An Overview of MedView

Youssef Ali Göran Falkman[*] Lars Hallnäs Mats Jontell[§]
Ulf Mattsson[§] Nader Nazari
Olof Torgersson
Chalmers University of Technology and Göteborg University
S-412 96 Göteborg, Sweden

Abstract

We give an overview of the MedView project and discuss background, current status, and future directions. MedView is a joint project with participants from oral medicine and computer science. The overall aim of the project is to develop models, methods, and tools to support clinicians in their diagnostic work. An important part of this is to be able to efficiently analyze and learn from the monumental amount of information being gathered in clinical records. In the MedView project, clinical data is continuously collected into a large knowledge base of formalized patient examinations. The structure of the knowledge base is based on a formalization of health-care processes and clinical knowledge in oral medicine harmonized within the network SOMNET (Swedish Oral Medicine Network). A number of tools have been built which enable users to extend, view, and analyze the contents of the knowledge base. The system permits immediate analysis of information based on the formal model used. It also well suited for education of dental students. Furthermore, it also provides a basis for distant consultations and generates a solid foundation for multicenter trials and activities.

1 Introduction

The MedView project was initiated in 1995 when some researchers at the clinic of Oral Medicine, faculty of Odontology, Göteborg University, and the department of Computing Science at Chalmers University of Technology, got together and started to discuss their respective research interests. They soon found that their interests had an intersection: The odontologists were looking for ways to use

---

[*] Department of Computer Science, University of Skövde
[§] Clinic of Oral Medicine, Faculty of Odontology, Göteborg University

computers to improve their daily work and research, and the computer scientists were looking for an area of knowledge on which to apply ideas on knowledge representation and exploration. The solution was obvious, start a joint project aimed at formalizing knowledge in the area of oral medicine, using the mentioned ideas on knowledge representation, and producing computer based tools for use in clinical and analytical work. MedView was born.

Since then several years have passed and many hours of work have been put into the project by various people. The aim of this paper is to give an overview of the work done in the project, its current status, and hint at directions for future areas of research.

Already at the outset one thing was clear: The project would only be of interest from a clinical point of view if it produced tools which improved the daily work of a clinician in oral medicine. Researchers from the clinic of oral medicine had been involved in several attempts to build computerized systems earlier, and deemed them as failures since they did not really do anything for them, neither as clinicians nor as a researchers. They were more or less just systems to put on a digital media what they already had on paper records or photo slides.

A first strategic decision was to not try to build yet another electronic medical record system, but to focus on knowledge gathering and analysis based on a formal description of the concept "examination". This led to the following things to be done, approximately in the given order:

1. Provide a formal framework and methodology to be used.
2. Formalize the knowledge to be gathered based on this methodology in a close cooperation between odontologists and computer scientists.
3. Develop tools for entering the information gathered at an examination into the knowledge base directly in the examination room.
4. Develop tools for viewing the contents of the knowledge base, both for use in the examination room and later for retrospective studies.
5. Develop tools for analyzing and exploring the knowledge base and for adding concepts built on top of the basic formal method.

Today, the first three steps are essentially finished, while the fourth and fifth of course are of the kind where there is always more to be done.

The rest of this article is organized as follows. In Section 2 we give a description of MedView from a medical point of view. Section 3 gives the theoretical model of MedView from a computer scientist's point of view, and mentions some of the areas of computer science to which MedView applies. Section 4 describes the current status of the project. Section 5, finally, contains ideas and directions for future work.

2 MedView and Oral Medicine

Diagnostic work and clinical decision-making are central items in every field of medical practice, where clinical experience, knowledge and judgment are the cornerstones of health care management. In order to achieve increased competence, the clinician is confronted with complex information that needs to be analyzed. There is considerable evidence that the unassisted human mind is challenged when exposed with multiple sets of data [13, 20, 44, 71]. Therefore, the clinician needs tools to improve analysis and visualization of data in the diagnostic and learning processes.

To support the human mind in extracting valuable patterns in clinical information, computer technology has been introduced in several areas of modern medicine with the aim to assist these cognitive processes. The systems provide a broad functionality, from distant consultations of individual patients to intelligent expert systems, where text and image information is collected and analyzed. In the elaboration of a computerized system, several critical problems have to be mastered in order to ensure that conclusions drawn are correct or justified. In this section, we discuss some issues we have confronted in the MedView project. We also describe how they are handled.

2.1 Clinical Experience and Diagnostics

2.1.1 Nomenclature and Definition of Clinical Information

The first step in the diagnostic process, illustrated in Figure 1 below, consists of gathering and storage of clinical information. In order to be meaningful for interpretation, these data must be recorded in such a way that they can be understood and interpreted in a precise manner by all members of the health care system. This means that a *formalized* and *harmonized* health care system is imperative.

The word “formalize”, in this context, means to establish and formally define basic health care activities that can provide an explicit structure for intelligent reasoning. This formalization is crucial in order to arrive at a correct diagnosis, based on an explicit definition [43, 106]. The term “harmonization” refers to the process of making the formalized activities adapted within a community [22, 29, 79, 82].

The demand and request for formalization and harmonization is certainly not new, but has frequently been associated with obstacles [43, 56, 86, 120]. Although several international attempts have been made to establish a congruent medical nomenclature not many have been successful.

Today, clinical data are frequently expressed in natural language using terms based on individual subjective assessments or interpretations not defined or harmonized within the health care system [3, 81]. Several terminologies exist, often developed within a specific medical discipline, but they are seldom widespread

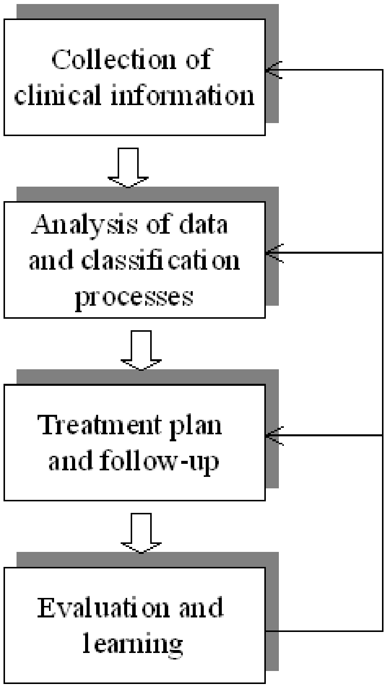


Figure 1: General description of the decision-making process. The process is a chronological sequence where each chain of events present obstacles that have to be controlled in the elaboration of a formalized and harmonized language.

and do not provide a useful international nomenclature system. Furthermore, most terminologies are not related to definitions of terms. Even when definitions exist they can be highly ambiguous. An example is the attempt to define oral leukoplakia. The latest definition reads “a predominantly whitish lesion which can not be diagnosed as any other definable disorder” [7]. This definition is closely related to the clinician’s ability to diagnose all other whitish lesions of the oral mucosa. Thus, to an inexperienced clinician the diagnosis of oral leukoplakia can involve almost any whitish lesion, while the experienced clinician will use it less often.

Evaluation of treatment care and scientific analysis is accordingly not meaningful when registered terms are not precisely defined. New computerized technologies will demand that strategies for clinical registration of data are developed in order to reduce these considerations. Currently, most systems used in the health care sector are not dealing with these problems, but are more focused on transportation and storage of data.

2.1.2 Analysis and Classification of Diseases

The second step in the decision-making process is the analysis of gathered data. The diagnostic process involves the clinician's ability to put the patient into a certain class or group [95, 114]. A diagnosis can be considered as a way of classifying clinical information to facilitate communication between health care providers and to assist in decisions of treatment strategies. The diagnosis is indeed only a common identity of a group of patients with similar clinical information profiles. Consequently, to define a disease it is essential that all patients have an identical information pattern, not shared by any patients who do not have the disease. This is rarely the case [72], and a diagnosis is often based on a description rather than an explicit definition.

The currently used diagnostic system has developed over several centuries. Diagnoses based on pathological anatomy have sometimes been replaced by diagnoses which reflect the introduction of physiology and laboratory research. Patho-anatomical diagnoses as, for example, 'gastric ulcer' was replaced by 'hyperacidity' to denote the patho-physiological dysfunction of this disease. Another problem concerning classification of diseases is that the extension of a disease may change over time along with new discoveries. Lichenoid contact reactions may serve as an example. The diagnosis oral lichen planus was recently split into oral lichen planus and lichenoid contact reaction [12]. However, this subdivision is not yet fully accepted which leaves the diagnostic system in a state of confusion where oral lichen planus may or may not include lichenoid contact reaction.

All in all, the diagnostic systems of today have different backgrounds and there are no rules to promote continuous modification to adapt to new scientific achievements. This lack of harmonization of clinical information will lead to significant problems when new information technologies are to be used in our health care system.

The quality of the analysis and classification process mentioned above is thus influenced by both initial steps in the decision-making process. First, the character and quality of input data will greatly influence our ability to perform subsequent analysis. Second, the classification process is in itself influenced by our ability to adopt adequate and reliable inclusion criteria from input data to a certain disease or diagnosis. Consequently, our knowledge, experience, and treatment strategies of various disorders will be based on conclusions from observations or studies that may not be comparable due to differences in nomenclature or diagnostic criteria. Obviously, it is essential to elaborate routines where these factors are controlled.

2.1.3 Visualization of Clinical Information and Learning

The third step in the decision-making process is the elaboration of treatment strategy and follow-up procedure which emanates as a result of the classification process (Figure 1). From the aspect of treatment strategy, it is common practice to record treatment rendered, but not the diagnostic basis for these treatment decisions. This practice may undervalue diagnosis, but also hamper feedback regarding the effectiveness of treatments relative to specific diagnoses [8].

The fourth and last step can be described as the way we draw conclusions and learn from performed therapies. These experiences are, within the medical community, generally presented as scientific articles or books in order to forward information to increase the knowledge of other clinicians.

Today, the conventional search in index-based volumes has been replaced by computerized databases available to all members of the scientific community. However, complex clinical information stored as images, concepts, videos, etc. is difficult to explore. Concept-based exploration of clinical databases has to be boiled down to a volume of information that is possible to handle. The potential risk with this process is that significant information may be overlooked if it is left outside the search profile. This situation arises especially where clinical hypotheses are to be tested and where the search profile, based on a combination of keywords, will not provide sufficient information. Data with significant interest to the scientist may therefore exist in the database but be left undetected.

In many respects we confront the same problem in clinical research. Information patterns, which may lead to new discoveries, are most likely concealed in large volumes of clinical information stored in non-transparent conventional patient records. Essential information which is not frequently encountered will escape detection as it is hidden in irrelevant information.

Consequently, tools are needed which can intelligently present large volumes of clinical data and where the capacity of the human brain to recognize significant patterns hidden in monumental amount of clinical information is maintained. Therefore, it is important that the capacity of our cognitive function to recognize relevant information and our ability to make rational verdicts is applied, an essential function that computers still are lacking. Computer technology can, however, visualize extractions of complex information as patterns which may initiate associations to new inquires, that may eventually lead to new knowledge.

2.2 What MedView Offers

In short, MedView offers a model for formalization and tools for knowledge gathering, visualization, and analysis of data. The tools are also aimed at improving the everyday work of clinicians in oral medicine. The formalization used was developed in close cooperation between participants from oral medicine and computer science, with the purpose of providing a model suited for both oral medicine and computerized storage and reasoning. The model of the health-care activities and medical expertise involved has evolved through collaboration within the Swedish Oral Medicine Network (SOMNET).

MedView is primarily aimed at increasing the speed by which we may obtain new and valuable information within the field of oral medicine. A formalization

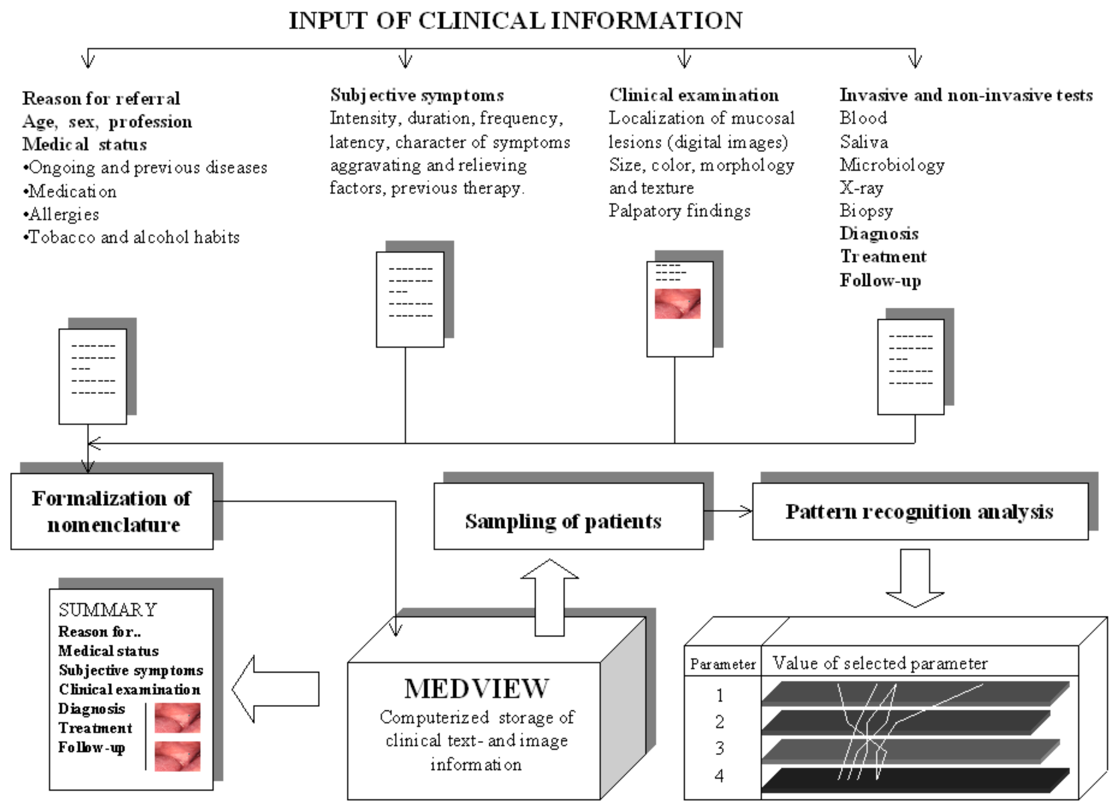


Figure 2: General description of MedView. MedView is used for formalized registration of clinical text- and image-based information into a knowledge base (top). The registered clinical information is synthesized into a readable text and displayed together with clinical images for each patient (bottom left). The contents of the knowledge base is subsequently used for analysis, evaluation, and learning (bottom right).

of clinical procedures and visualization of information provide a possibility for recognizing new trends and patterns otherwise hidden in large amounts of non-transparent clinical records. With MedView, the knowledge and intuition of the clinician can be combined with the potential of the computer to promote analysis and testing of hypotheses in a favorable environment. MedView is also well suited for educational purposes of dental students and in post-graduate training. It allows distant consultations and generates a solid foundation for multi-center trials and activities.

