

Domain Experts Tailoring Interaction to Users – an Evaluation Study

Helena Lindgren¹, Patrik J. Winnberg¹, and Peter Winnberg¹

¹ Department of Computing Science, Umeå University,
SE-90187 Umeå, Sweden
{helena, patrikw, pw}@cs.umu.se

Abstract. This paper presents ACKTUS, a modeling tool for developing knowledge-based systems for the health domain, and an evaluation study of the system. The main purpose of the evaluation was to investigate whether the functionality and interaction design of ACKTUS was sufficiently intuitive for the domain experts to contribute with knowledge and to model the interaction design of the three end users' applications. Another purpose was to evaluate the applicability of the activity assessment protocol AAIMA for analysis. The study design was qualitative and formative, using observations and interviews with users to collect data. Three medical experts and two experts in occupational therapy participated, providing expertise in four different domains. The participants increased their understanding and skills during the evaluation period leading to improved knowledge-based applications. The AAIMA protocol proved to be useful and the results are fed into ongoing development work on developing the adaptive functionality of the ACKTUS systems.

Keywords: Formative evaluation, interaction design, end-user development and adaptation, knowledge modeling, e-health

1 Introduction

Knowledge-based systems are to an increasing extent used by individuals for monitoring health and for supporting the daily work of health care professionals in the provision of care to individuals (e.g., [1]). In both cases it is of great importance to assure that the mediated knowledge has high quality (e.g. based in evidence-based medicine or other reliable sources) and that the interaction design of the system allows for tailored advice and personalized support in addition to transparency, in order for the end user to gain trust and confidence in the system's content and responses. For these purposes, the system ACKTUS (Activity-Centered Knowledge and interaction modeling Tailored to Users) is being developed, making use of Semantic Web-technology to model both the health-related knowledge and the interaction design of personalized knowledge-based systems [2], [3]. The knowledge engineering of such systems typically requires deep knowledge in the medical domain, knowledge of formal methods and an extensive amount of time to transform informal knowledge into formal structures (e.g., [4]). Because of this, the major aim

of the ACKTUS project is to facilitate the modeling process while also making the domain professional take the interaction design of the end user application into consideration. For the domain professional to be able to also model in what way the end user will interact with the system, functionality that supports this has been integrated and extended during development. Research focuses on how the design of the modeling environment allows the domain professionals to accomplish creating what they envision being the functionality and interaction design of the future support systems. This paper presents ACKTUS and the results from a formative evaluation study, conducted as part of the development of ACKTUS. The overall research question is to what extent the domain expert is able to grasp the structure of ACKTUS and how well the system supports the development of this understanding. A related research question is how well the system supports the domain expert in designing the interaction for other end users. These questions are addressed by investigating the participants' performance in modeling tasks in authentic work, before and after the interface re-design, which aimed to improve the understanding of the system. This was partly done by adopting and adapting an activity assessment protocol presented in [5]. Therefore, an additional central research question is to what extent the adapted protocol is useful for capturing knowledge and skill development in users.

1.1 Tools for Knowledge Representation and Interaction Design

Guidelines for clinical practice are being developed for the purpose of improving care. These typically come in natural language and are not straightforwardly translated into a guideline interpretable by a computer. In order to facilitate this process a number of representation languages have been developed by different research groups specifically for the purpose of representing clinical guidelines and treatment protocols. These languages typically allow for describing "task-network models", where sets of interacting decisions and actions are to be carried out in sequence or in parallel over a period of time. Some examples are Asbru, PROforma, GLIF, EON, GUIDE, and PRODIGY (see comparisons done by Peleg et al. [6] and Isern et al [7]). These languages have in common that they provide methods for regulating a care process with pre- and post-conditions attached to plans. Tasks can be defined and associated to different points in a plan, where the task may involve data collection to be made by a health care professional. At these points typically forms appear where the required information according to the underlying clinical guideline can be entered.

Evaluations have been made of these languages within applied implementation projects of clinical guidelines, e.g, [8], [9], [10]. It has been shown that it takes time for the domain expert to learn the underlying structure [8]. In some cases, the knowledge acquisition phase has been facilitated by developing simplified graphical user interfaces (GUI) on top of the original GUI such as in the case of using a MediaWiki for modeling Asbru plans [10]. Another approach to address this issue is described by Seyfang and coworkers [4]. They developed a semi-structured format for simplifying the process on top of Asbru. In another project, a UML activity flow diagram was used when PROforma showed limitations in modeling decision flows in a graphical structure in an easy way [9].

