information proves that physicians operate within a common context that can be represented by a set of clinical concepts.

Of the categories in SOAP format (Section 6.1) that were least present in the patient records we studied, were categories such as PLAN. However, our category TREATMENT is part of the category PLAN in SOAP, which proves that very often, information about DRUG THERAPY is sufficient as PLAN information.

6.4.5 To Work Within a Context

In addition to core information, such as PROBLEM STATEMENT, CLINICAL FINDINGS, REFERRALS, LABORATORY TEST ORDER, and DRUG PRESCRIPTIONS, several entries also contained information about FAMILY HISTORY and SOCIAL CIRCUMSTANCES (which includes work history). FAMILY HISTORY and SOCIAL CIRCUMSTANCES had their own sections in the record system, but physicians integrated this information in the encounter summary instead of using the dedicated sections for these categories of information.

Potential reasons for why physicians did not use the reserved sections to sort FAMILY HISTORY and SOCIAL CIRCUMSTANCES, are: (1) they did not want to exit from the patient record, but wanted to operate within one context, and (2) they did not feel like sorting their data in a structured manner. For a physician to move away from the context that he or she works within, seems like an unnatural way to work. Physician should have information available within the context they operate within. It is worth noting that the two categories, FAMILY HISTORY and SOCIAL CIRCUMSTANCES, do not belong naturally to any of the categories in the SOAP structure.

6.5 Conclusion

Based on what family physicians write into the record system, we have identified a set of concepts that classify and represents the patient-record content. The categories of concepts make up a hierarchy of classes, subclasses, and subsubclasses. Four main classes jut out from a root concept: (1) CASE HISTORY, (2) ASSESSMENT, (3) DIAGNOSE, and (4) TREATMENT. Information in the patient record did not follow the structure of the SOAP format; actually the encounter summaries did not have any regular order at all. However, most entries contained a set of core information related to concepts, such as patients' PROBLEM STATEMENT, CLINICAL EXAMINATION FINDINGS, LABORATORY TEST ORDER, DIAGNOSE, DRUG PRESCRIPTION, etc. The core information proves that physicians operate within a common context of information that can be represented by a set of clinical concepts. That we did not find any common sequence in the recorded information proves that physicians do not think and work in terms of a strict SOAP structure.

Family physicians enter information into the record as free text, in which they write extensive and verbose encounter summaries. The patient record is a chronological log of encounters. The entry represented the main unit of information. Even though entries were related to each other, in terms of a follow up or continuing encounter summary, most entries appeared as stand-alone encounter summaries. As a result of the episodic-oriented nature of the encounter summaries, redundant information was entered into the record. Physicians tended to repeat already known and recorded case history. The episodic-oriented entries seems to be in opposition to what physicians want—easy access to existing information and an overview of what physicians have previously asked and done in relation to the patients.

Chapter 7

Physicians' Perspectives on Patient-Record Systems

7.1 Introduction

This chapter reports on a survey of family physicians' personal opinions regarding the use of the patient-record system and its usefulness in patient care.

We have previously conducted two other surveys; one observation study of the primary-care process (Chapter 5) and one study of the contents of patient records (Chapter 6). From the observation study we learned that the computer-based patient-record system is an integrated tool in family physicians' everyday work. They both write and retrieve information from the patient record; where physicians are both information owners and information users. Thus family physicians are highly motivated to maintain the information in the patient records. From our study of the patient-record content, we learned that physicians write extensive encounter summaries. Family physicians are willing to spend a great deal of time documenting patient-specific information in their patient records. At this point, we would like to find out about family physicians own opinions regarding their use of the computer-based patient-record system, in terms of its usefulness during patient care, how they would like to characterize their patient-record systems, and if there are features that are missing or they would like to see improved with existing patient-record systems.

The survey will serve two important purposes: (1) to learn about physicians' own opinions regarding the use of record systems in everyday patient care; and (2) as part of building knowledge models for a computer-based patient-record system, knowledge elicitation deals with getting data and material needed for modeling. Surveys and structured interviews represent formal ways to extract usable knowledge, which then can provide us with useful information in the later phases of knowledge identification [167].

We wish to determine the record system's impact on patient care. With a system's impact on patient care, we mean to investigate if patient care is positively influenced by the patient-record system in providing physicians with useful and needed information during patient care; in such a way that physicians are provided with relevant information to their clinical decision-making process.

Our study comprises four specific areas related to family physicians use of the patient-record system: (1) do they consider the patient-record system as useful (i.e., what do they need and what do they get from the record system)?, (2) what do they like and what do they miss with existing patient-records systems, (3) what impact do they think the computer-based patient-record system has on them during the clinical encounter (i.e., do they feel tied up, restrained, or hampered by the structure of the system)? (4) what impact do they believe the record system has on clinical patient care, now and in the future?

Several of the questions in this survey are similar to those in other surveys found in the literature (for instance [57, 64, 78, 79, 89, 174]). However, performing a separate survey will supply us with idiosyncrasies related to Norwegian primary-care conditions, which are important for us to know about when we seek to improve the patient-record system for future use. "Constructing a feasible, workable decision support tool always implies that *the local context is built into the technique*. Inevitably, idiosyncratic, unique features of the specific site involved become embedded in the tool's script" [21].

7.2 Background

This survey is conducted in two phases. One phase has dealt with the preparation and performance of the survey, while the other phase has dealt with the data analysis. For part of our survey we have used a group of 4th year master students at the Norwegian University of Science and Technology (NTNU) in Trondheim. The students came from different departments, such as computer science, telecommunication, mathematics, economy and administration, bio-physics, and bio-medicine.

The students contributed to the survey through a course *Team of Experts* and the *DocVille* project (Section 7.2.1). In the course the students were responsible for examining various survey techniques, including the conjoint-analysis technique which is particular useful for investigating system properties and measure them against each other; developing a web portal, to publish project information and survey results on the web; making the questionnaires, after investigating various techniques; selecting tools for data analysis; and gathering necessary background literature and data for performing the survey. The students had no prior knowledge on research related to patient-record systems in Norwegian primary care.

7.2.1 *Team of Experts*

Team of experts[1] was a new and mandatory course, from spring semester 2001, for master-engineering students at NTNU. The course unit aims to train students through a problem-oriented and project-based approach, where they have to contribute with different ways to approach and solve problems. To make students more aware of their future role as team-players, they learn to solve multi- and inter-disciplinary problems, where each member of the team contributes with expertise from his or her engineering background.

---

[1] We refer to the course *Team of Experts* as *Experts in Team* (Eit) to match the Norwegian convention.

A *village* works on a specific subject, given by an employer, and comprises a set of expert teams (i.e., student teams). The expert teams are responsible for defining specific tasks within the overall subject of the project and organizing and managing themselves in the village. Each village has a facilitator available during the semester. The facilitator is not to involve directly in the students' process, that is, the facilitator is not to take any responsibility in the students' learning process. The facilitator is to assist and guide the students indirectly by asking open questions that the students themselves have to answer.

In this project, therefore, the role of the facilitator has been to try and influence the students in such a way that the content of the questionnaire would satisfy a certain level of quality, make sure the students reviewed background literature, and make sure that all important issues related to the survey were kept intact. However, as a digression, we experienced that the students overtook completely their own responsibility, as soon as they understood that we, as facilitators, would not help them in this regard.

DocVille

We invented a name for our village—*DocVille*. *DocVille* consisted of 20 students. The village's superordinate task was to conduct a survey among Norwegian family physicians. After 14 weeks, the students delivered a report; *Better computer-based patient records for primary care: a survey of use, needs, and barriers* [29].

At the start of the semester in January 2001, the students were given a description of the project subject—they were to answer a set of hypotheses regarding family physicians' use of computer-based patient-record system—with requirements to conduct a quantitative survey in order to answer the given hypotheses. Another requirement was that part of the survey had to include a conjoint analysis (conjoint analysis is described in the following subsection) .

The three hypothesis touched three areas of importance related to the computer-based patient-record system: (1) family physicians use of the patient-record system, (2) family physicians information needs during patient care, and (3) family physicians relationship to technology and the impact technology will have on patient care in the future.

Conjoint Analysis

Conjoint analysis (CJA) is an alternative to conventional survey-based research approaches. Conjoint analysis originated from mathematical psychology, where it had a strong theoretical foundation and proved to be a powerful analytical tool for surveying applications, studying, and eliciting individual preferences or judgments [38].

The method has gained widespread acceptance and use in health care, where it has been applied in several areas: eliciting patients' preferences in delivery of health services (e.g., optimizing an obsterical unit [67]), developing outcome measures, determining optimal treatments for patients [102, 103], evaluating alternatives within randomized controlled trials [63,81], and establishing patient's preferences in the physician-patient relationship. Conjoint analysis also has been used successfully in other domains such as market research and transport and environmental economics [162].

Conjoint analysis assumes that we can decompose a product in such a way that the product can be viewed as a bundle of attributes that combine into a set of scenarios or profiles. The conjoint-analysis method demonstrate peoples' willingness to trade among these attributes (characteristics), where it estimates the relative importance of different attributes. The meaningfulness of an attribute depends on other attributes' values. Attributes must have mutually exclusive values.

The number of scenarios increase with the number of characteristics and levels, only rarely can a questionnaire include all generated scenarios. Analysis (regression techniques) of results give utility values; the higher the value, the higher the utility. A high relative importance (the difference between maximum and minimum utility value) indicates respondent's sensitivity for change in attribute values. The most liked profile has the highest total utility value (the sum of utility values for each attributes). Because of the additive nature of these satisfaction values, we can play the *what if* game with them [67]. For example, *what if we combine time spent on entering data into the record system with a knowledge-based system; which of the two attributes are more important to the physicians?*

7.3 Materials and Methods

We chose to do two surveys, employing both quantitative and qualitative techniques. The quantitative survey contained two parts, where the first part included a conventional survey (i.e., a questionnaire), and the second part included a conjoint analysis. For the qualitative survey, we chose to do interviews over the telephone. In the telephone interviews we could go into depth on physicians' potential interesting and surprising answers.

Both surveys (quantitative and qualitative) had a common set of background questions with respect to age, gender, place of education, years of practice, type of clinic, type of patient-record system, etc.

7.3.1 A Telephone-Interview Guide

To assist in quality assurance of the interview guide we consulted NTNU's faculty staff from the Faculty of Medicine [2] Two pilot studies were performed, after which we estimated 30 minutes for every interview. The interview guide contained three parts (in addition to general background questions):

- Questions related to physicians' use of the computer-based patient record. For example, *What do you like about the computer-based patient-record system that you use?*; *Are there any tasks that were easier to carry out with the paper-based patient-record system?*; and *Do you feel that the record system restrains you while you work?*

---

[2] Department of Community Medicine and General Practice.

- Questions related to physicians' information use and needs. For example, *What information do you search for before an encounter?* and *Are there any functions that you would like to have that you do not presently have available?*
- Questions related to future perspectives on technology and the impact technology may have on patient care. For example, *How do you agree to the statement that "Better patient care has nothing to do with better patient-record systems"?* and *If you were to design a new patient-record system, what would be your focus in the new system?*

7.3.2 A Conventional Survey in Two Parts

The conventional survey had three sections: (1) a general background-question section, (2) a (conventional) questionnaire section, and (3) a conjoint-analysis section.

Table 7.1: Attributes and attribute values for conjoint analysis.

Attributes and attribute values	Explanation of attribute value
CPR time usage	
Very time consuming	
Time consuming	
Not time consuming	
Structure of CPR[a]	
Paper metaphor	A CPR system that is based on the paper metaphor
Problem orientation	A new structure for a CPR system, with a new way of presenting information and a new way of using it. Upon data entry, patient data will be related to sources and medical problems, in addition to date and time of entry
Patient data entry	
Keyboard and mouse	The traditional way of using keyboard and mouse for data entry and retrieval
Speech recognition	The users' voice will be recognized by the computer and it will automatically convert what the user says into text
Digital notebook	The user uses a pen for data entry as with traditional writing on paper
CPR as a knowledge system	
Includes a knowledge system	A knowledge system that may the physician with advice on assessment and treatment
Without a knowledge system	The current CPR system

---

[a] Computer-Based Patient-Record System (CPR).

Conjoint Analysis Profiles

Table 7.1 lists attributes, corresponding attribute values, and a description of the attribute values that we ended up using in the conjoint analysis profiles. The attributes reflect four properties of the record system that we saw as critical or valuable to measure up against each other; (1) time spent on using the record system, (2) alternative structures of the time-oriented patient record, (3) other ways to enter data into the record system, and (4) a possible integration of external knowledge sources with the patient-record system. With conjoint-analysis profiles we can find out which of these attributes physicians’ value the most when compared to each other.

We combined the attributes into nine different profiles (Table 7.2). Respondents had to rank the profiles from 1 to 8. One denoted a high degree of *disliking*, while eight denoted a high degree of *liking* (Table 7.3).

Table 7.2: Four of the nine conjoint-analysis profiles.

Alternative 1	
CPR[a] *time usage*	Not time consuming
Structure of CPR	Problem orientation
Patient data entry	Digital notebook
CPR *as a knowledge system*	Without a knowledge system
Ranking	1 2 3 4 5 6 7 8
Alternative 4	
CPR *time usage*	Time consuming
Structure of CPR	Paper metaphor
Patient data entry	Keyboard and mouse
CPR *as a knowledge system*	Includes a knowledge system
Ranking	1 2 3 4 5 6 7 8
Alternative 5	
CPR *time usage*	Time consuming
Structure of CPR	Problem orientation
Patient data entry	Keyboard and mouse
CPR *as a knowledge system*	Includes a knowledge system
Ranking	1 2 3 4 5 6 7 8
Alternative 9	
CPR *time usage*	Very time consuming
Structure of CPR	Problem orientation
Patient data entry	Keyboard and mouse
CPR *as a knowledge system*	Without a knowledge system
Ranking	1 2 3 4 5 6 7 8

---

[a] Computer-Based Patient-Record System (CPR).

The Questionnaire

Without counting the background questions, the conventional survey involved 30 questions. There were two types of questions, single questions and composite questions. With single questions, physicians had to rate their answers from one to five,

Table 7.3: Ranking scale for conjoint-analysis profiles.

I do not like they system very much									*very much*
Ranking:	1	2	3	4	5	$\boxed{6}$	7	8	

where one was *totally disagree* and five was *totally agree*. The composite questions had a rating scale from one to five, where one was *to a small extent* and five was *to a large extent*.

The number of composite questions ranged from four to five questions for each block. The questionnaire had five blocks of composite questions and nine single questions. The single and composite questions were mixed with regard to topics, so both single and composite questions touched the three topics: (1) physicians' opinions on their use and the usefulness of the patient-record system, (2) information use, and (3) information needs. The first group also had some questions related to the impact on the patient-record system on patient care, now and in the future.

7.3.3 Phone-Call and Survey-Distribution Procedure

The study was conducted during March and April 2001. A paper-based survey was distributed to 306 family physicians from all regions of Norway. The survey could also be completed electronically on a web portal designed specifically for the project (Section 7.2.1). Each respondent received a unique number and a password for logging onto the web portal. Respondents could start a session and continue a session whenever they wanted, just by logging onto the portal with the given username and password.

We used a random selection from a list of primary-care clinics provided by the Norwegian Medical Association[3]. This random selection of respondents included both the conventional survey and the telephone-interview survey.

It proved a tedious and time consuming task to make family physicians participate, especially in the telephone survey. In addition to spending a lot of time getting a hold of physicians during the day, just to ask if they would participate in a telephone interview, not many physicians had the time to answer a 30-minute interview. We, therefore, announced an invitation and a request for participation on the EYR mailing list (EYR is described in the following subsection) to increase the response rates in both surveys. We gave two written reminders and one electronic on the EYR mailing list.

To distinguish EYR respondents from the initial ones (the randomly selected ones) EYR respondents were given a universal username and password different (as opposed to personal ones that were given to the originally and randomly selected population). Implications of having respondents from two different populations will be discussed later (Section 7.5.2).

---

[3] Den norske lægeforening.

EYR

The Norwegian College of General Practitioners (NSAM) has a mailing list—EYR[4]—for Norwegian primary-care physicians. Topics of discussion on the list vary between technical questions, advice on medical practices, general exchanges on different views or opinions, announcements, etc. Members of EYR express an interest in the use of Internet and on the impacts technology may have on the primary-care profession. The list has been active since 1996 and holds 600 members (in 2001 [164]).

7.3.4 Analysis

To analyze data from the quantitative survey, we used SPSS Standard Version Release 11.0.0. For comparing nonparametric measures for bivariate correlations we used Spearman's rho. For nonparametric, two-independent samples (e.g., age) test we used Mann-Whitney U test. For comparison between interval-scaled variables such as age, we used independent two-sample t-test.

Data from the qualitative surveys (i.e., the telephone interviews) are organized into tables (the results we consider as the most important ones). Appendix A contains a glossary over terminology used in this chapter.

7.4 Results

7.4.1 Survey Response Rates

We performed 21 interviews over the telephone. Three of the respondents agreed on an interview based on our initial request, while the remaining 17 volunteered after we announced a request for participation on EYR.

Of the 306 physicians that received a paper-based survey 115 responded (i.e., a response rate of 38%), either through the web or by completing the paper form. After the announcement on EYR, we received another 52 web respondents, increasing the response rate to 58%.

There were minor differences, with regard to the background questions, among the telephone respondents, the conventional survey respondents (including the conjoint analysis), and those who responded after we announced on EYR (Table 7.4). We chose to combine the EYR responses with other responses on both surveys.

7.4.2 Self-Contradictory Answers over the Telephone

Many of the answers given in the telephone interview proved to be contradictory and it was hard for us to make any generalizations. What one physician reported as a positive feature (e.g., *The system gives good overview*, *The system gives good documentation when compared to paper*, *Encounter summaries are better and longer*), another reported as a negative feature (e.g., *The system makes it difficult to get overview*, *Overview was better on*

---

[4] In Old Norse mythology EYR was the goddess for medicine. EYR was also the name of the first Norwegian medical journal (1826-37).

Table 7.4: General-background statistics for the telephone interviews and the conventional survey (including EYR respondents).

	Telephone n=21	Conventional n=177
Age		
Years (mean)	44.0	43.3
Gender		
Female	34%	30%
Place of graduation		
Tromsø	16%	8%
Bergen	21%	23%
Oslo	37%	34%
Trondheim	16%	11%
Other	10%	24%
Record System		
ProfdocD	14%	31%
ProfdocW	10%	13%
WindMed	38%	37%
InfodocD	5%	5
InfodocW	24%	15%
Other	4.5%	2%
Paper	4.55%	1%

paper, *Difficult to find data because the record is too long*) of the record system. However, we have summarized some of the answers in Table 7.5 and Table 7.6.

Table 7.5 lists what we have called a *wish list* of what family physicians expressed they wanted improved in the current computer-based patient-record systems. The order in the list is random (i.e., number one is not rated as more important than number 11). Among the more interesting requests—because these are issues that we address in our research work—we can mention number one, two, and 10, which brings forward issues related to the need for a better overview, problem orientation, and integration with external knowledge sources, respectively.

Table 7.6 lists physicians' use and needs for information from the patient-record system during patient care. The physicians responded that they used and needed the record system during all phases of the encounter.

7.4.3 Quantitative Opinions on the Patient-Record Systems

The results from the conventional and conjoint analysis gave us more insight into how family physicians view their patient-record system and what impact they believe the record systems have on patient care.

Positive Opinions on the Record System

77% of the physicians responded that they felt they managed to use the record-system functions. 85 % of the physicians considered existing patient-record systems convenient to use. 84% disagreed when asked if the record system restricted them during

Table 7.5: Norwegian family physicians' wish list on future patient-record systems.

No	Requests for improvement in future patient-record systems
1	A more user friendly system, the current record is too long and difficult to follow
2	Problem and relation oriented
3	Screen is too little
4	More well-developed help functions
5	Certificates should already be in the CPR[a] and printable on laser
6	All outpost (e.g., letters, referrals, etc.) should be kept in the record
7	Rich set of filters for individual or patients-specific data (e.g., diabetes-related encounters last two years, last years encounters on anti-coagulant patients not concerning TT[b]; should also be possible to filter out examinations with patients that have cardiac insufficiencies and get answers to *what am I overlooking/not seeing/what do I normally do?*
8	Continuous generation of hyptertext (words interpreted semantically and marked and connected (as a hyperlink) to a network of databases / internet pages etc. (e.g., *Patient has a worsening of his asthma*—click asthma and chose)
9	To be able to make family links (when handling patients in a family)
10	Pedagogical tools (e.g., anatomic atlas, images), standardized patient information (e.g., dietary lists), evidence-based medicine
11	Digital communication with other clinics, hospitals, and emergency wards (e.g., discharges, resume, x-ray results directly into the record instead via paper as the case is at present)

---

[a] Computer-Based Patient-Record System (CPR).

[b] TT is an acronym for Trombo Test. The test is similar to International Normalized Ratio (INR).

Table 7.6: Physicians' use and needs for information from the patient-record system during patient care.

Pre encounter	Encounter	Post encounter
Search for information that give overview over patient's condition(s)	Read old encounter summaries	Complete the encounter summary if necessary
Find out if patient has been here before	Read patient's history	Complete unfinished referrals, requisitions, and conclusions
Find out how many times the patient has visited	Write messages regarding plan	Do patient-administrative tasks
Read last encounter summary	Do patient administration (e.g., print sick leaves and other applications)	Write letters to patient (e.g., laboratory results)
Search for family history	Search drug information and prescribe drug	Print drug prescription
Search for history	Search for laboratory samples and other examinations	
Read reason for encounter	Print laboratory test order	
Search for previous disease / illness	Write referrals	
Search for current medication	Perform financial issues	

patient care. 12 % of the physicians answered that they felt restrained by using a record system and that they did not handle it. In general, male physicians felt that they mastered the computer-based patient-record system better than female physicians did.

An Impact on Patient Care

Around 79% of the physicians believed that the patient-record system had a positive impact on patient care and 51% answered that they used the record system more when patients came frequently (with the same medical problem). Male physicians, more than female, felt that their skills in mastering the patient-record system had impact on patient care.

I manage completely the use of existing patient-record systems

Combining age, experience with the record system, and Question 1, the answers showed (Table 7.7) that those with more experience claimed they managed the record system better than those will less experience ($p<.001$). There is a significant difference in opinions between the genders; male physicians agree more when asked if they manage completely their patient-record systems than female physician do (p=.026, n=172, Mann-Whitney; means: 3.71 for female physicians and 4.01 for male physicians).

Table 7.7: Age and experience correlated (Spearman's rho) with Question 1:

I manage completely the use of CPR[a]	
Age	Experience with use of CPR system
.139	.293
p=.069	p=.000[b]
n=171	n=171

[a] Computer-Based Patient-Record System (CPR).

[b] Correlation is significant at the .01 level (2-tailed).

I have no objections in learning a new patient-record system

Combining age, experience with the record system, and Question 6, the answers showed (Table 7.8) that younger physicians have less objections in learning a new patient-record system than their elder colleagues have ($p<.01$). However, there is no significant difference in opinions between the genders when it comes to learning new patient record systems, both genders agree positively (p=.633, n=173, Mann-Whitney; means: 3.77 for female physicians and 3.90 for male physicians).

An improved system will increase my efficiency and improve patient care

Combining age, experience with the record systems, and Question 9, the answers showed (Table 7.9) that there is no significant difference in age and experience when physicians evaluate the impact of the record system on patient care. There was a slight tendency among the younger physicians to believe the record system had a positive

Table 7.8: Age and experience correlated (Spearman's rho) with Question 6:

I have no objections in learning a new CPR[a]

Age	Experience with use of CPR system
-.216	-.086
p=.004[b]	p=.264
n=172	n=172

---

[a] Computer-Based Patient-Record System (CPR).

[b] Correlation is significant at the .01 level (2-tailed).

impact on patient care (p=.090). However, there is no significant difference in opinions between the genders when it comes to evaluating the record systems' impact on efficiency and patient care, both genders agree positively (p=.396, n=172, Mann-Whitney; means: 3.50 for female physicians and 3.64 for male physicians).

Table 7.9: Age and experience correlated (Spearman's rho) with Question 9:

An improved CPR[a] *will increase my efficiency and improve patient care*

Age	Experience with use of CPR system
-.130	-.082
p=.090	p=.288
n=171	n=171

---

[a] Computer-Based Patient-Record System (CPR).

Adopting to New Technology

Regarding questions on learning and adopting a new technology, 67% uttered they would not mind, while 17% indicated they had some kind of inveighs against learning a new system. 16% indicated neither positive nor negative attitude. However, 55% of physicians believed they needed more time and education to adopt to new systems than compared to other occupation groups.

When asked if they thought the system contained bugs or lacked functionality, 25% disagreed, 23% had no opinion on the matter, and 52% believed so.

Genders and Communication

On all the questions (from 1 to 30), except number 17 and 18, female physicians' answers tended to be on the left side of the scale, while male physicians' answers tended to be on the right side of the scale. Thus female physicians **disagreed** more to questions and statements than their male counterparts.

Questions 17 and 18 have to do with information entry and retrieval during the clinical encounter: *When do you enter information into the record?* and *When do you seek information from the record?*, respectively. On question 17 and 18, female physicians **agreed** more than male physicians. Female physicians answered that they searched for information *before* the patient came into the office and wrote information into the record system *after* the patient had left the office. Male physicians distributed their information entry and search over all phases of the patient-care process (i.e., pre-encounter, *in situ* encounter (including anamnesis, assessment, diagnostics) and post-encounter).

Answers to question 17 and 18 also had the largest mean difference between the genders than the rest of the answers between male and female physicians (.58 and .51, respectively).

Information Use and Needs

Both female and male physicians gave high ranks to the answers on question 22-30, which had to do with what information they used mostly, and to what extent they believed medical knowledge, knowledge about the record system, abilities to communicate, and how qualities of a good record system can influence the quality of patient care.

In general, on questions regarding information use, the responses indicated that they used and entered patient-record information *to a large extent* (i.e., means > 4.7 for both genders) during patient care. 46% entered information during case history (anamnesis), 38% entered during assessment, 43% during diagnostics, and 57% when the patient had left the office. 55% sought information before the patient had entered the office, 43% sought information during case history, 47% during assessment, and 40% during diagnostics.

To the question *What kind of information do you use mostly?*, 64% responded previous disease and illness. 85% used information from the main patient record (the chronological log); 42% used previous diagnoses; 69% used previous and current drug prescriptions and referrals (discharges); and 71% used laboratory test results.

7.4.4 Time versus Structure

In the conjoint analysis we compared four critical attributes against each other. These were: (1) time spent using the record system, (2) structure of the patient record, (3) ways to enter information, and (4) integration of a knowledge-based system. The most critical attribute was the time spent using the computer-based patient-record system (Figure 7.1).

The values if Figure 7.1 show that time has an average value that is 72 times more valuable when compared to new functionality (1.3645 / 0.0019), or 4.1 (1.3645 / .3327) times more valuable when compared to new ways for data entry, or 4 (1.3465 / .3369) times more valuable when compared to an integrated patient- and knowledge-based record system

Table 7.10 demonstrates satisfaction values for the various attributes. For example, physicians wanted a knowledge-based patient-record system *more* than they wanted other ways to enter information into the patient record. With *knowledge system* they understood a system that would assist them during assessment and treatment. Physicians preferred the keyboard over voice recognition or electronic pen.

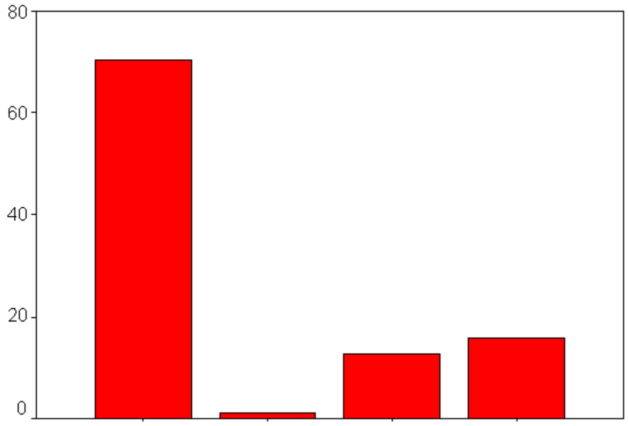


Figure 7.1: The four attributes' relative importance to each other. The y-axis is average importance and the x-axis is factor (or attributes). From left to right, the first bar shows CPR *time usage*, the second bar *Structure of* CPR, the third bar shows CPR *as a knowledge system*, and the bar to the right shows *Patient data entry*.

Table 7.10: Attribute-level satisfaction values.

Attribute and level	Satisfaction value[a]
CPR[b] time usage	
Very time consuming	-1.6018
Time consuming	0.24
Not time consuming	1.3645
Structure of CPR	
Paper metaphor	0.019
Problem orientation	-0.019
Patient data entry	
Keyboard and mouse	0.3369
Speech recognition	-0.1459
Digital notebook	-0.1910
CPR as a knowledge system	
Includes a knowledge system	0.3327
Without a knowledge system	-0.3327

---

[a] n=148.

[b] Computer-Based Patient-Record System (CPR).

7.5 Discussion

7.5.1 Physicians Have User Friendly Systems

A majority of physicians expressed satisfaction with their patient-record system. The system was easy to use and they managed to use it well, in terms of system functionality. The patient-record structure did not hamper physicians during patient care, probably because physicians could enter freely information into the record system (i.e., non-controlled data entry).

Physicians felt that the system contributed positively to patient care, and they answered convincingly that they believed a good computer-based patient-record system had a positive effect on patient care—improved systems would eventually improve patient care. Physicians also claimed that they would learn a new record system. However they believed they needed more education and training than other occupational groups, when it came to using computers and technical tools in everyday work.

These findings—that physicians felt, to a large extent, satisfied with their patient-record systems—are in accordance with what Fosse and Bærheim also found in their study [61]. After switching to a computer-based patient-record system, physicians had more efficient encounters, more time for the patient, easier drug prescriptions, more efficient financial routines, and possibilities to generate patient-specific profiles such as blood sugar expansion or laboratory result profiles.

7.5.2 Limitations on Methods

Limitations on our methods are evaluated with respect to the use of EYR respondents to increase response rates and students to perform parts of our survey.