2.2.1 Formalization and Harmonization

When elaborating the MedView system, great care was taken to determine what clinical information could be defined as useful and constitute the foundation in the knowledge base. The result from these considerations was standardized protocols for input of clinical information, where the nomenclature used was developed in close collaboration between the involved clinics. Case history and all clinical data are entered by use of predefined parameters from the mentioned protocols. Through this process a solid base for subsequent analysis and intelligible reasoning of results is obtained. The nomenclature and information structure is thus *formalized* and *harmonized* within the network. The formalized protocols have a logic interpretation (see Section 3.1), which make them suitable for automated reasoning in a computerized system. At the same time, they are simple enough to have an obvious intuitive reading needing no further explanation.

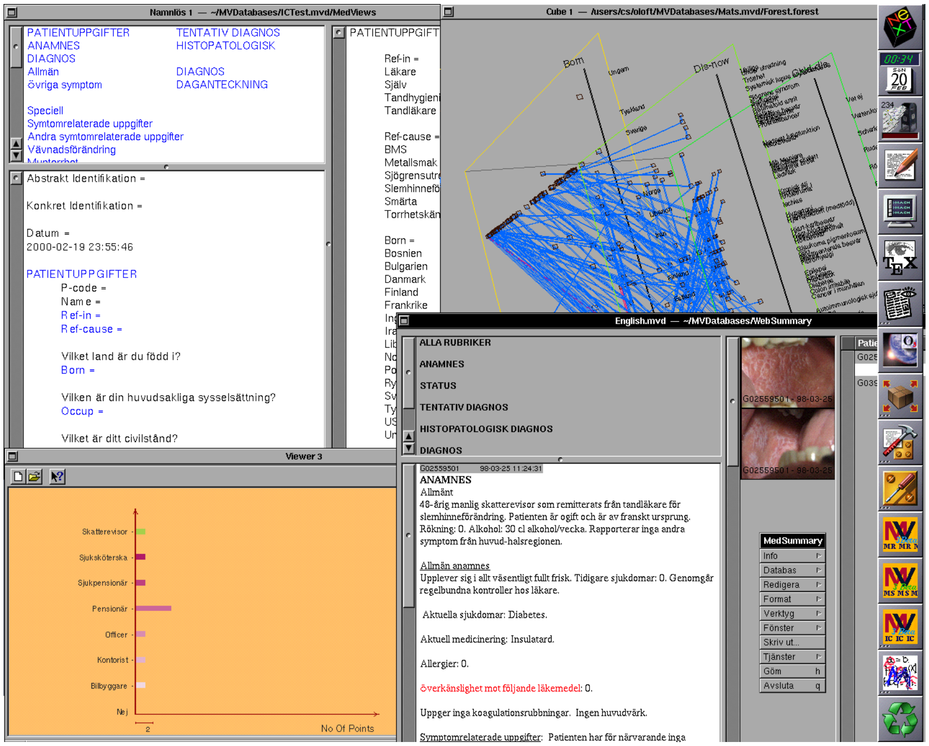
The protocols defined for collection of data are rather extensive including detailed interviews of disease history and protocols for clinical examinations. Existing mucosal lesions are described in terms of localization and clinical appearance. Mucosal lesions are also documented with digital video technique. This technique offers the advantage that the digitized images are immediately accessible in the knowledge base, both for analysis and for distant consultations. Results from biopsies, laboratory tests, and other invasive or non-invasive investigations are included, as are diagnoses, treatment modalities, and clinical outcomes of performed therapies. Additional information not included in the protocols but relevant for the present patient can be included as text.

2.2.2 Everyday Tools

MedView is not all about formalizing and analyzing data. It is also about changing and improving the everyday work of clinicians by providing tools facilitating clinical processes. In part, the development of these tools is a necessity for the success of knowledge gathering: In order to make it possible to collect data in an efficient manner, it must be possible to enter data into the knowledge base during examinations. To motivate this extra work, applications are needed that give immediate feedback in terms of enabling the use of entered data in ways that improve clinical procedures.

The registration of information based on the formalized protocols is done using a specialized input application described further in Section 4.3. Each examination corresponds to a record, including digitized images taken at the examination. Apart from the input application, there are several output applications or *viewers* designed for visualization of obtained information. The viewers are focused on analysis, interpretation, and evaluation, both of individual patients and of groups of patients selected from the knowledge base.

Since an extensive amount of information is collected for each patient, the effect of performing input during examinations is that all information about each



Abstrakt Identifikation =

Konkret Identifikation =

Datum =
2000-02-19 23:55:46

PATIENTUPPGIFTER
P-code =
Name =
Ref-in =
Ref-cause =

Vilket land är du född i?
Born =

Vilken är din huvudsakliga sysselsättning?
Occup =

Vilket är ditt civilstånd?

PATIENTUPPGIFT
Ref-in =
Läkare
Själv
Tandhygieni
Tandläkare

Ref-cause =
BMS
Metallsmak
Sjögrensutr
Slemhinneför
Smärta
Torrhetskän

Born =
Bosnien
Bulgarien
Danmark
Finland
Frankrike

G02559501 98-03-25 11:24:31
ANAMNES
Allmänt
48-årig manlig skatterevisor som remitterats från tandläkare för slemhinneförändring. Patienten är ogift och är av franskt ursprung. Rökning: 0. Alkohol: 30 cl alkohol/vecka. Rapporterar inga andra symptom från huvud-halsregionen.

Allmän anamnes
Upplever sig i allt väsentligt fullt frisk. Tidigare sjukdomar: 0. Genomgår regelbundna kontroller hos läkare.

Aktuella sjukdomar: Diabetes.

Aktuell medicinering: Insulatard.

Allergier: 0.

Överkänslighet mot följande läkemedel: 0.

Uppger inga koagulationsrubbningar. Ingen huvudvärk.

Symptomrelaterade uppgifter: Patienten har för närvarande inga

Figure 3: Some MedView applications. From top left: Input application, 3D viewer, bar chart viewer, summary application.

patient is immediately available in a well-organized searchable knowledge base. When an examination is completed, MedView can generate a summary where all the contained information is displayed as a readable text, with digitized images of the mucosal lesions shown simultaneously. The character of the generated summary is in its layout in most ways similar to the regular patient record encountered in daily practice. The application and the methods used for generating text are described further in Section 4.4. When a patient comes back for a follow-up the system can synthesize a full medical history together with associated images providing the clinician with all the needed background information. The time that can be gained by letting the system generate patient record text, instead of using the traditional method of dictating and typing the text, more than makes up for the extra time required to gather the information.

2.2.3 Analysis Tools

Once information is gathered the analysis and learning phases begin. Typically, these tasks are performed at the users desktop computer rather than in the examination room.

MedView permits selection of patients from the knowledge base according to any combination of parameters included in the registration protocols. A search may thus be of a simple nature as, for example, finding all patients with confirmed "diabetes mellitus". However, it can also be more complex with several criteria involved, as finding "female patients with bilateral lichenoid reactions in buccal mucosa, treated with local application of clobetasol". The search profile can be decided and directed by the user according to the objective or purpose for the analysis. The system identifies patients that fulfill the chosen criteria and displays them in various ways for instance in a traditional bar chart. A screen with several applications, among them a bar chart viewer, is shown in Figure 3.

The selection of patients can be subjected to pattern recognition analysis. An application enabling a three-dimensional display of a multivariate analysis, where the result is shown in a cube which may be rotated and viewed from different angles by the examiner is discussed in Section 4.7. Another, enabling clustering of patients regarded as similar, in some user-defined way, is described in Section 4.8. The hierarchical clustering of examinations is displayed in a three-dimensional tree. The main purpose of these applications is to visualize patterns in a group of selected patients. They are therefore mainly focused on the possibility of learning and testing various hypotheses within the created knowledge base.

2.2.4 Extensibility

The formal model used in MedView is such that the currently used formalization can be easily extended and modified over time. The protocols used can be seen as a first approximation of the needed knowledge structures. When we learn more, the protocols can be extended to collect more information and describe harmonized nomenclature for a larger part of oral medicine. In fact, this process is in progress all the time. When new values are needed to describe a particular attribute in a protocol, they can be added directly. To keep a high level of harmonization it is of course important to communicate such additions within the network.

Recently, new protocols have been developed which include formalization of new concepts like tooth-status. With these new protocols MedView can be used by a broader category of odontologists. Due to the properties of the formal model, see Section 3.1, old examinations are not made obsolete by the introduction of new protocols. It's just that some knowledge available in new examinations might be missing in the old ones. We hope that it will be possible to formalize and harmonize larger and larger parts of the health care activities within oral medicine in the same manner. Then, broaden the view to other parts of odontology.

The formal model of MedView also makes it easy to introduce more complex concepts based on the basic data collected through the common protocols. For instance, it might be desirable to introduce concepts that group together a number of possible values, or to express a new diagnosis in terms of common observations from a number of cases.

2.3 Discussion

MedView addresses the issue of learning from the complex data which originates from everyday clinical practice in the field of oral medicine. To enable this, MedView was designed with the ambition to meet the demands for formalization and harmonization in the decision-making process. In this section we discuss the project and put it in the context of other medical computer based systems.

2.3.1 Gathering Clinical Information

The first step in the decision-making process constitutes the gathering of clinical information. Conventionally, the clinician collects data, which are then summarized and written down in a non-transparent record. The registration and summary applications of MedView resembles the standard way we collect information in daily clinical practice and the comparison to the conventional record is so far obvious. However, all registrations in MedView are performed using protocols with a *formalized* nomenclature where data are stored in a computer. The relevance of individual terms and parameters in the protocols may be debated, but the stringent use of a formalized language creates a basis for reduced discrepancy in clinical registrations within the network.

If a formalized language can be adopted within a community or network, the question arises if the use of a computerized record will facilitate subsequent use of obtained clinical information. Several studies have documented the usefulness of computerized applications in this context, also within the field of odontology [19, 28, 34, 83, 94, 112]. The computerized records enable quick access to reference and educational information [65, 77] and enhanced storage of structured medical knowledge [53]. The MedView system, using both text- and image-based information, is in agreement with these studies. Computer records are also introduced for quality assurance, replacing the paper record. These systems are aimed at assessment and improvement of patient care at the time of treatment, thus building quality management into the caregiving process [83, 115]. However, to the best of our experience, the vast majority of commercial systems available for computerized patient records within odontology are mainly focused on *individual* patient care and the possibility to analyze the *entire* patient material and visualize clinical patterns is usually rather limited. MedView offers the advantage of *combining* the conventional computer record with the possibility for information visualization and analysis.

2.3.2 Analysis and Classification

The second and third steps of decision-making involve analysis and classification processes with subsequent treatment and follow-up (Figure 1). Computer systems may be characterized as *active* or *passive* in decision-making [30]. Passive support occurs when a computer facilitates access to relevant patient data or clinical knowledge for interpretation by the physician. Active support requires that the computer processes input data to a higher level of information, e.g., a diagnostic or expert system. Such models have been developed to enhance clinical security by facilitating the ability to draw conclusions from background knowledge and diagnostic hypotheses [50, 96]. Most studies in this field have been written within medical research, such as dermatology and a rather limited number within odontology [108]. In [108] it was pointed out that a problem with expert systems in general was the lack of accepted clinical terminology in the medical community.

Systems for computerized support in decision-making processes within odontology can coarsely be classified into four groups:

- The first group consists of studies where a system has been developed as an aid or tool in a very specific clinical situation such as design of removable dentures [25, 26, 47], artificial tooth form selection [100], objective assessment of mucosal lesions [67, 68], or surgical operations [31, 98, 99].
- In the second group, computers for decision-making are used in applications with a somewhat broader perspective. This includes applications where predefined criteria or a questionnaire are used as decision parameters for arriving at a correct diagnosis for an individual patient. Applications like this exist in endodontology [51, 73], oral radiology [15, 27, 113], and oral pathology [60, 61, 84, 97]. This group also includes computerized systems for evaluation of diagnostic performance and therapeutic decisions [35, 74] and studies on how decision analysis in general can be applied to dentistry [69, 70].
- The third group describes computerized expert systems in odontology with the characteristics mentioned above. Such systems have been elaborated in orthodontics [49, 102, 107], endodontology [37], oral pathology [36], cariology [10], and oral radiology [1].
- The fourth and last group consists of systems with the purpose of using the potential of neural networks for analysis of decision-making, therapy planning, and quality assurance [14, 105].

To conclude, the vast majority of these systems are mainly focused on treatment planning and decision-making for *individual* patients, rather than on the possibility to generate further *knowledge* through analysis of continuously obtained clinical information. However, attempts with this purpose are described [45, 48]. Furthermore, in oral medicine there are, to the best of our knowledge, very few papers [104, 121] published with the aim to use computer technology and database engineering for any of the above mentioned purposes or as a tool to increase clinical knowledge.

MedView is mainly a passive support system, primarily focused on facilitating pattern detection where hypotheses can be evaluated in a favorable environment.

The knowledge and intuition of the clinician can be combined with the potential of the computer to promote testing of hypotheses and augment analysis. However, the standardized collection of data definitely provides a future possibility for development of active expert systems.

2.3.3 Evaluation and Learning

The last step in the decision-making process is represented by evaluation and learning and to add to our knowledge and experience. However, the reliability of any analysis or learning process depends on the quality of input data and formalization of nomenclature [87, 119]. The same reasoning also applies to the usefulness of expert systems, which is impaired by incompleteness and inaccuracies of the databases. The need and demand to find appropriate standards and nomenclatures is therefore very important since discrepancies in these fields will always decrease the reliability of the systems [30]. Similar thoughts were expressed in [55] where the development of standardized computerized records as a tool in interdisciplinary communication is advocated. All in all, efforts to draw conclusions from performed observations are highly commendable, as long as we remember that the foundation for our analysis is never better than the quality of input data. Again, *formalization* represents the initial but also fundamental part for analysis, decision-making and harmonization and these processes are facilitated when aided by computer technology. A main purpose with MedView is therefore to act as a hypothesis generator.

The use of MedView may also be viewed from the view of education. An individual clinician may not encounter enough cases to develop adequate experience of a certain condition. A network, such as SOMNET, is a way to overcome this problem. All clinics within SOMNET have access to the knowledge base. A multi-center network, with the combined knowledge of individual clinicians creating a knowledge base founded on formalized criteria, increases our ability to reach useful information for education and learning [38, 64, 75, 76]. The formalized protocols generate a possibility for integrated research between clinics. Distant consultations of individual cases have been successful in several tele-medicine applications, among them MedView. Used right, computer technology is most certainly a valuable instrument to increase clinical experience and to promote learning within the field of oral medicine.

3 MedView and Computer Science

The MedView project involves several areas of computer science, mainly knowledge representation, formal reasoning systems, declarative programming, object-oriented programming and software development, artificial intelligence (AI), and human-computer interaction (HCI).

