Experiences with the Development, Implementation and Evaluation of Automated Decision Support Systems

Paul de Clercq, Arie Hasman

Department of Medical Informatics, Maastricht University, Maastricht, The Netherlands

Abstract
A framework for the implementation of guidelines -GASTON- was developed. Its functions range from the definition of guidelines (using a GLIF like approach) to the implementation of a DSS that can be coupled to existing information systems. This paper discusses experiences with the development of three systems in different domains. It is concluded that the toolbox corresponding to the framework could be successfully used to develop these systems.

Keywords:
Guidelines, decision support system, GLIF, evaluation, interfacing

Introduction
There have been numerous efforts to develop systems that support guideline-based care in an automated fashion, covering a wide range of clinical settings and tasks [1]. However, building systems that are both effective in supporting clinicians and accepted by them has proven to be a difficult task. In order to be successful, attention must be paid to various areas such as guideline representation, acquisition, verification and execution [2].

During the last years, various groups have worked on the development of generic methodologies for representing, acquiring and implementing computer-based guidelines [3]. This paper reports experiences concerning the development, implementation and evaluation of guideline-based systems that were created with one of these approaches. This approach, named Gaston, consists of a framework that facilitates all stages in the guideline development process, ranging from the definition of models that represent guidelines to the implementation of run-time systems that provide decision support [4]. The framework consists of 1) a guideline representation formalism that uses the concepts of primitives, Problem-Solving Methods (PSMs) and ontologies [5], 2) a guideline authoring environment that enables guideline authors to define guidelines, and 3) a guideline execution environment that translates defined guidelines into a more efficient symbol-level representation, which can be processed by an execution-time engine that forms the Decision Support System (DSS).

The aim of this paper is to examine and evaluate the possibilities of the Gaston approach. Questions commonly asked when developing and implementing computer-based guidelines such as 'how to represent different kinds of guidelines in a straightforward manner', 'how to facilitate guideline authors during the acquisition process', 'how to map concepts from a guideline to corresponding concepts in the real world' and 'how to support care providers in daily practice using guideline-based decision support systems'.

The remaining part of this paper addresses these questions by describing and discussing three systems that were developed for use in the specialties of family practice, critical care and psychiatry: 1) GRIF: a reminder system that provides automated feedback on test ordering in general practice [6], 2) CritICIS: a real-time critiquing system used in critical care environments such as Intensive Care Units [7] and 3) M-PADS: a psychopharmacological advisory system that provides decision support concerning the selection of the most suitable psychoactive drug [8].

Also, a fourth system was developed, aimed at the management of chronic diseases. However, this system and the results are described in more detail elsewhere [9].

When describing the GRIF system we focus on the impact of the system on its users (family physicians). The description of the M-PADS system focuses mainly on the guideline representation part. The CritICIS system description focuses on guideline verification, guideline representation and acquisition.

The next three sections describe each system in more detail. The paper ends with a general discussion on the use of these systems and their relation with the Gaston framework.

Automated feedback on test ordering in general practice

Overview
The GRIF system was developed with the aim of checking Family Physicians’ (FP) laboratory test ordering behavior. GRIF consists of five parts: a knowledge base, an order entry system, a module that provides reactive support (i.e. the advice), a module that provides passive support and a database.

Design and Implementation
The GRIF knowledge base contains 134 rule-based guidelines extracted from national and regional guidelines. The guidelines were entered into the knowledge base using the GRIF Knowledge Acquisition Tool (KA-Tool). A domain ontology was built, based on the International Classification of Primary Care (ICPC) [10]. Using concepts of this domain ontology, guideline authors could enter the guidelines in the KA-Tool (examples of the
workings and the user interface of the GRIF KA-Tool are shown elsewhere [6]).

Clinician researchers carried out a logical verification (detection of contradictions and conflicts). A structure verification of the knowledge base was also carried out. Finally, recommendations of the GRIF system were compared with comments of human experts, based on their interpretations of the guidelines [11].

The acquired rules are transferred to the GRIF DSS (which is also part of the Gaston framework), which provides the active support (e.g. generates the actual recommendations). The DSS reads the patient data and checks whether any of the rules fires and which feedback has to be provided (figure 1).