Adding EYR to the Sample Selection

Our rationale for announcing a request for participation on the EYR mailing list (Section 7.3.3) was to increase our response rates. We chose to combine EYR responses with the rest of the initial responses, instead of using the responses separately. Results from the background questions proved that EYR respondents did not deviate substantially from the remainder of respondents. However, since members on EYR use actively the Internet to work and communicate, they have already discovered the positive impact of this kind of technology. We risk the danger of having respondents that are particularly biased about technology, computers, and the Internet. The EYR physicians may have a positive attitudes toward technology and transfer this view onto the answers in our survey.

Using Students to Assess the Questionnaires

The use of students to make the questionnaire and the interview guide may influence the quality of the questions, especially since the students had no prior knowledge of computer-based patient-record systems. To avoid poor quality on the questions in the survey, we gave the students several directions that would help them to focus and to be able to study the circumstances necessary to conduct a survey that would produce useful end results. First, the students were given a literature list with references to other related surveys on physicians' information needs and uses [30, 64, 135, 174] and to other studies employing conjoint analysis [38, 63, 67, 81, 102, 103]. Second, the students were given three hypothesis that they had to answer in a final report. The three hypothesis were meant to guide the questions that the students selected for the questionnaires.

Physicians May Have Different Understanding of Concepts

The questions in the survey are not constructed with in mind—to extricate a potential difference in definitions of the various concepts covered in the questions, so, the questions do not take into consideration that physicians may have different understandings of the questions. There is no way for us to know if, for example, physicians understanding of the notion of *structure* is the same for all of them.

As said before, the results from the telephone interviews were contradictory. We have no way to unravel if these answers are due to physicians actually being different (with different opinions and experiences), or if they are actually due to physicians' different understandings of the various concepts in the survey. We, therefore, cannot make any hard conclusions based on results from the telephone interviews.

A Time Paradox

The results from the conjoint analysis are taken from the students' initial analysis and we have not been able to reproduce these results in our analysis. So, these results cannot be given too much attention. Since we cannot evaluate the method used for analysis, we chose to only present the findings and not suggest what the results indicate.

From the conjoint analysis we found that physicians value time more than any of the other attributes such as structure of the patient record, or other ways to enter information into the record, or a knowledge-based patient-record system to help assist them during patient care.

It is also noteworthy, that time being such a critical attribute, contradicts findings from our two other surveys (Chapter 5 and Chapter 6); where we found that physicians spend a significant portion of time writing encounter summaries. The results from the conjoint analysis say that physicians value time more than they value other potential, efficient ways of entering information into the record system. Structure is also ranked as less important than time, which is highly surprising given that what physicians claim they miss is record structure that enables them a better overview over information in the patient record.

7.5.3 All Information Useful but Lack of Overview

Of the various questions that physicians generate during patient care, a significant part of them can be partially answered with information from the patient records. We found that Norwegian family physicians use all the information in the patient record at all times during patient care. The information that physicians mostly use concern patient-specific information, such as previous case history, family history, previous disease and illness, laboratory test results, past and current diagnoses, past and current drugs, etc.

According to Ely et al. [56, 57] the questions that arise during consultation—about two questions for every three patients—three quarters of them have to do with patient care and three fifths of them have to do with patient-specific data. Ely et al. also found that one third of the questions relate to treatment and specific conditions, one quarter relates to diagnosis, and one quarter relates to prescription of drugs and medication.

7.5.4 Request for Problem-Oriented Patient Records

Although we did not perform an explicit survey on the need for a problem-oriented patient-record system; the physicians suggested themselves that the problem-oriented patient record would improve the structure, thereby, giving a better overview over information in the patient records.

Nylenna [129] reports that the greatest shortcoming and weakness of the Norwegian computer-based patient record relates to the lack of overview. Information on individual patient records (easily) becomes overly complex and unstructured. Nylenna further states that in situations important to have overview, it is often very likely that the patient has a comprehensive case history. Hence, it becomes more difficult to get a good overview. Nylenna also say that to distinguish between important and not so important information is generally difficult, and it is almost impossible in time pressured situations.

Medicine lacks a modern information infrastructure that rigorously and efficiently connects all those who produce and archive knowledge to all those who need the proper application of that knowledge [194]. Patient data in the patient record are needed in various different care situations by various different clinicians. We seek to understand what kind of information is needed in various kinds of care situations, in order to be able to provide a better structure to the records that facilitates a better overivew of information in them.

Tang and Hammond [182] suggests to provide efficient tools that help clinicians retrieve and understand data relevant to their decision-making tasks. Systems today are not designed to answer many of the common questions physicians ask with respect to patients' medical condition(s). At present there exists no satisfactory solution to the problem of finding those few categories of information most likely to be useful in any given situation among the mass of data available [174].

7.5.5 Female Physicians Favor Communication

If we evaluate the overall answers given by both genders on questions related to use and needs of information from the patient-record system—physicians use and need information during all phases of patient care. We take a little sidestep and consider the difference in answers between the two genders.

Female physicians answered that they search for information *before* the patient enters the office and write information into the record *after* the patient has left the office.

These results may indicate that female physicians value the communication between patient and physician in such a way that they try to minimize the use of the record system while the patient is in the office. Male physicians retrieve and write information during all phases of the clinical encounter.

7.6 Conclusion

Physicians have expressed that they manage the functionality of their record systems, that they do not feel hampered or restrained by the systems, that they believe the record systems have positive impact on patient care, and that they feel themselves open to learn new technologies and to employ these in clinical care. A majority of physicians believed that to improve the record systems will give direct benefits in patient care. Based on these findings, it seems that record systems in Norwegian primary care have an advantageous potential.

The positive attitudes physicians give through their answers gave us the impression that the record system is favored by most family physicians. However, physicians uttered clearly that present day's patient-record system holds bugs and deficiencies; the record system lacks ways to give physicians a good overview over its contents. Physicians requested a problem-oriented structure in order to improve access to patient information in the record systems.

The potential is there—to improve patient care through improved patient-record systems—because patient-record systems are considered useful and patient information is needed by physicians during all phases of patient care. The challenge is to maintain the positive features that exist already in the record systems and to improve on the negative ones. What physicians want is better overview over patient information, and they suggest themselves that a problem-oriented patient record would enable them to retrieve information in a more feasible way than what they can with existing time-oriented patient-record systems.

Part III

Method and Results

Chapter 8

Preliminaries to the Knowledge Model

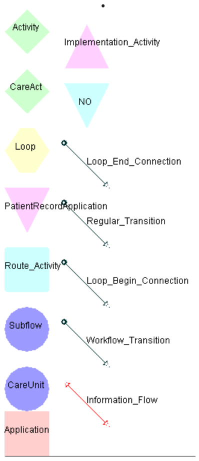
8.1 Introduction

This chapter introduces elements in the knowledge model that the consecutive chapters (Chapter 9–Chapter 12) do not describe in detail. Our design and implementation of the knowledge model is based on a set of requirements:

1. Family physicians want a better overview of the patient-record content than what they presently have in time-oriented patient records (this requirement was also confirmed in our empirical studies; Chapter 5—Chapter 7). We assume that a problem-oriented patient record provides a structure that gives a better overview than the time-oriented patient records
2. Patient-record systems should support a problem-oriented view of data that enables physicians to view patient-record content in relation to medical problems (an initial requirement from Section 1.1.3)
3. The problem-oriented view of patient data should *not* force physicians to work in formal ways (Section 1.1.3). With formal ways we mean a patient-record system that compels clinicians to enter data in a strictly controlled way into the record system. The requirement implies that the patient-record system should itself recognize its data input and be able to structure it. We chose to add process knowledge to the record system so that it can recognize various care situations and determine information relevant to those situations. Knowledge about primary-care activities, goals, purpose, and information needs from the patient record will enable the system to reason with its information content and how this information is used in various care situations
4. The information in the patient record should be possible to rank according to relevance to care situations. Relevance varies according to the change of care activity. Not all information have relevance at all times. A layered structure of information will give a better overview since all information is not presented at the same time. Furthermore, information in the record will always be accessible

via a layered architecture. The layers indicate a distance, or relevance, to the care situation. The closer a layer is, the more relevant the information is to the present care situation.

Figure 8.1: A figure legend which describes type of objects and arrows used as notations in the preceding graphs. The objects and arrows belong to two different process models: (1) a generic workflow process model and (2) a generic primary-care process model. Objects and arrows that belong to the workflow process model are: *Activity*, *Loop*, *Route_Activity*, *Subflow*, *Implementation_Activity*, *NO* (i.e., a dummy activity), *Application*, *Loop_Begin_Connection*, *Loop_End_Connection*, *Regular_Transition*, and *Workflow_Transition*. Objects and arrows that belong to the primary-care process model are: *CareAct*, *CareUnit*, *PatientRecordApplication*, and *Information_Flow*



8.2 Modeling the Primary-Care Process

We have modeled the primary-care activities and subprocesses that we reported in Chapter 5 and was summarized in Table 5.5. The next section describes the graph notations that we have used for the figures in this chapter and the preceding five chapters (Chapter 9-Chapter 13).

8.2.1 Graph Notation

Figure 8.2 and Figure 8.3 have been modeled in the Protégé-2000 system[1] (Section 2.5.5 and Section 3.4). The basic building block in Protégé are frames, where each frame is a first-class object with slots and facets that relate potentially to other frames.

---

[1] The Protégé-2000 system has been developed and is maintained as *open source* software by Stanford Medical Informatics (SMI) [177].

A resulting representation in Protégé is a nested frame-based network (Section 2.5.4), or a semantic network (Section 3.4.1).

The graphs presented in Figure 8.2 and Figure 8.3 are, therefore, a graphical visualizations of a nested frame-based systems (i.e., semantic network). In the figures, the nodes are instances of classes and the links are instances of relationships. The classes and relationships that we refer to are defined in an ontology, which represents the knowledge structure in a knowledge base (ontologies are described in more detail in Chapter 3).

In the knowledge base we incorporate process knowledge for the patient-record system. The process knowledge includes activities and subprocesses in the primary-care process, clinical concepts in the patient records, and concepts in an information model that supports a problem-oriented view of patient data (components in the ontology are described in more detail in Chapter 11 and Chapter 12).

Figure 8.1 illustrates the graph notations that we use in our figures. The objects and arrows are used to model activities and information flow between activities in processes. We model processes on two levels, either as a generic workflow process model or as a generic primary-care process model. We base our generic workflow process model on the meta model that is specified by the Workflow Management Coalition [187]. The generic primary-care process model is an instantiation of the workflow process model:

- Objects are either an activity or an application used in an activity
- An application is any tool that is used in an activity
- Arrows are information flows between activities or applications
- Each arrow is labeled with a *condition*, in terms of information, which has to be satisfied before transitioning from one activity to another
- The *CareAct* object is an instance of a class in the *CareActType* ontology discussed in Section 8.3.2 and illustrated in Figure 8.5. The relationship (i.e., an arrow) between an activity object and a patient-record object in the primary-care process model is an information flow. The information flow also has to be an instance of a class in the *CareActType* ontology.

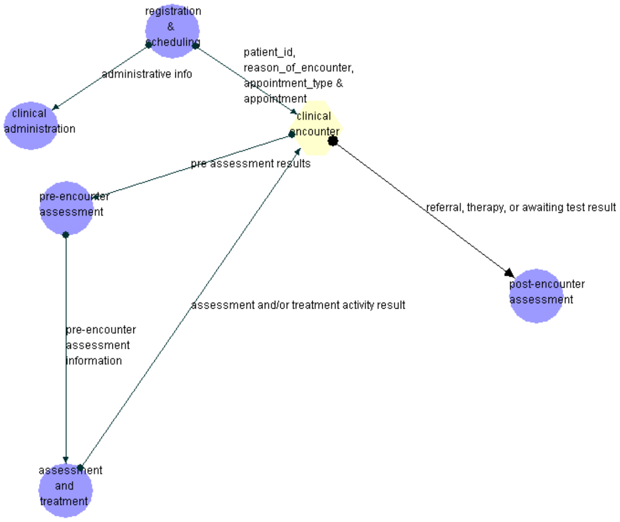
The purpose of modeling information flow between activities is not to automate workflow, but to indicate what information is produced from one activity to another, and that may potentially be used in another activity. Thus, we model information dependencies among activity nodes.

8.2.2 The Primary-Care Process

Activities in the primary-care process are divided into two types of activities, *administrative* activities and *clinical* (encounter) activities. We have focused on the latter group of activities, activities that are relevant for medical decision making; where the clinical activities are *dependent* on information from the patient-record system. We say that the clinical activities belong to the clinical encounter. The division of activities into two different types, administrative and clinical, is in accordance with the HL7 RIM, which was described in Section 4.6.2.

In Figure 8.2, the clinical encounter has been modeled as a loop (i.e., the yellow hexagon shape). The loop serves two purposes: (1) to support an intermission in the encounter due to for instance a local laboratory-test order that implies a time delay in the encounter (i.e., laboratory results must exist before assessment and treatment can continue) or (2) to support assessment and treatment processes that continue over several encounters

Figure 8.2: A generic model of the primary-care process. Circles are subprocesses and the hexagon is a loop. Arrows are information flows. See Section 8.2.1 for a more elaborate graph legend.



Each subprocess contains a set of activities and each activity is modeled with a dependency to a generic section of the patient record (Figure 8.3). The dependencies we have used are listed in Table 8.1 under the two columns of *Purpose* and *Information*. Table 8.1 summarizes information from Table 5.5 and other information that we have elicited based on the empirical studies reported in Chapter 5-Chapter 7. We will explain Table 8.1 in more detail in Section 8.3.

A dependency between an activity and the patient record can be viewed as an interaction; an activity has a purpose, the purpose of the activity specifies some need for information[2], and where the information for a specific activity can be found in the patient-record system.

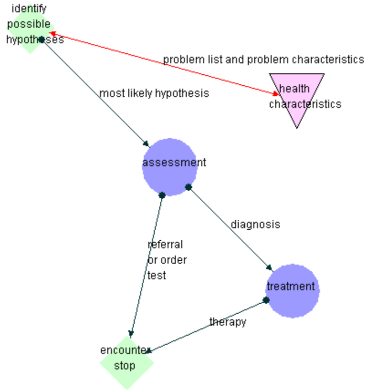
Figure 8.3 shows an activity—*identify possible hypothesis*—and its interaction with the patient-record system. The part of the patient-record system that is depicted in the figure represents a generic part of the patient record, which we have labeled *health*

---

[2] We do not consider an activity's information needs that are not found in the patient-record system.

characteristics. The *Problem list and problem characteristics* are names of the information flow or dependency between the activity and the patient record. An activity needs this kind of information from the patient-record system.

Figure 8.3: A subprocess—assessment and treatment—in the primary-care process. Circles are subprocesses, rectangles are activities, and the up-side down triangle is the patient-record application. Arrows are information flows. See Section 8.2.1 for a more elaborate graph legend.



8.3 Activity-Goal-Purpose-Information Needs

The *Information* column in Table 8.1 specifies an activity's information needs from the patient-record system. Information needs are specified mainly in terms of overview, where overview is represented by a level number. Overview 1 is level 1, overview 2 is level 2, and overview 3 is level 3.

The three levels of overview we use to relate an activity with its information needs. To know an activity's information needs, we need to know an activity's purpose. To know an activity's purpose, we need to know an activity's overall goal. Before we present an activity's overall goal, we first explain the three levels of overview that are used.

8.3.1 Overview and Problem Orientation

Physicians want (better) overview of the patient-record content. Part of providing overview in a patient-record system has to do with structure. This structure will be offered by a problem-oriented patient record. In addition to structure, physicians' overview depends on the care situation; the notion of overview is different when a patient first introduces a medical problem than when a patient has a medical history of 10 years with that same problem. As physicians say themselves (in Chapter 5.3.2, Chapter 6.4.3, and Chapter 7.4.2), part of having overview has to do with:

Statement 1 *Knowing what physicians have previously asked and done to a patient.*

We use Statement 1 to help us determine the impact of physicians' requirements of having an overview. We relate overview to the following:

- Overview is a dynamic concept that changes with an activity, because the overview statement (Statement 1) refers to knowledge about what physicians have *done* (i.e., actions, tasks, activities, etc.)
- Overview is related to a patient and what the patient has said. Thus we have to distinguish between what the patient has said and what other actors have said (i.e., what the physician has said). Information is *role* dependent. We have already defined two roles; (1) information owner and (2) information user (Section 5.2.3)
- Overview suggests some kind of sorting of information because physicians want to distinguish what patients have *said* (i.e., a statement) and what physicians have *done* (i.e., an action). We will use a set of categories (i.e., clinical concepts) to help us distinguish between different types of information. A set of clinical categories have been defined in Section 6.4.

8.3.2 A Problem-Oriented View

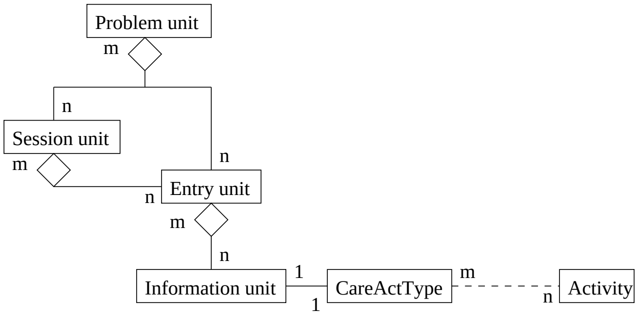
We have related part of having overview to a problem-oriented structure of information in the patient records. A problem-oriented view of patient data fulfills two requirements: (1) it supports the way physicians think and work (Section 1.2), and (2) it gives better overview of patient-record content. We have developed an information model that supports a problem-oriented view of data. The information model is illustrated in Figure 8.4. Important constructs in the model are:

- The *Information unit*: which is the basic unit of information in the patient record. Every information unit is labeled with one of the concepts in the *CareActType* ontology (Figure 8.5). In Table 6.12 we demonstrated how patient-record content was split into information units and labeled with a clinical concept
- The CareActType ontology: which represents the relationship between content in the patient record (i.e., an information unit) and a clinical *Activity* (i.e., care activity). For example, an activity such as *Perform clinical examination* will produce results such as *Clinical examination findings* that will be reported in the patient record. A concept in the CareActType ontology will represent both the activity and the result of the activity in terms its findings
- The Activity: which is part of the process model. The information model and the process model are related to the CareActType ontology that *bridge* activities in the primary-care process with patient-record contents. The concept *Clinical examination* refers to both the activity and the results of the activity. This *implicit relationship* between activity and the result of an activity is in accordance with how HL7 in RIM [184] defines it: "An act includes the 'results', 'answers' or informational 'procedure products' gained during the act"; the output of an

activity implies the activity itself, or an activity includes the output of the activity. The implicit relationship between activity and results of the activity, enable us to define a relationship between an activity and content in the patient record. An output from an activity, that is recorded in the patient record, relates to an activity in the primary-care process (that produces this output).

- The *Entry unit*: which is equal to the classic entry in current time-oriented patient records. An entry unit comprises a set of information units (as illustrated in Table 6.12) which make up the encounter summary
- The *Session unit*: which comprises a set of entry units, as demonstrated in Table 6.8. The entry units are related to each other. The relationship can vary. For example, a typical pattern in current patient records are three entries, where the first entry is a full encounter summary, the second entry contains laboratory results, and the third entry is a drug prescription. These entries are all related to the first entry. In our information model we can cluster these three entry units together into a session unit
- The *Problem unit*: which comprises a set of entry and session units that all relate to the same problem. Each information unit has a label and an association with an entry unit. However, an entry unit may or may not be associated with a problem unit.

Figure 8.4: A problem-oriented information model. Boxes are conceptual entities. The relationships between the entities are *composed_of* and *is_a*. The relationship between *CareActType* and *Activity* is marked with a dashed line to indicate an *implicit* relationship between the information model and the primary-care process model.



A distinct difference between, and the flexibility of our entry unit compared to the classic entry in a time-oriented patient records (Table 6.5) has to do with the labeling of information units. Every information unit is labeled with a CareActType, which represents the relationship between activity in the care process and patient-record content.

The set of CareActTypes, or clinical concepts[3], we use are listed in the ontology in Figure 8.5. We model classes in the CareActType ontology as *abstract* classes (i.e.,

---

[3] Clinical concepts and CareActType are used interchangeably in this text.

classes cannot have direct instances) to emphasize the relationship between care activities and the result of activities which are documented in the patient-record system. The classes in the ontology are based on the clinical concepts (Table 6.4) that we presented in Chapter 6.

In Figure 8.4, the relationship between CareActType and Activity has been drawn with a dashed line to indicate an *implicit* relationship between the information model and the primary-care process model. The link demonstrates the relationship that information units in the patient record have with activities in the primary-care process (e.g., an information unit with a CareActType *Clinical finding* relates directly to the activity *Perform Clinical Examination*).

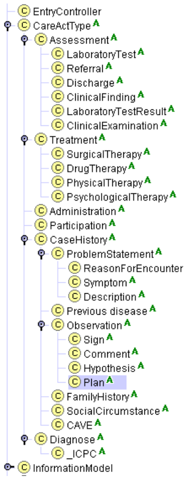


Figure 8.5: The CareActType ontology with clinical concepts from the patient record. The classes in the ontology are *abstract classes* (i.e., the *A* suffix).

8.3.3 Three Levels of Overview

If we integrate the three levels of overview used in Table 8.1 with our problem-oriented information model, we can define the three levels of overview: (1) overview 1 is overview of the patient-record contents, (2) overview 2 is overview of medical topics (or a patient's conditions), and (3) overview 3 is overview of a specific medical problem. We explain our definitions with the following:

1. Level 1 overview is *content overview*: To have overview over the patient-record content implies knowledge of medical concepts in the record. We use the medical concepts defined in Chapter 6 (Table 6.4). Overview over patient record content make up part of a physician's *memory picture* [128] and provides a foundation for further target search and problem solving (i.e., the next two levels of overview)

2. Level 2 overview is *medical condition overview*: To have overview of patient-record content gives a foundation for further exploration of a patient's medical condition and specific medical problems—to see problems in relation with other problems. A patient's *problem list* represents a projection of his or her medical condition (i.e., past and current medical problems). Medical condition overview refers to a set (or list) of medical problems

3. Level 3 overview is *medical problem overview*: To have overview of a patient's medical condition implies having an overview of a patient's past and current medical problems. Medical problem overview refers to overview over a specific problem and is equal to Nygren and Henriksson's third level of reading a patient record—reading for the purpose of specific problem solving. Nygren and Henriksson [128] present three different ways of reading a patient record (i.e., a medical record): (1) reading to get overview, (2) doing target search with specific clinical concepts in mind, and (3) reading for the purpose of specific problem solving.

8.3.4 An Activity's Purpose and Information Needs

The activity *identify possible hypothesis* in Figure 8.3 is listed in Table 8.1 under the subprocess *Assessment* with corresponding purpose—*Identify problem*—and information needs—*Overview 2-3* (from the patient record).

Overview 2 refers to a set of problems (i.e., a problem list). Overview 3 refers to a specific problem and information within that problem that relates concretely to the activity. For the activity of *identify possible hypothesis* a physician has to decide if a patient's *Reason for encounter* and *Problem statement* is related to an existing problem or a new problem. To make this decision, a physician needs overview of information on a patient's problem list (overview 2) and health characteristics (overview 3).

8.3.5 An Activity's Goals

An activity's purpose is guided by its goals. An activity's information needs are guided by its purpose, and an activity's relation to the patient record is guided by its information needs. Reinecke [153] gives the following definition of a physician's job: "A physician's job is to diagnose and to treat disease". We say that an activity's goal is one of two, either to diagnose (i.e., assessment), or to treat (i.e., given a diagnose).

For example, the purpose of *assessment* is to gather information and the purpose of *treatment* is to give some kind of therapy (given a diagnose). Thus, activities that belong to assessment in Table 8.1 have an overall goal to gather information so that a diagnose can be given, while activities that belong to treatment have an overall goal to give therapy of some kind.

8.3.6 Medical Decision-Making Activities

The subprocess *Laboratory test* in Table 8.1 comprises two activities: (1) *Specify laboratory test* and (2) *Evaluate laboratory results*. The laboratory test process actually includes more activities than these two, for instance activities such as *Specify laboratory test*, *Perform laboratory test*, *Receive laboratory results*, and *Evaluate laboratory results*. Because our focus is on activities related to the medical decision-making process, and how these activities depend on information in the patient record, we do not consider activities that are not dependent on information in the patient record (or activities that do not produce information that is stored in the patient record). Thus the activities listed in Table 8.1 are activities that are dependent on information from the patient record and they are vital to physicians' clinical decision-making process.

Other comments on Table 8.1 that are relevant to mention are given in the following:

- Assessment is both a process and a purpose. Until a diagnose exists, the assessment process goes on in a loop

- Assessment includes the subprocesses *Existing problem?, Perform clinical examination, Specify laboratory test, Specify referral* and *Evaluate discharge*

- The activities *Specify referral* and *Evaluate discharge* correspond to each other in the same way as *Specify laboratory test order* and *Evaluate laboratory results* do

- Information owner and user are the participants or actors responsible for activities or tasks. The information owners and users that we use are the ones that we presented in Chapter 5 (e.g., Figure 5.3, Figure 5.5, and Figure 5.4), such as *patient, physicians, clinician*, etc.

8.4 Summary

This chapter is a preliminary chapter to the next four chapters that deal with a more elaborate description of the methods used to develop the knowledge model for a primary-care patient-record system. As input to the knowledge modeling are a set of requirements. The requirements are a combination of the initial requirements from Section 1.2.1 and additional requirements that appeared as a result of the empirical studies we did—reported in Chapter 5, Chapter 6, and Chapter 7.

The knowledge model comprises a set of primary-care activities, a set of clinical concepts, a problem-oriented information model, and relationships among the care activities and patient-record contents. For each care activity we have identified the activity's overall goal, intention, and need for information. The need for information is related to information in the patient records.

Activities are essential in family physicians' decision-making process in the sense that they either produce or consume information that is related to the contents of the patient records. Based on a relationship between an activity and the record content we are able to create means for ranking information relevant to care activities. Moreover, the relationships among care activities and patient-record contents are represented in an ontology that we use to rank information according to relevance with a care activity.

Table 8.1: Primary-care activities, purpose, information needs, location, and participants (partly based on Chapter 5).

Subprocess	Activity	Purpose	Information[a]	Office[b]	Owner[c]	User[d]
Register & Scheduling	Identify & Register patient	Establish patient record	Patient ID & Demographics	Front Office	Clin	Clin
	Reason for encounter	Overview & Context	Reason for encounter	Front office	Clin	Clin
	Schedule appointment	Schedule appointment	Appointment type	Front office	Clin	Clin
			Appointment date	Front office	Clin	Clin
Pre-encounter assessment	Prepare patient	Prepare patient	Results	Lab	Clin	Phys
	Get into context	Get into context	Overview 1-2-3	Phys's office	Clin	Phys
Assessment	Receive patient's case history	Problem statement	Case history Overview 3	Phys's office	Phys	Phys
	Identify possible hypotheses	Identify problem	Overview 2-3	Phys's office	Phys	Phys
	Set diagnose	Diagnostics	Overview 3	Phys's office	Phys	Phys
Existing problem?	Determine context	Identify context	Overview 3	Phys's office	Phys	Phys
	Continuation?	Assessment	Overview 3	Phys's office	Phys	Phys
	Sequelae?	Assessment	Overview 2-3	Phys's office	Phys	Phys
	Sub-problem?	Assessment	Overview 2-3	Phys's office	Phys	Phys
Clinical examination	Perform clinical examination	Assessment	Overview 2-3	Phys's office	Phys	Phys
	Evaluate clinical findings	Assessment	Findings	Phys's office	Phys	Phys
Laboratory test	Specify laboratory test	Assessment	Lab spec	Lab	Phys	Clin
	Evaluate laboratory results	Assessment	Results	Phys's office	Clin	Phys
Referral	Specify referral	Assessment	Overview 2-3	Phys's office	Phys	Clin
Discharge	Evaluate discharge	Assessment	Overview 3	Phys's office	Clin	Phys
Treatment	Determine therapy	Treatment	Overview 3	Phys's office	Phys	Phys
Surgical therapy	Give therapy	Treatment	Overview 3	Phys's office	Phys	Phys
Drug therapy	Give therapy	Treatment	Overview 3	Phys's office	Phys	Phys
Physical therapy	Givey therapy	Treatment	Overview 3	Phys's office	Phys	Phys
Post-encounter assessment	Finalize encounter	Finalize encounter	Overview 1-2-3	Phys's office	Phys	Phys

---

[a] Information needs.

[b] Clinical office: (1) Physicians office (Phys's office), (2) Laboratory (Lab), and (3) Front office.

[c] Information owner: (1) Clinician (Clin) and (2) Physician (Phys) .

[d] Information user: (1) Clinician (Clin) and (2) Physician (Phys).

Chapter 9

Ranking of Information in the Computerized Problem-Oriented Patient Record

Ranking of Information in the Computerized Problem-Oriented Patient Record

Elisabeth Bayegan[a,b], Øystein Nytrø[a,c], Anders Grimsmo[b]

Norwegian University of Science and Technology, Trondheim
[a]Department of Computer and Information Science
[b]Department of Community Medicine and General Practice
[c]SINTEF Informatics

Abstract

We propose a framework for a problem-oriented patient record for general practice[1] and defend that the problem-oriented medical record represents an intuitive way to organize the patient record. By adding a layer of knowledge to the electronic patient record the record system is able to better utilize the information stored in the record. If a record system is process aware, having knowledge of work processes and is able to distinguish between different contexts in use, the system can provide relevant and useful information during the handling of patients' medical problems. Information is ranked according to its relevancy in a given context by using action patterns – traces. Traces give valuable indications of what is going on during the process of patient care. Decision frames represents relevant contexts based on the information in the record. Both decision frames and traces provide an environment in which more optimal medical decisions can be made.

Keywords:

Computerized Patient Records; Problem-Oriented Medical Records; Knowledge-Based Systems; Knowledge Representation

Introduction

Medicine, as a science, is highly uncertain and treating patients can be seen as a therapeutic experiment [2]. Physicians write down everything of significance to patient care in their patient records. The purpose is to be able to recognize and remember events of importance. The human mind has limited ability to recall and process complex information and make decisions under severe time constraints [3]. Physicians need a tool to help them gather, process, and retrieve information while working with patients and comprehensive patient records bacause:

- Physicians lose overview of the record content after a few years of a long-lasting patient-physician relationship.
- The patient record accumulates over time and will contain information on many medical problems.