The nature of the project is such that all the above areas are needed and have to be integrated with each other to produce high-quality software tools. These tools are then applied continuously in the everyday work of clinicians and researchers in oral medicine. In addition, hypotheses are directly testable since there is, and has been from the very beginning, an existing userbase. Both applications and knowledge models can be put to the test. If knowledge models cannot be understood and used by the medical experts involved, they are likely not to be of great value to the project. Likewise, applications developed can be introduced and tested. If an application does not provide a useful interface or a meaningful feature set it has to be modified. Examples of applications exist that have been developed but never used. On the other hand, MedRecords described in Section 4.3 has been in use for several years and can be said to be proven a good tool for its task.

A brief description of the basic theoretical model used for knowledge representation in MedView is given in Section 3.1. This model has been used as the basis for a programming system aimed at being the deductive engine of MedView. Currently, the model is not implemented in a uniform manner across applications as discussed in Section 4.2.1.

The various topics mentioned above relates to the MedView project in the following ways:

- Knowledge representation and formal reasoning systems. MedView is based on a theory of definitions [46]. This theoretic model with connections to logic and logic programming is used for all knowledge representation.

- Declarative programming. We believe that declarative programming is a very powerful tool for developing certain kinds of applications, such as symbol-manipulation, knowledge representation, intelligent reasoning etc. Furthermore, declarative programming come very close to the theoretical model used. Consequently, we are developing a declarative programming system [109] based on the theoretical model used in MedView.

- Object Oriented Programming (OOP). OOP is used as a tool in MedView to build applications. It is used since it, in our opinion, is the best existing paradigm for developing modern GUI based applications. The OOP tools used are interfaced with our own frameworks for integrating declarative program components. Thus, we can use OOP programming and declarative programming together and use each paradigm for the task where it's best suited.

- Artificial Intelligence. Knowledge based systems such as MedView is an important part of AI. In particular, case-based reasoning techniques have been studied [32]. Adding more AI-techniques is an area for future research.

- Human-Computer Interaction. The systems developed must interact well with clinicians, students, and researchers in oral medicine. Also, easy to handle administrative tools are needed. This makes studies in HCI an integral part of the project.

Note that both a sound theoretical basis and implementation of knowledge structures, and real working software solutions to be applied in daily work are equally important. The development of working high-quality software is necessary to influence the examination process so that knowledge can be collected and analyzed. We are also interested in using information technology to improve the healthcare process. A well-founded theoretical model is necessary, or the mentioned applications and the data gathered cannot be explained and analyzed in a meaningful manner.

3.1 Theoretical Model

The basic model of clinical information used in MedView is act-oriented. We think of explicit clinical information as resulting from acts of *defining* medical terms in various situations. A clinical diagnosis, an examination record and so on, can all be seen as definitions of collections of specific clinical medical terms [85].

The formalization of definitions as data structures that is used here is based on the idea that a definition generates a local logic, a reasoning model restricted to specific terms. These local logics are then the basis for reasoning using given formal clinical terms.

As a data structure, a definition D can simply be thought of as a collection of equations

$$D \left\{ \begin{array}{rcl} a_0 & = & A_0 \\ a_1 & = & A_1 \\ & \vdots & \\ a_n & = & A_n \end{array} \right. \qquad n \geq 0,$$

where terms, $a_0, \ldots, a_n$, are defined in terms of conditions, $A_0, \ldots, A_n$. The definiens of a term a, $D(a)$, is then the collection of conditions A that define a in D. The local logic of D consists of a relation $A_1, \ldots, A_n \vdash B$, that is, B follows from $A_1, \ldots, A_n$ according to D, where the two constituting rules are

- $A_1, \ldots, a, \ldots, A_n \vdash B$ if $A_1, \ldots, A, \ldots, A_n \vdash B$ for all A defining a in D,
- $A_1, \ldots, A_n \vdash a$ if $A_1, \ldots, A_n \vdash A$ for some A defining a in D.

The logic of D consists of these two rules together with ordinary rules of reasoning for given complex defining conditions built up from atomic terms. A full description is given in [46].

The model we use can be summarized as follows:

- formal clinical data are seen as definitions of clinical terms,
- reasoning is always local to given definitions, there is no single global logic for formal clinical reasoning.

As a concrete example of a definition we show a small part of an examination record:

$$D \begin{cases} status = direct \\ direct = mucos \\ direct = palpation \\ mucos = mucos\_site \\ mucos = mucos\_col \\ mucos\_site = 112 \\ mucos\_col = white \\ mucos\_col = brown \\ palpation = palp\_site \\ palp\_site = 112 \end{cases}$$

In D, the term *status* is defined by the term *direct*, which in turn is defined in terms of *mucos* and *palpation*. Thus, $D(direct) = \{mucos, palpation\}$. All these terms are part of the general structure of an examination record, which is shared by all examinations. In contrast, the term *mucos_col* is defined by the observed values *white* and *brown*, specific for this particular examination record.

4 Current Status

MedView has been developed in an iterative process through close collaboration between experts in oral medicine and computer science, using a mixture of contextual design [11], user oriented design, and logical analysis of the problem and required knowledge. Essentially, the analysis and design of the system can be divided into two sub-problems: knowledge representation and development of applications for gathering and exploring clinical data. Knowledge representation issues are discussed above. In this section we describe the status of the implemented system.

The system is currently in daily use in eight examination rooms at four different clinics. The examination rooms are equipped with a PC on a custom-built table, shown in Figure 4, and a digital video camera. The collected data are stored on a server.

A basic assumption underlying the design of MedView is that of separating the activities of entering information and viewing, or otherwise using, the entered information. The rationale behind this is that the cognitive tasks involved are very different. Thus, specialized applications, each described below, have been developed for each task.

Figure 4: Clinician working with MedView. The computer is placed on a custom-built table aimed at minimizing the interference with the communication between patient and clinician.

Information is collected in a critical situation, namely during examinations. For each examination, values for many different attributes describing an examination must be given. A good deal of effort was therefore put in early in the project to build efficient tools for knowledge gathering consistent with the underlying theoretical model.

4.1 Knowledge Base Contents

Currently, (March 2000) the knowledge base built in the MedView project contains approximately 1500 examination records covering more than 700 cases. The average growth rate is 20 new patients and 30 visits by previously examined patients a week. The main knowledge base is located at the clinic of oral medicine in Göteborg. The various clinics within SOMNET have local knowledge bases containing the examinations made at each clinic. The contents of these local knowledge bases are added regularly to the knowledge base in Göteborg so that the entire amount of data collected can be accessed through one common knowledge base. The clinics within SOMNET will have full remote access to this central knowledge base.

The contents of the knowledge base is mainly used in two ways. First, in the examination room to display the history of the patient under examination. Second, to perform analysis, learn from, and search for patterns in the knowledge base. The second task is typically performed on the clinician's desktop computer. So far, the most used analytical tool is The Cube discussed in Section 4.7. However, building more viewers for exploration, search, knowledge extraction, education, and so on, is an important area for current and future research. We present a number of more specific suggestions in Section 5.

4.2 Applications Overview

We briefly describe the applications currently used within MedView. Some more being under development but not yet taken into use within SOMNET are discussed in [110].

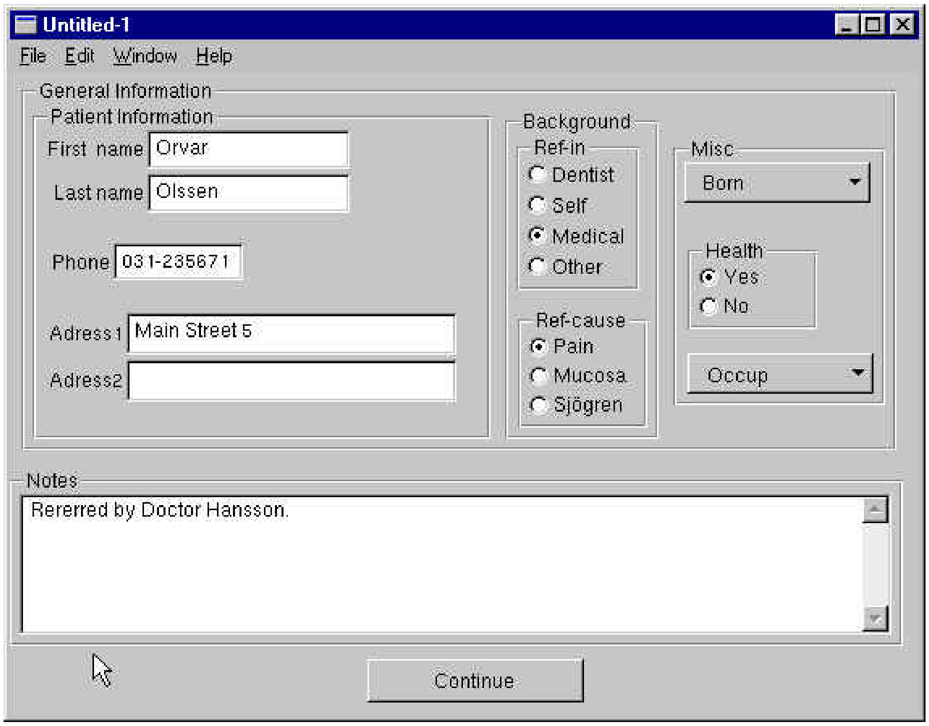
4.2.1 Background

Today, MedView runs on a combination of machines running Windows95/98/NT and some machines running OpenStep/Mach 4.2 and NextStep 3.3. Originally the system was developed using NextStep and GCLAII.

The operating system NextStep, used for the first versions of MedView, was chosen for its advanced GUI, networking capabilities, and object-oriented application development environment. The name NextStep usually denotes both the operating system and the object-oriented application development environment used to build applications for it. The application development environment later evolved to OpenStep available for several platforms including Windows NT.

GCLAII [4, 5, 62, 63] is a definitional programming language developed at the Swedish Institute of Computer Science, (SICS). Due to the similarity with the definitional model used to represent knowledge it was the natural choice for implementing the knowledge base and reasoning part of MedView. However, it was soon discovered that the performance of GCLAII was not sufficient for use in MedView. This led to the development of a simple object-oriented definitional machinery called DefinitionG.

DefinitionG implements the most important features of the definitional knowledge base model and can be subclassed if needed to add more. While DefinitionG lacks both purity, as a definitional representation of the knowledge base, and features, it has nevertheless been crucial for the development of *working* software solutions within MedView. Currently, we are in the process of replacing DefinitionG with a more fully-fledged object-oriented framework for definitional computing called Gisela [109]. One way to view Gisela is as a successor to GCLAII. Gisela has from the start been designed to better fit the demands on a definitional machinery to be applied within MedView.



General Information
Patient Information
First name Orvar
Last name Olssen
Phone 031-235671
Adress1 Main Street 5
Adress2

Background
Ref-in
○ Dentist
○ Self
● Medical
○ Other
Ref-cause
● Pain
○ Mucosa
○ Sjögren

Misc
Born
Health
● Yes
○ No
Occup

Notes
Rererred by Doctor Hansson.

Figure 5: A traditional form. Data is entered using a number of standardized widgets. Clicking the button labeled Continue will show the next screen.

Originally, MedView had only a small number of users, all active within in SOMNET. With the desire to increase the userbase, it became obvious that it was not possible to require the use of the rare operating system NextStep. Therefore, during 1998 and 1999 a transition was made to the Windows family of operating systems. The development platform NextStep was replaced by OpenStep and in some cases Java to enable the transition.

4.3 MedRecords

MedRecords (MR) is the *input* application used by clinicians to enter detailed formalized examination data during patient visits. Although the version of MR used today has some special features for use within MedView it is best seen as a general-purpose program useful for entering many different kinds of data.

The most common way today to design an application where data needs to be entered is to use forms [39]. The forms are typically built from objects such as text-fields, pull-down lists, and check-boxes. An example of such a form is shown in Figure 5. In MR we have developed a technique for entering data where the forms are replaced by a specialized text-editor, coupled with hypertext links for navigation, and easily scrollable text lists containing possible values.

The design goal behind MR was to create an unobtrusive, easy-to-use, space efficient, and scalable method for entering data, where the forms used can be created by users without requiring any programming knowledge. Here we describe the interaction technique and our experiences from it for entering information about a large number of patient examinations over a period of about two years.

4.3.1 Analysis

MR was conceived as a solution for entering data based on an analysis of the constraints given. The analysis describes a conceptual model of the act of entering data. It also lists external requirements describing the environment in which data is to be entered.

The conceptual model of the knowledge base used in MedView is that of a collection of definitions, where each definition describes one medical examination. Each such definition can be pictured as a collection of equations as described in Section 3.1. Thus, in the MedView setting, entering data is the act of *creating* a definition. Therefore, our goal was to support the act of *defining*, in a precise manner, a particular medical examination. MR aims to mirror this view of the knowledge base, while keeping a non-technical interface to the user.

The environment for which MR was developed introduces a number of requirements that had to be satisfied. Some of the more important ones are:

- Data is entered by the clinician him/herself while a patient is being examined.

- Each record in the knowledge base can have a large number of different attributes and each attribute can have a very large number of possible values.

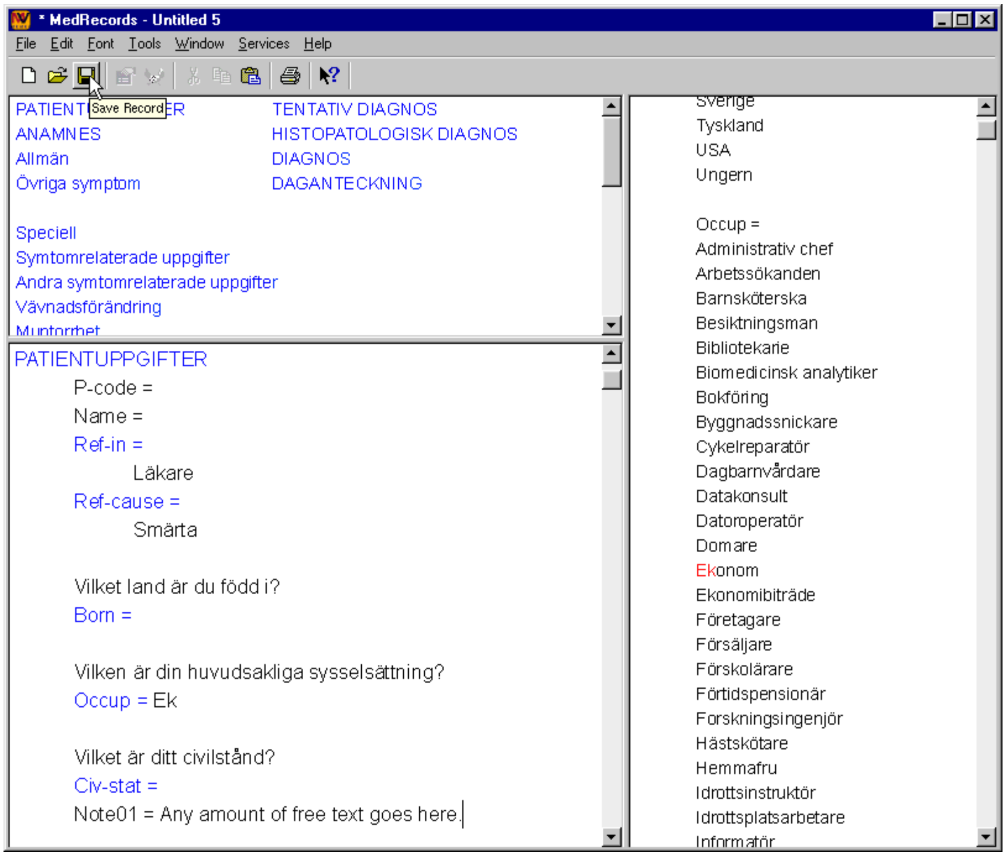
- Values for attributes are most often taken from formalized lists of valid values. However, free text and digitized images may also be included.