Results

The efficacy of the GRIF system was evaluated in a laboratory setting. In total 24 randomly selected FPs reviewed a random sample of 30 request forms they filled in earlier that year. After the request forms were updated by them (by removing or adding some tests) GRIF displayed critical comments about any remaining non-adherence to the guidelines. The number of requested diagnostic tests decreased with 17% (95% CI: 12-22%) due to the comments of the GRIF system. In addition, the fraction of tests ordered not in accordance with the practice guidelines decreased by 39% (95% CI: 28-51%). The FPs accepted 362 (50%) of the 729 recommendations. In the laboratory tests no patients were present. In the field test that followed, the patients could influence the test requesting behaviour of FPs. Eleven FPs in two regions in the Netherlands used GRIF instead of filling in the usual paper request forms from August 2000 to July 2001 during patient consultation. Usage behavior, the quality of the provided information and the fraction of recommendations that was accepted by the FPs were analyzed. During the intervention period, the FPs produced 2498 request forms with 10139 tests on it, using the GRIF system. Of the 2780 recommendations, the percentage of accepted recommendations varied between 3.4 and 8.3 percent dependent on the type of recommendation that was given. Advice that suggests removing a test because another - more appropriate or efficient - test was also requested and advice that suggest to request an alternative test were followed most frequently. The median time to generate, read and act on the presented feedback comments was 13 seconds. Entering (coded) medical patient data costs FPs a relatively large part of their patient consultation time.

The GRIF system is in operation for more than two years in 15 Family Practices in the Netherlands and is still being used in daily practice.

A real-time reminder system in Critical Care environments

Overview

The CritICIS system is a real-time reminder system that reminds ICU health care workers of inconsistencies between the carried out treatment and the corresponding guidelines. The objectives of the project were to provide decision support to health care workers in clinical and emergency care and to design a knowledge acquisition environment that enables ICU care providers to formulate, update and verify guidelines without the assistance of a knowledge engineer.

Design and implementation

Three different types of guidelines were used: 1) relatively simple rule-based guidelines, 2) multiple-step guidelines that consist of primitives (flowcharts) and 3) guidelines that use Problem-Solving Methods (PSMs), generic strategies to solve domain-independent stereotypical tasks [5]. An example of a guideline entered in the Gaston KA-Tool is shown in figure 2. The KA-Tool makes use of the IMPACT terminology, a set of medical terms describing the state of a patient in an Intensive Care Unit.
Entered guidelines are transferred to the CritICIS DSS. The DSS reads in the necessary patient data and compares the data with the guidelines. Whenever a guideline is not followed, the DSS sends a warning to the ICU care providers, similar to the warning given by the GRIF system (figure 1). The DSS has access to two sources of data: 1) a Patient Data Management System (PDMS) that holds clinical data such as prescribed drugs and established diagnoses, and 2) a patient monitoring system that broadcasts physiological data such as a patient’s blood pressure or heart rate.

Results
The CritICIS system contains guidelines aimed at both physicians and nurses. For a national study, in which the ICU participated, the results of certain laboratory tests had to be entered into the PDMS for a particular group of patients. During one year, the reminders that were given to the nursing staff whenever one or more relevant data items were missing in the PDMS database, were logged. The results show that the number of reminders per discharged patient did not decrease as a function of time, as one would expect. Interviews with the nursing staff revealed that they were not using CritICIS as a reminder system, but as an intelligent order entry form. Part of the staff deliberately did not check daily whether the data were complete since they knew that CritICIS would check at patient discharge whether data were missing and then would provide an opportunity for entering the required data. The fact that the developers of CritICIS intended to increase the acceptance of CritICIS by not only reminding the nursing staff but also giving them the means to enter the missing data on the spot resulted in this behavior. The advantage of this approach is that is does increase the system’s acceptance. The drawback however, is that the users start depending on the system, which increases the possibility of errors whenever the system is not functioning or has an incomplete knowledge base.