We suggest a framework for a computerized problem-oriented patient record (CPOPR) that is able to rank information according to its relevance in a given context. The CPOPR must be able to know how physicians work and think for it to be able to know what is relevant information. The proposed framework rely on the principle of embodying knowledge of what is in the record and how this information is used. The framework is based on four key topics:

1. A 2-layer knowledge representation with a facts layer and a (derived) knowledge layer. The knowledge layer comprises two models; a process and a relevance model that both can analyze action patterns in the patient record.
2. Contexts, that are provided by the record system defined as decision frames. Decision frames present ranked information, i.e. facts from the record, to be used in decision-making. The filtering of irrelevant information makes it easier for the physician to focus attention on information that is relevant.
3. Traces, which are patterns of actions that enable the record system to rank information according to its relevancy, i.e. an interpreted collection of events.
4. A problem-oriented model of information in the record. We are not referring to the classic association of problem-orientated medical record (POMR) introduced by Weed [5] in 1968, but a model in which everything recorded is associated with a medical problem, i.e. all events, facts, and

---

[1] Disclaimer: All references to electronic patient record systems are based on conditions for general practice in Norway. More than 90% of general practice physicians use electronic patient record systems today [1].

relations in the record. With a problem model[2] it is possible to rank information according to relatedness to neighboring problems and processes, or to determine irrelevancy.

The CPOPR system is *not* a decision-support system in traditional sense, i.e. using medical reasoning for the purpose of diagnosing [7] but is a helpful information system capable of analyzing, storing, and retrieving information valuable in the process of handling a patient's medical problems. Our aim is to provide only necessary, useful, and needed information from the patient record, extracting irrelevant information for the given situation.

This paper is organized into six sections. Section two presents the 2-layer framework and its two basic models: the process model and the relevance model; central concepts of these models are also described. Section three introduces the concept of decision frames. Decision frames are sub-contexts comprising relevant information from the patient record necessary for medical decision-making. Section four uses the framework to demonstrate its practical usefulness. Section five discusses future work and potential enhancements. Section six concludes the paper.

A Framework for CPOPR

The purpose of a helpful record system is to provide appropriate and needed information. Our proposed framework consists of two distinct levels: a facts level and a knowledge level [8]. The first level corresponds to what, traditionally, has been recorded in Computerized Patient Records (CPRs) – information on what has been heard, seen, thought, and done [9], thus, notes on *direct observations* [10]. The second level is a meta-layer to the first, i.e. what can be said about what has been heard, seen, thought, and done. The knowledge layer enables the record system to say something about its content and use, and is based on two interacting models:

1. A process model: Knowledge about a physician's work processes enables the record system to recognize where in the process the physician is and determine information needs at the current stage.
2. A relevance model: Information is ranked according to its relevance in a given context. The model makes it possible to present relevant information to the user when needed. Information can also have degrees of relevance in a given context.

Both models use the facts on direct observations as a foundation for saying something about where in the process the physician is and what the current information needs are. Through the process model the record system is capable of determining if the physician is in diagnostics or in the treatment phase of a patient's problem, if the handling of a medical problem is successful or not, and if there is a need to go back and look at other potentially relevant issues. The relevance model interprets action patterns in the patient record where the objective is to estimate what the information needs are and what would possibly be beneficial to present to the physician. Below, we will describe the two models and their sub-components.

A Process Model – Formalizing the Encounter

Encounters are initiated when a patient has a problem that needs to be handled by a physician. The physician's job is to diagnose and treat diseases [11] splitting the encounter into two modes of operation: *assessment and treatment.* Each problem is either a diagnostic problem – trying to determine what caused the problem – or a management problem – the cause is known but the effort is to choose the best option for managing the problem [12]. As a general rule, information needs are greater in the assessment phase than in the treatment phase. Figure 1 refers to a process model containing three, rough essential steps in an encounter:

1. A problem statement which is based on the set Φ of patient data, i.e. patient's complaints, symptoms, and signs.
2. A set A of actions used for handling a patient's medical problem.
3. An evaluation of the set O of outcome. The evaluation of outcome has a time delay (Δt) and will usually take place in the next occurring encounter.

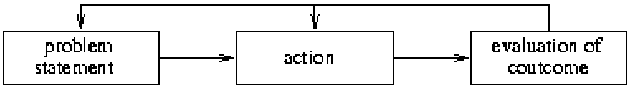


Figure 1 - A process model for episodes of care (EoC)

Referring to the process model in Figure 1 ; there is no clear division between the assessment and the treatment phase. The action state covers, or functions as a substitute for both phases. Feedback arrows indicate that the problem-solving process is iterative. An episode of care (EoC) [13] is a set of distinct encounters related to a specific medical problem. The patient's condition, after a set of actions, determines the next set of encounters. Effects of handling a problem leave the patient in one of three conditions: 1) The problem is cured, 2) The problem is persistent but the condition is either better or status quo, and 3) The problem is worse than before. If the patient's condition persists or gets worse it is necessary to go back and re-evaluate the problem statement and proceed with attainable actions. A re-evaluation requires access to and overview of previous information.

Defining "The Problem"

Problem is a nametag for the medical problem that needs to be handled by a physician. Common to all problems is that the problem has to be solved with medical knowledge. No matter the cause, let us define the reason for contact, or patient complaint [14] as *the problem*.

---

[2] The problem model is not a primary focus in this text. For a more detailed explanation on problem-oriented patient records, see [6].

A Relevance Model

A diagnosis emerges from a context in which decisions are made. A context consists of a combinatorial analysis of medical knowledge, tests, procedures, observations [4] on patients' current condition, and facts recorded in the record. An essential and necessary first step in medical decision-making is the linking of patient specific data and medical knowledge [12]. Computer systems can contribute by presenting relevant and useful information in this process.

Within a particular context, some information is more valuable than other. The challenge for a CPR system is to present relevant information to the physician. Information can be ranked according to its relevance in a context. From Figure 2, we have on one side of the scale all information, both relevant and irrelevant. On the opposite side we have only relevant information. In between, there are areas where irrelevant information is removed and where relevant information is more focused in a specific context.

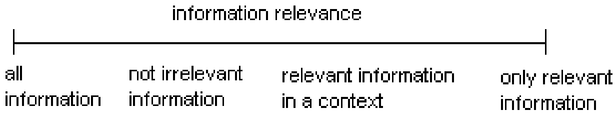


Figure 2 - Ranking of CPR content that is relevant to a medical problem

Traces and Action Patterns

Ranking of information is done through analyzing traces, or action patterns in the patient record. By analyzing these patterns it is possible to say something about the nature of the problem and potential relevant information. Traces depend on:

1. A process model to determine where in the process the physician is.
2. Information about remedies in use, which help to characterize the nature of the problem.

Through traces it is, therefore, possible to say something about the current problem status, where in the process of care the physician is, and what information is relevant at that time and for what purpose.

Remedies and Number of Encounters

A physician has a set of tools to his or her disposal when handling a patient's problems. Remedies constitute the action space, the set *A*, referred to in the process model. The objective of analyzing the remedies used is to classify the nature of the problem, or the type of situation, the physician is in. The remedies are listed in *Table 1* and can be classified into three groups depending on the situation they are used in[3]. The classes represent *type of situation*:

1. Omnifarious remedies are independent of type of situation.
2. Uncertain remedies are used when the nature of the problem is uncertain, they usually occur in the assessment phase.
3. Certain remedies are used when dealing with problems that are *certain*, to some degree, usually in the phase of treatment.

Table 1 - Remedies classified into type of situation

Omnifarious	**Uncertain**	**Certain**
Dialogue	Clinical exam.	Medication
Time	Lab. tests	Operation
Schema	Referrals	Physical Therapy

The number of encounters, together with the type of remedies in use, indicate more accurately the problem state, see *Table 2* A high frequency of encounters with short intervals indicate that the symptoms, signs, or findings do not give the physician accurate answers. If the physician is not sure of the problem's importance, or is not able to eliminate a potential danger, then encounters are most likely assigned with short intervals not to lose time[4].

Table 2 - Problem state as a combination of number of encounters and remedies in use

Remedies	**Frequency of Encounters**	
	High	Low
Uncertain	Problem uncertain, indication of importance	Problem uncertain, seem not important
Certain	Problem known and important, needs follow-up	Problem known, follow-up not important

Decision Frames

The environment, contributed by the CPR system, in which a decision is made is called a decision frame and consists of a collection of facts from the patient record. A decision frame is a sub-context containing patient specific data related to the given context. The presentation of relevant information from the patient record, essential for a medical decision at hand, is what we refer to as a decision frame. A decision frame has eight dimensions[5], these are:

---

[3] Observation is also an important remedy, but is not measurable in the same way as the ones listed in ***Table 1***.

[4] A serious problem will most likely also involve specialists, which can be inferred from the record.

[5] Based on the work done by Sadegh-Zadeh in [15] but modified and extended for our purpose.

p: Patient attributes, such as name, gender, age, family and social history, CAVE[6], etc. *p={Tom Smith, male, 55, Tom's mother has osteoporosis, Tom is an artist, Tom is a smoker since his early twenties, CAVE Tom is allergic to penicillin}.*

c: The set of clinical personnel that are authorized to read from and write data into the record.

$\Phi=\{\phi_1,\ldots\phi_m\}$ Patient data, e.g. a complaint, symptom or sign for which a diagnosis is sought. *Φ={Tom has morning stiffness, Tom has symmetric joint swelling in his hands, Tom has tested positive on serum rheumatoid factor, Tom has rheumatoid nodules, Tom has had swelling of three or more joint areas}.*

$D=\{\delta_1,\ldots,\delta_n\}$ The set of diagnoses *D={Tom has rheumatoid arthritis}.*

t: Time[7].

A The set of actions.

G: The set of goals to achieve.

O: The set O of outcome of handling a medical problem. *O={Problem cured, Problem better, Problem status quo, Problem worse}.*

A decision frame represents an instantiation of each of the eight dimensions and is further dependent on the knowledge the system has of a disease (or a medical problem), history of treatment, what remedies are in use, the patient's current condition, where in the process of care the physician is, and what information is relevant at that time and for what purpose. The values of the variables are a result of past and present values, e.g. $\Phi = \Phi_p + \Phi'$, respectively.

Using The Framework To Rank CPR Content

There are several factors that reinforce the disadvantages with existing implementations of CPRs. Interleaving medical problems, the accumulation of information regarding a specific medical problem, more than one physician handling the patient over time, and time in general; often there is a distance in time between the first visible symptoms and the problem evolved into a disease.

Clinical hypotheses [16],[17],[18] and the reasoning behind the medical decision can be visualized by looking more extensively on the events in the record, e.g. patient's complaints, symptoms, signs, actions, and the patient's condition:

- Patient data: Have complaints, symptoms, and signs changed ($\Delta\Phi$) and if they have to a better or worse condition?
- Actions: Remedies in use; the type-of -remedy that dominates the action set indicates the nature of the problem.
- Patient condition: The number of encounters indicates importance of the problem[8].

Using Traces To Select Relevant CPR Content

An example will help illustrate the impact of our proposed framework:

p: *{Ann Jones, female, 45}*

Φ: *{Troublesome and dry eyes, A feeling of dry mouth, Drink a lot to swallow food, Feeling tired, Burning in throat}*

A': *{Referral to rheumatologist; suspicion of Sjögren's syndrome}*

β_1 *{Eye complaint}*

A'': *{Discharge from rheumatologist: patient tested negative on Sjögren's syndrome*

$\Delta\Phi$: *{Pain in stomach, Lost 5 kg}*

A''': *{Referral to internist: gastroscopy}*

β_2: *{Dyspepsia}*

A trace for this example might look like:

1 Introduction of a problem list. A problem list is possible due to the problem-oriented model, e.g. D = {URTI[9], Family planning/oral contraception, Acute bronchitis, General weakness/tiredness, Eye complaint}.

2 Elimination of irrelevant problems, for instance D' = {URTI, Family planning, Acute bronchitis} is irrelevant in this context.

3 The problem being handled is of uncertain nature because:
   - Referrals to specialists.
   - No conclusive symptoms. A problem of uncertain nature requires more information while under decision-making.

4 The problem statement has changed, i.e. patient complaints and symptoms ($\Delta\Phi$) from one encounter to another, which indicates most likely that the condition has not improved. The patient has currently stomach complaints.

5 There is an expectation to receive a discharge on the referral to internist regarding gastroscopy.

6 It is possible to say that the physician is still in the assessment phase and has a greater need for information about what has previously been done.

7 The re-evaluation of the problems indicates a necessary re-evaluation of potential hypotheses.

Discussion

A helpful CPR system, as we have described, can contribute to solving hard and vague medical problems that have a large recorded history of events, interleaving problems and possibly secondary diseases to primary diseases. With such medical problems the amount of

---

[6] CAVE *lat. beware* used as a reminder for known allergic reactions.

[7] Time can be interpreted differently depending on the value of other dimensions but the topic is not in the scope of this text.

[8] An exception would be if the patient has hypochondria.

[9] Upper respiratory tract infection.

information grows progressively and it is a challenge to have a broad view of all the problems while still remain focused on what is relevant. A helpful CPR system that has knowledge of its content and use and knows how physicians work is therefore able to analyze and interpret the content of the record and present it to the user when needed.

Conclusion

Transition from paper-based patient records to CPR systems have resulted in efficiency gains in routine work, such as electronic billing, electronic prescriptions, referrals, electronic schemas in general, etc. Current CPR systems are nothing more than logs of continuous events and the information stored is, more or less, only accessible for the user who owns the information and knows where to search for it [19]. CPR systems have not taken advantage of modern computer methods, but now, health professionals have recognized the importance of strategic information systems where the focus is on new methods for information storage, retrieval, and analysis [20].

The introduced framework and its knowledge representation allow the CPR system to better utilize information residing in the record. Knowledge about physicians' work processes empowers the record system to recognize where in the process physicians are and determine relevant information needs at that current stage. The proposed helpful record system aims to introduce a seamless transition between the users' goals and intentions and what information the system itself can provide. Our objective with the presented framework is to contribute and indicate potentially relevant information, necessary for diagnostics and decision making – by being non-intrusive, just taking advantage of what already exists in the patient record.

Acknowledgement

We would like to thank Associate Professor Arild Faxvaag at The Norwegian University of Science and Technology, Department of Bone and Joint Disorders. Funding of this paper is by The Norwegian Research Council.

References

[1] Grimsmo A. *Elektronisk pasientjournal i allmennpraksis*. Samfunnsmedisinsk forskningssenter Surnadal, 6656 Surnadal, 1999.

[2] Shortliffe EH, Barnett GO. Medical Data: Their Acquisition, Storage, and Use. In Shortliffe EH, Perrault LE, Wiederhold G, Fagan LM, editors. Medical Informatics: Computer Applications in Heath Care and Biomedicine. 2nd edition. Springer Verlag, 2000.

[3] Weed LL. *Medical Records, Medical Education, and Patient Care*. The Press of Western Reserve University. Fifth Printing. 1971.

[4] Weed LL, Weed L. Opening the black box of clinical judgment: an overview. BMJ 1999: 319(1279).

[5] Weed LL. Medical Records that Guide and Teach. *The New England Journal of Medicine* 1968: 278(12).

[6] Bayegan E, Nytrø Ø. In defence of problem-oriented patient records. (submitted 2001).

[7] Altman RB. AI in Medicine: The Spectrum of Challenges from Managed Care to Molecular Medicine. Fifteenth National Conference on Artificial Intelligence 1998.

[8] Newell A. The Knowledge Level. *Artificial Intelligence* 1982: 18 pp. 87-127.

[9] Rector AL, Nolan WA, Kay S. Foundations for an Electronic Medical Record. *Methods of Information in Medicine* 1991: 30 pp: 179-188.

[10] Rector AL, Nolan WA, Kay S, Goble TA, Howkins TJ. A Framework for Modelling the Electronic Medical Record. *Methods of Information in Medicine* 1993: 32(2) pp: 109-119.

[11] Reinecke L. A Black Day for Clinical Judgment. bmj.com 1999. http://www.bmj.com/cgi/eletters/319/7220/1279/DC1#EL2 (downloaded September 2000).

[12] Weed LL. Clinical judgment revisited. *Methods of Information in Medicine* 1999:38 pp:279-86.

[13] Lamberts H, Hofman-Okkes I. Episodes of care: a core concept in family practice. *Journal of Family Practice* 1996:42 pp: 161-7.

[14] Severs M. Headings for Communicating Clinical Information from the Personal Health Record 1998, http://www.wpr.nhsweb.nhs.uk/about_epr/Paper25a.html (downloaded June 2000).

[15] Sadegh-Zadeh K. Fundamentals of clinical methodology 4. Diagnosis. *Artificial Intelligence in Medicine* 2000: 20 pp. 227-241.

[16] Bentsen BG, Bruusgaard D, Rytter Evensen Å, Fugelli P, Aaraas I (ed.). *Allmennmedisin - Fag og praksis*. Gyldendal Norsk Forlag, 1991.

[17] Sackett DL, Haynes RB, Tugwell P. *Clinical Epidemiology: A Basic Science For Clinical Medicine*. 2nd edition. Lippincott Williams & Wilkins Publishers, August 1991.

[18] Hjortdahl P, Hunskår S (ed.), Johannessen T, Holtedahl KA, Kirkengen AL, Nessa J. *Allmennmedisin - Klinisk arbeid*. Ad Notam Gyldendal AS, 1997.

[19] Grimsmo A, Nytrø Ø. Bedre informasjonsarkitektur og anvendbarhet av den elektroniske journal i primrhelsetjenesten. Et medisinsk-teknologisk samarbeid. Prosjektbeskrivelse for Norges forskinigsråd, 1997.

[20] Altman RB. Ten Grand Challenges for Medical Informatics. Western Journal of Medicine 1997.

Address for correspondence

Elisabeth Bayegan (bayegan@idi.ntnu.no)
http://www.idi.ntnu.no/~bayegan/
Associate Prof. Øystein Nytrø (nytroe@idi.ntnu.no)
Associate Prof. Anders Grimsmo
(anders.grimsmo@medisin.ntnu.no)

Chapter 10

A Knowledge-Based, Problem-Oriented Patient Record System

A Problem-Oriented, Knowledge-Based Patient Record System

Elisabeth Bayegan and Øystein Nytrø
Norwegian University of Science and Technology,
Department of Computer and Information Science and
Department of Community Medicine and General Practice,
NO-7491 Trondheim, Norway

Abstract. The concept of a problem-oriented patient record was presented in the late 1960s but has yet to gain wide acceptance. In this paper we suggest a distinction between the idea of problem orientation and the implementation of the idea. We argue that the problem-oriented patient record offers an intuitive and useful way to work with patient information. We show that the concept of problem-oriented patient records facilitates better care of patients by supporting continuity of care, removing redundant and confusing information, and enabling easy overview of and access to its content. We further propose a two-layer framework that has knowledge of its content and use and is able to better utilize information in the record by presenting relevant information to the user at a time when needed. Conceptually, this is done by adding a layer of knowledge to the patient record system: 1) Knowledge about physicians' way of thinking and working, 2) Their corresponding information use and need during patient care, and 3) Tools to determine information relevance in a given situation; such a knowledge-based system is able to reason with its content and use.

1 Introduction

Lawrence L. Weed introduced the Problem-Oriented Medical Record (POMR) in the late 1960s with intentions of improving the structure of medical records [1]. Current patient record systems make it difficult, if not impossible, to obtain history of illness from its earliest stages because of their episodic orientation (i.e. on a per-encounter basis). Without a complete record of a patient's problems, clinicians cannot track course of development and systematically learn from own experience. For efficient and high-quality patient care the patient record should provide a foundation for linking patient-specific data with medical knowledge, learning, and improving skills of medical practice. A problem-oriented patient record proposes a change from a reductionistic view to a holistic view – to see problems in connection with prior and possible future events.

Previous implementations of POMR such as the Problem-Oriented Medical Information System (PROMIS[1]) have yet to gain widespread acceptance. The early PROMIS system imposed constraints on user's input data - system design directed clinicians to become more complete, rigorous, and systematic in documenting clinical information - one of the most persistent complaints against the system. Moreover, PROMIS required change in all levels of health care delivery and forced replacement of existing practice rather than functioning as an alternative or supplement [2].

However, POMR as an idea offers an intuitive and useful way to work with the patient record by structuring information related to a patient's medical problem into a unit,

---

[1] Medical Center Hospital in New England, USA.

providing a context for dealing with medical problems, improving efficiency, and supporting continuity of care. Despite less successful experience with computerized POMR [3], we believe it possible to overcome some of the known deficiencies - such as enforcing strict and thorough data entry - with a knowledge-based approach to implementation.

2 Traditional Patient Records versus Problem-Oriented Patient Records

Many patient record systems that exist today contain mostly unstructured text – the text valuable in different circumstances of patient care (Table 1). To extract information from the record relevant for decisions that appear consumes time and because physicians work under severe time constraints [4], they have little time left to search for this information. Although, clinicians benefit from efficiency gains of the computerized patient record system, efficiency gains are not enough. A record system should also support physicians in geting overview of and access to information in the record.

Table 1 A record entry form a time-oriented patient record. Information in the record is not easily reusable because of the unstructured text.

> 131291JTA Hypertension treatment for several years, tried without medication for a period this spring, but was suffering from headaches and started again with TERNORMIN. BP 160/90 puls regular 62. Indicates a little dizziness, particularly when dark. Ophthalmoscope - no papilledema, bleeding or exudates, some minor blood vessel alterations.
>
> ---HYPERTENSION ESSENTIAL NOS
>
> ---HYPOTHYROIDISM
>
> RP: THYROXIN NA TBL 0.05 mg No:100 Pck.6 Reit:3
>
> Pain left side abdominal-thorax evenings after going to bed, increasing last half year, particularly after meals. Pressure. Variable dyspepsia for years. Clin.ex.: Cor-pulm NA. Abd: BZ feces NA
>
> ---STOMACH ACHE UNSPECIFIED
>
> Lab.test sent to hospital STOMACH ACHE UNSPECIFIED

2.1 Potential for Improving Existing Patient Record System

Below, we look at some limitations of an existing patient record system in order to come up with some requirements for improving a future patient record system. Table 1 is an excerpt from a Norwegian text-based patient record system[2]:

1. The text can be characterized as rich and extensive in words. Usually, the record contains several pages of text. Many implementations of patient record systems have time orientation [5] - recording events in order of appearance. Information is, more or less, only accessible to the author, or information owner, who knows where to search.
2. The patient record is built around a principal free-text area to document patients' complaints, symptoms, signs, clinical findings, diagnoses, prescriptions of medication, laboratory tests, and other information pertinent to the specific medical problem at hand. The free-text area contains a lot of information, but with very limited possibilities to extract parts of it.

---

[2] We base most of our references to record systems on experience from Norwegian primary care patient record systems. Typically, physicians themselves enter information into the patient record system during patient care. Therefore, we consider physicians as both information owner and information user, thus, simplifying issues related to security and privacy of information. In general, clinicians in primary care have read access to all patient-specific data from the patient record.

3. The record system offers separate windows to document patients' family and social history but to record this information requires extra effort in moving away from the main record. Physicians tend to integrate family and social history into the free-text area relevant for the medical problem being treated at the time of writing. The record system structure appears as rigid and seems to force physicians to work in non-optimal ways.
4. Interleaving medical problems impede overview and, thereby, also continuity of care. A chronological record mixes descriptions of different medical problems. The process of getting an overview of previous history of a medical problem or to view simultaneously documents related to a specific problem either consumes time or does not exist. Due to lack of overview and continuity of care, the record contains both redundant and confusing information, complicating work with the patient record unnecessarily. Evidence from a study we made of several patient records indicate that[3]:
    a. Many of the diagnoses used in the record relate to different manifestations of the same medical problem. The variety of diagnoses creates confusion and difficulties in getting an overview of a patient's major medical problems
    b. Much of the text documented is redundant, making the record lengthy. We consider the overhead a result of not having the possibility to view medical problems as a unit (e.g. as an Episode of Care (EoC)).
5. When activating the patient record, it highlights text from previous encounter, which implies that the clinician has to scroll backwards to recapitulate previous documents if treating a problem not related to the previous encounter.

Table 2 Example of a problem list and variations in diagnoses given to a patient. The problem orientation provides a structure for layering information; the problem name function as a high-level concept, while the set of diagnoses represent a lower-level description of the patient's condition.

Problem	**Variations in diagnose**
Myalgia neck/shoulder	Stiff neck
	Tension headache
	Headache
	Shoulder syndrome
	Tendinitis shoulder
	Myalgia neck
	Myalgia shoulder

2.2 Requirements for a Future Patient Record System

We relate requirements for a problem-oriented patient record to some of the limitations listed in previous section:

1. A problem-oriented patient record should present levels of relevant information. For lengthy records, not all information is useful at all times, information have degrees of relevance. Information in patient record should be ranked according to its relevance in a given situation, but at the same time leave other information easily accessible to the user if needed or wanted.
2. A problem-oriented patient record should also provide physicians flexibility when working with record information. Table 2 is an example of how information can be structured and divided into layers of relevant information; the ``problem" is a

---

[3] At the time of writing, preparation of results from study are in progress.

high-level concept, while the set of diagnoses represent a lower-level description of the patient’s condition.

3. A problem-oriented patient record system should not force physicians to work in formal ways.
4. EoC represents a core concept in general practice, initiated by the first encounter with a health care provider and completed with the last encounter. Immediate follow up from previous encounter of a problem secures continuity of care. In contrast to chronological records, an EoC-supported record system enables easy access to and overview of information for medical problems.
5. A system should adapt to clinician's situation (i.e. process awareness) by incorporating knowledge of clinicians' work processes into the record system.

3 Toward a Problem-Oriented Patient Record

One key to a successful implementation of a problem-oriented patient record is a system with knowledge of processes and abilities to rank information as relevant or irrelevant to a situation. We propose a two-layer framework for a problem-oriented patient record that aims to minimize the transition between how physicians work and how the record system forces the user to work:

1. A knowledge model: A record system with knowledge of its content and use empowers the record system to recognize where in the process physicians are and determine corresponding information use and need.
2. A relevance model: Information is ranked according to its relevance in a given situation (i.e. a medical problem), enabling the system to present levels of relevant information.

2.3 Ontologies for Building Knowledge Base

As one of our goals has been to design an architecture that enables the record systems to extract information that is more relevant than other from the complete patient record, our approach has been to identify and classify clinicians' interaction with the patient record.

We use the Protégé-2000 system to develop an ontology for a patient record system that link clinicians' work processes and activities with the use and need for actual clinical data in the patient record. Protégé-2000 comprise a set of knowledge-acquisition tools for developing ontologies, it supports a graphical tool, and has query abilities to assist in discovering the relationships between concepts and participating components [6].

Ontologies define a framework for organizing concepts in a knowledge base [5]. The purpose of our ontology has been to identify *informationsources* that can be ranked according to some criteria of relevance and comprises three sub ontologies.

1. **A process ontology** that identifies the primary care workflow; i.e. the way physicians work and their corresponding information need related to each activity
2. **A patient record ontology** that captures the content of the patient record; i.e. clinical terms
3. **An information ontology** that functions as a bridge between the process ontology and the patient record ontology. The information ontology implements the concepts of problem orientation.

Merging the three ontologies enable us to recognize the intersection between the workflow ontology, the information ontology, and the patient record ontology. *Informationsources* distinguish activities in the primary care process and clinicians' information need within that activity with corresponding clinical terms in the patient record (Table 3). *Informationsources* is a nametag for the set of patient-specific data that relates to an activity. The selection of relevant *informationsources* for each activity is in itself a ranking of the record content. Yet, a remaining challenge is to develop measures for ranking these sources.

Table 3 *Informationsources* for two activities. *Informationsources* comprise a set of information considered as relevant for an activity.

Action	Information Source
Determine assessment actions	Problem list
	Problem history
	Related problem list
Clinical examination	Clinical examination history
	Clinical examination history for related problems

4 Conclusion and Future Work

Among the most useful contributions of patient record systems is, in particular, the ability to support physicians in their decision-making process. The information in the patient record serves many purposes, but one of the most important ones is the linking of patient data and medical knowledge. We aim to assist physicians in their decision-making process by presenting relevant information and better utilization of information in the record system. We believe that the idea of problem-oriented patient records is still an intuitive and useful way to organize the patient record – despite previous less successful implementations of it. A problem-oriented patient record system provide a foundation for a well-structured patient record that 1) supports physicians' way of thinking and working, 2) removes redundant and confusing information, and 3) enables easier access to and overview of information in the record.

References

[1] L.L. Weed. Medical records that guide and teach. *The New England Journal of Medicine*, 278(12), March 21, 1968.

[2] H.P. Lundsgaarde, P.J. Fischer, and D.J. Steele. Human problems in computerized medicine. University of Kansas Publications is Anthropology Number 12, Lawrence, Kansas, 1981.

[3] M. Bainbridge, P. Salmon, A. Rappaport, G. Hayes, J. Williams, and S. Teasdale. The problem oriented medical record – just a little more structure to help the world go round? Clinical Computing Special Interest Group (CLICSIG) of the PHCSG, 1996.

[4] L.L. Weed. Medical records, medical education, and patient care. The Press of Western Reserve University, 5. printing ed., 1971.

[5] J.H. van Bemmel and M.A. Musen. Handbook of medical informatics. Springer, 1997.

[6] M. Peleg, I. Yeh, and R.B. Altman. Modeling biological processes using workflow and petri net models. *Bioinformatics* (in press).

Acknowledgment

Samson Tu and Mark A. Musen, Stanford Medical Informatics, Stanford University School of Medicine; Erik Naggum; and Anders Grimsmo, Department of Community Medicine and General Practice, Norwegian University of Science and Technology. Funding of this paper is by the Norwegian Research Council.