- When a new value is encountered it must be easy to add it to the list of valid values.

- The protocols or forms used are developed by the expert users themselves without requiring any programming knowledge.

- The layout of forms should be configurable by each user.

Since the act of entering data is separated from viewing data, MR was designed for entering data only. It is not intended for viewing examination records.



PATIENT ER
ANAMNES
Allmän
Övriga symptom

Speciell
Symtomrelaterade uppgifter
Andra symtomrelaterade uppgifter
Vävnadsförändring
Muntorrhet

TENTATIV DIAGNOS
HISTOPATOLOGISK DIAGNOS
DIAGNOS
DAGANTECKNING

PATIENTUPPGIFTER
P-code =
Name =
Ref-in =
Läkare
Ref-cause =
Smärta

Vilket land är du född i?
Born =

Vilken är din huvudsakliga sysselsättning?
Occup = Ek

Vilket är ditt civilstånd?
Civ-stat =
Note01 = Any amount of free text goes here

Sverige
Tyskland
USA
Ungern

Occup =
Administrativ chef
Arbetssökanden
Barnsköterska
Besiktningsman
Bibliotekarie
Biomedicinsk analytiker
Bokföring
Byggnadssnickare
Cykelreparatör
Dagbarnvårdare
Datakonsult
Datoroperatör
Domare
Ekonom
Ekonomibiträde
Företagare
Försäljare
Förskolärare
Förtidspensionär
Forskningsingenjör
Hästskötare
Hemmafru
Idrottsinstruktör
Idrottsplatsarbetare
Informatör

Figure 6: MedRecords Form. At the top left is the navigation area which is used to navigate into the appropriate part of the input view at the bottom. To the right is a list of values linked to the attributes in the input view.

4.3.2 Design

Following the analysis of entering information as the act of defining records, MR was designed to display partial definitions and to provide efficient techniques for completing them.

The user interface of MR consists of one window divided into three views as shown in Figure 6. At the top left is a navigation area, below it is the input view where data is entered, and to the right is a list of commonly used values. Apart from auxiliary windows for editing preferences and the like, all work is performed within this single window. The contents of each view is taken from template files in Rich Text Format (RTF). Thus, the contents of a form can be replaced completely without any modification to the program. Further, the layout of each view can be designed using all common features found in word-processors with respect to font, colors, tabbing, etc. In addition, each user may customize the layout.

The interaction paradigm is based on a small number of basic operations found in many applications. The input view at the bottom left works as a specialized text-editor. It displays an incomplete definition or a "form" which is edited when data is entered. The form contains arbitrary lead texts and a number of knowledge base attributes, each followed by an equals sign. The equals sign marks the beginning of an implicit input textfield where the value of the attribute is entered, see Figure 6. Only these implicit input fields in the input view may be edited by users. All other parts of the displayed text are fixed.

Navigating within the input view can be done by tabbing between the different attributes, scrolling, using standard navigation keys, or by following the links in the navigation view at the top. The navigation view typically displays links into all the main sections of the input view. Clicking a link has the expected effect: focus is moved to the corresponding area of the input view.

Values may be entered in several ways. First, by simply typing the value. As a value is being typed, the first matching value in the list to the right is highlighted. Pressing the completion key or clicking on the highlighted value inserts it into the form. Second, by following a link from an attribute to its value list to the right and clicking on the desired value. The value is then inserted into the input view. Additional documents related to a record, such as images, may be included by dragging and dropping them on the input view.

Thus, MR is based on a simple flow of actions from navigation view, to input view, to value list view and back. All the actions involved are simple and well known to most users. Data may be entered in any order and all attributes in the input view are instantly accessible via the navigation view.

4.3.3 Discussion

MR has evolved through a continuous interaction between users and developers. Starting from when the first prototype was stable enough, it has been in use to enter data during clinical examinations. First, a prototype was built to test the concept. Based on the success of the prototype, a more complete application was built together with an editor that is used to create new forms. Today, a third version, which runs on Windows and Mac OS X systems, is in use.

MR has been used as the input application to enter data for all the records in the MedView knowledge base. All this data has been entered by the clinician performing the examination while talking to and examining the patient. The interaction paradigm based on well-known components such as keyboard, mouse, hypertext links, drag & drop, and ordinary text editing works very well.

Current forms consist of about 100 attributes and a large number of values, e.g., lists of different drugs and diseases. The navigation tools are sufficient although some fine-tuning of the systems scrolling behavior is called for.

Compared to traditional form-based interfaces we believe that MR scales very well. Having 100 different readily available attributes in one screen poses no problem. Displaying traditional forms for the same amount of attributes would require navigating between many different screens, typically in some fixed order.

A recent form covering more than 1000 different attributes in the area of oral medicine has been tested. The value lists associated with this form contains more than 12 000 distinct values. The initial experiments with this form indicate that MR works as well as with smaller forms. Another aspect is the simplicity with which new forms may be created. New forms are created using InterfaceMaker, see Section 4.5, an editor comparable to HTML-editors. No programming is required.

The success of MR shows that focusing on simplicity and long-time usefulness instead of elaborate GUIs can be a good thing indeed. Testing the concepts of MR on a large number of different kinds of forms remains an area for future work.

4.4 MedSummary

The first, and so far most used, knowledge base *viewer* is MedSummary (MS). MS is used in conjunction with MR in the examination room, but also to display detailed information during analysis of the material in the knowledge base.

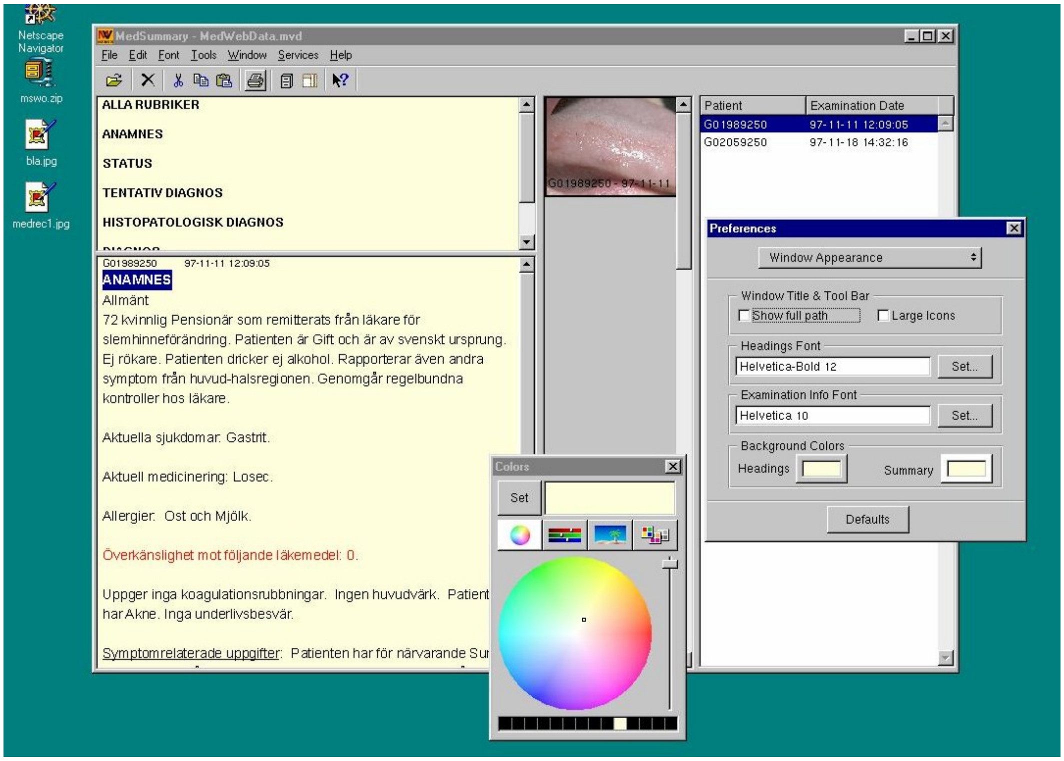
The view of the knowledge base presented by MS is that of a textual summary of one or more examination records together with any associated images as shown in Figure 7. The purpose is to display in a format suitable for viewing the information collected in MR. While it is possible to view data using MR it is not a recommendable way to learn what an examination record is all about. Instead of showing the form or screen used to *collect* data, we use Natural Language Generation (NLG) to generate from the collected data a comprehensible summary of all or parts of the examination(s).

4.4.1 Working with MedRecords and MedSummary

As mentioned earlier, we make a clear distinction between input applications and viewers. While this may sound obvious, the electronic medical record systems we have encountered use the same display to input information and to view it. Consequently, the displays used are optimized neither for entering nor for viewing information.

When working with MR and MS these two activities are separated. New examination data is entered with MR, the contents of existing examination data is viewed using MS. The main window of MS shown in Figure 7 contains a listing of selected examinations to the right, thumbnails of images in the middle and the generated medical record text to the left. Clicking on a thumbnail image will show it full-sized in a separate window. Different texts can be generated by selecting between the headings shown at the top left.

When a previously examined patient comes back for a follow-up, the user



ALLA RUBRIKER

ANAMNES

STATUS

TENTATIV DIAGNOS

HISTOPATOLOGISK DIAGNOS

G01989250 97-11-11 12:09:05

ANAMNES

Allmänt

72 kvinnlig Pensionär som remitterats från läkare för slemhinneförändring. Patienten är Gift och är av svenskt ursprung. Ej rökare. Patienten dricker ej alkohol. Rapporterar även andra symptom från huvud-halsregionen. Genomgår regelbundna kontroller hos läkare.

Aktuella sjukdomar: Gastrit.

Aktuell medicinering: Losec.

Allergier: Ost och Mjölk.

Överkänslighet mot följande läkemedel: 0.

Uppger inga koagulationsrubbningar. Ingen huvudvärk. Patient har Akne. Inga underlivsbesvär.

Symptomrelaterade uppgifter: Patienten har för närvarande Su

Patient	Examination Date
G01989250	97-11-11 12:09:05
G02059250	97-11-18 14:32:16

Preferences

Window Appearance

Window Title & Tool Bar

☐ Show full path ☐ Large Icons

Headings Font: Helvetica-Bold 12 [Set...]

Examination Info Font: Helvetica 10 [Set...]

Background Colors: Headings Summary

[Defaults]

Figure 7: MedSummary: main and preferences windows. Different texts may be generated by selecting from the headings at the top left. Clicking on a minimized image will show it full-sized in a separate window.

can create a suitable background text. To do this, the user selects the desired previous examinations, and then clicks on a heading to show a summary together with existing images. Once a summary has been created it can be edited as any ordinary text-document if necessary. The text can then be printed and used for things like providing a detailed medical history if the patient is sent to another clinician.

The text-generation used is very flexible and can be adjusted easily both with respect to contents and formatting. Thus, different users may have different summary texts based on the same database if desired. Apart from values of attributes that allow free text, generating summaries in different languages poses no special problems.

4.4.2 Natural Language Generation

Natural Language Generation is the activity of generating text from some kind of sources. A good overview of the area can be found in [23, 24]. In principle, there are two approaches to generation, the *deep* and *shallow* approach. A deep system builds on a deep understanding of linguistics whereas shallow systems use simpler methods to generate text. The advantage of deep text-generators is that they are more domain independent and thus can be applied to various areas with relative ease. However, building a deep system requires a lot of knowledge and resources. Shallow systems are typically specialized for a particular task and need not be more complicated than the task demands. On the other hand, they are less reusable for another task. Some deep systems are described in [9, 42, 80, 92], examples of shallow systems can be found in [17, 40, 91]. Discussions of the two approaches can be found in [17, 23, 89, 90].

Typically, a NLG system is divided into three phases [88]

- Content Determination
- Sentence Planning
- Surface Realization

performed in sequence. Thus, the system first decides what the text should contain, then plan the general structure at sentence level, and finally, realize the desired structure into text. Other approaches are used as well. For instance, [78, 103] propose an integrated constraint-based method that performs all three activities at the same time. The RAGS project [18] is an attempt to develop a reference architecture for NLG systems.

4.4.3 Text Generation In MedSummary

The main focus during the development of MS and the text-generation used has been to create a very flexible system where users can experiment with different texts without having any linguistic expertise. Thus, from a NLG point of view the system is a basic shallow system. Close to a simple mail-merge system, it can be classified as a slot-and-filler, or canned-text with knowledge base references system [17].

Although the text-generator used is very simple it can be clarifying to describe it using the three standard phases mentioned above. Examination summaries have a structure based on the formalization of examinations used in MedView. An examination record forms a tree structure with top level nodes representing the different main tasks from which information is gathered at examinations. Text can be generated for all tasks performed at an examination, for a particular task, or any desired combination of tasks. It is also possible to generate a text covering several examinations. In terms of the phases above:

- Content planning. Depending on the user's choice it is decided what parts of a text template should be used in the resulting text and for which examinations summary text should be created.
- Sentence planning. Depending on which attributes of an examination record have values, it is decided which sentences of the selected template should be included in the text. Sentences for which values are missing are omitted.

- Surface Realization. Depending on the values for attributes in the database particular text-fragments are selected and used to fill slots in sentence templates.

The text-generator takes as input a *template* describing the texts to generate. This template consists of a number of files providing (i) an RTF template text with slots to be filled in depending on the values for attributes in an examination record, (ii) a file describing the connections between slots in the template text and attributes in the knowledge base, (iii) a file that classifies the attributes of the knowledge base into a number of groups, (iv) a file that defines the text-fragments to use as slot-fillers for attribute values. The last file does not simply list value-text pairs, but allows some slightly more complex substitution patterns as well.

The text generator parses the template files into a number of definition objects. Most notably, each attribute gets its own definition object describing text fragments for all possible values of the attribute. Formatting information is kept from the RTF template. To modify the look of generated texts each user may freely change all formatting attributes, font, color, aligning etc. without affecting the actual contents of summary texts.

With the syntax currently used, part of a template could be:

> **§DISEASE HISTORY§**
> Age year old Sex $Occup$ who is referred by $RefIn$ because of a $RefCause$. The patient is $CivStat$ and comes originally from $Born$. $Checkup$.