Besides these languages, which are primarily developed for knowledge engineering purposes, we are not familiar with any generic knowledge modeling system that also allows the domain expert tailor the interaction with the end user application in a similar way as the system in focus for our evaluation study.

1.2 ACKTUS and the Application Domains in Focus for the Study

ACKTUS is being developed for the purpose of providing a modeling environment for domain experts who also may be potential end users of knowledge-based systems in the health domain. Another purpose of the system is managing knowledge integrated into knowledge-based systems. One goal is to design the system in such a way that end users are able to contribute to the knowledge in interaction with the system. Therefore, it is not expected that the end users or the domain experts participating in the knowledge modeling are familiar with formal models such as logic or formal ontologies. The task-network modeling languages described in the previous section are developed for modeling clinical practice guidelines and treatment protocols, mainly to support a care process within a care providing organization or between organizations. By contrast, in the current application projects, ACKTUS is used for synthesizing several knowledge sources into a support system in domains where the domain knowledge is incomplete and evolving. Another difference from the task-network languages is that the support systems developed using ACKTUS are not necessarily aimed at health care professionals. The end users in the current application projects are, besides health professionals, mining and construction workers and older persons in their homes. Therefore, ACKTUS also incorporates knowledge such as rules-of-thumb and best practice knowledge to be integrated apart from the knowledge expressed in clinical practice guidelines and other evidence-based medical knowledge sources. Preference orders among knowledge sources, level of expertise in knowledge domains and other contextual factors are taken into account in the reasoning about e.g., diagnosis in a patient case, or about adjustments in a work environment for preventing injuries.

The Semantic Web-application ACKTUS consists of a service-oriented architecture and integrates a formal terminology and information model in the form of an RDF/OWL ontology. The ontology captures 1) components used for tailoring interaction with the resulting knowledge applications, 2) components of argument-based reasoning and 3) components for modeling the user agents as actors in a reasoning process [2]. The ontology components are used by the domain experts for designing the content and interaction flows, and they put constraints on the modeling activity in order to support the domain expert in the modeling tasks. The outcome is a set of instantiated components and their relations, which function as structured tools for the end user in the interaction with the end user application. In addition to the end user application the ACKTUS architecture allows a developer, with sufficient permissions, to write applications that can access and manipulate the information using readily available Internet technologies like HTTP.

The key functionalities of ACKTUS are the following (Figure 1):

1. Integrate and model *knowledge sources*

2. Translate the content of knowledge sources into semi-formal *argumentation schemes*
3. Implement the schemes as formal *rules*
4. Define *reasoning contexts* to be activated by interaction protocols
5. Define and model *interaction protocols*
6. Define and model *assessment protocols* as sets of interaction protocols
7. Create rules for assessment flows between assessment protocols and interaction protocols
8. Apply and evaluate the composed system through end user system interfaces and with fictive patient or user cases