Chapter 11

The Helpful Patient Record System: Problem Oriented and Knowledge Based

The helpful Patient Record System:
Problem Oriented And Knowledge Based

Elisabeth Bayegan, MS[1] and Samson Tu, MS[2]

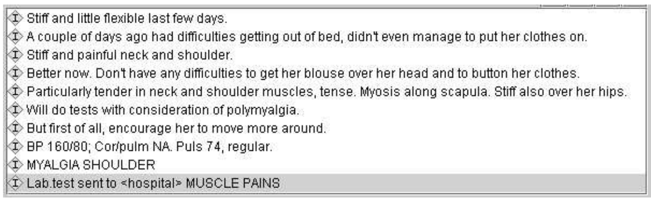
[1]Norwegian University of Science and Technology, Department of Computer and Information Science and Department of Community Medicine and General Practice, Trondheim, Norway
[2]Stanford Medical Informatics, Stanford University School of Medicine, Stanford, CA

ABSTRACT

In contrast to existing computerized patient record systems, which merely offer static functionality for storage and presentation, a helpful patient record system is a problem-oriented, knowledge-based system which provides the clinician with situation-specific information from the patient record, relevant to the activity within the patient care process. We suggest extending the data model of current patient record systems with (1) knowledge for recognizing and interpreting care situations, (2) knowledge of how clinicians work and what information they need, and (3) means to rank information according to its relevance in a given situation. We present a framework that enables representation of three prerequisite features for a future helpful patient record system: the primary care workflow process, the problem-oriented information model, and means to identify relevant information to the care process and medical decisions.

INTRODUCTION

The linking of patient-specific data with medical knowledge is crucial for efficient and high-quality patient care. However, the task of extracting patient-specific data that are both useful and necessary for medical decisions is not a trivial task. Typically, today's patient record information model is time oriented; patient data are related to a unique patient ID and date of entry. Such a rigid model prevents easy access to and overview over information in the record, for instance to view patient data in relation to a specific medical problem.



Stiff and little flexible last few days.
A couple of days ago had difficulties getting out of bed, didn't even manage to put her clothes on.
Stiff and painful neck and shoulder.
Better now. Don't have any difficulties to get her blouse over her head and to button her clothes.
Particularly tender in neck and shoulder muscles, tense. Myosis along scapula. Stiff also over her hips.
Will do tests with consideration of polymyalgia.
But first of all, encourage her to move more around.
BP 160/80; Cor/pulm NA. Puls 74, regular.
MYALGIA SHOULDER
Lab.test sent to <hospital> MUSCLE PAINS

Figure 1 A record entry from a time-oriented patient record. Information in the entry is not easily reusable because of the unstructured text.

The patient record contains a considerable amount of information valuable in different circumstances, but as the situation is today, most of the recorded clinical information in patient record systems[a] is unstructured text. Record entries (Figure 1) are composed of descriptions of patient's symptoms, signs, complaints, reason for encounter, personal, family, and social history, examination findings, etc. In addition, patients usually present more than one problem; so individual notes may contain information on many interleaving problems.

It is time-consuming to extract information from the record relevant to decisions that must be made during patient care. Clinicians work under severe time constraints leaving little time left for them to search for information. Although, clinicians benefit from efficiency gains of computerized patient record systems, efficiency gains are not enough:

- Clinicians lose overview over record content after a few years of a long-term patient-clinician relationship
- The patient record accumulates over time and will contain information on many medical problems that, eventually, will have some relationship to one another.

The Problem-Oriented Patient Record

Lawrence L. Weed introduced the Problem-Oriented Medical Record (POMR) in the late 1960s with intentions of improving the structure of medical records [1]. Implementations of POMR such as the Problem-Oriented Medical Information System (PROMIS[b]) have yet to find widespread acceptance. Of the most persistent complaints against PROMIS are constraints on input data - the system was designed to direct clinicians to become more complete, rigorous, and systematic in documenting clinical information. PROMIS required change in all levels of health care delivery and forced replacement of existing practices, rather than function as an alternative or supplement [2].

The Helpful Patient Record System

POMR as a concept offers an intuitive and useful way to work with the patient record by structuring information

---

[a] Most references to patient record system are based on experience from Norwegian primary care patient record systems.
[b] Medical Center Hospital.

related to a patient's medical problem into a unit, providing a context for dealing with medical problems, improving efficiency, and supporting continuity of care. Despite less successful experience with computerized POMR, we believe it is possible to overcome some of the known deficiencies - such as enforcing strict and thorough data entry - with a knowledge-based approach to implementation. We suggest extending the current patient record data model with knowledge that enables the system to reuse information that is already in the record, in situations where it is useful and needed. Our framework incorporates three features:

- **Process knowledge** of clinicians' work processes empowers the record system to recognize at what stage in the process they are and to determine relevant information needed at that stage. Our objective is not to take part in diagnostics of health problems but to contribute and indicate potentially relevant information for decision-making by taking advantage of information that already exists in the patient record. Adding process knowledge enables the system to adjust to the user and not vice versa.
- **A problem-oriented information model**, which structures related information into a problem unit. In contrast to POMR, our sense of a problem-oriented patient record (POPR[c]) has no mandatory binding to the classic SOAP[d] format. The model is flexible and has no absolute restriction that every entry in the record must relate to a problem; some entries have no natural relation to a medical problem, while other entries are obvious problems and will, as a consequence, relate to a problem. Flexibility of the model lies in the fact that every record entry consists of a set of information units each labeled with a clinical heading, instead of linking record units to medical problems. The model handles uncertainty in early stages of a patient's medical problem and grouping of problems as sub problems, by making it possible to link record entries at a later stage, if and when they appear as (sub) problems.
- **Relevance ranking of a set of *information sources*.** For each activity in the primary care process we identify a set of information sources as relevant to that activity. The ranking that is done is based on: (1) *problem orientation* which abstracts information relevant to a medical problem and 2) a set of *clinical headings* that link information from the patient record to activities in the process model. Both provide a robust tool for navigating among information in the patient record.

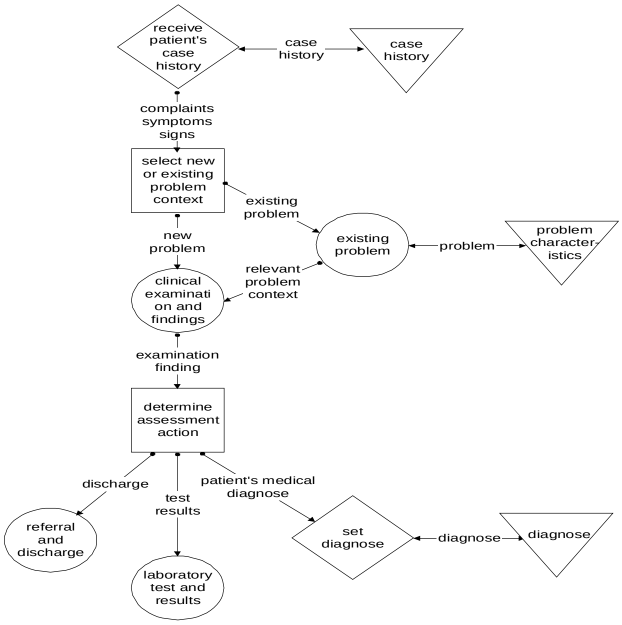


Figure 2 Modeling of a generic assessment (sub) process to determine interaction with the patient record system. Diamond shapes represent activities, circles sub processes, squares decision boxes, and triangles information sources. Information sources contain information from the patient record system relevant to the related activity.

METHODS

To develop our framework of process knowledge, a problem-oriented information model, and means to rank information in the patient record, we have used the Protége-2000 system. Protége-2000 is a set of knowledge-acquisition tools for implementing ontologies and knowledge bases. Besides being a tool for defining declarative domain ontologies, Protége-2000 supports a graphical tool, has query abilities, and produces machine-interpretable data structures [3]. We use Protége-2000 to make an ontology and to model the primary care process and corresponding interaction with the patient record system. The ontology comprises problem-oriented concepts, clinical headings, and definitions of information sources.

Modeling The Primary Care Process

As our aim is to extract from the complete patient record information that is relevant to patient-care situations our system requires knowledge of how to recognize and interpret these care situations and knowledge of what information is useful and necessary to each of them. We model the primary care process for the purpose of acquiring knowledge of each activity in the process and corresponding interaction with the patient record system, which further provide us with insight on information needed by clinicians for each activity in the

---

[c] We use POPR instead of POMR to distinguish our conception of a problem-oriented record from the classic POMR and SOAP format.

[d] Subjective, Objective, Assessment, and Plan.

care process. Moreover, a system that is able to recognize and interpret situations and provide with useful and relevant information to each situation will minimize the traditional gap between the way clinicians work and how the system compels the user to work.

The primary care process is modeled with an ontology but also represented graphically (Figure 2). Conceptually, we can stratify ontologies; lower levels define generic concepts while higher-level strata have more specialized concepts [4]. Generic concepts for our process ontology are based on the workflow model proposed by the Workflow Management Coalition [5] and modified for our purpose. The generic workflow model is applicable to all kinds of processes, from business processes to different processes in the medical domain [6]; it defines generic concepts, such as processes, activities within each process, necessary tools and data for each activity, participating actors, and transition requirements from one activity to another.

A graphic representation of the primary care process enables us to visualize an activity and corresponding interaction with the patient record, so as to identify a set of *information sources* for each activity (see Figure 2). Information sources are shown as triangles in our graph; presenting an activity's set of relevant information from the patient record.

Formalizing Patient Data

Part of the ontology comprises a set of clinical headings to formalize patient data in the record. One of NHS Clinical Heading Project's objectives has been to identify a limited number of headings that will perform the jobs that are needed for clinical communication and navigation and that can be agreed upon by every participant involved in clinical communication [7]. The result is a well-established framework of clinical headings compatible with both HL7 RIM subject areas [8] and CEN TC251 standardization work on electronic patient record communication. *Clinical Headings* refer to a set of words and phrases that clinicians use to name sections of their communications (e.g. symptoms, examination findings, test results, diagnosis, etc.). Clinical headings have no formal syntax or definition, so our ontology represents the formal data structure needed to strictly define concepts in the patient record.

Most of the content in the patient record is free text, despite this fact there is an interesting feature that can be used to our advantage when structuring the content of the patient record. Through our study of electronic patient records, it was evident that clinical headings match our expectations for encoding patient data; to match clinicians' information need in each activity with specific content in the patient record. Sentences, or part of sentences, correlate easily to a clinical heading and clinical headings again parallels directly with activities in the primary care process. As a result, we do not need to encode every element of the patient record it is sufficient to encode a set of sentences, which relate to a clinical heading. Thus, clinical headings function as a bridge between information needed and the process activity it is required within.

The Problem-Oriented Information Model

Figure 3 illustrates the problem-oriented information model and show how clinical headings act as a bridge between patient record content (*information units* in the figure) and activities - linking concrete information from the patient record with specific activities in the primary care process.

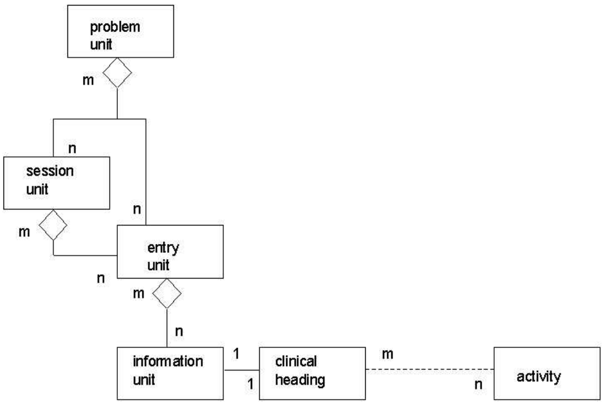


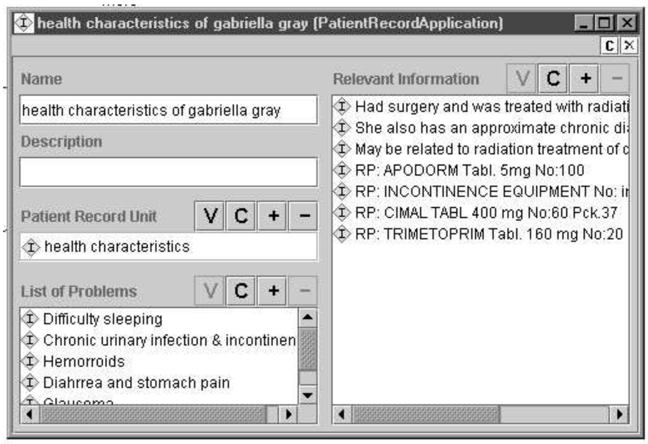
Figure 3 The problem-oriented information model. The basic unit is the *information unit.* Each information unit is labeled with a clinical heading, which bridge patient record content to activities in the process model.

As seen in figure 3, we have maintained the classic entry in time-oriented patient records (*entry unit* in our model) but have added two other concepts such as *session unit* and *problem unit* to the information model. An entry (see figure 1) contains a summary of problems, actions, clinical examinations and findings, laboratory tests ordered, etc. Even though the underlying information model is problem oriented the strength of our model lies in the labeling of information units, each information unit is labeled with a clinical heading, which in turn correlates to activities in the primary care process. The difference between classic record entries and our entry unit is that an entry unit now consists of a set of information units that each are labeled with a clinical heading that specify its content and purpose. The capturing of content in each information unit enables us to reuse them in different circumstances. Information units are the basic building block of our information model and can be a set of sentences, a sentence or part of a sentence. The clinical heading that labels the information unit defines the content of it.

Furthermore, a problem unit comprises a set of sessions or entries, or both. An entry unit contains a set of information units that each belong to different clinical headings such as symptoms, family history, clinical findings, test results, diagnose, medical treatment, etc. A set of related entry units make up a session unit (i.e. a discharge entry has a corresponding referral entry for which the referral was specified). An information unit, an entry unit, or a session unit, may belong to more than one entry unit, session unit, or problem unit, respectively. As a consequence, the model is flexible, an entry may function as a separate unit or it may be related to one or several problems.

Ranking Information Sources

For each activity in the primary care process a set of information sources are identified. Information sources comprise relevant information needed by clinicians for a particular activity, presented in levels according to their situational relevance. Currently, we use topical relevance as criteria for determining relevant information sources for an activity (relevance criteria for clinical headings will be discussed). The underlying problem-oriented information model - combined with a set of clinical headings - enables our framework to present these information sources. Ranking of information as relevant is based on:



Name: health characteristics of gabriella gray

Patient Record Unit: health characteristics

List of Problems:
- Difficulty sleeping
- Chronic urinary infection & incontinen
- Hemorroids
- Diahrrea and stomach pain
- Glaucoma

Relevant Information:
- Had surgery and was treated with radiati
- She also has an approximate chronic di
- May be related to radiation treatment of c
- RP: APODORM Tabl. 5mg No:100
- RP: INCONTINENCE EQUIPMENT No: i
- RP: CIMAL TABL 400 mg No:60 Pck.37
- RP: TRIMETOPRIM Tabl. 160 mg No:20

Figure 4 An information source for a specific activity. *List of Problems* and *Relevant Information* show two distinct levels of relevant information. Selecting a problem from the list will display problem-specific information.

1. The problem-oriented information model that structures information from the patient record related to a medical problem within a unit. To organize information into problems is in itself an abstraction of relevant information.
2. Clinical headings that encode the patient record content. Labeling of information units into headings, combined with the problem orientation, provide the record system with a robust tool for navigating in patient data.

RESULTS

To evaluate our framework, we have split the evaluation in two: (1) evaluation of clinical headings and the problem-oriented information model and (2) evaluation of the process model, activities, and their attendant set of information sources.

Transformation of a Time-Oriented Patient Record

To evaluate the problem-oriented information model, we were particularly interested in discovering how well our model would fit time-oriented entries into problem units and if the set of clinical headings was sufficient to encode the patient record content. We transformed a time-oriented patient record[e] to a problem-oriented patient record - using our problem-oriented information model. Our problem-oriented knowledge base contains 267 information units, 112 record entries, 26 medical problems, and 46 diagnoses. The set of problems belong to three different patients. 13 clinicians - including physicians, nurses, and laboratory assistants - have been treating these three patients over a period of 9 years, from primo 1991 to mid 2000. Results from the transformation demonstrated that:

- Clinicians do not enter information into the record in a strict order; entries could be classified into a set of finite clinical headings.
- Clinical headings were sufficient labels of information units; they encoded patient data and linked directly to activities in the process model
- Some entries did not naturally relate to a medical problem, while others had an obvious relation to a medical problem. Our model was flexible and permitted the option of an entry to function as a separate unit, or relate to one or several problems.
- The record contained redundant information, which we interpreted as:
    - An argument for the problem-oriented patient record. The problem-oriented patient record provides a context in which continuity of care is supported, so as to prevent both redundant information and repeated actions
    - Emphasis on what type of information clinicians need; we will use this fact to develop ranking criteria for information as relevant or not to a situation
    - An indicator that information which clinicians needed during patient care, *was* in the record but they did not have the time to search for it.

---

[e] The time-oriented patient record system we used includes more than 1000 patient records over a 9-year period and around 30 clinicians operating (reading/writing) on those records.

Simulation of the Primary Care Process

We conducted a simulation of our process model to see how patient-specific encounters were captured and what information was determined as relevant to each activity within the encounter (i.e. information sources). For a set of patient encounters, we simulated step-by-step activities in the care process and evaluated the set of information sources (i.e. a simulation of the set of queries performed to abstract patient data relevant for each activity). We performed 17 simulations, which resulted in 31 information sources, such as the one illustrated in Figure 4. Figure 4 demonstrates an information source with information relevant to "assessment of myalgia shoulder". There are two levels of information presented in the information source: *List of Problems* and *Relevant Information*. A third level is accessible by selecting a concrete problem from the problem list. Each problem contains a set of relevant information to that specific problem.

DISCUSSION AND FUTURE WORK

The evaluation of our work so far has been formative with intents to validate and ensure that the three components in our framework are sound. Ultimately, we want to evaluate our work in a clinical setting in terms of a system that is able to present a problem-oriented patient record capable of following the work process of clinicians and provide patient-specific information from the record that is useful and necessary for medical decisions to be made. Before reaching this point - a fully implemented system - there are several intermediary steps that have to be addressed and completed. One next step is to continue our simulation with a group of clinicians to clearly show our framework's potential and feasibility; as a first step we have to demonstrate and validate the set of information sources as both relevant and useful to clinicians.

On the development side there is great potential for extending our knowledge model with knowledge and ranking criteria that enable the system to do more specific ranking than has been expressed. However, the three levels of information that our information sources now are able to present represent a substantial improvement from the chronological record provided in many patient record systems today. The three levels of information can be further ranked in degrees of relevance to a specific situation. Right now, there is no ranking among clinical headings, for instance there are no options to rank among family history, clinical findings, or prescribed medication. When a clinical heading - i.e. prescribed medication - is ranked as relevant to an activity, the whole set of data for that clinical heading is show (e.g. both past and present medication). We will continue to work on heuristics that enable our system to rank among clinical headings according to problem and situation. A further challenge is to decide what knowledge is sufficient for the record system to be able to present a list of relevant problems instead of presenting the whole problem list.

CONCLUSION

Although POPR as a concept offers an intuitive structure and organization of patient records, implementing a solution that will gain widespread acceptance among clinicians is another matter. Most decision-support systems up until now tend to - inscribed - outline who or what the relevant elements are and what their respective roles consist of [9]. Introduction of a system into an existing practice should be a process of continual negotiation rather than enforcing strict change. As mentioned in the introduction, our knowledge-based approach to system implementation is suggested as an alternative to meet some of the challenges that exist in developing decision-support systems for health care. We argue in favor of a knowledge-based problem-oriented system as the key to a successful implementation of a patient record. Problem orientation is the foundation for a well-structured record which supports the way clinicians work, removes redundant and confusing information, and enables easier access to and overview over information. A system that has knowledge of its content and use is better able to integrate with continual practice by being non intrusive.

REFERENCES

1. Weed LL. Medical records that guide and teach. MD Computing, 10(2): 100-114, March/April 1993.
2. Lundsgaarde HP, Fischer PJ, and Steele DJ. Human problems in computerized medicine. University of Kansas Publications in Anthropology Number 13, Lawrence, Kansas, 1981.
3. Musen MA. Medical informatics: searching for underlying components. Methods of Information in Medicine. In press.
4. Beale T. Archetypes constraint-based domain models for future information systems. Technical report, 2000.
5. The Workflow Management Coalition Specification. Workflow management coalition interface 1: process definition interchange process model. Technical report, 1999.
6. Quaglini S, Stefanelli M, Lanzola G, Caporusso V, and Panzarasa S. Flexible guideline-based patient care flow systems. Artificial Intelligence in Medicine, 22:65-80, 2001.
7. Kay S et al. Headings for communicating clinical information. Technical report, Salford Research Group, NHS Research report, Spring 2001.
8. Beeler G et al. HL7 Reference Information Model. Version V 01-12 (1/2/2002), ModelID: RIM_0112. HL7 Version 3 Standard.
9. Berg M. Rationalizing medical work: decision-support techniques and medical practice. The MIT Press, 1997.

Acknowledgement

Mark A. Musen, Stanford Medical Informatics, Erik Naggum, Anders Grimsmo, NTNU, Department of Community Medicine and General Practice and Øystein Nytrø, NTNU, Department of Computer and Information Science.

Chapter 12

Ontologies for Knowledge Representation in the Computer-Based Patient Record System

Ontologies for Knowledge Representation in a Computer-Based Patient Record

Elisabeth Bayegan, Øystein Nytrø and Anders Grimsmo
Norwegian University of Science and Technology
Department of Computer and Information Science
Department of Community Medicine and General Practice
NO-7491 Trondheim, Norway
{*bayegan,nytroe*}*@idi.ntnu.no; grimsmo@medisin.ntnu.no*

Abstract

In contrast to existing patient-record systems, which merely offer static applications for storage and presentation, a helpful *patient-record system is a problem-oriented, knowledge-based system, which provides clinicians with situation-dependent information. We propose a practical approach to extend the current data model with (1) means to recognize and interpret situations, (2) knowledge of how clinicians work and what information they need, and (3) means to rank information according to its relevance in a given care situation. Following the methodology of second-generation knowledge-based systems, that use ontologies to define fundamental concepts, their properties, and interrelationships within a particular domain, we present an ontology that supports three prerequisite features for a future helpful patient-record system: a family-care workflow process, a problem-oriented patient record, and means to identify relevant information to the care process and medical problems.*

1. Introduction

In contrast to existing computerized patient-record systems, which merely offer static applications for storage and presentation, a *helpful* patient-record system is a problem-oriented, knowledge-based system which provides the clinician with situation-sensitive information from the patient record, relevant to the activity within a patient-care process. The linking of patient-specific data with medical knowledge is crucial for efficient and high-quality patient care. However, the task of extracting patient-specific data that is both useful and necessary for the medical decision to be made is not a trivial task. The most common information model in patient-record systems today has time orientation—patient data relates to a unique patient ID and date of entry. Such a rigid architecture does not allow for easy access to and overview over information in the record, for instance, to view patient data related to a medical problem.

Table 1. Extract from a patient record.

> 131291JTA Hypertension treatment for several years, tried without medication for a period this spring, but was suffering from headaches and started again with TERNORMIN. BP 160/90 puls regular 62.Indicates a little dizziness, particularly when dark. Ophthalmoscopy: no papilledema, bleeding or exudates, some minor blood vessel alterations
> —HYPERTENSION ESSENTIAL NOS
> —HYPOTHYROIDISM
> RP: THYROXIN NA TBL 0.05 mg No:100 Pck.6 Reit:3
> Pain left side abdominal-thorax evenings after going to bed increasing last half year, particularly after meals. Pressure.Variable dyspepsia for years. Clin.ex.: Corpulm NA. Abd: BZ feces NA
> —STOMACH ACHE UNSPECIFIED
> Lab.test to hospital STOMACH ACHE UNSPECIFIED

Most of the recorded clinical information in current patient record system[1] is unstructured text, containing a lot of information valuable in different circumstances (Table 1). Information entries contain description of patients symptoms, signs, complaints, reason for contact, results from clinical examination, incomplete discharge information, etc. Clinicians also tend to integrate family history and social circumstances in the text area. In addition, patients usually present more than one problem while seeing a clinician, so the text contains information on many interleaving problems. It is time consuming to extract information from the record that is relevant for decisions that appear during patient care. Because clinicians work under severe time constraints, there is little time left for clinicians to search for relevant information. Although, clinicians benefit from

---

[1] For example, record entries in Norwegian family-care patient-record systems.

efficiency gains of the electronic patient-record systems, efficiency gains are not enough:

- Clinicians lose overview of record content after a few years of a long-term patient-clinician relationship.
- The patient record accumulates over time and will contain information on many medical problems that, eventually, will have some relationship to one another.

Previous efforts in making intelligent software systems for clinicians have focused on expert systems that provide alerts and reminders, automate or aid in diagnostics. Expert system's practical application have been relatively poor because they attempt to apply medical knowledge out of its context [6]. Expert systems have usually been stand-alone systems and not integrated as part of the patient-record system, they also had poor support for organizational factors such as process and workflow knowledge. These experience have resulted in a shift in development of computer systems for clinical decision making in such a way that they now aim to provide with simple, but value-added diagnostic assistance. Experience in previous system development has proved that diagnostic system—in general—cannot replace clinicians' medical knowledge.

We present a practical approach for improving the patient-record architecture. We extend, conceptually, an existing patient-record data model with a level of knowledge; knowledge that enables the patient-record system to interpret situations in the care process and provide means for ranking information according to its relevance in these care situations. Ability to rank information implies knowledge about its content and use and means for recognizing situations. We, therefore, have added knowledge about clinicians' work processes and activities, their problems to be solved for each activity (an activity's goal or intention), and what information they need during each activity. The knowledge level is proposed through a two-layer framework for a problem-oriented patient record, a fact layer and a knowledge layer. The knowledge layer comprises two models:

a knowledge model of the record system's content and use, which empowers the record system to recognize where in the process physicians are and determine corresponding information need. We present an ontology that integrates a workflow model (a process ontology) with a problem-oriented knowledge model (an information ontology) and a model of the patient-record content (a content ontology). The content ontology functions as a bridge between the workflow and information ontology (that contains patient data) and identify a set of *information sources*, through the linking of each activity's information needs with contents in the patient record.

a relevance model that ranks information according to its relevance in a given care situation. The integration of the three ontologies in the knowledge model identify a set of *information sources* that can be ranked according to some criteria for ranking. Information relevance depends on the care situations and we propose to present information as *levels of relevance* according to different care situations. Information may have degrees of relevance directed by activity (and corresponding goal or intention), a specific medical problem, patient-specific data, and a utility function (for instance, that specify relevance between medical problems). The ranking of information presented in this paper is based on relevance to activity and type of problem (problem orientation).

Figure 1 illustrates the relationship between the knowledge and the relevance model. The three ontologies that the knowledge level comprises identify a set of information sources from the patient record and is considered relevant to a specific care situation. The relevance model ranks these information sources (patient-specific data) according to some relevance criteria. The end result contains a set of ranked information sources from the patient record, thus relevant information to a specific care situation from the patient record.

The three ontologies (i.e., process, information, and content) have been drawn as knowledge bases because they, in the figure, also contain *domain structure and concept instances*. A knowledge base contains an ontology and an associated set of instances [1]. The instances make up the leafs of the classification tree (i.e., the ontology). For example, the knowledge base in the middle of the figure—the patient knowledge base—contains the information ontology, which specify the *knowledge structure*, and patient-specific data; the ontology and the patient data together make up the *patient-record knowledge base*; thus our knowledge base does not include problem-solving or inference mechanisms.

2. An Epistemology of Patient Records and Clinical Data

From a historical perspective, back in the early days of building knowledge-based systems, McCarthy and Hayes [5], in 1969, divided the non-trivial task of representing intelligence in computer systems in two: epistemologic adequacy and heuristic adequacy. Epistemologic adequacy is the study of what kind of facts are available of the world and how to draw conclusions from these facts (i.e., how to represent the world in such a way that solutions to a problem follow from facts expressed in the representation). Heuristic adequacy is the study of mechanisms that decide what to

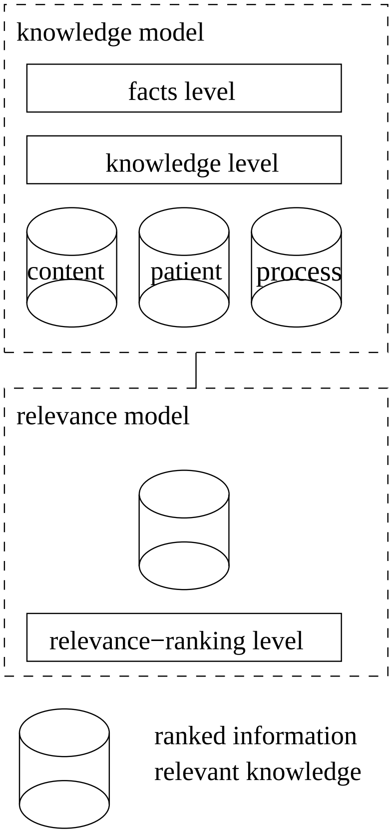


Figure 1. Relationship between the knowledge model and the relevance model.

do and how to solve a problem. In other words, the former view emphasizes knowledge required to solve a problem, the latter view emphasizes processing mechanisms required to bring the epistemic knowledge into real. We focus, in this paper, on the epistemologic adequacy for a patient-record system.

From a philosophical perspective, the complexity of the world is overwhelming and *informaticians* need advice in deciding what in the world to attend to and what to ignore. Designing a knowledge representation involves making a set of decisions about how and what to see of the world (i.e., a conceptualization). A conceptualization is a way of conceiving the world and deciding what to model in a knowledge representation. Davis et al. [3] further relate knowledge representation to ontologies, by stating that, the selection of a representation means making a set of ontological commitments. The ontology determines what we can see of the world, bringing some parts of the world into focus while ignoring others [16]. Moreover, ontology development is a way of reducing the complexity of the world to model; ontologies represent the domain structure of knowledge and enables knowledge sharing [2].

From a pragmatical perspective—after more than 30 years of semi-successful experience in building knowledge-based system—principles for a new discipline have emerged—a discipline for second-generation knowledge-based systems [8]. The second-generation engineering discipline includes knowledge modeling, ontology design and development, and problem-solving methods for applying this knowledge. The emergence of second-generation knowledge-based systems provide more explicit and maintainable data structures—ontologies—for encoding and applying (clinical) knowledge [7]. The knowledge-based system communities now seek practical solutions of how to integrate knowledge in computer systems and ontologies define a framework for organizing concepts in a knowledge base [14]. Methods of knowledge engineering—what to encode and what to say about the world—have evolved, such as CommonKADS [12] and the Protégé-2000 system [9, 13]. These methods introduce a practical approach on how to incorporate knowledge into a representation. We have used both CommonKADS and Protégé in our work, to develop an ontology and a knowledge base for a patient-record system.