The CritICIS system also contains a number of medical guidelines to detect drug interactions, contraindications and side effects (represented by means of PSMs). The guidelines were executed whenever a physician changed an item in the PDMS database. When the same reminder was issued more than once for a certain patient, it was marked as ‘hidden’ and not shown again. Over a period of 6 months, the DSS was executed 16,340 times. During those 16,340 runs, the DSS issued one or more reminders 2,928 times. The total number of reminders was 3,753 of which 2,731 were not shown to the physician (marked as ‘hidden’). Of the 1,022 reminders that were given, 583 were ignored, 224 were classified as ‘correct’ and 215 were classified as ‘incorrect’. All the guidelines that led to an incorrect reminder were updated in order to decrease the number of false reminders. Some of these guidelines were already examined during a retrospective study using data of earlier admitted patients [7]. However, this study showed that it is still possible to get incorrect reminders as a result of new patient data and changing policies or guidelines.

We also measured the satisfaction with CritICIS of the three intensivists that used the system already two years. A questionnaire, based on the IBM computer usability satisfaction questionnaire [12], was used. The questionnaire consisted of items, relating to usability, training and support, user satisfaction, behavioral changes and usefulness. Furthermore, the questionnaire contained a number of open questions concerning opinions about and experiences with the system. The results were discussed with the intensivists.

All intensivists judged the system workable. The user interface was generally regarded as ‘intuitive’ and easy-to-use. The intensivists differed in opinion on the issues of productivity and effectiveness, as some stated that it ‘slowed down the process of entering patient data’. Also support and training were sufficient and they were satisfied with the system.

In general, the intensivists did not believe that the use of critiquing systems such as CritICIS would automatically change their behavior, especially because of the amount of data to be entered. They did state, however, that they would be willing to encode more information in the PDMS for the purpose of decision support. Also, a combination of critiquing and pro-active decision support would be favorable for them.

All intensivists strongly agreed that systems such as CritICIS are useful and that similar systems must be implemented in other departments.

Other comments of the intensivists concerned issues related to completeness, local adaptation and interfacing. They stated that to improve the acceptance of the system in daily practice, it is necessary that the guideline knowledge base at least must contain those guidelines that cover the daily routine of the ICU. They further stated that a systematic procedure is mandatory that facilitates entering new guidelines or updating existing ones. Hospital organizational bodies must support this procedure. Also, they want to use (inter)national guidelines as a basis, from which they must be able to make local adaptations that fit their own institution.

In addition to the current critiquing approach, they suggested a more pro-active approach. Currently, the CritICIS system is implemented as a critiquing system that warns physicians whenever a guideline is not followed. A pro-active approach would enable them to ask the system for advice regarding certain complications, treatments or differential diagnoses. Finally, they suggested that it must be possible for CritICIS to send reminder-related data back to the PDMS if the intensivist agrees to a reminder. For example, whenever a reminder states that the prescription of a certain medication is not advisable, it must be possible to inform the PDMS that this medication must be stopped immediately.

The CritICIS system is in operation since 2001 at the ICU of the Catharina Hospital, Eindhoven, the Netherlands.

A Multidisciplinary Psychoactive Drug Selection Advisory System

Overview
The Multidisciplinary Psychoactive Drug Selection –advisor system (M-PADS) is a decision support system developed for selecting the most appropriate psychoactive drug in order to treat psychiatric patients [13]. It contains guidelines and PSMs that solve a number of tasks, varying from relatively simple tasks that
address drug interaction, to more complex ones that process ‘deep knowledge’ (using a semantic network).

**Design and Implementation**

A neuropsychopharmacological domain ontology relevant to rational psychoactive drug selection was defined, which combines all involved knowledge domains, required to support the psychoactive drug selection task. Furthermore, a knowledge analysis was performed to describe the psychoactive drug selection task by means of a clinical algorithm and to model in a semiformal way the specification of this task. The psychoactive drug selection task can be viewed as a modular task, which consists of the execution of different subtasks. Each (sub)task was modeled by means of a PSM. The KA-Tool loads the domain ontology, relevant PSMs and primitives, and creates a user interface that enables guideline authors to define guidelines that describe the psychoactive drug selection task.

After entering the guidelines, the KA-Tool generates a knowledge base, which is interpretable by the M-PADS DSS. The DSS is activated when the clinician selects the psychoactive drug selection task. First the clinician has to enter the established DSM-IV diagnosis of the patient. A list of possible drug-therapy options is now generated. The clinician may now enter a concurrent medication (e.g. Zantac) and/or a special disease state (e.g. hypotension), and/or a special patient group (e.g. elderly) of the patient. The decision support system generates on request a list of possible contraindications. The level of each contraindication is mentioned (e.g. absolute/relative). The clinician can now order the required monitoring activities for the relative contraindication.