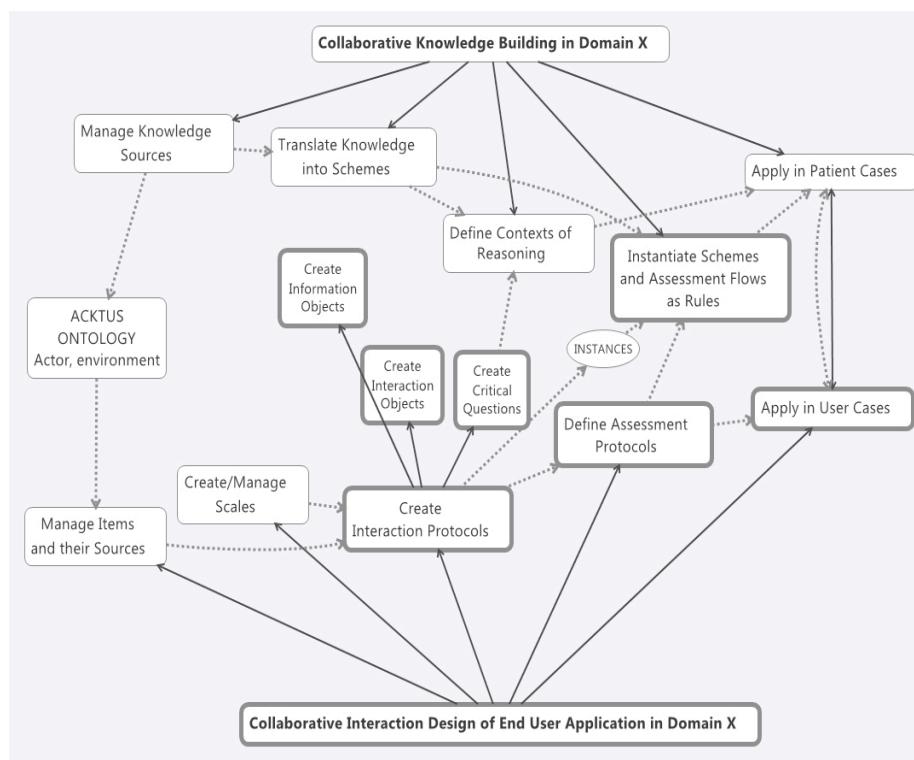


Fig. 1. Overview of ACKTUS functionality in terms of major tasks and their subtasks (indicated by arrows). Dotted arrows represent a flow of activity. Tasks focused in our evaluation is marked.

The different functionalities are used to different extent in the application domains, due to their different characteristics. In the dementia domain, where care provision is highly dependent on evidence-based clinical practice guidelines, the tasks 1-3 are crucial and the validation of the work needs to be thoroughly done [11]. Since diagnostic reasoning is performed differently by novice and expert physicians [12] (e.g., in diagnosing dementia diseases) the notion of *reasoning context* is used (Task 4). The reasoning contexts are defined by the domain expert, and frame the topics to

investigate at different levels of activity, and may be activated following the individual end user's need for support. The end user is a health professional responsible for diagnosing patients with a suspected dementia disease.

In the domain of occupational health and environmental medicine, less evidence-based knowledge is available, making the tasks 1-3 less focused. Furthermore, a large portion of the information is embedded in the work environment and the motivations for using parts of the application are rooted in the end user's current life situation. Here a larger emphasis is put on the interaction design and flow for the purpose of generating interest and motivation in the end user for taking work-related health issues seriously (Tasks 5-8). The same applies to the occupational therapy domain, where the therapist designs (tailors) the interaction to be accomplished by the client subjected to rehabilitation assessment and intervention [13]. The end users of these applications are private persons, with different life situations motivating the use of a tailored support system. These applications may be supplementary and may share knowledge between agents, such as in the case when the occupational therapist's client is a person developing, or is already diagnosed with an early stage of dementia, or is a person living with a person suffering from dementia.

In this evaluation study we focus on the tasks 5-8 since these are about equally important to the participants, who work in different application projects in our study during the evaluation period. The study then supplements an earlier study focusing the modeling of clinical guidelines in the dementia domain using the tasks 1-3 and 8 [3].

2 Methods

A qualitative and formative user-centered evaluation of the end-user graphical user interface of the ACKTUS system was performed using a general action research methodology. The evaluation was performed in two phases, representing two iterations of the development of the system as a whole, including the user interface design. In the first phase the methods activity analysis and cognitive walkthrough were used to form a baseline for user observations and interviews. A re-design was performed before the second phase based on the results of the first phase. During the second phase further evaluations were made using observation of and interviews with users. The methods used and the procedure is described in more detail in this section.

The evaluation study followed an action research methodology, which is a collective and participatory research method to achieve change or learning. In accordance with action research methodology, the project was carried out in cycles of planning, acting, observing the results in a natural environment and reflecting upon them, all the while involving those responsible for and affected by the practice to be changed [14].

The system and its context were analyzed using activity theory as base for analysis [14], generating a structured view of the workflow and the actions possible to perform in the system. Tasks were chosen from this set of actions for the users to perform in the user evaluation sessions. The chosen actions did not cover all aspects of the user interface, but instead those that were most relevant to the users in their ongoing modeling work using the system. The results from the user observations were then

analyzed using an activity assessment protocol adapted from Lindgren [5], order to identify where in the interaction problems arose and to what degree the user managed to perform independent work with the system. In the following subsections the activity analysis and the user evaluation study are further described.