2.1. Protégé-2000

The Protégé-2000 system contains a set of knowledge-acquisition tools for developing ontologies and knowledge bases. The Protégé-2000 system supports a graphical tool and has query abilities to assist in discovering the relationships between concepts and participating components; the system also produces machine-interpretable data structures [8]. We have used the Protégé-2000 system to develop an ontology for a patient record system that link clinicians' work processes and activities with use of actual clinical data in the patient record. The ontology defines the knowledge structures in our knowledge base of patient-specific data. We further use this knowledge base to produce information sources that presents a set of ranked information from the record relevant to a specific activity in the care process.

3. Ontological Components

As our aim is to extract from the complete patient record information that is more relevant than other in a specific situation, our approach has been to identify and classify clinicians' interaction with the patient record. The process has been three fold; first we identified physicians' activities within the family-care process, then we identified their corresponding information needs related to each activity, and finally, we matched physicians' information needs with patient-specific data in the record. Our ontology comprises three sub ontologies:

a process ontology that defines the family-care workflow process, activities, corresponding participants for each activity, and their interaction with the patient-record system.

a content ontology that formalizes clinical terms in the patient record system and functions as a bridge between the process ontology and the information ontology. The ontology links concepts in the process ontology with concepts in the patient-record ontology (which make up information sources) through its set of *care act types*.

an information (or patient-record) ontology that defines the underlying information model of the patient-record system. The information ontology represents a model of patient-record data, which further, supports a problem-oriented view of patient data.

3.1. An Ontology for a Family-Care Process

The Workflow Management Coalition has defined a workflow model, initially for optimizing business processes, but which has gained acceptance in other fields, such as the medical domain [10, 11]. The generic workflow model has been translated into a workflow ontology, which has already demonstrated to model successfully biological processes [10]. The workflow ontology defines generic concepts, such as processes, activities within each process, necessary tools and data for each activity, participating actors, and transition requirements from one activity to another. We have used the workflow ontology to model a stereotypical family-care process, for the purpose of acquiring knowledge of each activity in the process and each activity's corresponding interaction with the patient-record system. The family-care process provides us with insight into clinicians' information needs for each activity in the care process.

Through our field study of clinicians, we identified some idiosyncrasies of the family-care process that we have used as guidelines when developing the family-care process ontology:

extensive use of the patient-record system: due to family-care physicians being both information *owner* and *user*. Family-care physicians use the patient record during all phases in the care process and have a high motivation for entering information into a patient record.

a generic family-care process: that comprises three main phases (i.e., assessment, diagnostics, and treatment). Each phase include a limited set of activities. Both phases and activities remain the same for each patient.

a limited set of remedies: Family-care physicians have a limited set of remedies, which can be classified into three categories: (1) *omnifarious*, for remedies used during all phases and activities, (2) *assessment*, for remedies used during the assessment phase (including diagnostics, and (3) *treatment*, for remedies used during the treatment phase.

traces: Each instance of a generic family-care process can be characterized by a patient's unique set of medical problems and conditions. For example, *Perform clinical examination* represents a generic activity in the family-care process. The result of a clinical examination—such as *Cor-pulm NA. Abd: BZ feces NA* (taken from Table 1)—represents a patient-specific result due to his or her specific and unique medical condition. The relationship between an activity and the result of an activity (its content) is projected as content (traces) in a patient record. The content in a patient record contains traces (action patterns) that represent the uniqueness of a patient's condition and, therefore, also represent the instantiation of a family-care process.

We model processes on three levels. The generic workflow process represents the highest level of modeling, while the second layer models a generic family-care process. The third and lowest level of modeling represents instances of the generic family-care process, where patients (and their medical condition) represents the specific instances of the process. A patient record defines the content of patient-specific process, thus, we can say that the activity in a family-care process has a projection of itself in a patient record. The patient record contains the content of each instance of a family-care process and further functions as "glue" between each encounter episode between patient and physicians (or clinician).

3.2. An Ontology for Patient-Record Content

The content ontology contains a set of concepts—care-act types—that define the content of the patient record and which further relate to activities in the process model. The content ontology functions as a bridge between the process ontology and the information ontology by linking an activity with patient-specific information.

Many existing patient records contain free text (unstructured text)—with varying degrees of structured text (semi-formal templates)—and there exists no consistent, uniform, or homogeneous use of clinical terms. We wish to obtain a way to standardize the use and interpretation of clinical terms, so, that the meaning of clinical terms will not be lost or misinterpreted during communication and navigation. Clinicians use clinical terms different from each other, or they use them to mean different things. Since we want to link clinicians' information need (identified in the workflow process) with specific clinical terms in the patient record; we need a data structure to establish a formal syntax for the clinical information in the patient records. We want the content ontology to serve the purpose of structuring the content of the patient record through its formal definitions of essential concepts. Our data structure makes it possible to integrate clinical terms, represented by care-act types, with the process ontology and the information ontology.

The NHS Clinical Headings Project's main objective has been to identify a limited number of headings that will perform the jobs that are needed for clinical communication and navigation and that can be agreed upon by every participant involved in clinical communication. The project produced a well-established framework of clinical headings compatible with HL7 and CEN TC251's standardization work on electronic patient records. *Clinical Headings* refer to a set of words and phrases that clinicians use to name sections of their communications (e.g., symptoms, examination findings, test results, diagnosis, etc.). Clinical headings have no formal syntax or definition, therefore, our content ontology represents the formal data structures needed to strictly define concepts in the patient record.

The content ontology is based on the clinical-headings framework, but we have extended and modified it for our objective—to integrate with the process and information ontology—for the intention of ranking information and presenting it to the clinician when useful and needed. The ontology (Figure 2) comprises four abstract classes and 19 sub-classes, which captures clinical information regarding: (1) a patient's health characteristics, (2) clinicians' set of actions, (3) time modifiers for clinical information, not included in the above two classes, but still needs to be recorded, and (4) different role views (e.g., time or problem orientation) that have been identified as necessary and useful for communicating clinical information. In the rest of this text, clinical headings now refer to clinical concepts, terms, and care-act types in the content ontology (akin classes and sub-classes).

Two classes (shown in Figure 2), *Assessment* and *Treatment*, subsume majority of the patient record content. Assessment encompasses all clinical terms that describe results of actions (e.g., *Record case history* or *Perform clinical examination*) and which further specifies the state of a patient (i.e., the results of the action, thus the content in the patient record). The class *Treatment* includes clinical terms that describe anything clinicians have done, based on assessment actions, to or for a patient regarding therapy (given a patient's diagnosis).

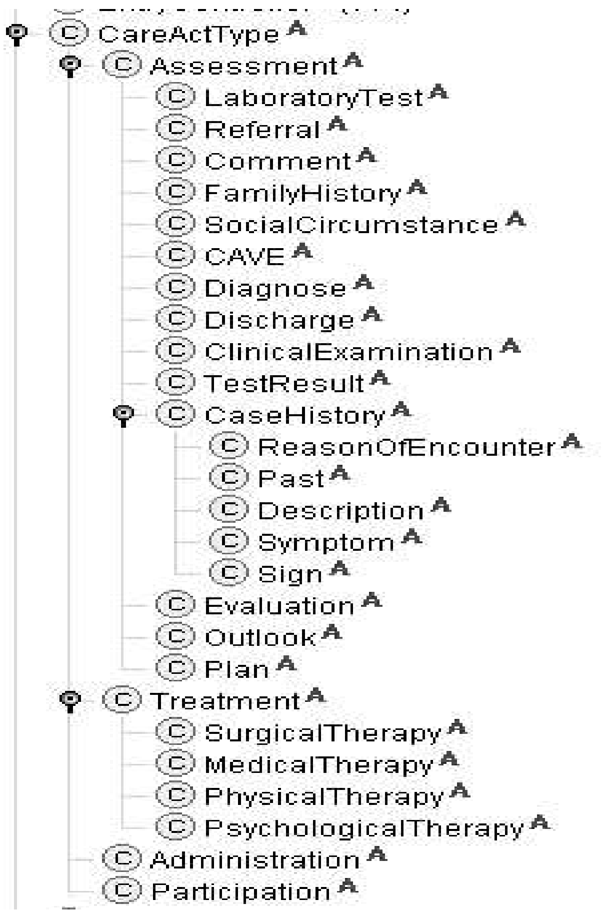


Figure 2. The content ontology demonstrating *CareActType***—a concept linking a specific activity with corresponding content (of the activity) in the patient record.**

3.3. An Information Ontology and Problem Orientation

We model a patient record's information architecture through an ontology that contains concepts for an information model. Instances of the information model make up a specific patient record. A set of patient records make up a patient-data knowledge base that adheres to the structure of the information ontology.

The information ontology is designed to represent different views on patient data. The purpose of the information ontology is to model the underlying information architecture for the patient record system. In most current patient-record systems, lack of a rich data representation make it difficult to view patient data other than chronological (i.e., time-oriented view) because patient data is only linked to a unique ID and date of entry. However, it is should be pos-

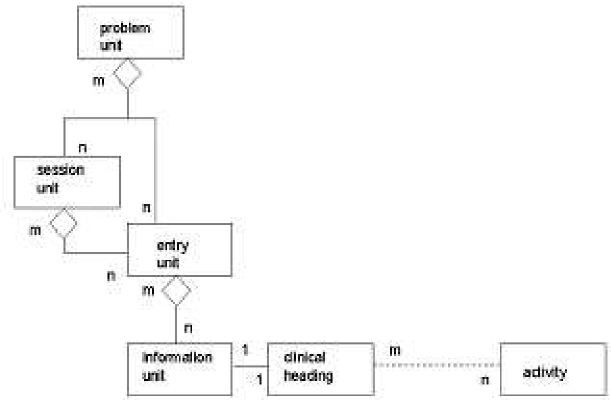


Figure 3. An information model that supports a problem-oriented view of patient data.

sible to view record content in additional ways, for instance problem oriented. A time-oriented record contains interleaving medical problems and to extract data, pertinent to a specific medical problem, consumes usually more time than physicians can afford to spend, during clinical encounters.

Lawrence L. Weed [15] introduced the problem-oriented medical record as a concept for improving structure of patient records. The concept offers an intuitive and useful way to organize the patient record because it accommodates the way clinicians work. A problem-oriented patient-record system provides a foundation for a well-structured patient record that correlates information in the patient record to a medical problem.

Although, the idea of problem orientation was introduces in 1968, implementations of problem-oriented patient records have yet to gain wide-spread acceptance. Among the most persistent complaints against problem-oriented medical records is the requirement on structuring patient-data entries [4].

Our information ontology contains problem-oriented concepts (Figure 3). These are *information unit*, *entry*, *session*, and *problem unit*. An information unit represents the smallest unit of information; this unit can be any clinical heading or care-act type (e.g., *History, Social Circumstance, Symptom, etc.*). An entry comprises a set of information units. An entry in the record has a responsible clinician attached to it, in addition to a time and date stamp. The entry simulates a classic entry in existing records, but we add knowledge about its content (i.e., a label such as a care-act type). A session is a high-level concept of an entry, in the sense that a session contains a set of entries. It is useful to combine a set of entries that relate to the same situation, for instance, entries that relate to the same diagnose. A problem unit is a collection of entries and sessions that all relate to the same medical problem.

The problem-oriented idea is an example of how information can be structured and divided into layers of relevant information; the *Problem* is a high-level concept, while the set of diagnoses represent a low-level concept. Diagnose is the essential descriptor in existing episodic-oriented patient-record systems.

4. Results

The three ontologies may function independently and by themselves as separate models and ontologies. However, merging them enables us to identify a set of information from the patient record (information sources) that are relevant to an activity within the process of care. The content ontology is a bridge between the process ontology and the information ontology that links concepts from information-source concepts to patient-specific data in the patient record. Information sources represent both high-level clinical concepts and clinical terms that are, either directly or indirectly, complementary to specific content in the patient record.

To evaluate our models, ontologies and knowledge bases, we have split the evaluation into two: (1) evaluation of clinical headings (thus the content ontology or care-act types) and the problem-oriented information model and (2) evaluation of the process model through simulations of patient-specific encounters.

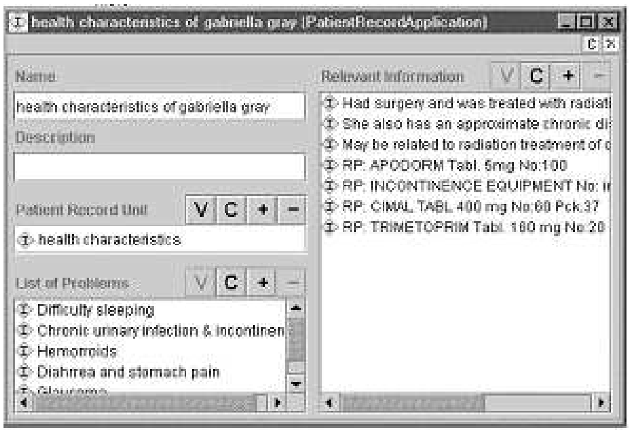
4.1. The Problem-Oriented Model

We transformed a time-oriented patient record into a problem-oriented patient record, using our problem-oriented information model to evaluate how well our model fit time-oriented entries into problem units. Our problem-oriented knowledge base contains 267 information units, 112 record entries, 26 medical problems, and 46 diagnoses. The set of problems belong to three different patients. 13 clinicians—including physicians, nurses, and laboratory assistants—have been treating these three patients over a period of 9 years, from primo 1991 to mid 2000. Results from the transformation demonstrated that:

1. Entries could be classified into a set of finite clinical headings.

2. Clinical headings were sufficient labels of information units; they encoded patient data and linked directly to activities in the process model.

3. Some entries did not naturally relate to a medical problem, while others had an obvious relation to a medical problem.



health characteristics of gabriella gray (PatientRecordApplication)

Name

health characteristics of gabriella gray

Description

Patient Record Unit

health characteristics

List of Problems

- Difficulty sleeping
- Chronic urinary infection & incontinen
- Hemorroids
- Diahrrea and stomach pain
- Glaucoma

Relevant Information

- Had surgery and was treated with radiati
- She also has an approximate chronic di
- May be related to radiation treatment of c
- RP: APODORM Tabl. 5mg No:100
- RP: INCONTINENCE EQUIPMENT No: n
- RP: CIMAL TABL 400 mg No:60 Pck:37
- RP: TRIMETOPRIM Tabl. 160 mg No:20

Figure 4. Relevant knowledge presented during simulation of patient-specific care process.

Although, clinicians did not enter information into the record in a strict order or that some entries did not relate to any medical problem, our problem-oriented model proved flexible and permitted the option of an entry to function as a separate unit, or relate to one or several problems.

4.2. Simulation of the Family-Care Process

To evaluate the usefulness and validity of our process model, we conducted a simulation to see how patient-specific encounters were captured and what information was determined as relevant to each activity within the encounter (i.e. information sources). Information sources, like the generic family-care process can be instantiated into patient-specific information sources that determines the exact set of information from a patient's record as relevant to the patient-specific process in concern.

For a set of patient encounters, we simulated step-by-step activities in the care process and evaluated the set of information sources (i.e. a simulation of the set of queries performed to abstract patient data relevant for each activity). We performed 17 simulations, which resulted in 31 information sources, such as the one illustrated in Figure 4. Figure 4 demonstrates an information source with information relevant to *assessment of shoulder myalgia*. There are two levels of information presented in the information source: *List of Problems* and *Relevant Information*. A third level is accessible by selecting a concrete problem from the problem list. Each problem contains a set of relevant information to that specific problem.

5. Discussion and Future Work

The presented framework has introduced two models: a knowledge model and a relevance model. The relevance model is for ranking a set of information sources, while the knowledge model comprises an ontology that represents activities, actors, and their relationships with information sources. The knowledge level enables the identification of a set of information sources. The relevance model, as presented in this paper, involves ranking on the level of problem orientation and care-act types, thus the identified set of information sources has been ranked according to its relevance to a problem and its relevance to activity. However, ranking of information can be done according to an additional utility function based on medical knowledge, that is, based for instance on relevance between different medical problems. We do not currently use a utility function because there is a trade off between the system's level of knowledge and the effort of adding medical knowledge to the system (i.e., the epistemologic adequacy). A working hypothesis has been that the patient record contains a lot of useful and needed information that can be utilized before adding medical knowledge to the system. This implicit knowledge is relevant in different situations during patient care, but the challenge is to provide the record system with knowledge to facilitate this information (i.e., awareness of its content and use), which has been the focus of our presented framework.

In addition, the current version of our content ontology, depends on manual labeling of clinical data. However, we believe that we will be able to label automatically clinical terms in current patient records. Several findings indicate the possibility of parsing the patient records and label them according to the clinical headings in the content ontology (care-act types):

1. Physicians have a limited set of remedies to their disposal (used in assessment), which thereby, limit the domain of discourse.
2. Physicians use a de-facto notation for writing narratives such as (from Table 1) *Clin.ex.: Cor-pulm NA* for annotating finding from a clinical examination, *EEE: hospital 19941212. Cataract operation left eye. Uncomplicated* for hospital discharge, or *X-RAY: 19920210 hospital US abdomina. Norm.findings. X-ray thorax. Norm. findings* for x-rays (a type of hospital discharge). The de facto notation can function as heuristics when classifying patient-record data into care-act types (clinical headings).
3. The free-text contains a set of semi-structured text (from Table 1) such as *Lab.test to hospital STOMACH ACHE UNSPECIFIED* indicating an laboratory test order, *RP: THYROXIN INA TBL 0.005 mg*

No:100 Pck:6 Reit:3 for prescription of medication, or *—HYPERTENSION ESSENTIAL* for diagnose. The semi-structured text also helps as heuristics when labeling information as care-act types.

6. Conclusion

Our current ontology defines a generic family-care process that can be instantiated into patient-specific processes, a set of clinical terms that defines the content of a patient record, and an information model that enables the content model to link concepts from a patient record to activities in the process model. The link between concepts in the patient record with activities is represented as a set of information sources. The linking of patient-specific data to activities in the care process produces an instantiation of the family-care process and information sources into patient-specific processes that has a projection of its activities in the patient record. The projection can be read from the patient record as traces or action patterns.

The presented ontology and corresponding knowledge base for information ranking enables the patient record system to better utilize residing information in the record. Conceptually, we have added a layer of knowledge to the existing facts level in the record. Knowledge about clinicians' work processes empowers the record system to recognize where in the care process the clinician is and determine relevant information sources to specific activities, thus a more *helpful* record system that minimizes the gap between the intuitive user interaction and that required by the system. Our aim is not to take part in clinicians' diagnostics of patients' health problems but to contribute and indicate potentially relevant information—necessary and useful for diagnostics and decision making—by being non intrusive; taking advantage of what already exists in the patient record system.

Acknowledgment

Samson Tu and Mark A. Musen, Stanford Medical Informatics, Stanford University School of Medicine, Erik Naggum, and Bjørn Remseth. Funding of this paper is by the Norwegian Research Council.

References

[1] R. B. Altman, M. Bada, X. J. Chai, M. W. Carillo, R. O. Chen, and N. F. Abernethy. Riboweb: An ontology-based system for collaborative molecular biology. *IEEE Intelligent Systems and Their Applications*, 14(5):68–76, September/October 1999.

[2] R. Chandrasekaran, J. Josephson, and V. Benjamins. What are ontologies, and why do we need them? *IEEE Intelligent Systems*, 14(1):20–26, January/February 1999.

[3] R. Davis, H. Shrobe, and P. Szolovits. What is a knowledge representation? *AI Magazine*, 14(1):17–33, 1993.

[4] H. P. Lundsgaarde, P. J. Fischer, and D. J. Steele. *Human Problems in Computerized Medicine*. University of Kansas Publications in Anthropology Number 13, 1981.

[5] J. McCarthy and P. J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In Meltzer and Michie, editors, *Machine Intelligence*, volume 4, pages 463–504. Edinburgh University Press, 1969.

[6] S. Montani and R. Bellazzi. Intelligent knowledge retrieval for decision support in medical applications. In V. P. et al., editor, *MEDINFO 2001*, Amsterdam, 2001. IMIA, IOS Press.

[7] M. Musen. *Knowledge Engineering and Agent Technologies*. IOS Press, Amsterdam, 2002.

[8] M. A. Musen. Medical informatics: Searching for underlying components. *Methods of Information in Medicine*, In press.

[9] N. Noy, R. Fergerson, and M. Musen. The knowledge model of Protégé: combining interoperability and flexibility. In *2nd International Conference on Knowledge Engineering and Knowledge Management (EKAW'2000)*, Juan-les-Pins, France, 2000.

[10] M. Peleg, I. Yeh, and R. B. Altman. Modeling biological processes using workflow and petri net models. *Bioinformatics (accepted December 2001)*.

[11] S. Quaglini, M. Stefanelli, G. Lanzola, V. Caporusso, and S. Panzarasa. Flexible guideline-based patient care flow systems. *Artificial Intelligence in Medicine*, 22:65–80, 2001.

[12] G. Schreiber, H. Akkermans, A. Anjewierden, R. de Hoog, N. Shadbolt, W. V. de Velde, and B. Wielinga. *Knowledge Engineering and Management—The CommonKADS Methodology*. Massachusetts Institute of Technology, 2000.

[13] S. Tu, H. Eriksson, J. Gennari, Y. Shahar, and M. Musen. Ontology-based configuration of problem-solving methods and generation of knowledge-acquisition tools: Application of PROTÉGÉ-II to protocol-based decision support. *Artificial Intelligence in Medicine*, (7):257–289, 1995.

[14] J. van Bemmel and M. Musen, editors. *Handbook of Medical Informatics*. Springer, 1997.

[15] L. L. Weed. Medical records that guide and teach. *MD Computing*, 10(2):100–114, March/April 1993.

[16] R. A. Wilson and F. C. Keil, editors. *The MIT Encyclopedia of the Cognitive Sciences*, chapter Computational Intelligence. The MIT Press, 1999.

Part IV

Synopsis

Chapter 13

Discussion

13.1 Introduction

In this chapter we discuss the most important topics from each of the four papers in Chapter 9–Chapter 12: (1) *Ranking of Information in the Computerized Problem-Oriented Patient Record* [13], (2) *A Knowledge-Based, Problem-Oriented Patient-Record System* [12], (3) *The Helpful Patient Record System: Problem Oriented and Knowledge Based* [15], and (4) *Ontologies for Knowledge Representation in the Computer-Based Patient-Record System* [14], respectively.

Although the empirical studies were performed first, we did not formalize their discoveries until a later stage in our process. Here we try to make clear how specific empirical findings have concretely affected some of the framework components' characteristics. The empirical results have not changed any elements in the papers. However, they have contributed to some of the components in the framework in the first paper (Chapter 9) have matured in the last paper (Chapter 12), which we will talk about in this chapter.

To demonstrate one of the earlier notions in the framework—decision frames—we show some of the results that we have from a simulation of the prototype. With decision frames, we want to illustrate how a structure and relevance ranking of information is the foundation for what is presented to the physician, as part of a frame and relevant to their decision-making process. We conclude with a demonstration of decision frames and relevance ranking, before we summarize the chapter.

13.2 A Two-Layer Framework for the Patient-Record System

The two-layer framework (Figure 13.1) has evolved from its earliest version in Chapter 9 to its latest version in Chapter 12. The framework now includes specific components such as the problem-oriented information model, the CareActType ontology of clinical concepts representing patient-record content, and process knowledge of care activities and subprocess, participants, relationships between activities and patient-record content, etc.

Our ideas for a two-layer framework was influenced by the work done in the two projects PEN&PAD (Section 4.5.5) and GALEN (Section 4.2). They introduced a two-layer framework for the electronic medical record, where the first level was a facts level and the second level was a meta level. The facts level contained the *direct observations* of what physicians had *heard, seen, thought, and done*. The meta level contained what can be said about what has been heard, seen, thought, and done. The meta level was intended to enable the record system to say something about the first level. Thus, the framework's second layer distinguished between *the clinical conversation* (i.e., what physicians have heard, seen, and thought) and *the medical decisions* (i.e., what physicians have done).

When we separate the facts from the meta level we get two immediate benefits: (1) the framework remains *faithful* to the direct observations and (2) the framework distinguishes between the various interpretations that physicians can do, based on the facts level. Facts are the input to the encounter processes and is the foundation for any interpretation that a physician does, and many interpretations may exist. The interpretation again reflects a physician's selection of actions. Once an interpretation is selected, the encounter process will evolve and change according to the present interpretation—assessment and treatment will reflect the interpretation.

13.2.1 The Usefulness of a Two-Layered Framework

To use an example from Chapter 9, *Ranking of Information in the Computerized Problem-Oriented Patient Record* [13]; in a situation where a patient has symptoms such as *troublesome and dry eyes, a feeling of dry mouth, drink a lot to swallow, feels tired, and has a burning throat*, the family physician might suspect Sjögren's syndrome. As a result of this hypothesis, the family physician considers to refer the patient to a rheumatologist for further assessment on possible Sjögren's syndrome. *Referral to the rheumatologist with assumption of possible Sjögren's syndrome* reflects an interpretation the family physician does based on the patient's problem description.

Furthermore, in the example given in [13], let us assume that the family physician's initial interpretation led to a dead end—the discharge from the rheumatologist indicated that the patient tested negative on Sjögren's syndrome. The family physician then chooses to go back to the patient's initial problem statement (with possible additional and new statements) and focuses on other symptoms[1], for instance, *pain in stomach, she lost 5 kg the last few months*. These symptoms will lead to other interpretations and another set of actions.

For a record system to follow physicians' change of interpretations in a care process and to determine information as relevant or not to the care situations, it is necessary for the system to be able to distinguish between the problem statement and the interpretations based on these facts. By separating between the direct observations (i.e., the facts level) and the interpretations based on these facts (i.e., the meta level), we can distinguish and rank these interpretations according to the care situation [13]. If a family physicians chooses to go back to the initial problem statement, the first set of actions (i.e., hypothesis of Sjögren's syndrome and the actions based on this hypothesis) may not be considered as relevant for the second interpretation, but the problem statement may still be relevant.

---

[1] Another alternative would be to use time, time to see how the patient's condition evolves. *Time* is a useful remedy in primary-care (Chapter 9 and Section 13.3).

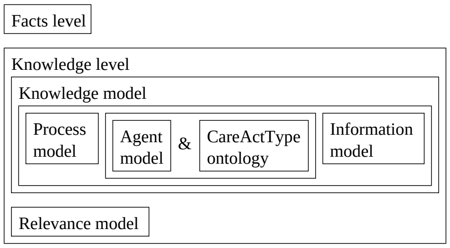
13.2.2 Components in the Two-Layer Framework

The two layers are mirrored in our framework with accompanying components (Figure 13.1):

- The facts layer, which contains the patients' problem statement. The facts layer reports on the direct observations of what patients have said and what physicians have heard, seen, thought, and done
- The knowledge layer, which contains two models:
  - The knowledge model, which contains three models:
    - A process model of the organization, tasks and activities, goals, purpose, participants, and information needs
    - A concept model, which contains two models: (1) an agent model and (2) a CareActType ontology
    - An information model, which supports a problem-oriented view of patient data
  - The relevance model for ranking a set of *information sources*. Information sources are a set of information from the patient record that is determined as relevant to a care activity. The relevance model contains the set of rules that enables the record system to rank information sources as relevant to a care activity.

In the following sections, we will discuss in more detail how the separation into two levels is visible in other components in the framework (e.g., the agent model and the CareActType ontology).

Figure 13.1: The two-level framework and components in *The Knowledge Model*.



13.3 An Agent Model

One of the components that has evolved most from Chapter 9 to Chapter 12 is the process model presented in Figure 1 in Chapter 9. The original model was a three-component process model of the clinical encounter, where the fist component was based on the patient's problem statement, the second component was based on the set of actions done by the family physicians, and the third component was based on the evaluation phase, which would determine the next problem-statement phase and the physicians' possible next set of actions.

The process model is renamed to an agent model because the model's main outcome is the separation of input and output information to the encounter process. Input and output information are agent dependent. Since, input and output are agent dependent, they are also task dependent because tasks are agent dependent. The agent model is illustrated in Figure 13.2.

Several of the properties in the process model are still reflected in the agent model:

- The process model makes a distinction between information owner and information user; the patient is the information owner of the problem statement and the physicians is the information user. The physician then overtakes the role of information owner once the problem statement is recorded in the patient record, and later used as input to his or her decision-making process

- The process model viewed the set of remedies that family physicians' have to their disposal as a limited set. The limited set of remedies enable us to categorize them into three classes: (1) omnifarious, (2) uncertain, and (3) certain. The two latter classes reflect remedies that characterize either assessment or treatment activities, and could be renamed to assessment and treatment remedies, respectively. The first class of remedies—*omnifarious*—categorizes remedies that are used in both assessment and treatment activities, and can, therefore, not be used to interpret care situations

- The process model was based on the notion of traces. Traces are action and they are represented as semi-structured text in current patient records. Traces reference some action that has been done in a clinical encounter. Examples of action patterns are *Lab.test order to* hospital *HYPOTHYROIDISM*, ****SICK LEAVE NOTE for RHEUMATOID ARTHRITIS L88 100*, and *Referral*hospital *for CARCINOMA BASOCELLULARE CUTIS*. Traces enable the record system to say something about the actions that have been done, such as the ordering of a laboratory test, or the writing of sick leaves, or the writing of a referrals, in the examples just given above. Interpreting traces helps us make assertions on what the physicians have done and what phase[2] the physicians are working within

- The process model viewed the clinical encounter as repetitive (i.e., problem statement, actions, and evaluation) and containing a set of few activities (i.e., the limited set of remedies), which make it easy for us to model the primary-care process in terms of generic subprocesses and activities. Even though the clinical encounter is easy to model, much of what makes every encounter unique is projected in the patient record (as summaries of a patient's problems and medical condition). The encounter summaries contribute with the *content* to the actions, where the activities are generic, but the combination of them (as a result of the patient's problem and medical condition), make up the uniqueness of each

---

[2] When speaking of phases a physician is working within, we mean either assessment activities or treatment activities.

encounter. The fact that the content to each process lies in the patient record further demonstrates the valuable impact the record system has in primary-care processes. Physicians can use the record system to get into context for a patient encounter, which further enables the physician to provide continuity in a patient's care process

13.3.1 Agents Added to the Process Model

The agent model, depicted in Figure 13.2, demonstrates actors participating in the clinical encounter such as *patient* (or family member, relative, or a care taker), *physician*, and *clinicians* (e.g., nurse, laboratory assistant, specialist physician, etc.).