Now, if part of the definition of an examination is

```
Occup = Lärare.
Ref-in = Tandläkare.
Ref-cause = Slemhinneförändring.
Civ-stat = Gift.
Born = Sverige.
Checkup = Ja.
```

and the value-text maps include the following:

```
Occup:
Lärare = teacher.

Ref-in:
Tandläkare = a general dentist.

Ref-cause:
Slemhinneförändring = mucosal lesion.
```

```
Civ-stat:
Gift = married.

Born:
Sverige = Sweden.

Checkup:
Ja = Attends medical check-ups regularly.
```

the generated text for DISEASE HISTORY becomes:

> **DISEASE HISTORY**
> 58 year old female teacher, who is referred by a general dentist because of a mucosal lesion. The patient is married and comes originally from Sweden. Attends medical check-ups regularly.

Since most values in the MedView knowledge base are in Swedish they have to be given a translation to generate English text. However, if values had been given a neutral language independent coding instead, it would still have been necessary to translate from these codes into English text.

4.4.4 Implementation

MedSummary is written in Objective-C. The text templates used are parsed into a number of definition objects, which were developed as an early part of the Gisela project [109]. There are two versions of the generator, one that uses an RTF template and produces output in RTF format and one that uses an HTML template and produces HTML output. The HTML generator makes it very simple to produce summary texts for web publication, see Section 4.9. The performance of the text-generator is quite sufficient, the desired summary text is displayed immediately.

4.4.5 Discussion

In [90], the term *automatic text generation* (ATG) is used to refer to any computer program that automatically produces texts from some input data. ATG systems are then divided into NLG systems and *template* systems. A template system is defined as a system that simply manipulates character strings using little, if any, linguistic knowledge. From this point of view, the current MS application should be seen as a template system. However, we find it useful to discuss the system in the light of NLG and we are moving towards including more NLG techniques into the system.

As stated earlier, ease of use by non-experts has been deemed more important than producing optimal text quality or using linguistically motivated methods.

The template files forming the basis for generation should best be seen as the user interface to the system for content management. By this we mean that it should not be regarded as *the representation* of the framework used but rather as an *interface* to enter information into the system from which a suitable internal representation can be built. This internal representation could be something with a deeper basis in NLG than the current system. If a more sophisticated system should be introduced it must not be at the expense of the possibility for users to design their own summary texts.

It is interesting to note that several choices made in the development of MS are essentially orthogonal to the approaches suggested for NLG systems. We discuss some issues below. We also mention where NLG techniques would be appropriate in MedSummary and related systems.

Creating a Corpus The corpus-based approach [23] advocates that the first step in the construction of a NLG system is to build a corpus of example texts. This corpus, which should cover the full range of texts the system will produce, is then analyzed for linguistic and information content. Our approach has instead been to build a system where the users, through experiments with given tools, can decide the texts themselves. Actually, an initial prototype for MS was built using something of a corpus-based approach. A number of templates were extracted in collaboration with a domain-expert and realized into an application. However, it was soon discovered that a system that required the assistance of a programmer to alter the contents of summary texts was not appropriate. Therefore, the current system where texts can be continuously refined was developed. By now, after a couple of years of use, it would probably be possible to take a number of generated texts from the system, check them for errors, and use them as a corpus.

Flexibility It is often argued that a major advantage of sophisticated NLG over template systems is that deeper systems are more flexible and easier to maintain. Exactly why this is the case is not always clear. It is interesting that in the development of MS we have selected to use a simple template approach to achieve great flexibility. Of course, this is related to the fact that it is necessary that the *end-users* themselves can modify what the text generated from examination data should be. The text-files used as templates for text-generation are simple enough to be modified by end-users. To expect that they would be able to easily control the workings of a sophisticated NLG system is not realistic.

Text quality Systems building on linguistic knowledge are generally able to produce text of higher quality compared to template based systems. Whether this higher quality is needed depends on what the texts should be used for and on the complexity of the generated texts.

The structure of medical record text is typically very static and uses a rather formal language. Furthermore, there is no need to produce text with great variation. On the contrary, too much variation might be disturbing since clinicians reading the texts expect them to follow certain patterns. Most of the texts generated by MS are read once in the examination room and then discarded. It is also more important that the summary is displayed immediately than that text quality is optimal. This indicates that for the MS application domain a template-based approach is sufficient.

Since the templates used for generation have been refined repeatedly for several years, the quality of the generated texts is in most cases sufficient for their task. In case further refinements are needed, the text may be edited by the user.

Recently, we have put together a web application where summaries, together with images, for selected patients can be viewed. Text can be generated in either English or Swedish. Adding more languages is a simple matter of modifying the text templates. However, in this context it is not possible for users to create new templates. It could therefore be appropriate to use a more complex system since it will be maintained by experts and not by end-users.

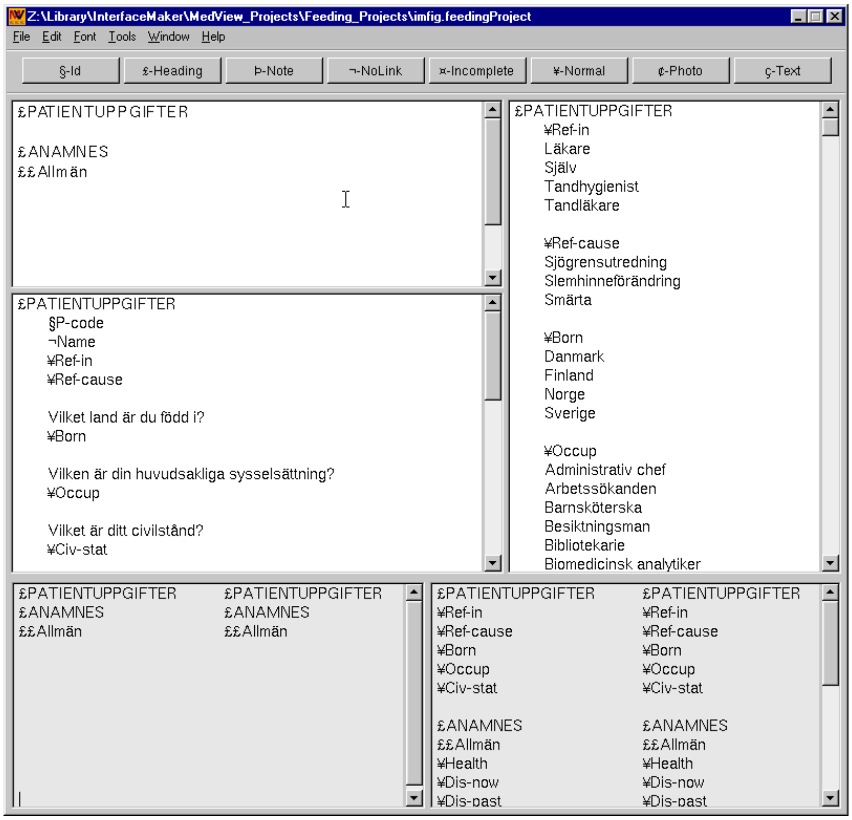
Hybrid Approaches While a deep NLG system does not appear to be needed in MS, using a *hybrid system* would be quite useful. Several hybrid systems have been developed which combine templates with deeper NLG techniques [16, 57, 93].

One obvious technique being a candidate for inclusion in a future version of MS is *aggregation*. Aggregation is used to combine related phrases and sentences together in a linguistically correct manner. Some basic aggregation can be performed in an experimental text-generator we have implemented using the Gisela framework [110]. In MS sentences are either included or omitted depending on whether all attributes needed to generate the sentence have values or not. In the Gisela-based generator the choice is made at a higher level; this, among other things, allows the combination of two sentences into one in certain cases.

Finally, we note that *multi-modality* is of increasing importance in document generation. We need to be able to include diagrams, tables, and other graphics into patient summaries. The images displayed along with the generated text in MS are as important for the clinician as the generated text. Support for tables is present in the Gisela-based generator mentioned above. Creating fully multi-modal documents is an interesting challenge for the future.

4.5 InterfaceMaker

Both MR and MS are developed to allow that the contents of the forms or protocols used and the text generated is completely replaceable without changing the application itself. To aid in creating new forms there is a tool called InterfaceMaker (IM). IM is similar to HTML-editors. The user writes the various texts of the new form and adds tags to create links, see Figure 8. IM also supports the creation of text templates for MS.



```
£PATIENTUPPGIFTER

£ANAMNES
££Allmän
```

```
£PATIENTUPPGIFTER
      §P-code
      ¬Name
      ¥Ref-in
      ¥Ref-cause

      Vilket land är du född i?
      ¥Born

      Vilken är din huvudsakliga sysselsättning?
      ¥Occup

      Vilket är ditt civilstånd?
      ¥Civ-stat
```

```
£PATIENTUPPGIFTER
      ¥Ref-in
      Läkare
      Själv
      Tandhygienist
      Tandläkare

      ¥Ref-cause
      Sjögrensutredning
      Slemhinneförändring
      Smärta

      ¥Born
      Danmark
      Finland
      Norge
      Sverige

      ¥Occup
      Administrativ chef
      Arbetssökanden
      Barnsköterska
      Besiktningsman
      Bibliotekarie
      Biomedicinsk analytiker
```

```
£PATIENTUPPGIFTER     £PATIENTUPPGIFTER
£ANAMNES              £ANAMNES
££Allmän              ££Allmän
```

```
£PATIENTUPPGIFTER     £PATIENTUPPGIFTER
¥Ref-in               ¥Ref-in
¥Ref-cause            ¥Ref-cause
¥Born                 ¥Born
¥Occup                ¥Occup
¥Civ-stat             ¥Civ-stat

£ANAMNES              £ANAMNES
££Allmän              ££Allmän
¥Health               ¥Health
¥Dis-now              ¥Dis-now
¥Dis-past             ¥Dis-past
```

Figure 8: InterfaceMaker: main window.

IM is more of an administrator tool than a user tool. The general methodology developed within MedView suggests that new forms or protocols are created only when the contents have been formalized and harmonized. Thus, creating new forms should be done by authorized persons only, at the time when a new protocol is to be adopted within the user community.

4.6 Basic Visualizations

The very first visualization of the knowledge base developed was an application that shows ordinary 2D views of data in the knowledge base. The user can view data in a scatter-plot as shown in Figure 9 or as a bar chart. Values for any number of attributes can be displayed simultaneously to let the user look for interesting groups of patients. In Figure 9 the upper left corner shows a cluster of patients born in Sweden who have been referred by their general dentists because of a mucosal lesion.

The application also allows the user to view only a restricted part of the knowledge base by first making a selection based on any combination of the attributes in the knowledge base. Furthermore, details for any particular dot in

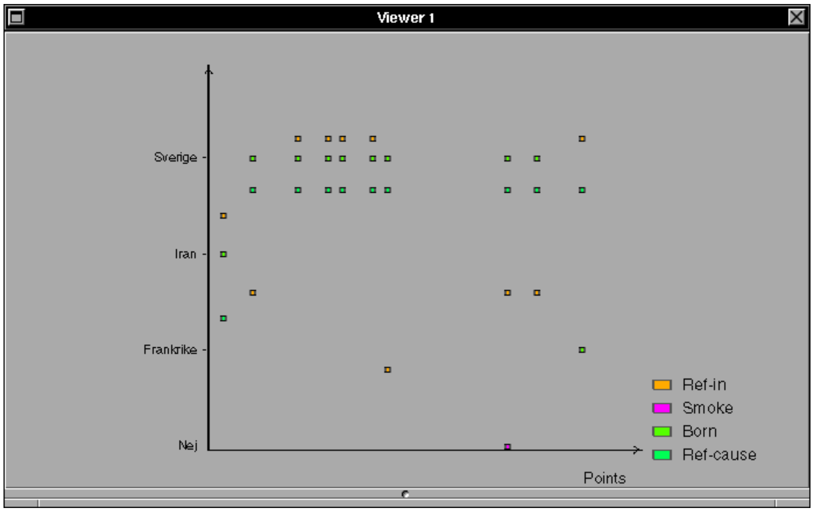


Figure 9: A scatter-plot showing values in the knowledge base.

the display can be shown. So far, the 3D visualizations discussed below have been preferred by the clinicians. We will investigate the reasons for this and ways to build better 2D viewers in the future.

4.7 The Cube

The most used visualization of the knowledge base so far is called The Cube. The Cube has been developed to enhance the clinician's ability to intelligibly analyze the patient material within the knowledge base and to allow for pattern recognition and statistical analysis. The Cube is based on the idea of using parallel coordinates [52] as a solution to multidimensional data analysis [6, 21]. The visualization of parallel coordinates is discussed in [111, 117, 118].

The starting point was the formalization of the notion of a clinical examination as a definition. A formal examination is seen as a set of definitions of specific examination terms. An excerpt of such a definition was shown in Section 3.1.

Clearly, for a given collection of examinations, such a term can be viewed using a simple scatter plot with the x-axis as a sort of time line, e.g. ordered by examination date, and with the values of the term on the y-axis. Thus, if we want an overview of the total set of terms it is natural to think in terms of multiple parallel diagrams (this is similar to the scatter matrix of [21]). This view was then generalized into dynamic 3D parallel diagrams with support for direct manipulation, a 3D cube. The idea is similar to the concept of 3D parallel coordinates, e.g., the casement displays used in [111]. The reason for using 3D was that the notion of 3D parallel diagrams was conceptually very natural from a clinical point of view; it seemed to be a natural model of the raw material of clinical experience. The idea of investigating such a 3D object to learn from

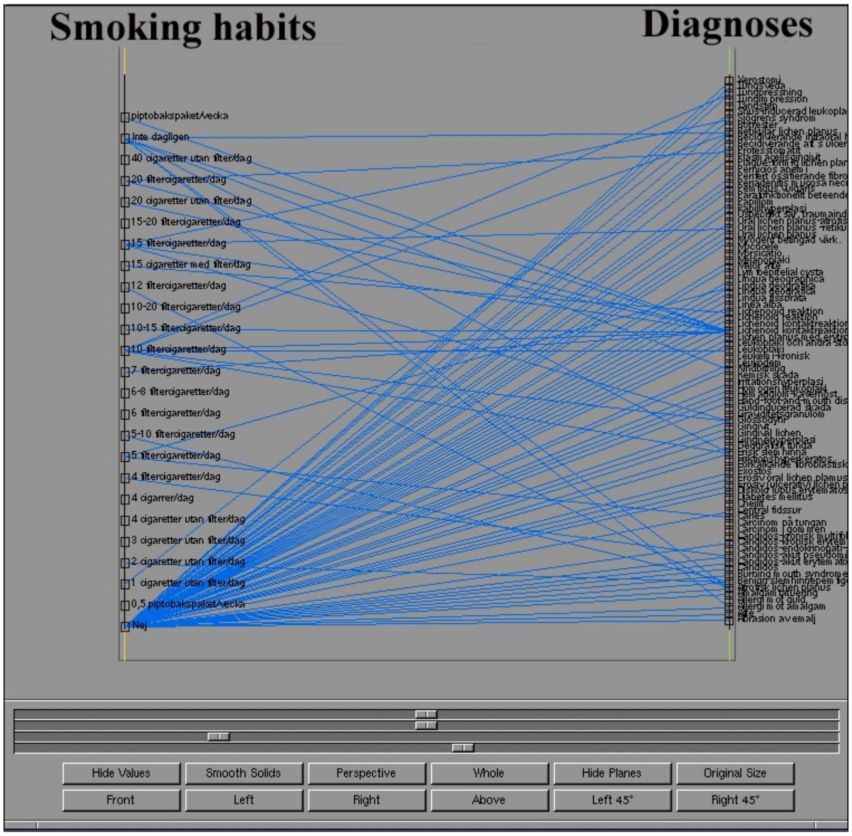


Figure 10: All diagnoses related to the number of cigarettes/day.

clinical data was very appealing.