2.1 Activity Analysis

The options and available actions of the system were analyzed using activity theoretical models. The work that can be performed by a user was divided into two overarching activities on an abstract level, which are carried out by actions of different levels of concreteness. The two main activities to be executed in interaction with the system were identified as: 1) *Collaborative knowledge building in domain X*, 2) *Collaborative interaction design of end user application in domain X* (Figure 1). Firstly, the activity of collaborative knowledge building describes how medical health professionals input knowledge into the ACKTUS system interpreted into different levels of complexity. Secondly, how the knowledge is presented and interactively mediated to an end-user is modeled in the collaborative interaction design activity.

These activities are achieved through a number of possible actions using the ACKTUS system as a mediating tool. For instance, knowledge building is achieved by defining knowledge sources and structuring the knowledge into schemes and rules, while the interaction design includes actions such as managing scales, questions and assessment protocols.

Activity is performed in communities of actors; in this case, the collaborative nature of building knowledge and structuring the interaction flow for utilizing that knowledge. A single user will rarely build the knowledge from scratch, as they will typically work in relation to others' ongoing and finished work already entered into the system. The interaction will also develop over time as the knowledge in the system and conditions develop. For instance, the interaction will look different for the first users who enter the first pieces of knowledge than for later users. From this set of activities and actions, a task analysis was developed, listing all options and actions that the interface makes available, from using buttons and controls to e.g. reviewing information that is presented. This task analysis was expanded upon for use in the cognitive walkthrough segment of the evaluation.

AAIMA Protocol for Assessment. A modified version of an instrument for assessing ability in performing computer-supported activity (protocol for Assessment of Autonomy in Internet-Mediated Activity - AAIMA) was used [5] (Table 1). The activity analysis was adapted into a protocol based on this model for use in the observations of users in determining what level of autonomy and complexity a user achieves in performing an action with the system as a tool. Each activity and action is described with three levels of complexity according to the model [5].

For actions, the levels are organized as follows. On the lowest, most basic level (denoted Level 0), the user must understand the purpose of the task and be able to find the section of the interface where the task may be performed. Without an understanding of the purpose of the action, it cannot be executed in a goal-driven manner, which means it could not be considered an action (since actions are goal-based). That is, if a user were to randomly click in the interface and happen to

perform the goals of the action by chance, he is still not described as having performed or accomplished the action. On a higher level, Level 1, the user is able to edit information that is already entered into the system with an understanding of the object's components (i.e. that a rule needs premises and a conclusion to function).

Table 1. AAIMA-protocol [5] adapted to the study. We omit here the columns showing needs and motives and limit the table to showing the main activity and four sub-actions in focus.

Act. level	Activity description
Activity	Collaborative Interaction Design of End User Application in Domain X
Level 0	Understanding the purpose of the interaction design but does not participate
Level 1	Contributing by defining basic units of information
Level 2	Contributing by designing assessment flows
Level 3	Structuring interaction flow and framework
Action	Model Interaction Protocols
Level 0	Finding interaction protocols, understanding their relation to items and scales, understanding the difference between information and interaction objects and critical questions
Level 1	Editing existing interaction protocols, reviewing definitions
Level 2	Creating new interaction protocols with appropriate scales or questions
Level 3	Full understanding of the role of interaction protocols in structuring interaction flows, reviewing relevance of scales, etc.
Action	Model Assessment Protocols
Level 0	Navigating to assessment protocols and understanding their structure and purpose
Level 1	Editing existing assessment protocols, defining concepts
Level 2	Creating new assessment protocols
Level 3	Reviewing an assessment protocol's structure by testing it
Action	Instantiate Assessment flows as Rules
Level 0	Finding rules, understanding scheme structure and how their content could be fully expressed in rules
Level 1	Editing existing rules and understanding the purpose of its individual components (premises, conclusion, critical questions)
Level 2	Creating new rules, specifying their components and relation to a scheme
Level 3	Understanding the function of rules and their role in reasoning, evaluating whether a set of rules encompass all the content of a scheme
Action	Apply in User Cases
Level 0	Navigating to user cases, understanding their relation to assessment protocols and rules
Level 1	Understanding the function of characteristics such as role, motive etc.
Level 2	Following assessment flows in a purposeful manner according to a specific user case
Level 3	Able to effectively utilize user cases to evaluate the validity of the interaction flows of the system, assessing relevance of assessment protocols and questions to a certain type of user