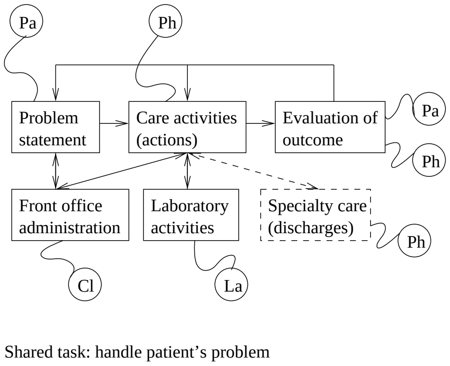
> An agent is an executor of a task. It can be human, computer software, or any other "entity" capable of executing a task. In the agent model, the capabilities of each agent are described. The model can also be used to represent constraints on an agent, such as norms, preferences, and permissions that apply to the agent. For example, a constraint might be an organization rule: A specific decision-making tasks should not be performed by a computer (Schreiber et. al, 1994) [168].

"Furthermore, it lists the communication links between the agents in carrying out a task ... The purpose of the agent model is to understand the roles and competences that the various actors in the organization bring with them to perform a shared task" [167].

We have extended the agent model to include two more tasks, than the original process model, front-office tasks and laboratory tasks. The agent model contains the following components:

- A problem statement (Figure 5.5) is given by the patient and functions as input to the physician's set of actions. The problem statement is performed by the patient, which is the agent for this task
- A set of care activities (actions) that a physician determines to do, these belong to a set of limited remedies discussed in Section 13.3. The actor for this task is the family physician
- A task of evaluating the outcome of the actions in response to a patient's problem statement. This task functions as new, or additional, input to either the patient's problem statement or to the physician's set of actions. Both patient and physician can be actors to this task
- A set of front-office tasks (Figure 5.3) that contribute directly to the clinical encounter or physicians' medical decision-making process. As was demonstrated in Chapter 5, family physicians are dependent of information from the front office (e.g., *Register & Schedule patient*, *Record patient's reason for encounter*, *Record patient demographics*, *record discharge information* etc.)
- A set of laboratory tasks (Figure 5.4).

Figure 13.2: The agent model. The squared boxes are tasks performed by agents. Attached to each box is a circle which is the agent responsible for the tasks. *Pa* is short for *Patient*, *Ph* is short for *Physician*, *Cl* is short for *Clinicians*, and *La* is short for *Laboratory assistant*. The links between the tasks are the communication (of information) between agents.



13.3.2 Agents Reflected in the CareActType Ontology

We distinguish between the direct observations (the patient's problem statement and physician's observations) and the interpretations that are done on these observations (assessment, diagnostics, and treatment actions). This distinction between the input to a process and the actions based on the input is also reflected in our CareActType ontology (Figure 8.5). The CareActType ontology has four main classes that jut out from the root concept :THING[3]: (1) case history, (2) assessment, (3) diagnosis, and (4) treatment. The former class represents the information input to the care process, while the three latter classes represent the actions done in the care process (i.e., based on interpretations).

13.3.3 Agent Communication and Information Flow

Communication between the various agents is reflected the roles of information owner and user. The agent model in Figure 13.2 may not show it clearly but there is relatively little communication and information flow between the various agents and their tasks (Figure 5.3–Figure 5.6 and Table 8.1 illustrates in more detail the information flow between agents and their activities). The clinical encounter activities are, more or less, independent from each other. Apart from information from the patient, the clinical encounter depends mostly on information from the front office, the laboratory, and specialty care (in terms of discharges).

Through the agent model we are able to: (1) separate problem statement (e.g., what is heard) from care activities (e.g., what is done). The CareActType ontology is constructed with this in mind—to distinguish the various information sources (i.e., the various agents that participate in the clinical encounter), (2) demonstrate that there

---

[3] There are other subclasses under the root concept :THING as well, but they are not considered as main concepts relevant to physicians' clinical decision-making process.

is little information flow, or communication, between the various agents and clinical offices, (3) characterize the encounter with a generic pattern, where there are few clinical activities and subprocesses, and (4) adhere a set of care activities to a limited set of remedies.

13.4 Decision Frames and Information Ranking

In [13] we refer to decision frames as "the environment in which a decision is made . . . and consists of a collection of facts from the patient record". Decision frames are the ultimate presentation of ranked information.

Originally, decision frames were introduced as a set of eight dimensions (e.g., patient demographics, responsible clinicians, problem statements, diagnoses, time, actions, goals, outcome of actions). These dimensions are made up of information from the classes in the CareActType ontology because the CareActType ontology classifies the content in patient records. Decision frames comprise extracted information from the patient record, relevant to care activities. Classes in the CareActType ontology represent completely the patient-record information, where the classes relate activities with patient-record content and, therefore, also make up the content of decision frames.

The set of categories from the CareActType ontology that make up a decision frame vary with the phase a physician is working within. For example, if a physician is working in an assessment activity *Perform clinical examination*, a decision frame will include classes from the CareActType `class Assessment`, such as `class LaboratoryTest`, `class Referral`, `class ClinicalFinding`, etc. The classes refer to both an activity and patient-record content and as a result, the names of classes are a combination of both. If a physician is working with a treatment activity *Give drug therapy* classes from the CareActType `class Treatment`, such as `class SurgicalTherapy`, `class DrugTherapy`, etc. will be selected as part of a decision frame.

13.4.1 Patient-Encounter Processes

Figure 8.2 illustrated the generic primary-care process model. We have instantiated the primary-care process, in terms of eight patient-specific encounters for Gabriella Gray (Figure 13.3), such as 030895, 231095, 110396, 120396, 140396, 190396, 210396, and 290396. Figure 13.3 shows a patient-specific encounter process as a blue circles. The circles are subprocesses (Figure 8.1 contains a more elaborate graph notation) and contain per se other subprocesses and activities (Figure 13.4). An encounter process (i.e., a blue circle) is equal to an entry in the patient record (or entry units in the problem-oriented model in Figure 8.4).

We have filled in information that we have from the (time-oriented) patient record. For those activities and flows we do not have information on, we leave them as they are (with generic information). Figure 13.5 shows another encounter process—*primary care workflow 110396*—where flows such as *lab.test* and *test results* have not been filled in with patient-specific information.

If we choose one of the subprocesses in Figure 13.5, for example *assessment & treatment*, another subprocess is revealed, as the one in Figure 13.7. From this figure we

Figure 13.3: Patient-specific encounters for Gabriella Gray. Blue circles are subprocesses (Figure 8.1).

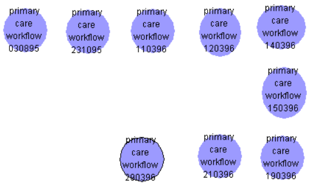
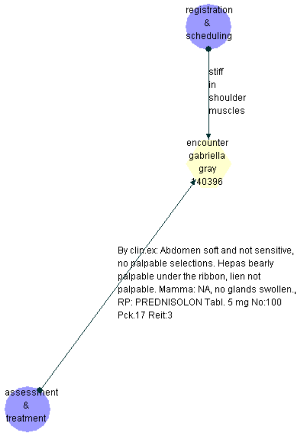


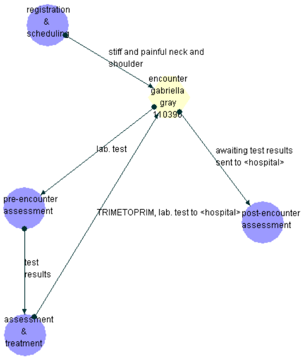
Figure 13.4: A patient-specific encounter of Gabriella Gray dated March 14, 1996. Blue circles are subprocesses, while the yellow hexagon is a loop. Arrows are information flow (Figure 8.1 contains a more elaborate figure legend).



By clin.ex: Abdomen soft and not sensitive, no palpable selections. Hepas bearly palpable under the ribbon, lien not palpable. Mamma: NA, no glands swollen., RP: PREDNISOLON Tabl. 5 mg No:100 Pck.17 Reit:3

see patient specific activities such as *identify possible hypothesis* (for Gabriella Gray's problem statement) and *encounter stop* in green diamond-shaped boxes. For the activity *identify possible hypothesis* an association with the patient record has been made with *health characteristics of Gabriella Gray*. The part of a patient record that is relevant to an activity is illustrated as an upside-down triangle in pink. The arrows demonstrate information flows between activities and patient records (for a more detailed description of the graph notation, see Section 8.2.1).

Figure 13.5: A patient-specific encounter of Gabriella Gray dated March 11, 1996. Blue circles are subprocesses, while the yellow hexagon is a loop. Arrows are information flow (Figure 8.1 contains a more elaborate figure legend).



13.4.2 Problem-Oriented Patient Records

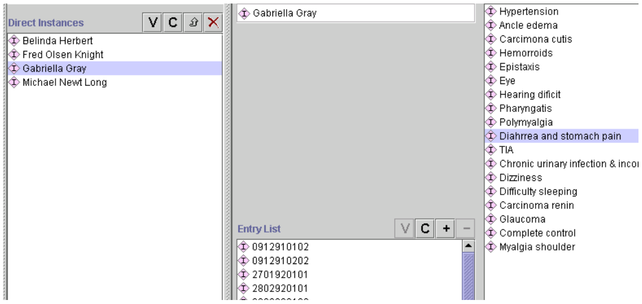
We have represented time-oriented patient records as problem-oriented patient records. The problem-oriented model (Figure 8.4) has been implemented in terms of information units, entry units, session units, and problem units. Figure 13.6 demonstrates Gabriella Gray's problem list. The middle part of the figure shows the entry list containing the set of entries (and sessions) that Gabriella Gray has in her patient record. The entry list equals the time-oriented patient record (or chronological log) of original encounter events. The figure does not show any session units for Gabriella Gray (she does not have any). On the right side of Figure 13.6 Gabriella Gray's problem list is shown with 18 different problems such as hyptertension, ancle edema, carcimona cutis, hemorroids, epistaxis, etc.

13.4.3 Ranked Information

On the left side of Figure 13.7 we present relevant information to the activity *Identify possible hypothesis*. The relevant information has been extracted from Gabriella Gray's patient record. The set of relevant information is called *Health characteristics of Gabriella Gray*, which is shown as a pop-up window that presents information related to *Health characteristics* (e.g., *List of problems* and *Relevant information*).

In Figure 13.7 we show three levels of relevant information:

Figure 13.6: A problem list for Gabriella Gray. The left side of the picture lists the four patients that we have problem-oriented patient records for, where Gabriella Gray is one of the patients. In the middle is Gabriella Gray's entry list, which corresponds to the time-oriented patient record, and on the right we see her medical problem list.



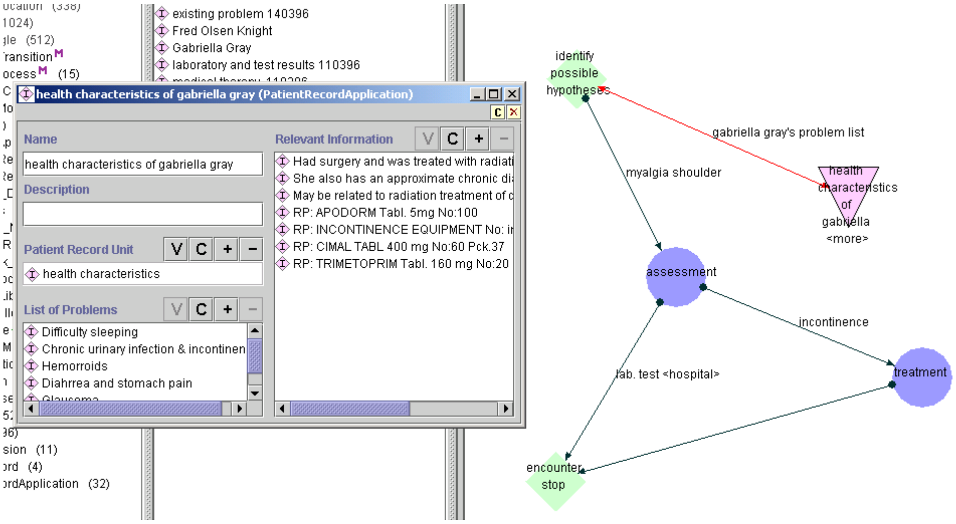
Belinda Herbert
Fred Olsen Knight
Gabriella Gray
Michael Newt Long

Gabriella Gray

0912910102
0912910202
2701920101
2802920101

Hypertension
Ancle edema
Carcimona cutis
Hemorroids
Epistaxis
Eye
Hearing dificit
Pharyngatis
Polymyalgia
Diahrrea and stomach pain
TIA
Chronic urinary infection & incoı
Dizziness
Difficulty sleeping
Carcinoma renin
Glaucoma
Complete control
Myalgia shoulder

Figure 13.7: Relevant patient-record information to a patient-specific activity. Blue circles are subprocesses, green diamonds are activities, and pink triangles up-side down are patient-record applications (various sections of the generic patient record). Arrows are information flow (Figure 8.1).



existing problem 140396
Fred Olsen Knight
Gabriella Gray
laboratory and test results 110396

health characteristics of gabriella gray (PatientRecordApplication)

Name: health characteristics of gabriella gray

Description:

Patient Record Unit: health characteristics

List of Problems:
Difficulty sleeping
Chronic urinary infection & incontinen
Hemorroids
Diahrrea and stomach pain

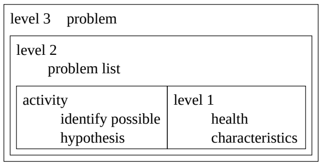
Relevant Information:
Had surgery and was treated with radiati
She also has an approximate chronic di
May be related to radiation treatment of c
RP: APODORM Tabl. 5mg No:100
RP: INCONTINENCE EQUIPMENT No: ii
RP: CIMAL TABL 400 mg No:60 Pck.37
RP: TRIMETOPRIM Tabl. 160 mg No:20

- Level 1: *Relevant information* (which include *Previous illness/disease and Current medication*) to *Health characteristics*, which again is relevant to activity *Identify possible hypothesis*
- Level 2: *Problem list* of Gabriella Gray
- Level 3: By selecting one of the problems in the problem list information for a *specific problem* will show up

The three levels have also been illustrated conceptually in Figure 13.8. The three levels are meant to provide information with varying degree of relevance (or distance) to a specific activity. Hence the three levels are dynamic according to an activity, that is, the content of the three levels vary with the care activity.

For instance, another activity than *Identify possible hypothesis* such as *Perform clinical examination*, will have different information in the three levels of information presented to the activity. Relevant information to *Perform clinical examination* will be examination findings (and type of examinations) as level 1, *Problem list* as level 2, and *Problem* as level 3.

Figure 13.8: Three levels of ranked information relevant to the activity *Identify possible hypothesis*: (1) activity-specific information (level 1), (2) problem list (level 2), and (3) problem-specific information (level 3).



13.5 Summary

In this chapter we have presented the most essential components in our two-layer framework. The two levels are a facts level and a knowledge level. The focus in this thesis is on components in the knowledge level that enables the patient-record system to recognize its input information, relate it to care activities, and thereby present the physicians (or other clinical users) with relevant information to the various care activities in the primary-care process.

As part of our attempt to provide a structure to the patient-record system that enables physicians overview over record content, our information model supports a problem-oriented view of patient data. Problem orientation structures information in the record in conformity with a unit—a unit which organizes information that are relevant to the same medical problem.

In the presented framework, we combine a problem-oriented structure with relevance ranking to offer a structure that gives physicians a better overview than what is given in existing time-oriented patient-record systems. The framework is based on relationships between patient-record information and care activities. Information relevance varies according to care activity and we use the relationship between care activity and patient-record content to rank information relevance.

The combination of the problem-oriented information structure and the set of clinical concepts enables a ranking of information according to: (1) type of information (i.e., the concept), (2) medical problem (i.e., a problem list or a specific problem), and (3) the relationship between type of information, medical problem, and care activity. Each activity in the care process, therefore, has a set of relevant information sources associated with it, where each information source is again divided into three levels of relevance according to: (1) activity-specific information, (2) a patient's problem condition (i.e., problem list), and (3) problem-specific information for each problem in the list.

Chapter 14

Evaluation and Future Work

14.1 Introduction

In this chapter we evaluate the components in the framework, and the knowledge-engineering methods that we have used to construct the knowledge components. We also present future work for some of the components that we evaluate.

The components in the framework are the problem-oriented information model that we have used to re-represent time-oriented patient records, the CareActType ontology, and the primary-care process model. The knowledge components are meant to provide the record system with means to: (1) provide physicians with better overview through a problem-oriented structure and (2) recognize, encode, and relate patient-data input to physicians' various care situations. In Chapter 1 we introduced a set of objectives:

O1. To extend current patient-record systems with a problem-oriented view of patient data that facilitates a better structure of the patient record system, that again will give better overview over information in the record.

O2. To add process knowledge to the record system as a way to overcome the challenge of forcing the user to work with and enter data into the system in a controlled way. A system that has knowledge of its content and use will be able to reason with its knowledge. If a patient-record system can recognize care situations it can present to physicians patient-specific information that is relevant to the clinical encounters. Thus, the user gets better access to the information in the record system; while at the same time, qualities, that family physicians already appreciate in their current record systems, are kept intact.

The created knowledge base contains a set of problem-oriented patient records that enable us to evaluate the problem-oriented information model. We assume that the structure provided by the problem orientation gives a better overview than the nonstructure provided else wise (i.e., time orientation), because the problem-oriented structure enables a sorting of information according to medical problem units. Overview is facilitated through the extraction of information into units and a navigation among units according to medical problem names, which are features that are not available in time-oriented patient records. It is therefore reasonable to assume that a problem-oriented patient-record system provides a better overview than a time-oriented patient-record system.

14.1.1 Requirements to Change

Instead of evaluating the aspect of better overview in the problem-oriented information model, we evaluate additional requirements that the model imposes on the user. The problem-oriented information model will introduce an additional set of requirements, which will demand a change in the users habits. We call these requirements *requirements to change*.

Too many changes in a system's way of compelling the user to work will increase the possibility of the user rejecting to accept the new system (or its added functionality) [8]. We have aimed to minimize the requirement to change the user's way of working with the system, because by introducing changes, we risk the chance of the user not accepting the system's new features. The evaluation of requirements will be based on:

1. What are the specific requirements that are introduced by the problem-oriented information model
2. How many are the requirements to change? The requirements should be as few as possible, because we want to minimize the risk that the user rejects the system due to its new changes
3. A comparison of the requirements to change with previous systems' requirements. We compare with earlier systems' requirements that have proved to not be accepted by the users.

14.1.2 Representation of Patient-Record Information

The CareActType ontology contains a set of concepts that classify and represent the patient-record content to make possible a ranking of information according to care situations. For this to be possible, the record system must be able to: (1) recognize and encode its input data and (2) relate input data to the activities in the primary-care process. The set of clinical concepts, therefore, serve two purposes:

- They encode patient-record information
- They represent a relationship between activities in the patient-care processes and the patient-record information.

Encoding of information is based on what physicians already enter into the record system. The clinical concepts reflect how physicians presently work with the record system. We have already demonstrated in Table 6.12 how the concepts represent clinical information in the patient records.

The relationship between activities in the care process and the patient-record information will be demonstrated through testing of the knowledge base, where we can simulate patient-specific encounters. Patient-specific encounters are concrete primary-care processes that contain a subset of generic care activities. We evaluate the concepts in the CareActType ontology with respect to how they enable us to present relevant information to physicians during the various care activities in the primary-care process.

14.2 Levels of Overview

Three different levels of overview are possible with the problem-oriented information model in combination with the CareActType ontology:

1. The problem orientation structures information in patient records according to problem units. A list of medical problems projects the patient's current and past medical condition(s). The problem orientation provides overview through reference to a list of medical problems that represent the content in patient records
2. The set of concepts in the CareActType ontology classify patient-record content into categories of clinical information. These categories function as a tool for common understanding and navigation of information in patient records. The ontology provides overview through reference to clinical concepts that represent the content in patient records
3. For each medical problem in the problem list, information is also classified according to clinical concepts in the CareActType ontology. The overview given by the combination of problem orientation and CareActType concepts provides an overview on a finer-grained level than the two preceding levels of overview.

14.2.1 Problem Orientation

When compared to the time-oriented information model, the problem-oriented information model differs with respect to two features: (1) an entry or session unit is associated with a problem unit, where each problem unit is labeled with a problem name and (2) each entry or session unit is composed of a set of information units, where each information unit is labeled with a CareActType concept.

Problem Units

The problem-oriented information model does not require that every entry unit has to be associated with a problem name. Some entries do not belong naturally to a problem. We discovered during the re-representation of time-oriented to problem-oriented patient records that some entries typically have no associations to medical problems (i.e., these entries did not contain a diagnose). For example, prescription of eyedrops (e.g., *RP: KLORAMFENIKOL EYEDROPS No: 10ml* and *RP: TIMPILO EYEDROPS No:2 x 5 ml Pck: 8 Reit: 3*), prescription of drugs for allergy (e.g., *RP: ZYRTEC Tabl. 10mg No:100 Pck. 33 Reit: 3*), or prescription for pollen or hay fever were commonly entries with no specific diagnose. Thus, these entries would not be related to any specific medical problem.

If an entry or session unit is not associated with a problem unit, the entry unit is similar to a regular entry in a time-oriented patient record. Except for one difference; every information unit within an entry unit is still represented with a CareActType concept label. With entries that are not associated with a problem unit, they become stand-alone entries in the record. However, since these entries have labeled information units, the information units in those entries can be extracted and related to care activities because of their labels. Relevance ranking of patient information is based on both problem orientation and the labeling of information units. So, information units, in entries not related to problem units, can still be ranked according to their relevance to care situations.

Problem Names

The names we have used to label problem units in the knowledge base—the problem names—are based on ICPC codes for symptoms and diagnoses. ICPC may not satisfy the needs for a standardized vocabulary of problem names to label problem units. If a problem unit is going to be a recognized unit, we may probably need standards for problem names and problem lists. Ongoing research on problem lists such as [31, 32] can eventually lead to standardized vocabularies and terminologies for problem names and problem lists. However, in the mean time, we see the ICPC as the temporary coding scheme for labeling problem units for primary-care patient records.

Labeling of Problem Units

The labeling of problem units is done manually, in the same manner as diagnoses are given manually today. Family physicians select a diagnose code from the ICPC collection provided by the patient-record system. We do not see the manual labeling of problem units as an obstacle because most family physicians are used to giving a diagnosis for a majority of the patient encounters. From a survey based on Swedish family physicians, “about 76% of all GPs stated that they ’almost always’ classified some disease in relation to an encounter” [123]. That most encounters entries included a diagnose is in accordance with findings in our study of patient-record contents (Section 6.3.1). If encounter entries do not contain a diagnose, they are likely to relate to a previous encounter summary that contains a diagnose; such an entry belongs to a session unit (Table 6.8).

In addition to the requirement of manual labeling of problem units, there are also some issues that we do not deal with in the problem-oriented information model:

- Existing units such as episodes of care [90, 91] have not yet been defined, but the problem-oriented information model is fully extendible for defining such units. Episodes of care depend on temporal data, and more complex definitions and knowledge of how to determine when a problem is active or not. We have not gone into the study of these aspects, because we have focused on the process knowledge required to implement our objectives and not on the temporal and medical knowledge required to determine problem units’ active or inactive status

- The problem-oriented model does not contain distinct SOAP concepts like the original problem-oriented medical record presented by Weed (Section 1.1.3 and Section 8.3.2). But the four main concepts in the CareActType ontology can be mapped to the SOAP format; *CaseHistory* maps to the *Subjective* SOAP format, *Assessment* maps to the *Objective* and *Assessment* SOAP formats, and *Treatment* maps to the *Plan* SOAP format.

14.2.2 Classification of Patient Data

When evaluating concepts in the CareActType ontology, there are at least two aspects to take into consideration: (1) how useful are the concepts in classifying and representing information in patient records; and (2) how useful are the associations that they represent between activities in the care process and patient-record contents (i.e., how useful are they to rank patient-record information as relevant to care situations):

1. The classes of concepts in the CareActType ontology are created based on clinical concepts that we extracted during the study of information categories in existing patient records (Chapter 6, Table 6.4); the clinical concepts reflect what family physicians presently write into the records

2. Concepts in the CareActType ontology define a relationship between patient-record content and care activities in the process model. The knowledge base that we have implemented can demonstrate a three-level ranking of information according to a set of rules that we have defined for the knowledge model. The rules are based on the information given in Table 8.1; each activity is defined with a purpose, information needs, and an association to the patient-record content through overview levels 1-2-3:

    (a) The rules we use, to associate activities with patient-record content (i.e., through CareActType concepts), are easy to change if necessary. For example, a present rule for the activity *Identify possible hypothesis* (Figure 13.7) associates the activity with clinical concepts categories such as *Previous illness/disease* and current *Drug Therapy* (Figure 8.5). We can, for instance, remove the association of *Identify possible hypothesis* with *Drug Therapy*. By removing the latter association, current medication will not show up as relevant information (i.e., level 1 in Section 13.4.3) in Figure 13.7

    (b) The set of CareActTypes can also be changed and modified according to how well they classify and extract patient information to the various activities. For this modification to take place, it requires that the concepts be tested out in a clinical setting

For future work we suggest that the evaluation of CareActType concepts be done among family physicians. We believe it is necessary to evaluate the concepts among a representative group of Norwegian family physicians, especially, since the organization of concepts in the CareActType ontology deviates from the well-recognized SOAP format and the concepts are intended to be used in a problem-oriented patient-record system.

Another task for future work has to do with the evaluation of CareActType concepts' compatibility with other clinical concepts, for instance, hospital-record concepts (e.g., [160]). If we want to prepare for electronic communication and compatibility between primary, secondary, tertiary care, etc., it will be necessary to be compatible with ontologies that represent other health information systems.

The prototype remains to be tested in a clinical setting. For future work is the implementation that is required for the prototype to be tested in a clinical setting. If we do not take into consideration security and privacy issues, the prototype is still far from a stage where it can be brought into a clinical setting. For example, the prototype lacks a user interface, a database management system, functions for data operation, etc., or integration with existing patient-record systems.

The knowledge we added, with components in our framework, to the patient-record system is process knowledge of primary-care activities, activities' goals and purposes, participants, and corresponding information needs for each activity. We have not incorporated any medical knowledge, for instance, knowledge about how medical problems relate to each other. Adding medical knowledge, in the future, is not trivial. However, as the problem-oriented patient record with process knowledge will eventually take part in clinical settings, the problem unit may contribute (through evidence-based medicine) in record systems that include machine learning algorithms [112]. If the record system also has medical knowledge, stand-alone entry units can be sorted under a problem unit by the record system, thereby, avoiding stand-alone entry units (if this is desirable; it may not be necessary to classify absolutely all information in the record into units).

Rules for Classification

Information units have been labeled manually when entered into the knowledge base. It is possible to automate the process of labeling information units. While studying the patient-record content we found heuristics that can enable us to automate the process of labeling units. Thus, we believe that it is possible for a patient-record system to encode its information units.

Half of the patient-record information is semi-structured text. The other half is free text which can be classified into categories such as such as REASON FOR ENCOUNTER, SYMPTOMS, PROBLEM DESCRIPTION, CLINICAL EXAMINATION, HYPOTHESIS, COMMENTS, FAMILY HISTORY, and SOCIAL CIRCUMSTANCE, etc. As already mentioned in Chapter 6, much of the free text has a de facto structure and so we can use this notation to identify what categories the free text belongs to. In addition to the de facto notation, we can use the following heuristics to aid the automation process of classifying patient-record information into CareActType categories:

- Information on REASON FOR ENCOUNTER can be extracted from the daily schedule (i.e., part of the appointment system (Appendix A))

- Information on SYMPTOMS can be defined as a list of the most used symptoms in the free texts, such as fever, dizziness, stomach ache, pain in foot, pain in shoulder, etc.

- Information on PREVIOUS ILLNESS / DISEASE can be extracted because it usually refers to diagnostic names, or it is referenced as *past* information
- Information on FAMILY HISTORY can be extracted because it usually contains a reference to some family members (e.g., sister, brother, uncle, aunt, mother, father, etc.)
- Information on WORK HISTORY can be extracted because it often contains a description of the job, type of profession, job title, etc.

Concepts such as SOCIAL CIRCUMSTANCES, COMMENTS and PROBLEM DESCRIPTION refer to more challenging labels to automatically encode or classify.

We are presently in the process of automating the labeling process[1]. So far, we have been able to extract around 50% of the text in patient records and classify them according to concepts in the CareActType ontology; these are promising test results which indicate that information units in Norwegian patient-record entries can be labeled according to CareActTypes concepts.

Templates and Archetypes

By integrating process knowledge into the record system, our approach is based on the assumption that the system can itself recognize and encode its own input and content. We have categorized and labeled information units, as a first step in the approach. The focus up to this point has been to determine the various categories and to see the potential impact on the information model that these categories have had on the extraction of information.

Once information units are classified, they are still free-text information units with a label. For future work, the free-text units can be further formalized. The templates presented by openEHR [16, 17] as *archetypes* seems like a sensible way to proceed. For example,the measurement of arterial blood pressure [25] is recorded as a text string like *BP 160/95*. Once this text string is recognized as a clinical finding (and a specific type of clinical examination), the text string can be formalized according to templates for blood pressure like $BP < systolicpressure >/< diastolicpressure > mmHg$ or $BP < systolicpressure >–< diastolicpressure > mmHg$. Archetypes include detailed semantics and we suggest to do the formalization process of information units in a top-down fashion. We suggest for the system first to recognize the content according to a category, then to formalize it further into a specific notation such as a template.

14.3 The Knowledge Engineering Methods

The CommonKADS is a suite of knowledge modeling tools. CommonKADS tools range from six different types of models (Section 2.5.3) that intend to capture different aspects of knowledge, to templates for knowledge models based on previous experiences in building knowledge systems, to libraries of problem-solving methods . “The

---

[1] Results from this project are reported in [58]. The project is a student project, fall 2002, at NTNU, Department of Computer and Information Science under the supervision of Associate Professor Øystein Nytrø .

aim of CommonKADS is to fill the need for a structured methodology for KBS projects by constructing a set of engineering models built with the organization and the application in mind" [168]. To construct our knowledge models, our ontologies, and our knowledge base, we used the Protégé-2000 system .