**Results**

By using the methods, described in the previous section, it was possible to develop a psychopharmacological advisory system based on explicit models of the neuropsychopharmacological domain and the problem solving method related to the psychoactive drug selection task. These explicit models integrate the clinical pharmacological, pathophysiological and pharmacotherapeutic knowledge required to support rational psychoactive drug selection. At the moment, M-PADS provides patient specific advice, based on up to date knowledge to treat major depressions. Since the knowledge base is organized in a modular fashion with declarative and procedural knowledge separated, it can be easily expanded or modified so that the knowledge remains up to date. The M-PADS has opportunities to build reusable and explainable knowledge based systems for pharmacotherapy.

**Discussion**

Although a lot of progress is being made in the area of guideline-based decision support, these decision support systems are still not implemented on a large scale. One of the largest problems is that the medical community is very heterogeneous by nature. Numerous medical specialties exist, each with their own types of guidelines, intended users and information systems. Local institutions usually have their own local customs and regulations, which demand that it must be possible to ‘override’ national guidelines with local adaptations. Also, interfacing decision support systems with third-party systems (e.g. Electronic Patient Records) as well as with the system’s users (e.g. care providers) usually requires a lot of effort and resources due to a lack of standardization.

The Gaston approach limits the amount of time and resources by developing an architecture that can be used to implement a large range of guideline-based decision support systems. The experience, described in this paper, shows that Gaston can be used to develop systems that differ in application domain (e.g., family practice, critical care, psychiatry, chronic disease management), application environment (e.g., FP information system, PDMS) and application users (e.g., FP, physician, nursing staff).

In the projects, custom-developed domain ontologies were used (partly based on existing terminologies such as ICPC and IMPACT). In order to improve standardization aspects, it might be more favorable to use standard terminologies such as UMLS [14] or SNOMED [15] for all projects. On the other hand, it is important that guideline authors in local institutions ‘recognize’ their own concepts. For example, a guideline may refer to the medication Fluoxetine Hydrochloride that is defined in a standard domain ontology, which might be better known in some specialties by its brand name Prozac. In Gaston, mappings tools were utilized to reuse similar concepts in various projects.

The Gaston KA-Tool was used to acquire all guidelines, varying from the relatively simple rule-based guidelines (GRIF, CritICIS) to more complex guidelines and PSMs (CritICIS, MPADS). Similar to the development and application of domain ontologies, it is important to reach a balance between standardization and easy-of-use. Defining multiple user interfaces in the Gaston KA-Tool, based on the underlying guideline representation model, made it possible to reuse the KA-Tool in all projects. The results from the CritICIS system show that local adaptation and versioning aspects are very important. Although the Gaston tools contains methods that facilitate 1) overriding national guidelines with local adaptations and 2) updating local guidelines by guideline authors without the assistance of knowledge engineers, the organization of institution must ensure that these tasks are also carried out. If not, too much false reminders may be given, which will decrease the system’s acceptance dramatically.

In all projects, the Gaston execution engine was used as a DSS and interfaced with existing patient information. Two of the systems (GRIF and CritICIS) are used in daily practice. In all projects, the main bottleneck was interfacing the execution engine with the external patient information systems. The fact that we developed the domain ontologies with the terminology of the patient information system in mind (e.g., the IMPACT ontology was used in CritICIS as well as in the PDMS) simplified the interfacing between the patient information systems and the Gaston execution engine. Separating the Gaston execution engine into multiple components each performing a different task (e.g., guideline inference, system interfacing and user communication) increased the reusability of the Gaston execution engine in multiple application domains and settings. This approach has also been recognized by other groups, which started developing execution engines for use in multiple domains [16].
In conclusion, although the number of systems that were developed using the Gaston approach is still limited, the first experiences and results are very promising. The fact that Gaston covers the entire guideline development and implementation process and is supported by a number of generic tools related to the various phases in that process is one of the key elements that made it possible to reuse the approach in various projects. Although still a number of problems have to be addressed, especially related to standardizing, interfacing, organization and local adaptation, the foundation of Gaston is strong enough to extend it further.

References