On Level 2, the user is also able to add new knowledge into the system, with an understanding of the knowledge's interaction and relation to other pieces of information in the system (i.e. that an *assessment protocol* contains a set of *interaction protocols* or other assessment protocols). For the third and highest level, a complete understanding of the action and its context is required, and being able to review and analyze the relevance and validity of the knowledge that exists in the system.

If a level of any task causes problems (*breakdowns*), but the user manages to complete them under the guidance of a more experienced peer, the task is said to lie within the user's zone of proximal development (ZPD) [16]. This means the task has the possibility of being performed independently by the user within the near future through training and development. The notion of *breakdown* is used to describe the situation when the user is interrupted in completing the task, possibly due to limitations in the user interface design or the individual's understanding of the task.

2.2 User Evaluations

User evaluation sessions were carried out with five domain expert users in total. Two were experts in the field of environmental health, a domain in which the ACKTUS system is currently being applied (ACKTUS-EKF). These were chosen as they are a part of the project and have a say in its development, and represent the user group consisting of the medical professionals that will enter and structure the system's knowledge. They are also specialized in the field of the other user group, the end-user construction and mining workers. In addition, they are specialized in two supplementary sub-fields of environmental health, with a corresponding division of responsibility in the knowledge modeling process. A third medical professional, involved in the development of the dementia diagnosis subsystem (ACKTUS-Dementia), was chosen for the same reasons and participated in the second phase of the evaluation. This user had participated earlier by modeling clinical guidelines and rules corresponding to the tasks 1-3 described in previous section [3]. Two additional users, experts in the occupational therapy (rehabilitation) domain, participated. In their cases the evaluation was part of an introductory demonstration for the purpose of initiating a process of knowledge and interaction modeling using the subsystem ACKTUS-Rehab.

The users were asked to either continue tasks that they were to perform in the system in the course of their work, or perform certain predetermined tasks if they currently had nothing to work on. The areas covered by their work and the tasks were decided from what was determined to be most relevant to their work in the system at the time and included 1) managing *interaction protocols*, 2) managing *assessment protocols* and 3) creating/managing *assessment flows* using *rules*. The users were also asked questions about their experience with the interface, what was problematic or what they would like to change, etc. The users who tested the Rehab subsystem were given example tasks to finish that demonstrated how the system could be applied to their domain. Following this, they continued to model their own chosen examples with the same purpose, having gained understanding of the purpose of the system. The same procedure was done when the dementia expert was introduced to the tasks.

While the three medical professionals had experience with parts of the system, they were not familiar with all the tasks. The ACKTUS-EKF users were at a beginner level in regards to managing rules, and the ACKTUS-Dementia user had not attempted to define assessment protocols or use rules to define assessment flows before.

During the first phase of evaluation, the parts of the interface in focus for evaluation were in an initial, basic stage of development. Based on the users' performance and the other methods of evaluation, design suggestions were generated and implemented. Significant updates were made to the interfaces of interaction and assessment protocols in particular, and some updates were made to the interface for creating rules. These updated versions of the interface were used by the participants in the second phase, generating further design alterations to be implemented in the future.

The users' interactions with the system and comments during the session were recorded on video. A total of 12 hours of active modeling work was recorded part from recorded interviews. The actions mediated by the interface (clicks, typing, scrolling etc.) were then transcribed and analyzed using the aforementioned activity protocol. The transcription was divided into segments according to the actions of the protocol that the user was trying to achieve at a given time. Each action was then scored for each user by what level of complexity and with what degree of autonomy that an action was completed. Comments and suggestions were also noted and assembled as design suggestions.

3 Results

The results from the evaluation study involving domain expert users are further described in the following subsections with a focus on learning and developing understanding of the system detected and interpreted using the AAIMA instrument. The outcome of using ACKTUS in terms of developed end-user applications is described and aspects concerning redesign are motivated.