Of the various models that CommonKADS suggest (listed in Section 2.5.3), we have implemented the *organization model*, *the task model*, *the agent model*, and *the communication model*. Communication is modeled as information flow between primary-care activities or between an activity and parts of the patient-record system. The organization model, the task model, and the agent model are combined into one comprehensive process model. Remaining components from the CommonKADS suite that we have not implemented are the following:

- Inference knowledge: which is modeled in terms of operations on domain knowledge, domain knowledge which is captured in, for instance, the primary-care process model. Inference knowledge is similar to first-order logic rules that describe how domain knowledge can be combined in order to infer new knowledge. With inference modeling the aim is not to describe sequence or purpose of the inferences, only to describe input and output specifications and reference to domain knowledge that is needed for the inference

- Expertise knowledge: which is modeled in terms of input and output to problem-solving methods (i.e., control knowledge). With expertise-knowledge modeling the aim is to reduce the gap between the domain knowledge and the various problem-solving mechanisms that are needed to perform required functions in the prospective application. In Section 2.5.2 we presented a four-step approach to implementation of second-generation knowledge-based systems, where the mapping between the selected problem-solving methods and the domain knowledge (in terms of ontologies and knowledge bases) is one of the four steps. The mapping process is the main effort in expertise model construction

- Design knowledge: which is modeled in terms of architecture design, application design, and platform design. The Protégé-2000 system provides components that deal with these areas, in terms of back-end, front-end, and utility plug-ins (Section 2.5.5).

Both the CommonKADS methodology and the knowledge acquisition tools in Protégé provide means to help reduce the cognitive and abstruse tasks related to knowledge-system construction. Due to our initial modeling, we are not able to evaluate our method with something to compare against. As a digression, what would be interesting to test out, would be the reimplementation of the MYCIN system (Section 2.3.1) using the CommonKADS methods and Protégé tools. A reimplementation of the MYCIN system would enables us to determine and evaluate the impact of change in knowledge-engineering methods from first-generation to a second-generation knowledge system. We would then see explicitly the impact and usefulness of knowledge modeling when compared to just first-generation expert systems. We would expect to see significant changes in the representation formalism (in opposition to plain rule-based representation), extended domain knowledge, organization, task, and agent knowledge integrated into the system, and to see "newMYCIN" as an integrated part of a clinical information system.

14.4 Summary

Requirements on the user that are induced by the problem-oriented information is limited to one. The user has to associate an encounter summary (i.e., an entry unit) with a problem name, much in the same way as the user now associates a diagnose to almost every encounter summary that is written. If the user does not associate an encounter summary with a problem name, the encounter summary will appear as a stand-alone entry in the patient record, like an entry in the time-oriented patient record. However, there is an essential difference between the problem-oriented information model that we have introduced and the time-oriented models. Every information unit in the problem-oriented information model is labeled with a clinical concept. The clinical concepts represents an association to activities in the care process. Even though, encounter summaries are not associated with a problem name, their information units are still related to relevant care situations through their labels.

The labeling of information units is manually done, but we have presented heuristics that make it possible to automate the labeling process of information units. The results we have so far are promising, on extracting information units, recognize them, and classify them.

We suggest in the future to extend the framework components with medical knowledge that will make it possible to implement a more advanced utility function for ranking information in the patient-record to care activities in the primary-care process. For instance, once the problem-oriented patient record system is in clinical use, the ability to associate various medical problems with each other, we assume, will be considered a highly useful and relevant extension to the patient-record system.

Chapter 15

Concluding Remarks

The work presented in this thesis is based on the use of knowledge-elicitation techniques to investigate how we can create knowledge models for a problem-oriented patient record. Moreover, we have used these models to demonstrate how a knowledge base of patient records can give a better overview over and access to information, in terms of improved structure, than time-oriented patient records.

We divide the contribution of this thesis into three: (1) empirical studies, (2) a knowledge model, and (3) a knowledge base. Results from the studies can be used in more than one way. However, the empirical studies are used with their direct contribution to the knowledge models and as recommendations for the continuing research and implementation of a future patient-record system.

The clinical setting we refer to is Norwegian primary-care practice and Norwegian patient-record systems, but the contributions are generalizable to other clinical settings as well.

Contribution from Empirical Studies

We have carried out three empirical studies: (1) an observation study of family physicians in their clinical settings, (2) a study of patient-record contents, and (3) a quantitative and a qualitative survey among family physicians, with respect to their opinions on use and usefulness of patient-record systems in primary-care practices. Contributions from these studies, in addition to their inherent empirical values, are:

- **A primary-care process model of care activities.** The care activities play an important role in family physicians' medical decision-making process, in terms of providing and retrieving information necessary for clinical decisions that must be made during patient encounters. Each care activity is modeled with respect to its overall goal and purpose, participants involved in the activity, and information needs from the patient-record system

- **An ontology of clinical concepts.** The clinical concepts classify and represent the patient-record contents. To keep intact how physicians work with the record system, the concepts are based on what information family physicians presently write into the patient records. The concepts are formal headings for communicating clinical information. To communicate information implies that the sender and the receiver have a common understanding of what to communicate, the concepts, therefore, function as a tool for common understanding and navigation of clinical information in primary-care patient records.

A Problem-Oriented Patient-Record Model

A problem-oriented information model supports a problem-oriented view of patient data. Hence, information in the patient record is structured according to problem units. A problem unit comprises patient data that relates to the same medical problem. In a problem-oriented patient record, overview is facilitated through problem orientation.

In addition to supporting problem orientation, the information model is based on associations among information units and clinical concepts from the ontology. Every information unit is labeled with a clinical concept. If information units are not associated with a problem unit, they are still associated with activities in the primary-care process through their concept labels. Because of the relationship between concept and activity, the information model is able to rank patient-record contents according to relevance to both a problem unit and a decision-making activity. Overview is, therefore, provided on three levels: (1) through the problem orientation, (2) through the clinical concept ontology, and (3) through the associations represented by the clinical concepts, which make possible a ranking of patient-record information relevant to care activities in the process model.

A Knowledge Base of Patient Records

The knowledge base comprises three knowledge components. The components are the problem-oriented information model that we have used to re-represent time-oriented patient records into problem-oriented patient records; the CareActType ontology of clinical concepts; and the primary-care process model. The knowledge components provide the record system with means to: (1) support a problem-oriented structure and (2) recognize, encode, and relate patient data to physicians' decision-making activities.

15.1 Results

With the knowledge base we can simulate patient-specific encounter processes. A patient-specific encounter process represents a collection of decision-making activities from the clinical encounter. For each activity that takes place in the clinical encounter, we present relevant information that is extracted from the patient record. Three levels of relevant information is demonstrated for each care activity:

- Patient information that is specifically ranked for the decision-making activity
- A list of medical problems which projects the patient's medical condition of past and current problems

- Problem-specific information for each problem in the problem list mentioned above

Patient-record content can be ranked according to several criteria for ranking; we have demonstrated two ways of ranking: (1) a problem-oriented view of data which sorts information in the record according to medical problems and problem lists, and (2) a set of clinical concepts that sort information according to concepts, in addition to their inherent relationship to decision-making activities.

The knowledge representation can be extended with other means for relevance ranking. For example, by extending the knowledge representation with medical knowledge, a more advanced utility function can be used to rank the record information. Once the problem-oriented patient-record system is in clinical use, the ability to associate various medical problems with each other, we anticipate will be considered a useful and helpful extension to the patient-record system.

15.2 Recommendations for a Helpful Patient-Record System

Family physicians are generally satisfied with the existing record systems *except for the lack of overview that they find in the time-oriented patient records*. Our attempt to provide a better overview, as we have presented, is through an information architecture that supports both a problem-oriented view of patient data and means for relevance ranking of information.

In addition to the contributions the empirical studies gave to the creation of knowledge models, we can summarize, from them, an additional set of recommendations and requirements for a future patient-record system:

1. The primary-care process habits certain characteristics that make it a feasible process to model:
    (a) The primary-care process is repetitive and has a few and limited set of subprocesses and activities; only 13 subprocesses and 23 activities, which are not many. The process is, therefore, easy to model, in terms of a generic process
    (b) The uniqueness in each encounter is projected in the patient record as a summary of the clinical encounter. Given the repetitive nature of the primary-care process, each patient encounter is still unique. It is the patient, the patient's medical condition, and physicians' selection of activities that make every encounter unique
    (c) The patient-record system is a *content provider* and holds much of what physicians need of patient-specific information. Given the limited set of generic activities, it is the lookup in the patient records that fills the encounter process with content on what has been done (i.e., history of encounter summaries)
    (d) Physicians have access to the patient record at all times during patient care. A large portion of the patient care takes place in the physicians office, and

there is little information flow between the various clinical offices (i.e., the physician office, the laboratory and the front office)

2. The patient-record system is a *context provider* that helps physicians to switch contexts from one patient to another, or to switch context from one encounter to another. In a pre-encounter assessment phase family physicians use the patient-record system to get into context. In this phase, the record system can contribute significantly with a record structure that enables physicians to get an overview over contents in the patient record (e.g., clinical concepts, medical problems, and relevant information to the care activities)

3. Physicians want an easy-to use record system with few functions that match directly with the clinical tasks in patient care. Family physicians consider their record systems as user friendly and nonrestrictive in the clinical encounters. Physicians also use *all* the information they enter into the record system. We can explain the user satisfaction with the way that primary care is organized. For example, the patient-record system is integrated into the care process, where physicians have access to the record system at all times; physicians are highly motivated to enter information into the record system because they can use this information whenever they need it; and the system does not restrain physicians when they work (i.e., there is no controlled data entry)

4. Physicians claim that time is a critical factor. However, family physicians spend willingly a significant part of the encounter to write detailed and long narratives. Time measured up against the benefit of being able to use information in the patient record seems to be less important

5. Physicians think that a good record system can contribute specifically in patient care. Family physicians state that they have no reluctance toward learning new record systems or letting new technology play a greater impact on the patient-care process than now. These statements give high hopes that physicians will employ future record systems; as long as developers strive to maintain what physicians already express they appreciate about their record, and just add extensional features that physicians express they need.

Part V

Appendices

Appendix A

Glossary

This chapter contains a glossary of words, terms, expressions, and vocabulary used in the thesis. Entries that are marked by * are taken from Bentzen's [20] WONCA International Dictionary for Family Practice.

Anamnesis: A preliminary case history of a patient

Appointment system: The system used by a physician to plan and regulate the timing of patient encounters

*** Assessment:** (Syn. evaluation, examination, or opinion) has several meanings depending on the context in which it is used. In family practice assessment is often used in the following clinical situations:

- **General or complete assessment:** A standardized procedure to determine the physical, mental, and social well-being of the patient with appropriate investigations, including a complete record of findings and advice to the patient
- **Specific or partial assessment:** Includes a history and detailed examination which relates to a specific diagnosis or problem with appropriate investigations, and including a complete record of findings, and advice for the patient
- **Functional assessment:** The measurement, both objectively and subjectively, over a stated period of time of a person's ability to perform and adapt to his environment

Attribute : An inherent characteristic

Case history: See MEDICAL HISTORY

Category : Any of several fundamental and distinct classes to which entities or concepts belong; a division within a system of classification

CAVE: (in patient records) avoid, escape, elude, evade. Used to indicate a known allergic reaction or allergy

Class : A group, set, or kind that share common attributes

Classification : The act or process of classifying; a systematic arrangement in groups or categories according to established criteria

*** Clinic:** An or part of an establishment for examining and treating patients

Clinical examination: An examination of a patient (where the patient presents a clinical problem) by a physician

*** Clinical outcome:** specifies clinical status as a consequence of the course over time of a health problem, treatment, experiment, or other medical intervention

*** Coding system:** A system that allocates codes to objects, concepts, terms, or any entities (e.g., health problems, procedures, symptoms) using a finite set of numeric or alphanumeric identifiers (codes)

Concept : A unit of thought constituted through abstraction on the basis of properties common to a set of objects

Constraint : The state of being checked, restricted, or compelled to avoid or perform some action

*** Computer-based patient-record system:** (Syn. Electronic Medical Record (EMR)) contains the history of the patient and electronically transferred information from hospitals, laboratories, and other physicians

*** Consultation:** (Syn. encounter) The seaking of advice, where a physician responds to a patient’s reason for encounter, fears, ideas, expectations, or health problems. We characterize the consultation as a dynamic interaction between a physician and a patient with the aim of establishing a common agenda based on personal relationship and mutual trust in order to meet a patient’s needs. The consultation usually takes place on the physician’s premises

*** Diagnostic index:** A system in general practice recording diagnosis, date of presentation, patients’ name (or number), age, and gender. The index is useful when retrieving medical records for cohorts of patients with similar health problems and facilitates follow-up

*** Diagnostic procedure:** procedure used to arrive at a diagnosis. Can include the taking of the history and the performance of a physical examination, but usually refers to additional diagnostic procedures such as laboratory or radiologic procedures

Dictionary : A reference book containing words usually alphabetically arranged along with information about their forms, pronunciations, functions, etymologies, meanings, and syntactical and idiomatic uses; structured collection of lexical units, with linguistic information about each of them

*** Discharge:** (from hospital) the process by which the patient is sent home from hospital

*** Discharge summary:** A resume of the condition of a patient, course of illness, management, and recommendation for treatment after hospitalization

*** Drug:** Any substance, organic or inorganic, used for medical purposes. In addition to its chemical designation it will have an approved name, the generic name, but it may also have several proprietary names given by the manufacturer

*** Drug prescribing:** The physician's right to order drugs to the patient from a pharmacy

*** Encounter:** (Syn. consultation) any professional interchange between a patient and a health-care provider, be this provider a single professional or a health-care team. One or more health issues (problems or diagnoses) may be dealt with at each encounter. When addressing more than one health issue during one encounter, this encounter relates to more than one episode of care

Direct encounter: An encounter in which face-to-face meeting of patient and professional occurs. We further divide a direct encounter into:

Office encounter: (e.g., surgery encounter, consultation) A direct encounter in the health care provider's office or surgery

Home encounter: (e.g., house call, home visit) A direct encounter occurring outside the office (clinic)

Hospital encounter: A direct encounter in a hospital setting. We count one encounter for each patient visit and further divide hospital encounters into:

In-patient encounter: A direct encounter with a patient admitted to the hospital.

Out-patient encounter A direct encounter with a patient not admitted to the hospital, either in the emergency room or in the out-patient clinic (polyclinic)

Encyclopaedias : A work that contains information on all branches of knowledge or treats comprehensively a particular branch of knowledge in articles arranged alphabetically often by subject

Entity : Something that has separate and distinct existence and objective or conceptual reality

Facet : Any of the definable aspects that make up a subject (as of contemplation) or an object (as of consideration)

*** Family physician:** (Syn. general practitioner (GP), family practitioner, family doctor) a medical practitioner who provides primary and continuing care to patients and their families within their community. WONCA further defines the family physician as, "The physician who provides care for both sexes of all ages, for physical, behavioural, and social problems"

*** Family practice:** (Syn. general practice) a specialized branch of medical practice provided by FPs / GPs. The use of the terms general practice and general practitioner has come under some criticism. Many medical practitioners in the primary health care prefer the terms family physician and family medicine in order to emphasize the recognition of their branch of medical practice as a speciality in its own right

Follow-up encounter: An encounter between patient and physician in which followed up of an episode, previously initiated, occurs

*** Home visit:** (Syn. house call) the GP's / fp's visit to a patient at his own home rather than seeing the patient in the surgery. In the Western world the number of home visit has declined in the last decade (better communications and transport facilities) and thus a very important source of information about the patient, his family, and living conditions is in danger of being lost

Indirect encounter: An encounter in which no physical or face to face meeting between patient and health care professional occurs. We sub divide these encounters by the mode of communication (e.g., by telephone, letter, or through a third party)

Glossary : A collection of specialized terms with their meanings

Hierarchy : A ruling body of clergy organized into orders or ranks each subordinate to the one above it; the classification of a group of people according to ability or to economic, social, or professional standing; the group so classified; a graded or ranked series

Language : The words, their pronunciation, and the methods of combining them used and understood by a community

Lexicon: A book containing an alphabetically arrangement of the words in a language and their definitions: dictionary. The vocabulary of a language

*** Medical history:** comprises not only earlier and actual diseases but also hereditary disposition, habits, family relations, work, and social status. An elaborate case history is an important tool in diagnostics and treatment, but in general / family practice a problem focused history may be sufficient to solve a problem

*** Medical record:** A file of information relating to contacts in personal health care, comprising data on health status together with personal identifying data including administrative and economic data

Name: Designation of an object by a linguistic expression

Nomenclatures: A system or set of terms or symbols especially in a particular science, discipline, or art; System of terms that is elaborated according to preestablished naming rules

Object: Any part of the perceivable or conceivable world

Ontology: A branch of metaphysics concerned with the nature and relations of being; a particular theory about the nature of being or the kinds of existents

*** Patient record:** A comprehensive record aiming at gathering all health data or information regarding a given patient whichever health care provider has supplied the information

*** Physician assistant:** A person with a limited training in specific medical tasks. He assists the physician performing routine medical work

Polyclinic: out-patients' clinic.

*** Practitioner:** undertakes nursing duties in the practice, which may include home visits

*** Primary-health care:** (Syn. primary care). WHO (Alma Ata 1978) defined primary health care as follows: "Primary health care is essential health care made universally accessible to individuals and families in the community by means acceptable to them, through their full participation and at a cost that the community and country can afford. It forms an integral part of the country's health care system, of which it is the nucleus, and of the overall socio-economic development of the community". Primary health care can be delivered by PHC nurses, physicians or health professionals with a shorter medical training.

ALMA ATA DECLARATION a conference sponsored by WHO in 1978 in Alma Ata, a city in central Asia, and then in the USSR, produced a declaration defining primary health care and made it the basis of health and health care

*** Problem-oriented patient record:** A recording system used to describe the part of the physicians clinical examination of the patient where the findings should be described objectively

*** Problem-oriented medical record:** A medical record in which the patient's history, physical findings, laboratory results, etc. are organized to give a cumulative record of problems. This distinguishes it from the chronological record where encounters are organized in a time sequence. The management of each individual problem dealt with over the successive encounters may be described within each record entry according to the SOAP grid

*** Process:** (Syn. procedure). In medical care constitutes the actions undertaken by a physician to promote the health of his patient. Includes preventive and administrative activities, investigation, diagnosis, treatment, rehabilitation, and cooperation

*** Receptionist:** An office worker who receives patients at the GP's/FP's practice and has administrative responsibilities. Is often the first contact with the health care system

Reason for encounter: The agreed statement of the reason(s) why a person enters the health care system, representing the demand for care by that person. The terms written down and later classified by the provider clarify the reason for encounter and consequently the patient's demand for care without interpreting it in the form of a diagnosis. The reason for encounter should be recognized by the patient as an acceptable description of the demand for care (according to ICPC-2 [186])

*** Referral :** The process by which the responsibility for part or all of the care of a patient is temporarily transferred to another health care provider. Patients may be referred for a specific service, a general opinion, or for other reasons

*** Secondary care :** One of two levels of referred care (secondary and tertiary). Usually refers to care provided by a broadly skilled specialist such as a general surgeon, general internist or obstetrician, to whom the patient may be referred by the GP/FP

*** Sick leave :** Absence from work or study due to illness either certified by a physician or self certified by the individual

*** Sign :** in medicine is an objective finding during a physical examination of a patient

Slot: A place or position in an organization or sequence

*** Symptom:** Any expression of disturbed function or structure of the body and mind by a patient. Cough, pain, and tiredness are symptoms

Systematics: The science of classification; a system of classification; the classification and study of organisms with regard to their natural relationships

Taxonomy: The study of the general principles of scientific classification; especially : orderly classification of plants and animals according to their presumed natural relationships

Term: Designation of a defined concept in a special language by a linguistic expression

Terminology: Nomenclature as a field of study; the technical or special terms used in a business, art, science, or special subject; Set of terms representing the system of concepts of a particular subject field

*** Tertiary care:** One of two levels of referred care: secondary care, tertiary care. Usually refers to highly specialized care provided in a hospital setting such as neuro surgery or heart surgery to which the patients usually are referred by a hospital consultant and only occasionally by the GP/FP

*** Therapeutic index:** in general practice a system recording treatment (drugs) by date of presentation, patient's name (or number), age, and gender. The index is useful when retrieving medical records for cohorts of patients with similar treatment and facilitates follow-up

*** Time of encounter** The time at which the encounter occurs:

> **Encounter during scheduled hours:** Encounter which occurs during usual or posted working hours of the health care providers

Encounter during unscheduled hours: Encounter which occurs during times other than the usual working hours of the health care providers but excluding night encounters

Night encounter: Encounter made during ”night hours” as defined by the health care providers or the health care system

*** Treatment outcome:** The result of a medical or surgical intervention usually assessed after a set period of time

Vocabulary : A list or collection of words or of words and phrases usually alphabetically arranged and explained or defined: lexicon; Dictionary containing the terminology of a subject field.

Bibliography

[1] [http://www.opengalen.org/], August 1 2002.

[2] Russ B. Altman. Informatics in the care of patients: Ten notable challenges. *The Western Journal of Medicine*, 116(2):118–122, 1997.

[3] Russ B. Altman. AI in medicine: The spectrum of challenges from managed care to molecular medicine. AI *Magazine*, 20(3):67–77, 1999.

[4] Russ B. Altman, Michael Bada, Xiaoqian J. Chai, Michelle Whirl Carillo, Richard O. Chen, and Neil F. Abernethy. Riboweb: An ontology-based system for collaborative molecular biology. *IEEE Intelligent Systems and Their Applications*, 14(5):68–76, September / October 1999.

[5] Michael A. Bada and Russ B. Altman. Computational modeling of structural experimental data. *Methods in Enzymology*, 317:470–491, 1999.

[6] Mike Bainbridge, Peter Salmon, Ann Rappaport, Glyn Hayes, John Williams, and Sheila Teasdale. The problem oriented medical record — just a little more structure to help the world go round? Technical report, Clinical Computing Special Interest Group (CLICSIG) of the PHCSG, [http://www.schin.ncl.ac.uk/phcsg/conferences/cam96/mikey.html], June 2000 1996.

[7] Elisabeth Bayegan. The Relationship between Workflow and Buisness Process Re-engineering. 45073 computer science projects, Norwegian University of Science and Technology, Department of Computer Systems and Telematics, Trondheim, 1996.

[8] Elisabeth Bayegan. Business Process Reengineering—A study of Norwegian BPR projects. Master's thesis, Norwegian University of Science and Technology, Department of Computer and Information Science, Trondheim, 1997.

[9] Elisabeth Bayegan. Mediators: (Intelligent) Middleware for Information Systems. Technical Report ISSN: 0802-6394 04 / 02, Norwegian University of Science and Technology, Department of Computer and Information Science, Trondheim, Norway, 1998.

[10] Elisabeth Bayegan. Computational intelligence and knowledge represenation. Technical Report ISSN: 0802-6394 05 / 02, Norwegian University of Science and Technology, Department of Computer Science and Information Technology, Trondheim, Norway, 2000.

[11] Elisabeth Bayegan. Extended databases: Deductive, active, and temporal databases and their integration. Technical Report ISSN: 0802-6394 06/02, Norwegian University of Science and Technology, Department of Computer Science and Information Technology, 2000.

[12] Elisabeth Bayegan and Øystein Nytrø. A problem-oriented, knowledge-based patient record system. In György Surján, Rolf Engelbrech, and Peter McNair, editors, *Health Data in the Information Society—Proceedings of MIE2002*, volume 90 of *Technology and Informatics*, pages 272–276, Netherlands, 2002. IOS Press.

[13] Elisabeth Bayegan, Øystein Nytrø, and Anders Grimsmo. Information ranking in computerized problem-oriented patient records. In Vimla Patel et al., editor, *MEDINFO 2001*, Amsterdam. 2001 IMIA, IOP Press.

[14] Elisabeth Bayegan, Øystein Nytrø, and Anders Grimsmo. Ontologies for knowledge representation in a computer-based patient record. In *Proceedings 14th International Conference on tools with* ARTIFICIAL *I*ntelligence (ICTAI 2002)*. IEEE Computer Society.

[15] Elisabeth Bayegan and Samson Tu. A helpful patient record system: Problem oriented and knowledge based. In *Bio-medical Informatics: One Discipline* AMIA *2002 Annual Symposium Proceedings*. American Medical Informatics Association.

[16] T Beale, A Goodchild, and S Heard. EHR design principles. Technical report, openEHR Foundation, [http://www.gehr.org], November 22 2001.

[17] Thomas Beale. Archetypes constraint-based domain models for future-proof information systems. Technical report, The Good Electronic Health Record, [http://www.gehr.org, 2001.

[18] Bent Guttorm Bentsen, editor. ICPC, ICHPPC-2-D*efined*, IC-P*rocess*-PC—K*lassifikasjoner og definisjoner for primærhelsetjenesten*. TANO, 1991.

[19] Bent Guttorm Bentsen, Dag Bruusgaard, Åsa Rytter Evensen, Per Fugelli, and Ivar Aaraas. *Allmennmedisin—Fag og praksis*, chapter Diagnostikk. Gyldendal Norsk Forlag, 1991.

[20] Niels Bentzen, editor. WONCA I*nternational Dictionary for General/Family Practice*. In Press.

[21] Marc Berg. *Social Science, Technical Systems, and Cooperative Work, Beyond the Great Divide*, chapter Formal Tools and Medical Practices: Getting Computer-Based Decision Techniques to Work, pages 301–330. Lawrence Erlbaum Associates, Hillsdale, NJ, 1997.

[22] Eta S. Berner, George D. Webster, Alwyn A. Shugerman, James R. Jackson, James Algina, Alfred L. Baker, Eugene V. Ball, C. Glenn Cobbs, Vincent W. Dennis, Eugene P. Frenkel, Leonard D. Hudson, Elliott L. Mancall, Charles E. Rackley, and O. David Taunton. Performance of four computer-based diagnostic systems. *The New England Journal of Medicine*, 330(25):1792–1796, June 23, 1994.

[23] Elisa Bertino, Barbara Catania, and Gian Piero Zarri. *Intelligent Database Systems*. ACM Press (Addison-Wesley), 2001.

[24] Lawrence Birnbaum. Rigor mortis: A response to Nilsson's logic and artificial intelligence. *Artificial Intelligence*, 47:57–77, 1991. American Elsevier.

[25] Jan G. Bjålie, Egil Haug, Olav Sand, Øystein V. Sjaastad, and Kari C. Toverud. *Menneskekroppen—Fysiologi og anatomi*. Universitetsforlaget AS, 1998.

[26] Synnøve Ness Bjerke, Trine Bjøro, and Sonja Heyerdahl. Psykiatriske og kognitive aspekter ved hypotyreose. *Tidsskrift for Den Norske Lægeforening*, 121(20):2373–2376, 2001.

[27] Howard L. Bleich. Lawrence L. Weed and the problem-oriented medical record. *MD Comuting*, 10(2):70–71, March / April 1993.

[28] Ronald J. Brachman and Hector J. Levesque, editors. *Readings in Knowledge Representation*, 95 First Street, Los Altos, California 94022, 1985. Morgan Kaufmann Publisher, Inc.

[29] Jan Kristoffer Brenne, Victor Thanh Dieu, Christine Fløysand, Lars Christian Hegde, Per Magnus Løvold, Stein-Arne Sivertsen, Anne Brettingen, Janike Raaen Einerkjær, Marte Fodstad, Line Rørstad Jensen, Martin Thorsen Ranang, Vegard Steinsholt, Bård Støve, Tore Bruland, Cecilie Fløsand, Ingunn Grytting, Frida Letnes, Espen Sigvartsen, and Carina Stjernholm. Bedre elektroniske pasientjournaler i primærhelsetjenesten—en undersøkelse av bruk, behov og barrierer. Technical report, Norwegian University of Science and Technology, April 24 2001.

[30] Philip J.B. Brown and Peter Sönksen. Evaluation of the quality of information retrieval of clinical findings from a computerized patient database using a semantic terminological model. *Journal of the American Medical Informatics Association*, 7(4):392–403, Jul / Aug 2000.

[31] Steven H. Brown. No free lunch: Institutional preparations for computer-based patient records. In Nancy M. Lorenzi, editor, *Journal of the American Medical Informatics Association*, pages 486–490, Philadelphia, November 1999. AMIA'99, Hanley & Belfus Inc., Medical Publishers.

[32] Steven H. Brown and Dario A. Guise. Cross-institutional reuse of a problem statement knowledge base. In *21st Annual Symposium on Computer Applications in Medical Care*, pages 151–155, 1997.

[33] James R. Campbell, Paul Carpenter, Charles Sneiderman, Simon Cohn, Christopher G. Chute, and Judith Warren. Phase II evaluation of clinical coding schemes: Completeness, taxonomy, mapping, definitions, and clarity. *Journal of the American Medical Informatics Association*, 4(3):238–251, May / June 1997.

[34] B. Chandrasekaran, Todd R. Johnson, and Jack W. Smith. Task-structure analysis for knowledge modeling. *Communications of the* ACM, 35(9):124–136, September 1992.

[35] R. Chandrasekaran, J.R. Josephson, and V.R. Benjamins. What are ontologies, and why do we need them? *IEEE Intelligent Systems*, 14(1):20–26, January / February 1999.

[36] Vinay K. Chaudhri, Adam Farquhar, Richard Fikes, Peter D. Karp, and James P. Rice. Open knowledge base connectivity 2.0.3—proposed. Technical report, SRI International and Stanford University, Artificial Intelligence Center of SRI International and Knowledge Systems Laboratory of Stanford University, April 1998. [http://protege.stanford.edu/doc/users_guide/index.html].

[37] CH Cheng and RE Levitt. Contextually changing behavior in medical organizations. In *American Medical Informatics Association (AMIA)*, November 2001.

[38] Vijit Chinburapa, Lon N. Larson, Merrie Brucks, JoLaine Draugalis, J. Lyle Bootman, and Christopher P. Puto. Physician prescribing decisions: The effects of situational involvement and task complexity on information acquisition and decision making. *Social Science and Medicine*, 36(11):1473–1482, 1993.

[39] CG Chute. The Copernican era of healthcare terminology: A re-centering of health information systems. In *AMIA Fall Symposium*, pages 68–73, 1998.