4.7.1 Defining The Cube

An examination record consists of a number of examination terms (attributes). The user must first decide how many 3D parallel diagrams should be used, and which attribute should be displayed in which diagram.

The total set of attributes in the knowledge base is displayed in a panel and the user simply selects the desired attributes from this list. An attribute can also be marked to be used as the unit on the z-axis.

4.7.2 Viewing The Cube

The Cube consists of a number of planes, one for each attribute that was selected when The Cube was defined. These planes are presented along the x-axis. The z-axis is typically used as a timeline, i.e., as an ordering on the examination identification attribute, but an arbitrary attribute can be used as the unit on this axis. The y-axis then lists the values for the attribute of the corresponding plane in some order, e.g., in alphabetical or numerical order. In any case, each examination is represented by a line connecting individual values in the different attribute planes. If an examination has more than one value for an attribute, the values are connected with a line in the plane. A picture of The Cube with two planes can be seen in Figure 10.

The user can observe The Cube from any desired direction, either by dragging or rotating The Cube with the mouse, or by using the controls at the bottom of the main window.

The appearance of The Cube can be changed in various ways: a strictly parallel projection can be used, the elements of The Cube, e.g., the lines, points, and planes, can temporarily be hidden, the user can change the colors of the elements, elements can be set to be transparent, etc.

If the user finds some lines, i.e., examinations, particularly interesting, these lines can be selected and then opened in a separate window for closer inspection. Similarly, if some lines are blocking the view of others, these lines can temporarily be removed. To get a summary of a number of examinations, the user can select the lines and then open the corresponding examination files in MS.

It is also possible to get statistics about a selected plane: for each value in the plane, the number of examinations with that value is displayed using an ordinary bar chart.

4.7.3 Grouping of Attributes

When data from thousands of examinations are displayed in The Cube, the display will be filled with lines and it may be difficult to recognize clinically meaningful patterns. To solve this problem two techniques can be applied: either showing only a subset of the knowledge base based on a selection made before defining The Cube, or grouping values into classes in a hierarchical manner. For example, a number of diseases can be grouped into viral diseases. Such classifications of attribute values reduce the complexity of the display and facilitate the detection of interesting patterns.

Groups can be created and stored in a library for future use. From a theoretical point of view, a group is simply a definition relating values to groups. Examples of existing simple groups are a division between smokers and non-smokers and between patients with oral lichen planus and patients which do not have oral lichen planus. Combining various groups gives new interesting patterns to explore.

4.7.4 An Example

The Cube is used for finding patterns and correlations. The typical question posed is "How does a certain set of attributes relate to each other for the entire patient material?" or simply viewing a single attribute. If the patient material is homogenous from the aspect of parameters chosen in the analysis, the lines will appear parallel to each other within The Cube. Heterogeneity and outliers

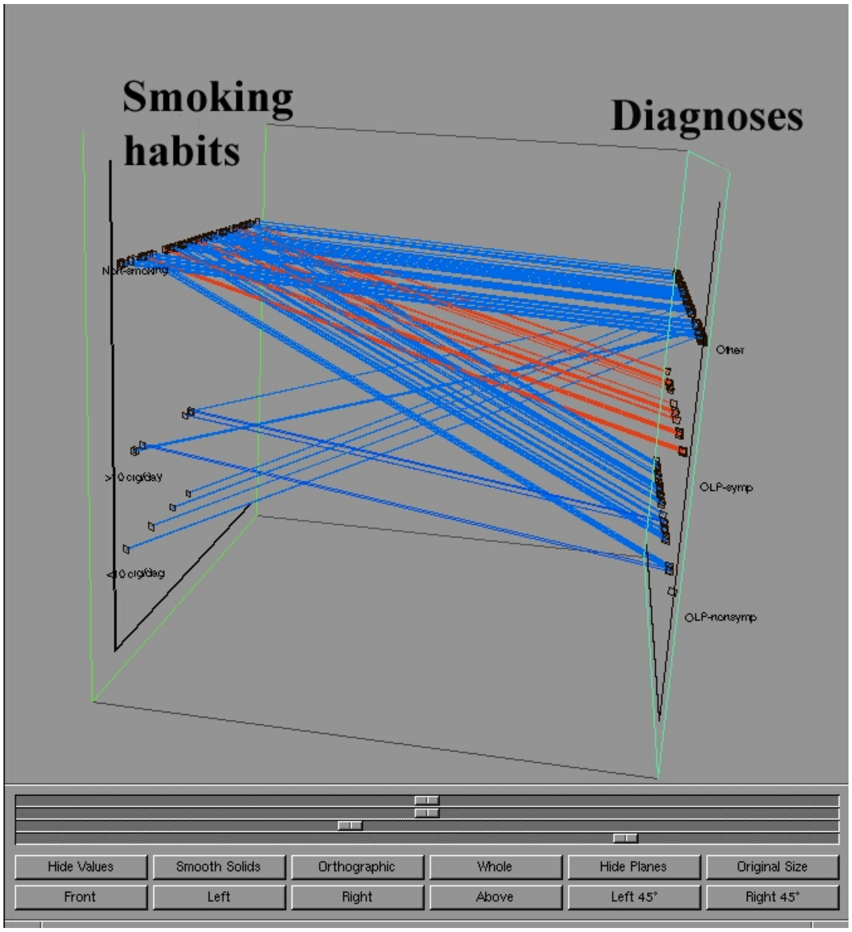


Figure 11: The picture in Figure 10 has been simplified by grouping values.

in the patient material for a certain parameter will, consequently, cause the lines to diverge from each other in the corresponding plane. By using the various selection possibilities, a step-wise procedure may be performed where hypotheses are continuously refined.

Oral lichen planus (OLP) is a disease with unknown etiology that effects the oral mucosa. In its most severe form, the disease presents with erosions and ulcerations, which interfere with, for example, eating of citrus fruits and spicy food. Some of the OLP lesions transform into a malignant disease of the oral mucosa.

In this example, The Cube was used to examine drug and smoking habits for symptomatic (ulcerated) and non-symptomatic (non-ulcerated) OLP, information that has not previously been reported. A cube with two planes was defined: on the first plane the smoking habits of the patients were presented and on the other plane the different diagnoses of the patients were displayed, see Figure 10. The display was then simplified by classification of smoking habits into three groups: non-smokers, patients smoking less than 10 cigarettes/day, and patients smoking more than 10 cigarettes/day, and by changing the color for the two different forms of OLP, see Figure 11.

It was revealed that patients with symptomatic OLP (OLP-symp) were non-smokers (100%) compared to patients affected by non-symptomatic OLP (OLP-nonsymp; 81% non-smokers; 11% more than 10 cigarettes/day; 8% less than 10 but more than 1 cigarette/day). The opposite was found for medication where only 47% of the OLP-nonsymp used drugs compared to 65% of the OLP-symp.

These findings raise thoughts about how different factors may influence the development of the two clinical forms of the disease. The reported observations have now to be statistically evaluated and further investigations by using The Cube have to be conducted to examine if patients with OLP-symp take other types of drugs than patients with OLP-nonsymp.

4.7.5 Discussion

A basic metaphor in MedView is that clinical experience can be viewed geometrically as a space of interconnected atomic points of knowledge. Using The Cube in clinical practice has shown that the tool works well conceptually, as an implementation of this idea.

It is interesting to note that, although 2D visualizations such as scatter-plots and bar charts are more obvious and may appear to be easier to understand, the clinicians prefer working with The Cube. One reason for this is that it is the better tool for viewing the collected clinical experience. It is more of a visualization of the space of interconnected points forming clinical experience.

The grouping, or aggregation, of values has been proven very useful to achieve better results with The Cube.

In the future we need to add more tools for direct manipulation of the, often very complex, displays. Examples of such tools are better and more powerful tools for selecting and, temporarily, discarding or hiding various elements of The Cube, and methods and algorithms for minimizing problems with disorientation and occlusion of elements. Parallel diagrams take practice for users to comprehend [101]. Therefore, the work on a methodology for clinical use of The Cube will be extended and carried further.

4.8 SimVis

Similarity assessments play an important role in most cognitive activities. For example, a clinician examining a patient wants to know if there are previous examination records that are similar to the current one, hoping that these might help him, or her, in the diagnosis of the new patient. However, before we can ask for “similar examination records” we must define what “similar” means. SimVis is a tool designed to allow and encourage clinicians to classify and cluster clinical data in different ways, i.e., a tool that enables them to interactively construct and try out new similarity measures.

Much effort have been spent on studying similarity measures within the medical domain, especially in the area of case-based reasoning (CBR) [41]. This includes work on using CBR-techniques in the retrieval of images from image databases [66] and knowledge mining [54]. In [2] clustering was used to find higher conceptual structures of medical data.

4.8.1 Definitional Similarity Measures

Since examination records are given as definitions, it follows that we must first study how to measure the similarity between these definitions in order to be able to classify clinical data.

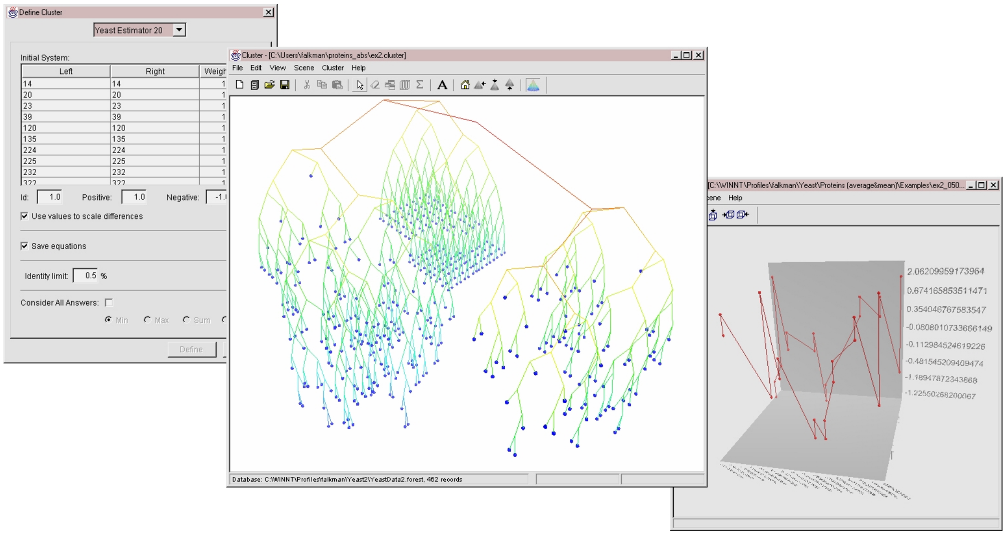
A similarity measure consists of a definition, E, a computation method, M, and a number of definitions, $D_1, \ldots, D_n$. One may think of E as a set of test points, a number of properties of $D_1, \ldots, D_n$, on which the similarity measure is based. The result of the application of M to E and $D_1, \ldots, D_n$, the similarity value, is a new definition, V, which describes both structural and computational similarities between $D_1, \ldots, D_n$. The computation of V is really only the first step in a more general estimation process. If an interpretation of V cannot immediately be found, it can be used as the starting point for further estimation. The result of this second step can, if necessary, be used as the starting point for a third step etc. The process usually ends when V equals a test-definition, S, indicating that the interpretation of V is clear to us. More on definitional similarity measures can be found in [32].

All parts of the similarity measure, including the data structures, the computation method and the estimation process, are given as definitions. Since all parts of the model are given as definitions, the user can, in principle, use the output of the model, i.e., the similarity value, as an input in any other part of the model.

Through experimentation with SimVis, a similarity measure can be found that, for instance, captures the characteristics of a certain diagnosis. This measure can then be used for finding records with this diagnosis, and, indeed, the similarity measure can be said to define this diagnosis.

4.8.2 The SimVis Tool

The general framework for computing similarity measures can be used as a basis for different information visualization models, where each model gives a different visual interpretation of the underlying similarity measures. What is required is a mapping from the results of the estimation process to the visual model. In SimVis a visualization model based on a three-dimensional hierarchical clustering is used. With SimVis, a clinician can interactively construct a similarity measure between examination records, apply the measure to a knowledge base of records, and visualize the resulting classification.



Initial System:

Left	Right	Weigh
14	14	1
20	20	1
23	23	1
39	39	1
120	120	1
135	135	1
224	224	1
225	225	1
232	232	1
322	322	1

Id: 1.0 Positive: 1.0 Negative: -1.

☑ Use values to scale differences

☑ Save equations

Identity limit: 0.5 %

Consider All Answers: ☐

● Min ○ Max ○ Sum

Database: C:\WINNT\Profiles\falkman\Yeast2\YeastData2.forest, 462 records

Figure 12: SimVis: the panel for constructing similarity measures (left), the visualization of clusters (middle), and the visualization of individual similarity values, i.e., cluster points (right).

SimVis consists of three modules, which are shown in Figure 12. The first module is used for constructing similarity measures and estimations. On the basis of the similarity values, a three-dimensional hierarchical clustering is created, visualized, and examined using the second module of SimVis. The similarity values themselves can be examined in detail using the third module.

Constructing Similarity Measures To construct a similarity measure, the user starts off by choosing one of many predefined computation methods. Each method has its own characteristics and parameters. These can, for instance, be the weights assigned to the different attributes of the examination records, if the length of the estimation process should be taken into account or not etc.

The user then defines which attributes of the examination records that should be taken into account. It is also possible to save the similarity measure for future use.

Visualization Model The user can apply the current similarity measure (or one saved from a previous session) to the knowledge base. From the resulting similarity matrix, a hierarchical cluster is constructed. A 3D visualization of the cluster can then be examined using the second module of SimVis (the middle window in Figure 12). To facilitate the exploration of data, clusters can be visualized in different ways: various parameters controlling the visualization could be modified, color codes could be used, the dynamics of the computation could be simulated by animating the construction of the clusters. If the user finds some sub-clusters particularly interesting, these can be selected and then opened in a separate window for closer inspection. Similarly, if some sub-clusters are blocking the view of others they can be temporarily removed.

The underlying similarity measures can be analyzed in detail as well using 3D parallel diagrams (the right window in Figure 12). The details of this visualization are described in Section 4.7.

4.8.3 Discussion

The theory for similarity measures underlying SimVis could be used as a basis for different classification and visualization models, not just hierarchical clustering. An alternative clustering could be the self-organizing map [58]. Apart from being tested in MedView, SimVis has also proved itself useful in the area of functional genomics, where it has been applied to problems connected to the analysis of expression data from proteome analysis of yeast [33].