[40] CG Chute, PL Elkin, DD Sheretz, and MS Tuttle. Desiderata for a clinical terminology server. In *AMIA Fall Symposium*, 1999.

[41] Christopher G. Chute, Simon P. Cohn, Keith E. Campbell, Diane E. Oliver, and James R. Campbell. The content coverage of clinical classifications. *Journal of the American Medical Informatics Association*, 3(3):224–233, May / June 1996.

[42] J.J. Cimino. Review paper: Coding systems in health care. *Methods of Information in Medicine*, 35(4 / 5):273–84, 1996.

[43] J.J. Cimino. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods of Information in Medicine*, 37:394–403, 1998.

[44] WJ. Clancey. The epistemology of a rule-based expert system—a framework for explanation. *Artificial Intelligence*, 20:215–251, 1983.

[45] Paul D. Clayton. The state of clinical information systems after four decades of effort. In *Yearbook of Medical Informatics 2001*, pages 333–337.

[46] Enrico Coiera. *Guide to Medical Informatics, the Internet and Telemedicine*. Chapman & Hall Medical, Great Britain, Oxford, 1997.

[47] The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nature Genetics*, 1:25–29, 2000.

[48] Randall Davis, Bruce G. Buchanan, and Edward H. Shortliffe. Production rules as a representation for a knowledge-based consultation program. *Artificial Intelligence*, 8:15–45, 1977.

[49] Randall Davis, Howard Shrobe, and Peter Szolovits. What is a knowledge representation? *AI Magazine*, 14(1):17–33, 1993.

[50] Myriam Deveugele, Ansem Derese, Atie van den Brink-Muinen, Jozien Bensing, and Jan De Maeseneer. Consultation length in general practice: cross sectional stud in six european countries. *BMJ*, 325(7362):472–478, August 2002.

[51] Richard S. Dick, Elaine B. Stern, and Don E. Detmer, editors. *The Computer-Based Patient Record: An Essential Technology for Health Care; Revised Edition*. Committee on Improving the Patient Records, Division of Healht Care Service, Institute of Medicine, National Academy Press Washington, D.C., 1997.

[52] Leaper DJ, Horrocks JC, Stainiland JR, and deDombal FT. Computer assisted diagnosis of abdominal pain using estimates provided by clinicians. *Brittish Medical Journal*, 4:350–354, 1972.

[53] e-jounal of the Association of Medical Directors of Information Systems. The informatics review. [http://www.informatics-review.com/decision-support/index.html], July 11 2001.

[54] Norsk elektronisk legehåndbok. Revmatologi, 2000.

[55] R. Elmasri and S.B. Navathe. *Fundamentals of Database Systems*, chapter Deductive Databases, pages 801–836. Addison-Wesley, 3rd edition, 2000.

[56] Robert B. Elson and Donald P. Connelly. Computerized patient records in primary care: Their role in mediating guideline-driven physician behavior change. *Arch Fam Med*, pages 698–705, 1995.

[57] John W Ely, Jerome A Osheroff, Mark H Ebell, George R Bergus, Barcey T Levy, M Lee Chambliss, and Eric R Evans. Analysis of questions asked by family doctors regarding patient care. *BMJ*, 319:358–361, August 7 1999.

[58] Henrik Fagerholt. Helseinformatikk: Tolking av journaltekst i elektroniske pasientjournaler. Norges teknisk-naturvitenskaplige universitet, Institutt for datateknikk og informasjonsvitenskap, Prosjektarbeid SIF8096P2.

[59] A.A.A. Fernandes, M.H. Williams, and N.W. Paton. A logic-based integration of active and deductive databases. *New Generation Computing*, 15(2):205–244, 1997.

[60] P.A. Flach. From extensional to intensional knowledge: Inductive logic programming techniques and their application to deductive databases. In B. Freitag et al., editor, *Transactions and Change in Logic DBs*, number LNCS 1472, pages 356–387, Heidelberg, 1998. Springer-Verlag Berlin.

[61] Anette Fosse and Anders Bærheim. Nsam-seminaret 1997—premisser for fremtidens datajournal. *Utposten—EDB-spalten*, 515(5), 1997. [http://www.uib.no/isf/utposten/1997nr4/utp97419.htm].

[62] E Galeazzi, A Rossi Mori, F Consorti, and A Errera. A cooperative methodology to build conceptual models in medicine. In *Fifteenth International Congress of the European Federation for Medical Informatics (MIE)*, pages 280–284, Thessaloniki, Greece, 1997.

[63] Howard N. Garb. A conjoint measurement analysis of clinical predictions. *Journal of Clinical Psychology*, 39(2):295–301, March 1983.

[64] Martin Gardner. Information retrieval for patient care. *BMJ*, 314(7085):950–960, 29 March 1997.

[65] Michael R. Gensereth and Steven P. Ketegpel. Software agents. *Communications of the ACM*, 37(7):48–53, July 1994.

[66] Carole Goble, Sean Bechhofer, Danny Solomon, Alan Rector, Anthony Nowlan, and Andrzej Glowinski. Conceptual, semantic and information models for medicine. In *Information Modelling and Knowledge Bases VI, IOS Press*, Amsterdam, 1995.

[67] Mary Anne Graf, Diane D. Tanner, and William R. Swinyard. Optimizing the delivery of patient and physician satisfaction: A conjoint analysis approach. *Health Care Management Review*, 14(4):34–43, 1993.

[68] Anders Grimsmo. Elektronisk pasientjournal i allmennpraksis. Samfunnsmedisinsk forskningssenter Surnadal, 6656 Surnadal, September 1999.

[69] Anders Grimsmo, Fredrik Langballe, Marit Kromberg, Arnt Ole Ree, Betty Pettersen, Berit Otnes, Jon Erik Finnvold, Elin Sæther, Jan Tvedt, Johan Myking, and Jim J. Yang. Definisjonskatalog for allmennlegetjenesten. Technical report, Kompetansesenter for informasjonsteknologi i helsevesenet A / S, 1996.

[70] Anders Grimsmo and Øystein Nytrø. Bedre informasjonsarkitektur og anvendbarhet av den elektroniske journal i primærhelsetjenesten. Et medisinskteknologisk samarbeid. Prosjektbeskrivelse for Norges forskningsråd, Juli 1997.

[71] Tom Gruber. Toward principles for the design of ontologies used for knowledge sharing. Technical report, KSL Technical Report 93-04, Knowledge Systems Laboratory, Stanford University, 1993.

[72] W. Edward Hammond and James J. Cimino. *Medical Informatics: Computer Applications in Health Care and Biomedicine*, chapter Standards in Medial Informatics. Springer-Verlag, second edition, November 2000.

[73] J. Han, Y. Huang, N. Cercone, and Y. Fu. Intelligent query answering by knowledge discovery techniques. *IEEE Transactions on Knowledge and Data Engineering*, 8(3):390–373, June 1996.

[74] Per Kristian Hilden and Anne-Lise Middelthon. Kvalitative methoder i medisinsk forskning—et etnografisk perspektiv. *Tidsskrift for Den Norske Lægeforening*, 122(25):2473–6, 2002.

[75] Per Hjortdahl, Steinar Hunskår, Terje Johannessen, Knut Arne Holtedahl, Anne Luise Kirkengen, and John Nessa. *Allmennmedisin—Klinisk arbeid*, chapter 1.1-1.4. Ad Notam Gyldendal AS, 1997.

[76] I. Horrocks, A. Rector, and C. Goble. A description logic based schema for the classification of medical data. In F. Baader, M. Buchheit, M.A. Jeusfeld, and W. Nutt, editors, *Reasoning about structured objects: knowledge representation meets databases. Proceedings of the 3rd Workshop KRDB'96*, pages 24–28, 1996.

[77] Edward H.Shortliffe and G.Octo Barnett. *Medical Informatics: Computer Applications in Health Care and Biomedicine*, chapter Medical Data: Their Acquisition, Storage, and Use. Springer-Verlag, second edition, November 2000.

[78] Medical Records Institute. Medical records institute's survey of electronic health record trends and usage. [http://www.medrecinst.com/], 1999.

[79] Medical Records Institute. Fourth annual MRI survey of electronic health record trends and usage. [http://www.medrecinst.com/resources/survey2002/index.shtml], 2002.

[80] Villy Johnsen, Hans-Jakob Haga, and Knut Helgetveit. *Metodebok for revmatologisk praksis*. Norsk revmatologisk forening, 1999. [http://www2.legeforeningen.no/yf/revma/mb/index.shtml].

[81] F. Reed Johnson, William H. Desvousges, Melissa C. Ruby, David Stieb, and Paul de Civita. Eliciting stated health preferences: An application to willingness to pay for longevity. *Medical Decision Making*, 18(suppl.), 1998.

[82] Stef Jossten and Sjaak Brinkkemper. Fundamental concepts for workflow automation in practice. Technical report, Centre for Telematics and Information Technology, University of Twente, P.O.Box 217, 7500 AE ENSCHEDE, The Netherlands, March 1995.

[83] Daniel Jurafsky and James H. Martin. *Speech and Language Processing—An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice-Hall Inc., Upper Saddle River, New Jersey 07458, 2000.

[84] Stephen Kay. Headings for communicating clinical information. Technical report, Salford Research Group, NHS Research Report, Spring 2001.

[85] Elpida Keravnou and Nada Lavrač. AIM portraits: tracing the evolution of artificial intelligence in medicine and predicting its future in the new millenium. *Artificial Intelligence in Medicine*, 23:1–4, 2001.

[86] E.-H.W. Kluge. Medical narratives and patient analogs: The ethical implications of electronic patient records. *Methods of Information in Medicine*, 38:253–9, 1999.

[87] Karen M. Kudla and Marjorie C. Rallins. Snomed: A controlled vocabulary for computer-based patient records. *Journal of American Health Information Management Association*, 69, 1998. [http://www.ahima.org/journal/feature/feature.9805.2.html].

[88] K.A. Kuhn and D.A. Giuse. From hospital information systems to health information systems - problems, challenges, perspectives. *Methods of Information in Medicine*, 40(4):275–287, 2001.

[89] Manon M. Kuilboer, Johan van der Lei, Arthur M. Bohnen, and Jan H. van Bemmel. The availability of unavailable information. In *AMIA Fall Symposium*, 1997.

[90] H. Lamberts and I. Hofmans-Okkes. Episode of care: a core concept in family practice. *Journal of Family Practice*, 42:161–7, 1996.

[91] Henk Lamberts and Malrice Woods, editors. ICPC I*nternational* C*lassification of* P*rimary* C*are*. Oxford medical Publications.

[92] Jayne Lawrence. Headings for communicating clinical information. Technical report, SCIPiCT Project Team, NHS Wales, May 2001.

[93] Rober S. Ledley and Lee B. Lusted. Reasoning foundations of medical diagnosis. *Science*, 130(3366):9–21, July 3 1959.

[94] Doug Lenat, George Miller, and Toshio Yokoi. CYC, WordNet and EDR: Critiques and Responses. *Communications of the* ACM, 38(11):45–48, November 1995.

[95] Douglas B. Lenat. Ontological versus knowledge engineering. *IEEE Transactions on Knowledge and Data Engineering*, 1(1):84–88, March 1989.

[96] Douglas B. Lenat. CYC: a large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38, November 1995.

[97] Hector J. Levesque and Ronald J. Brachman. A fundamental tradeoff in knowledge representation and reasoning. In *Proceedings of the CICSI/SCEIO Conference 1984*, Ontario, May 1984.

[98] David L.Sackett, R.Brian Haynes, and Peter Tugwell. *Clinical Epidemiology—A Basic Science For Clinical Medicine*, chapter Clinical Diagnostic Strategies. Little, Brown & Company, Boston, second edition, 1991.

[99] Henry P. Lundsgaarde, Pamela J. Fischer, and David J. Steele. *Human Problems in Computerized Medicine*. University of Kansas Publications in Anthropology Number 13, 1981.

[100] Yves A. Lussier and Michael Bourque. Comparing snomed and icpc retrieval accuracies using relational database models. In *AMIA Fall Symposium*, 1997.

[101] Shwe MA, Middleton B, Heckerman DE, Henrion M, Horvitz EJ, Lehmann HP, and Cooper GF. Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base I. The probabilistic model and inference algorithms. *Methods of Information in Medicine*, 30(4):241–255, 1991.

[102] Arne Maas and Lukas Stalpers. Assessing utilities by means of conjoint measurement: An application in medical decision analysis. *Medical Decision Making*, 12:288–297, 1992.

[103] Arne Maas and Peter Wakker. Additive conjoint measurement for multiattribute utility. *Journal of Mathematical Psychology*, 38:86–101, 1994.

[104] Kirsti Malterud. *Kvalitative metoder i medisinsk forskning*. Tano-Aschehoug AS, 1996.

[105] Kirsti Malterud. Kvalitative metoder i medisinsk forskning—forutsetninger, muligheter og begrensninger. *Tidsskrift for Den Norske Lægeforening*, 122(25):2468–72, 2002.

[106] Ronni T. Marshak. Workflow white paper—an overview of workflow software. *Workflow Computing Report*, 16(X):15–42, 1994.

[107] John McCarthy and Patrick J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In Meltzer and Michie, editors, *Machine Intelligence*, volume 4, pages 463–504. Edinburgh University Press, 1969.

[108] A.T. McCray and S.J. Nelson. The representation of meaning in the UMLS. *Methods of Information in Medicine*, 34:193–201, 1995.

[109] Eiving Meland, Hanne Ellekjær, Bjørn Gjelsvik, Alf Kimsås, Jostein Holmen, and Irene Hetlevik. Medikamentell forebygging av hjerte- og karsykdommer i allmennpraksis. *Tidsskrift for Den Norske Lægeforening*, (22):2643–2647, 2000.

[110] [http://www.webster.com/].

[111] Marvin Minsky. *The Psychology of Computer Vision*, chapter A Framework for Representating Knowledge. McGraw-Hill, New York, 1975.

[112] Tom M. Mitchell. *Machine Learning*. McGraw-Hill Companies, Inc., 1997.

[113] C. Mohan, G. Alonso, R. Guenthoer, and M. Kamath. Exotica: A research perspective on workflow management systems. *Data Engineering Bulletin*, 18(1):19–26, March 1995. [http://www.almaden.ibm.com/cs/exotica/exotica_papers.html].

[114] Gianpaolo Molino. *From Clinical Guidelines to Decision Support*, pages 3–12. Lecture Notes in Artificial Intelligence. Springer-Verlag, Berlin Ehidelberg, 1999.

[115] Stefania Montani and Riccardo Bellazzi. Intelligent knowledge retrieval for decision support in medical applications. In V. Patel et al., editor, *MEDINFO 2001*, Amsterdam, 2001. IMIA, IOS Press.

[116] Robert Müller. The CLINICON framework for context representation in electronic patient records. In *AMIA Fall Symposium*, 1997.

[117] M.A. Musen, R.A. Fergerson, W.E. Grosso, N.F. Noy, M. Crubezy, and J.H. Gennari. Component-based support for building knowledge-acquisition systems. In *Conference on Intelligent Information Processing (IIP 2000) of the International Federation for Information Processing World Computer Congress (WCC 2000)*, Beijing, 2000.

[118] Mark Musen. *Knowledge Engineering and Agent Technologies*. IOS Press, Amsterdam, 2002.

[119] Mark A. Musen. Scalable software architecture for decision support. *Methods of Information in Medicine*, 38:229–38, 1999.

[120] Mark A. Musen. Medical informatics: Searching for underlying components. *Methods of Information in Medicine*, In press.

[121] Mark A. Musen, Samson W. Tu, Amar K. Das, and Yuval Shahar. EON: A component-based approach to automation of protocol-directed therapy. *JAMIA*, 3:367–388, 1996.

[122] Allen Newell. The knowledge level. *Artificial Intelligence*, 18:87–127, 1982.

[123] Gunnar Nilsson, Hans Åhlfeldt, and Lars-Erik Strender. Computerisation, coding, data retrieval and related attitudes among swedish general practitioners—a survey of necessary conditions for a database of diseases and health problems. *International Journal of Medical Informatics*, 65:135–143, 2002.

[124] Nils J. Nilsson. Logic and artificial intelligence. *Artificial Intelligence*, 47:31–56, 1991. Elsevier.

[125] Natalya Fridman Noy and Carole D. Hafner. The state of the art in ontology design—a survey and comparative review. *American Association for Artificial Intelligence AI Magazine*, Fall 1997.

[126] N.F. Noy, R.W. Fergerson, and M.A. Musen. The knowledge model of Protégé: combining interoperability and flexibility. In *2nd International Conference on Knowledge Engineering and Knowledge Management (EKAW'2000)*, Juan-les-Pins, France, 2000.

[127] N.F. Noy and D.L. McGuinness. Ontology development 101: A guide to creating your first ontology. Technical report, Stanford University, Stanford, California, 94305 USA, 2001.

[128] E. Nygren and P. Henriksson. Reading the medical record. I.: Analysis of physicians' way of reading the medical record. *Computer Methods and Programs in Biomedicine*, (39):1–12, 1992.

[129] Magne Nylenna. Gode journaler er god medisin. *Tidsskrift for Den Norske Lægeforening*, 112(28):3560–4, 1992.

[130] SCHIN-Sowerby Centre of Health Informatics at Newcastle. Prodigy—pracitcal support for clinical governance. University of Newcastle and NHS Executive, Newcastle, UK. www.prodigy.nhs.uk.

[131] National Library of Medicine. Fact sheet—UMLS Metathesaurus. [http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html], August 4 2002.

[132] National Library of Medicine. Fact sheet—UMLS Semantic Network. [http://www.nlm.nih.gov/pubs/factsheets/umlssemn.html], August 4 2002.

[133] National Library of Medicine. Fact sheet—UMLS SPECIALIST Lexicon. [http://www.nlm.nih.gov/pubs/factsheets/umlslex.html], August 4 2002.

[134] National Library of Medicine. Fact sheet—unified medical language system. [http://www.nlm.nih.gov/pubs/factsheets/umls.html], August 4 2002.

[135] Vimla L. Patel, Andre W. Kushniruk, Seungmi Yang, and Jean-Francois Yale. Impact of a computer-based patient record system on data collection, knowledge organization, and reasoning. *Journal of the American Medical Informatics Association*, 7(6):569–585, Nov / Dec 2000.

[136] Chris Pearson. Headings for communicating clinical information. Technical report, NHS Information Authority, March 2000.

[137] Mor Peleg, Iwei Yeh, and Russ B. Altman. Modeling biological processes using workflow and petri net models. *Bioinformatics (accepted December 2001)*.

[138] Asunción Gómez Pérez and V. Richard Benjamins. Overview of knowledge sharing and reuse components: Ontologies and problem-solving methods. In V.R. Benjamins, B. Chandrasekaran, A. Gómez-Pérez, N. Guarino, and M. Uschold, editors, *IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5)*, pages 1–1 – 1–15, August 2 1999.

[139] NHS Clinical Headings Project. Headings for communicating clinical information flyer. [http://www.nhsia.nhs.uk/headings], January 2001.

[140] Silvana Quaglini, Mario Stefanelli, Giordano Lanzola, Vincenzo Caporusso, and Silvia Panzarasa. Flexible guideline-based patient care flow systems. *Artificial Intelligence in Medicine*, 22:65–80, 2001.

[141] Miller RA, Pople HE, and Myers JD. INTERNIST-1—an experimental computer-based diagnostic consultant for general internal medicine. *New England Journal of Medicine*, 307:190–209, 1982.

[142] A. Rector and I. Horrocks. Experience building a large, re-usable medical ontology using a description logic with transitivity and concept inclusions. In *Proceedings of the Workshop on Ontological Engineering, AAAI Spring Symposium (AAAI'97). AAAI Press*, Menlo Park, California, 1997.

[143] AL Rector. Coordinating taxonomies: Key to re-usable concept representations. In *Fifth conference on Artificial Intelligence in Medicine Europe (AIME)*, pages 17–28, Pavia, Italy, 1995. Springer.

[144] AL Rector, S Bechhofer, CA Goble, I Horrocks, WA Nowlan, and WD Solomon. The GRAIL concept modelling language for medical terminology. *Artificial Intelligence in Medicince*, (9):139–171, 1997.

[145] AL Rector, A Gangemi, E Galeazzi, and A Rossi-Mori. The GALEN CORE model schemata for anatomy: Towards a re-usable application-independent model of medical concepts. In *Twelfth International Congress of the European Federation of Medical Informatics (MIE)*, pages 229–223, Lisbon, Portugal, 1994.

[146] A.L. Rector, A.J. Glowinski, W.A. Nowlan, and Angelo Rossi-Mori. Medical-concept models and medical records: An Approach Based on GALEN and PEN&PAD. *Journal of the American Medical Informatics Association*, 2(1):19–35, Jan / Feb 1995.

[147] A.L. Rector, W.A. Nolan, and S. Kay. Foundations for an electronic medical record. *Methods of Information in Medicine*, 30:179–188, 1991.

[148] AL Rector, WA Nowlan, S Kay, CA Goble, and TJ Howkins. A framework for modelling the electronic medical record. *Methods of Information in Medicine*, 32(2):109–119, 1993.

[149] AL Rector, JE Rogers, and P Pole. The GALEN high level ontology. In *Fourteenth International Congress of the European Federation for Medical Informatics (MIE)*, Copenhagen, Denmark, 1996.

[150] AL Rector, WD Solomon, WA Nowlan, and TW Rush. A terminology server for medical language and medical information systems. In *IMIA WG6*, Geneva, May 1994.

[151] Alan Rector. Clinical terminology: What is it so hard? *Methods of Information in Medicine*, 38:239–252, 1999.

[152] Alan Rector. AIM: a personal view of where I have been and where we might be going. *Artificial Intelligence in Medicine*, 23:111–127, 2001.

[153] L. Reinecke. A black day for clinical judgment. *bmj.com*, 5 December 1999.

[154] Elane Rich and Kevin Knight. *Artificial Intelligence*. McGraw-Hill, second edition, 1991.

[155] Brachman RJ and Schmoltze JG. An overview of the KL-ONE Knowledge Representation System. *Cognitive Science*, 9:171–216, 1985.

[156] JE Rogers, C Price, AL Rector, WD Solomon, and N Smejko. Validating clinical terminology structures: Integration and cross-validation of read thesaurus and GALEN. In 1998, editor, *AMIA Fall Symposium*, pages 845–849.

[157] JE Rogers and AL Rector. Terminological systems: Bridging the generation gap. In *AMIA Fall Symposium*, 1997.

[158] JE Rogers, WD Solomon, AL Rector, P Pole, P Zanstra, and E van der Haring. Rubrics to dissectioins to GRAIL to classification. In *Fifteenth International Congress of the European Federation for Medical Informatics (MIE)*, pages 241–245, Thessaloniki, Greece, 1997.

[159] Cornelius Rosse, José L. Mejino, Bharath R. Modayur, Rex Jakobovits, Kevin P. Hinshaw, and James F. Brinkley. Motivation and organization principles for anatomical knowledge representation—the digital anatomist symbolic knowledge base. *Journal of the American Medical Informatics Association*, 5(1):17–40, Jan / Feb 1998.

[160] Hallvard Lærum, Gunnar Ellingsen, and Arild Faxvaag. A survey of physicians' use of electronic medical records systems in hospitals. *BMJ*, 323:1344–1348, 2001.

[161] Stuart J. Russell and Peter Norvig. *Artificial Intelligence—A Modern Approach*. Prentice-Hall International, Inc., 1995.

[162] Mandy Ryan and Shelley Farrar. Using conjoint analysis to elicit preferences for health care. *British Medical Journal*, 320:1530–1533, June 3 2000.

[163] P.R.F. Sampaio and N.W. Paton. Deductive object-oriented database systems: A survey. Department of Computer Science, University of Manchester.

[164] Hogne Sandvik. Analysis of an e-mail debate about a general practice list system. *Tidsskrift Norske Lægeforeningen*, 121(30):3509–3512, 2001.

[165] Leiv S.Bakketeig and Per Magnus. Hva er en forskningsprotokoll og hvorfor er den nødvendig? *Tidsskrift for Den Norske Lægeforening*, 122(25):2464–7, 2002.

[166] Guus Schreiber. *Pragmatics of the Knowledge Level*. PhD thesis, Univeristy of Amsterdam, Department of Social Science Informatics, Roetersstraat 15, NL-018 WB Amsterdam, The Netherlands, October 1992.

[167] Guus Schreiber, Hans Akkermans, Anjo Anjewierden, Robert de Hoog, Nigel Shadbolt, Walter Van de Velde, and Bob Wielinga. *Knowledge Engineering and Management—The CommonKADS Methodology*. Massachusetts Institute of Technology, 2000.

[168] Guus Schreiber, Bob Wielinga, Robert de Hoog, Hans Akkermans, and Walter Van de Velde. CommonKADS: A comprehensive methodology for KBS development. *IEEE Expert*, 9(6):28–37, December 1994.

[169] W.B. Schwartz, R.S. Patil, and P. Szolowits. Artificial intelligence in medicine. where do we stand? *New England Journal of Medicine*, 316:685–688, March 12 1987.

[170] Martin Severs. Headings for communicating clinical information from the personal health record. [http://www.wpr.nhsweb.nhs.uk/about_ epr/Paper25a.html], January 1998. Downloaded June 2000.

[171] Martin Severs and et al. Towards an information standard for organising clinical communications. Technical report, NHS Information Authority, [http://www.nhsia.nhs.uk/headings/pdf/PositionPaper.pdf], March 2000.

[172] Yuval Shahar. Timing is everything: Temporal reasoning and temporal data maintenance in medicine. In *AIMDM'99*, Lecture Notes in Artificial Intelligence 1620, pages 3–12. 1999.

[173] Derek Sleeman. Book review: Knowledge engineering and management: The commonKADS methodology. *Artificial Intelligence in Medicine*, 24:97–102, 2002.

[174] Richard Smith. What clinical information do doctors need? *BMJ*, 313(7064):1062–1068, 26 October 1996. [http://www.bmj.com/cgi/content/full/313/7064/1062/a].

[175] Ian Sommerville. *Software Engineering*. Addison-Wesley, 6th edition.

[176] Kent A. Spackman and Keith E. Campbell. Compositional concept representation using SNOMED: Towards further convergence of clinical terminologies. In *AMIA Fall Symposium*, 1998.

[177] [http://smi.stanford.edu/].

[178] Mario Stefanelli. Artificial intelligence for building learning health care organizations. In W. Horn et al., editor, *AIMDM'99*, LNAI 1620, pages 13–29, Berlin Heidelberg, 1999. Springer-Verlag.

[179] Margaret-Anne Storey, Mark Musen, John Silva, Casey Best, Neil Ernst, Ray Fergerson, and Natasha Noy. Jambalaya: Interactive visualization to enhance ontology authoring and knowledge acquisition in protégé. In *Workshop on Interactive Tools for Knowledge Capture, K-CAP-2001*, Victoria, B.C. Canada, October 20 2001.

[180] R. Studer, V.R. Benjamins, and D. Fensel. Knowledge engineering: Principles and methods. *Data and Knowledge Engineering*, 25(1-2):161–197, 1998.

[181] Bob Sugden, Ian Purves, Nick Booth, Peter Johnson, and Samson Tu. The PRODIGY knowledge architecture for chronic disease management in primary care. Technical report, Stanford Medical Informatics, Stanford University, USA and Sowerby Centre for Health Informatics at Newcastle, Newcastle University, UK, 1999.

[182] Paul C. Tang and W. Ed. Hammond. *The Computer-Based Patient Record: An Essential Technology for Health Care; Revised Edition*, chapter A Progress Report on Computer-Based Patient Records in the United States. Committee on Improving the Patient Records, Division of Healht Care Service, Institute of Medicine, National Academy Press Washington, D.C., 1997.

[183] Paul C. Tang and Vimla L. Patel. Major issues in user interface design for health professional workstations: summary and recommendations. *International Journal of Bio-Medical Computing*, (34):139–148, 1994.

[184] HL7. HL7 reference information model. Technical report, HL7, 2001.

[185] The UMLS Project. Technical report, 2002.

[186] WONCA International Classification Committee. *International Classification of Primary Care ICPC-2*. Oxford Medical Publications, second edition.

[187] Work Group 1. Workflow management coalition interface 1: Process definition interchange process model. Technical Report Document Number WfMC TC-1016-P, The Workflow Management Coalition, October 29 1999.

[188] Samson W. Tu and Mark A. Musen. A flexible approach to guideline modeling. In *AMIA Annual Symposium*, Washington, D.C., 1999.

[189] Samson W. Tu and Mark A. Musen. From guideline modeling to guideline execution: Defining guideline-based decision-support services. Technical report, Stanford Medical Informatics, Stanford University School of Medicine, Stanford, CA, USA, 2000.

[190] J.H. van Bemmel and M.A. Musen, editors. *Handbook of Medical Informatics*. Springer, 1997.

[191] Lawrence L. Weed. Medical records that guide and teach. *MD Computing*, 10(2):100–114, March / April 1993.

[192] L.L. Weed. Medical records that guide and teach. *The New England Journal of Medicine*, 278(12), March 21 1968.

[193] L.L. Weed. *Medical Records, Medical Education, and Patient Care*. The Press of Western Reserve University, fifth printing edition, 1971.

[194] L.L. Weed. New connections between medical knowledge and patient care. *BMJ*, 315(7102):231–240, 26 July 1997.

[195] L.L. Weed. Clinical judgment revisited. *Methods of Information in Medicine*, 38:279–86, 1999.

[196] L.L. Weed and L. Weed. Opening the black box of clinical judgment—an overview. *BMJ*, 319(1279), 13 November 1999.

[197] Robert A. Wilson and Frank C. Keil. *The MIT Encyclopedia of the Cognitive Sciences*. The MIT Press, 1999.

[198] Robert A. Wilson and Frank C. Keil, editors. *The* MIT *Encyclopedia of the Cognitive Sciences*, chapter Computational Intelligence. The MIT Press, 1999.

[199] Terry Winograd and Fernando Flores. *Understanding Computers and Cognition—A New Foundation for Design*. Addison-Wesley, 2.printing edition, November 1987.

[200] John W Wly, Jerome A Osheroff, Mark H Ebell, M Lee Chambliss, Daniel C Vinson, James J Stevermer, and Eric A Pifer. Obstacles to answering doctors' questions about patient care with evidence: qualitative study. *BMJ*, 324:1–7, 23 March 2002.

Index