Compared to the Cube and the simple 2D visualization tested, SimVis is a step towards a more active system, that is, a system performing tasks for the user, not just an exploration tool. The measures constructed could be used for exploration of data in the knowledge base.

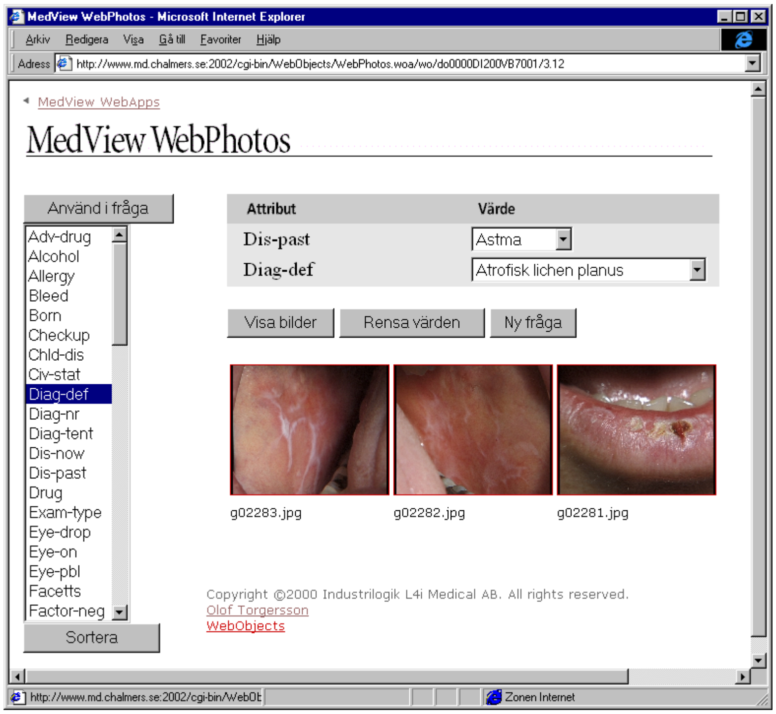
In the future SimVis will be extended into a general case-based reasoning (CBR) system [59, 116] that should provide assistance to clinicians within the field of oral medicine.

4.9 Web Applications

The real treasure of MedView is the knowledge base being built. An obvious step to give clinicians and researchers worldwide access to the data collected is to use the Internet. We are currently considering various web-applications that would allow exploration of the knowledge base using a Web-browser.

To setup a static website with information from the knowledge base would not be very interesting. Instead, we will build dynamic web-applications. Since essentially all the code written so far can be reused, the main problem will be design of an appropriate web-based user-interface.

So far, we have built two simple web applications: One that makes it possible to view patient summaries in the same way as in MedSummary, and one to search the knowledge base for images, see Figure 13. In the search for images each image is indexed by all the information collected at the examination when the image was taken. For example, a query might be “Find all lesions with Mucos-colr red and white and Mucos-txtur plaque”. This is possible since images taken at an examination are part of the total knowledge collected at the examination and the knowledge base can be searched for examinations matching any combination of attribute-value pairs.



http://www.md.chalmers.se:2002/cgi-bin/WebObjects/WebPhotos.woa/wo/do0000DI200VB7001/3.12

MedView WebPhotos

Attribut	Värde
Dis-past	Astma
Diag-def	Atrofisk lichen planus

Copyright ©2000 Industrilogik L4i Medical AB. All rights reserved.

Figure 13: Searching for images over the Internet.

5 Future Directions

We believe that MedView is a project that is worth continuing. The foundation for building a large knowledge base in the field of oral medicine has been laid down. Tools that have been put to the test at more than 1500 examinations have been developed and proven useful. Some analysis tools are in use, although in a smaller circle of users. However, to find areas that would be interesting to investigate further is not hard. We mention some of these in no particular order below.

5.1 Foundations

The theoretical model of MedView as a knowledge base containing definitions of examinations is not expected to change. However, some details may need further attention. Examples of such details are the way values are built-up and used. Today all values are atomic. This means that a value, say “2 times a day”, is represented by the atom `'2 x/day'`. There are several reasons for this: First, the basic definitional model used does assume that all values are atomic. Second, `'2 x/day'` does not require any special knowledge about atoms, terms etc. which makes it easier to understand for clinicians. Third, DefinitionG used in some applications to model definitions does only allow atomic values. Of course, an atom like this, which does possess an inherent structure, would be better represented as a compound term that makes it easy to access the components `2` and `x/day`, say `times_day(2)`.

We are currently working on using the Gisela framework as a replacement for DefinitionG as the tool for computing with definitions in MedView. Gisela is a much more flexible tool for programming with definitions than DefinitionG. As such, it is also less efficient and it remains to be seen how much more work is needed before it will reach the level where it can be used as *the* deductive engine and knowledge representation language of MedView.

With Gisela in place, it will be time to look further into knowledge representation and more advanced computations over basic data and knowledge structures built on top of it. Examples are defining new diagnoses based on data in the knowledge base, searching for patterns or similar cases, building a set of useful query filters, such as looking for patients with specific properties instead of examinations etc. Our belief is that Gisela will provide a definitional framework to do the things we need in a sufficiently clean and efficient manner.

5.2 Collaboration

There are several directions in which MedView can benefit from collaborations of various kinds. Some collaborative efforts, considered or ongoing, are mentioned here.

First, extending SOMNET is considered as a way to both increase the expert knowledge within the network, and speeding up knowledge gathering through a larger userbase. Since SOMNET is not about MedView only, this has to be a process where clinics are gradually assimilated. On a related note is letting general practitioners use MedView tools for evaluation and testing. The next extension would be to create an international network building a common knowledge base. Such a venture would of course demand serious efforts in the formalization and harmonization phases.

An international collaboration with Eastman Dental Institute in London is being initialized. There are strong relations between MedView and the work done at Eastman, both theoretically and practically. We hope that this will bring up interesting research opportunities.

Yet another important thing would be to have better cooperation with experts in information visualization, database mining, and pattern recognition. Finally, to further develop the NLG used in MedSummary collaboration with experts in computational linguistics is needed. We hope to be able to start work on this in the near future.

5.3 Applications

As we have mentioned several times before, we are only at the beginning of building tools for exploration of the information that has been collected over the years in MedView. Some ideas for future tools are:

- Database management. The definitional knowledge base model used in MedView needs computerized tools to monitor entered values, add value filters, corrections and so on. Some experiments in this direction are mentioned in [110].

- MedRecords with expert knowledge. Currently, MedRecords simply collects data. An interesting extension would be to add an intelligent agent that aids the user. The agent could provide suggestions for values, verify that entered values are consistent in some manner, or simply rearrange the value-lists so that the values deemed most likely by the agent occurs at the top.

- Combining MedRecords with graphical input devices. Although the basic paradigm used in MedRecords works very well in most cases it is sometimes better to enter data through a graphical user interface. To add a plug-in architecture that would allow various extra input methods would not be very complicated.

- Improved Visualizers. The name Med*View* indicates that viewing visually what is in the knowledge base is an important part of the project's goals. What the best tools for visualizing various aspects of the collected knowledge should be needs some serious work.

- Interactive Distant Consultations. Currently the members of SOMNET send patient information, including images, to each other via email. A better approach would be to build tools for real-time communication using audio/video such that the expert asked for advice can view the patient directly. With both parties having access to the common MedView knowledge base similar cases could be viewed and discussed in relation to the current patient.

- Educational tools. Using the collected material in MedView for education is an obvious application. Educational tools could be of various kinds and directed at different groups, students, graduate students, practitioners, researchers etc. We are currently investigating the possibilities of building an Electronic Handbook of Oral Medicine. An idea of this handbook would be to combine general rules given by experts with actual examples from the MedView knowledge base.

- Web Tools. Related to the above is accessing the MedView database using the World Wide Web. A web application for MedView could combine several of the suggestions above in a MedView web portal. By logging into this the user would have access to distant consultations, image search, patient summaries, electronic handbooks and tutorials, and so on. Once the basics of the various functions are in place, allowing access using the Internet is essentially a matter of programming and user interface design.

- Searching for patterns. Related to the need for visualizations aiding the user is automated database searches. A data-mining program could be constructed to search the knowledge base for patterns that could be reported to an expert for further evaluation. This approach is the dual to letting the user search for patterns using visualizations and direct manipulation.

The list could be made much longer but we stop here for now.

References

[1] L. M. Abbey. An expert system for oral diagnosis. *J Dent Educ*, 51:475–480, 1987.

[2] F. Alte da Veiga. Structure discovery in medical databases: A conceptual clustering approach. *Artificial Intelligence in Medicine*, 8:473–491, 1996.

[3] R. Armstrong, J. Lesiewicz, G. Harvey, L. Lee, K. Spoehr, and M. Zultak. Clinical panel assessment of photodamaged skin treated with isotretinoin using photographs. *Arch Dermatol*, 128:352–356, 1992.

[4] M. Aronsson. Methodology and programming techniques in GCLA II. In *Extensions of logic programming, second international workshop, ELP'91*, number 596 in Lecture Notes in Artificial Intelligence. Springer-Verlag, 1992.

[5] M. Aronsson. *GCLA, The Design, Use, and Implementation of a Program Development System.* PhD thesis, Stockholm University, Stockholm, Sweden, 1993.

[6] D. Asimov. The grand tour: A tool for viewing multidimensional data. *SIAM J. Sci. Stat. Comp.*, 1(6):128–143, 1985.

[7] T. Axell, J. Pindborg, C. J. Smith, and I. van der Waal. Oral white lesions with special reference to precancerous and tobacco-related lesions: conclusions of an international symposium held in Uppsala, Sweden, May 18-21 1994. International Collaborative Group on Oral White Lesions. *Journal of Oral Pathology*, 25(2):49–54, 1996.

[8] J. D. Bader and D. A. Shugars. A case for diagnoses. *J Am Coll Dent*, (64):44–46, 1997.

[9] J. Bateman. Enabling technology for multilingual natural language generation: the KPML development environment. *Natural Language Engineering*, (3):15–55, 1997.

[10] D. K. Benn, D. D. Dankel, D. Clark, R. B. Lesser, and A. B. Bridgewater. Standardizing data collection and decision making with an expert system. *J Dent Educ*, 61:885–894, 1997.

[11] H. Beyer and K. Holtzblatt. *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann: San Francisco, CA, 1998.

[12] J. Bolewska, H. Hansen, P. Holmstrup, J. Pindborg, and M. Stangerup. Oral mucosal lesions related to silver amalgam restorations. *Oral Surg Oral Med Oral Pathol*, 70:55–58, 1990.

[13] B. Brehmer. Dynamic decision making human control of complex systems. *Acta Psychol Amst*, 81:211–241, 1992.

[14] M. R. Brickley, J. P. Shepherd, and R. A. Armstrong. Neural networks: a new technique for development of decision support systems in dentistry. *J Dent*, 26:305–309, 1998.

[15] S. L. Brooks. Computed tomography. *Dent Clin North Am*, 37:575–590, 1993.

[16] B. Buchanan, J. Moore, D. Forsythe, G. Carenini, and S. Ohlsson. Using medical informatics for explanation in a clinical setting. Technical Report 93-16, Intelligent Systems Laboratory, University of Pittsburg, 1994.

[17] S. Busemann and H. Horacek. A flexible shallow approach to text generation. In E. Hovy, editor, *Proceedings of the Nineth International Natural Language Generation Workshop (INLG'98)*, pages 238–247, 1998.

[18] L. Cahill, C. Doran, R. Evans, C. Mellish, D. Paiva, M. Reape, D. Scott, and N. Tipper. Towards a reference architecture for natural language generation systems. Technical Report ITRI-99-14, University of Brighton, March 1999.

[19] J. Chasteen. A computer database approach for dental practice. *J Am Den Assoc*, 123:26–33, 1992.

[20] V. Chinburapa, L. Larson, M. Brucks, J. Draugalis, J. Bootman, and C. Puto. Physician prescribing decisions: the effects of situational involvement and task complexity on information acquisition and decision making. *Soc. Sci. Med.*, 36:1473–1482, 1993.

[21] T. Chomut. Exploratory data analysis in parallel coordinates. Research report, IBM Los Angeles Scentific Center, 1987.

[22] J. J. Cimino, P. D. Clayton, G. Hripcsak, and S. B. Johnson. Knowledge-based approaches to the maintenance of a large controlled medical terminology. *J. Am. Med. Inform. Assoc.*, 1:35–50, 1994.

[23] R. Dale and E. Reiter. Building applied natural-language generation systems. *Journal of Natural Language Engineering*, 3:55–87, 1997.

[24] R. Dale and E. Reiter. *Building Natural-Language Generation Systems.* Cambridge University Press, 2000.

[25] J. C. Davenport and P. Hammond. The acquisition and validation of removable partial denture design knowledge. I. Methodology and overview. *J Oral Rehabil*, 23:152–157, 1996.

[26] J. C. Davenport, P. Hammond, and P. Hazlehurst. Knowledge-based systems, removable partial denture design and the development of RaPiD. *Dent Update*, 24:227–233, 1997.

[27] D. N. Davis and D. Forsyth. Knowledge-based cephalometric analysis a comparison with clinicians using interactive computer methods. *Comput Biomed Res*, 27:210–228, 1994.

[28] M. Diehl. Developing the american dental association concept model for the standard computer-based oral health record. *J Am Coll Dent*, 62:30–32, 1995.

[29] C. Dragula and G. Burin. International harmonization for the risk assessment of pesticides results of an IPCS survey. *Regul. Toxicol. Pharmacol.*, 20:337–353, 1994.

[30] R. B. Elson and D. P. Connelly. Computerized decision support systems in primary care. *Prim Care*, 22:356–384, 1995.

[31] G. Enislidis, I. V. Wagner, O. Ploder, and R. Ewers. Computed intraoperative navigation guidance–a preliminary report on a new technique. *Br J Oral Maxillofac Surg*, 35:271–274, 1997.

[32] G. Falkman. Similarity measures for structured representations: a definitional approach. In E. Blanzieri and L. Portinale, editors, *EWCBR-2K, Advances in Case-Based Reasoning*, Lecture Notes in Artificial Intelligence. Springer–Verlag, 2000. To appear.

[33] G. Falkman, J. Norbeck, L. Hallnäs, and A. Blomberg. The use of similarity measures and three-dimensional hierarchical clustering for the analysis of expression data from proteome analysis of *Saccharomyces cerevisiæ*. Presented at *Bioinformatics'99, Lund, Sweden, June 1999*, June 1999.

[34] C. Farr. Million-dollar files: adding up the benefits of computerized patient records. *Dent Today*, 14:72–79, 1995.

[35] A. R. Firestone, D. Sema, T. J. Heaven, and R. A. Weems. The effect of a knowledge-based image analysis and clinical decision support system on observer performance in the diagnosis of approximal caries from radiographic images. *Caries Res 32*, 32:127–134, 1998.

[36] F. J. Firriolo and B. A. Levy. Computer expert system for the histopathologic diagnosis of salivary gland neoplasms. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*, 82:179–186, 1996.

[37] F. J. Firriolo and T. Wang. Diagnosis of selected pulpal pathoses using an expert computer system. *Oral Surg Oral Med Oral Pathol*, 76:390–396, 1993.

[38] D. Fontaine, C. Riou, and C. Jacquelinet. An intelligent computer-assisted instruction system for clinical case teaching. *Methods Inf Med*, 33:433–445, 1994.

[39] M. R. Frank and P. Szekely. Adaptive forms: An interaction paradigm for entering structured data. In *Proceedings of the 1998 International Conference user interfaces*, San Fransisco, CA USA, 1998.

[40] S. Geldof and W. V. de Velde. An architecture for template based (hyper)text generation. In *Proceedings of the 6th European Workshop on Natural Language Generation-EWNLG'97*, pages 28–37, 1997.

[41] L. Gierl, M. Bull, and R. Schmidt. CBR in medicine. In *Case-based reasoning technology: From foundations to applications*, volume 1440 of *Lecture Notes in Artificial Intelligence*, pages 273–297. Springer-Verlag, 1998.

[42] E. Goldberg, N. Driedger, and R. Kittredge. Using natural-language processing to produce weather forecasts. *IEEE Expert*, 9(2):45–53, 1994.

[43] W. T. Gossen, P. J. Epping, and I. L. Abraham. Classification systems in nursing: Formalizing nursing knowledge and implications for nursing information systems. *Methods Inf Med*, 35(1):59–71, 1996.

[44] L. Green and M. Becker. Physician decision making and variation in hospital admission rates for suspected acute cardiac ischemia. A tale of two towns. *Med. Care*, 32:1086–1097, 1994.

[45] T. A. Gregg and D. H. Boyd. A computer software package to facilitate clinical audit of outpatient paediatric dentistry. *Int J Paediatr Dent*, 6:45–51, 1996.

[46] L. Hallnäs. Partial inductive definitions. *Theoretical Computer Science*, 87(1):115–142, 1991.

[47] P. Hammond, J. C. Davenport, and A. J. Potts. Knowledge-based design of removable partial dentures using direct manipulation and critiquing. *J Oral Rehabil*, 20:115–123, 1993.

[48] R. M. Hammond and T. J. Freer. Application of a case-based expert system to orthodontic diagnosis and treatment planning: a review of the literature. *Aust Orthod J*, 14:150–153, 1996.

[49] R. M. Hammond and T. J. Freer. Application of a case-based expert system to orthodontic diagnosis and treatment planning. *Aust Orthod J*, 14:229–234, 1997.

[50] V. Hasselblad and D. McCrory. Meta-analytic tools for medical decision making: a practical guide. *Med Decis Making*, 15:81–96, 1995.

[51] J. J. Hyman and W. Doblecki. Computerized endodontic diagnosis. *J Am Dent Assoc*, 107:755–758, 1983.

[52] A. Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1:69–91, 1985.

[53] R. A. Jenders, M. Morgan, and G. O. Barnett. Use of open standards to implement health maintenance guidelines in a clinical workstation. *Comput Biol Med*, 24:385–390, 1994.

[54] I. Jurisica, J. Mylopoulos, J. Glasgow, and H. Shapiro. Case-based reasoning in IVF: Prediction and knowledge mining. *Artificial Intelligence in Medicine*, 12(1):1–24, 1998.

[55] D. Kalra. Electronic health records the European scene. *Bmj*, 309:1358–1361, 1994.

[56] M. A. Kamrin. Environmental risk harmonization federal/state approaches to risk assessment and management. *Regul Toxicol Pharmacol*, 25:158–165, 1997.

[57] A. Knott, C. Mellsih, J. Oberlander, and M. O'Donnell. Sources of flexibility in dynamic hypertext generation. In *Proceedings of the International Workshop on Natural Language Technique*, pages 64–71, 1996.

[58] T. Kohonen. *Self-organizing maps.* Springer-Verlag, 2nd edition, 1997.

[59] J. L. Kolodner. An introduction to case-based reasoning. *Artificial Intelligence Review*, 6(1):2–34, 1992.

[60] I. Kramer, N. el Labban, and S. Sokodi. Further studies on lesions of the oral mucosa using computer-aided analyses of histological features. *Br J Cancer*, 29:223–231, 1974.

[61] I. R. Kramer. Computers in clinical and laboratory diagnosis. *Int Dent J*, 30:214–225, 1980.

[62] P. Kreuger. GCLA II: A definitional approach to control. In *Extensions of logic programming, second international workshop, ELP91*, number 596 in Lecture Notes in Artificial Intelligence. Springer-Verlag, 1992.

[63] P. Kreuger. GCLA II: A definitional approach to control. Licentiate thesis, Chalmers University of Technology, 1992.

[64] M. J. Lincoln, C. W. Turner, P. J. Haug, J. W. Williamson, S. Jessen, R. M. Cundik, and H. R. Warner. Iliad's role in the generalization of learning across a medical domain. In *Proc Annu Symp Comput Appl Med Care*, pages 174–178, 1992.

[65] H. J. Lowe, B. G. Buchanan, G. F. Cooper, and J. K. Vries. Building a medical multimedia database system to integrate clinical information an application of high-performance computing and communications technology. *Bull Med Libr Assoc*, 83:57–64, 1995.

[66] R. Makura and T. Makura. MACRAD: Radiology image resource with a case-based retrieval system. In M. Veloso and A. Aamodt, editors, *Case-based reasoning research and development: Proceedings of the first international conference, ICCBR-95*, volume 1010 of *Lecture Notes in Artificial Intelligence*, pages 43–54. Springer-Verlag, 1995.

[67] U. Mattson, G. Heyden, A. Chodorowski, T. Gustavsson, M. Jontell, and F. Bergquist. Computer analysis in oral lichenoid reactions. *Acta Odont Scand*, 52:86–92, 1994.

[68] U. Mattson, A. Jönsson, M. Jontell, and J. Cassuto. Digital image analysis (DIA) of colour changes in human skin exposed to standardized thermal injury and comparison with laser doppler measurements. *Comput Methods Programs Biomed*, 50:31–42, 1996.

[69] A. M. McCreery and E. Truelove. Decision making in dentistry. Part I: A historical and methodological overview. *J Prosthet Dent*, 65:447–451, 1991.

[70] A. M. McCreery and E. Truelove. Decision making in dentistry. Part II: Clinical applications of decision methods. *J Prosthet Dent*, 65:575–585, 1991.

[71] J. Merz, M. Small, and P. Fishbeck. Measuring decision sensivity: A combined Monte Carlo-logistic regression approach. *Med Decis Making*, 12:189–196, 1992.

[72] N. Mohl and R. Ohrbach. Clinical decision making for temporomandibular disorders. *J Dent Educ*, 56:823–833, 1992.

[73] B. D. Monteith. Computerized expert system for the diagnosis of pulp-related pain. *Int J Prosthodont*, 4:30–36, 1991.

[74] R. Mulligan and G. J. Wood. A controlled evaluation of computer assisted training simulations in geriatric dentistry. *J Dent Educ*, 57:16–24, 1993.

[75] E. Munch. P.A.I.S., a personal medical information system. A comprehensive medical knowledge base. *Hno*, 42:355–373, 1994.

[76] L. Nicholson and I. Beaulieu. Interactive multimedia learning systems in oral medicine: experiences at the Faculty of Oral Medicine of Laval University. *J Can Dent Assoc*, 63:819–821, 1997.

[77] C. Nielson, C. S. Smith, D. Lee, and M. Wang. Implementation of a relational patient record with integration of educational and reference information. In *Proc Annu Symp Comput Appl Med Care*, pages 125–129, 1994.

[78] J. Nivre and T. Lager. Constraint-based text realization. Submitted to Natural Language Engineering, 1999.

[79] S. P. Nolan. The search for standards. *J .Heart .Valve. Dis.*, pages 7–9, 1995.

[80] Penman. The Penman documentation. Technical report, USC/Information Sciences Institute, 1989.

[81] D. Perednia, J. Gaines, and A. Rossum. Variability in physician assessment of lesions in cutaneous images and its implications for skin screening and computer-assisted diagnosis. *Arch Dermatol*, 128:357–364, 1992.

[82] P. E. Petersen, L. B. Christensen, I. J. Moller, and K. S. Johansen. Continuous improvement of oral health in Europe. *J Ir Dent Assoc*, 40(4):105–107, 1994.

[83] L. C. Peterson, D. S. Cobb, and D. C. Reynolds. ICOHR: intelligent computer based oral health record. *Medinfo*, 2, 1995.

[84] R. W. Priddy and L. Yip. ORPAMS: a data-management system for oral pathology. *Oral Surg Oral Med Oral Pathol*, 61:590–596, 1986.

[85] A. L. Rector and W. A. Nowlan. The GALEN project. *Computer Methods and Programs in Biomedicine*, 45:75–78, 1993.

[86] J. L. Reginster. Harmonization of clinical practice guidelines for the prevention and treatment of osteoporosis and osteopenia in Europe a difficult challenge. *Calcif Tissue Int*, 59:24–29, 1996.

[87] Y. Reisman. Computer-based clinical decision aids. A review of methods and assessment of systems. *Med Inf*, 21:179–197, 1996.

[88] E. Reiter. Has a consensus NL generation architecture appeared, and is it psycholinguistically plausible? In *Proc of the Seventh International Workshop on Natural Language Generation (INLGW-1994)*, pages 163–170, Kennebunkport, Maine, USA, 1994.

[89] E. Reiter. NLG vs templates. In *Proc of the fifth European Workshop on Natural-Language Generation*, Leiden, The Netherlands, 1995.

[90] E. Reiter. Shallow vs. deep techniques for handling linguistic constraints and optimisations. In *Proceedings of the KI-99 Workshop on May I Speak Freely: Between Templates and Free Choice in Natural Language Generation*, 1999.

[91] E. Reiter, C. Mellish, and J. Levine. Automatic generation of technical documentation. *Applied Artificial Intelligence*, 9, 1995.

[92] E. Reiter, C. Mellsih, and J. Levine. Automatic generation of on-line documentation in the IDAS project. In *Proceedings of the Third Conference on Applied Natural Language Processing (ANLP-1992)*, pages 64–71, 1992.

[93] E. Reiter, R. Robertson, and L. Osman. Types of knowledge required to personalise smoking cessation letters. In *Artificial Intelligence and Medicine: Proceedings of AIMDM-1999*, pages 389–399, 1999.

[94] P. R. Rhodes. The computer-based oral health record exploring a new paradigm. *J Calif Dent Assoc*, 22:29–33, 1994.

[95] D. Rizzi. Medical diagnosis. *Ugeskr Laeger*, 153:694–697, 1991.

[96] G. Rolfe. Science, abduction and the fuzzy nurse an exploration of expertise. *J Adv Nurs*, (25):1070–1075, 1997.

[97] J. L. Rudin. DART (diagnostic aid and resource tool): a computerized clinical decision support system for oral pathology. *Compendium*, 15, 1994.

[98] S. Seipel, I. V. Wagner, S. Koch, and W. Schneider. Three-dimensional visualization of the mandible: a new method for presenting the periodontal status and diseases. *Comput Methods Programs Biomed*, 46:51–57, 1995.

[99] S. Seipel, I. V. Wagner, S. Koch, and W. Schneider. A virtual reality environment for enhanced oral implantology. *Medinfo*, 1995.

[100] P. N. Sellen, D. C. Jagger, and A. Harrison. Computer-generated study of the correlation between tooth, face, arch forms, and palatal contour. *J Prosthet Dent*, 80:163–168, 1998.

[101] B. Shneiderman. *Designing the User Interface: Strategies for Effective Human-Computer Interaction.* Addison-Wesley Longman, Inc, Reading, MA, USA, 3rd edition, 1998.

[102] J. H. Sims-Williams, I. D. Brown, A. Matthewman, and C. D. Stephens. A computer-controlled expert system for orthodontic advice. *Br Dent J*, 163:161–166, 1987.

[103] H. Somers, B. Black, J. Ellman, L. Giardoni, T. Lager, A. Multari, J. Nivre, and A. Rogers. Multilingual generation and summarization of job adverts: The TREE project. In *Proceedings of the Fifth Conference on Applied Natural Language Processing*, 1997.

[104] S. T. Sonis and K. A. Costello. A database for mucositis induced by cancer chemotherapy. *Eur J Cancer B Oral Oncol*, 31B:258–260, 1995.

[105] P. M. Speight, A. E. Elliot, J. A. Jullien, M. C. Downer, and J. M. Zakzrewska. The use of artificial intelligence to identify people at risk of oral cancer and precancer. *Br Dent J*, 179:382–387, 1995.

[106] J. Stempczynska and E. Kacki. Problems of knowledge acquisition automation in medical expert systems. *Medinfo*, 1:857–860, 1995.

[107] C. D. Stephens, N. Mackin, and J. H. Sims-Williams. The development and validation of an orthodontic expert system. *Br J Orthod*, 23:1–9, 1996.

[108] S. E. Stheeman, P. F. van der Stelt, and P. A. Mileman. Expert systems in dentistry: Past performance—future prospects. *J Dent*, 20:68–73, 1992.

[109] O. Torgersson. Gisela—a framework for definitional programming, 2000.

[110] O. Torgersson. MedView and Gisela, 2000.

[111] J. W. Tukey. *Exploratory Data Analysis.* Addison-Wesley, Reading, MA, USA, 1977.

[112] J. A. Valenza. Medical risk report: improving patient management and record keeping through a problem-oriented approach. *J Gt Houst Dent Soc*, 65:46–48, 1994.

[113] P. F. van der Stelt. Computer-assisted interpretation in radiographic diagnosis. *Dent Clin North Am*, 37:683–696, 1993.

[114] V. Velanovich. Bayesian analysis in the diagnostic process. *Am J Med Qual*, 9:158–161, 1994.

[115] I. V. Wagner and W. Schneider. Computer based decision support in dentistry. *J Dent Educ*, 1991.

[116] I. D. Watson. An introduction to case-based reasoning. In I. Watson, editor, *Progress in case-based reasoning*, volume 1020 of *LNAI*, pages 3–16. Springer-Verlag, 1995.

[117] E. J. Wegman. Hyperdimensional data analysis using parallel coordinates. *Journal of the American Statistical Association*, 85(411):664–675, 1990.

[118] E. J. Wegman. The grand tour in k-dimensions. In *Computing Science and Statistics. Statistics of Many Parameters: Curves, Images, Spatial Modes. Proceedings of the 22nd Symposium on the Interface*, pages 127–136, New York, 1992. Springer-Verlag.

[119] S. C. White. Decision-support systems in dentistry. *J Dent Educ*, 60:47–63, 1996.

[120] P. J. Whitehouse, C. G. Sciulli, and R. M. Mason. Dementia drug development use of information systems to harmonize global drug development. *Psychopharmacol Bull*, 33:129–133, 1997.

[121] N. Zhizhina, A. A. Prokhonchukov, I. M. Rabinovich, and V. Pelkovskii. A computerized automated system for the differential diagnosis and treatment of oral mucosal diseases. *Stomatologiia*, 77(1):55–61, 1998.