3.1 Using the AAIMA Protocol for Detecting Learning and Skill Development in the Use of ACKTUS

A progress in ability to use ACKTUS for their modeling purposes was seen in the users participating in the study. The results from the AAIMA protocol analysis are summarized in Table 2, which provides an overview of the results. In the table each action and activity is marked with the highest level (a number between 0-3) that the user managed to achieve over the course of 1-3 evaluation sessions during each phase. Additionally, for each activity level, how automatized each action was performed (i.e. to what extent the action caused breakdowns) was also noted, but this is excluded from the table for brevity's sake. The ZPD notation in the table designates that the user was able to complete the action at the given level of complexity, but only with some aid or guidance from the evaluator, indicating that the activity may be performed independently within the nearest future (i.e., the next session). The

differences in performance between the two evaluation phases can be seen in Table 2. In the following description we use the term *participants* to distinguish the professionals participating in the evaluation study from the future end users of the resulting applications.

Table 2. Overview of results from analyses using the activity protocol in Phase 1 and 2. U3 and U5 did not participate in the first phase and U5 was new to the system in Phase 2.

Activity/action	U1 PH	U2 PH	U3 PH	U4 OT	U5 OT
<i>PHASE 1</i>	<i><60 min</i>	<i><60 min</i>		<i><60 min</i>	
Collaborative Interaction Design of End User Application in Domain X	2	2 (ZPD)	N/A	2 (ZPD)	N/A
Manage Interaction Protocols	2	2	N/A	2 (ZPD)	N/A
Manage Assessment Protocols	3	3 (ZPD)	N/A	3 (ZPD)	N/A
Instantiating Assessment flows as Rules	1	0	N/A	N/A	N/A
Apply in User Cases	N/A	N/A	N/A	N/A	N/A
<i>PHASE 2</i>	<i>102+150min</i>	<i>60+63 min</i>	<i>45+44 min</i>	<i>41 min</i>	<i>40 min</i>
Collaborative Interaction Design of End User Application in Domain X	3 (ZPD)	3 (ZPD)	3	3 (ZPD)	3 (ZPD)
Manage Interaction Protocols	3	2	3	3	2
Manage Assessment Protocols	3	3	3	3 (ZPD)	3
Instantiating Assessment flows as Rules	2 (ZPD)	2 (ZPD)	3	3 (ZPD)	2 (ZPD)
Apply in User Cases	3	3	3	3	2

Phase 1. In the first phase of observations, none of the participants reached Level 3 of the overall activity, that is, structuring an interaction flow and framework, since they did not manage to construct rules for instantiating assessment flows. This was also technically limited at the time of evaluation. They did reach activity Level 2, in that they were able to organize components of knowledge and interaction components in the system to create assessment protocols. As such, what was limiting the interaction at this stage seems to have been the understanding of and application of rules. The only action that reached Level 3 in the first phase, full understanding and reviewing contents, was *managing assessment protocols* (Figure 2). Level 3 caused problems for one user however, mostly concerning understanding what objects would appear when testing the protocol and how to choose which would appear (i.e. adding questions to the protocols). To summarize the first phase, it seems that the problems for the users were in the understanding of certain actions more than their execution. That is to say,

the users were able to edit and create new objects (critical questions etc.) but had trouble with distinguishing different kinds of objects, or understanding the objects' role and relationships in interaction flows (e.g. when an interaction object would come into play). Obtaining a full impression or overview of how things relate and function appeared to be problematic at this stage. Therefore, the system's presentation of these aspects was worked on before Phase 2, for instance, in the labeling of controls and objects, titles, explanations of concepts and natural structure, progression and mappings, etc. (some results are shown in Figures 3 and 4).

Screeningformulär

Namn
Vibration

Beskrivande text

Extra information: Ingen information vald

Uppdatera

- Dina besvär - Screeningformulär
 - o Vita fingrar - Screeningformulär
 - Har du upplevt att dina fingrar blivit vita på detta sätt i samband med kyla eller fukt?
 - o Andra symptom i fingrar - Screeningformulär
 - Har du upplevt domningar, pinnningar, nedsatt känsel för temperatur eller beröring i dina fingrar?
- Dina maskiner - Screeningformulär
 - o Hur länge har du arbetat med vibrerande maskiner?
 - o Vilka är de tre vanligaste maskinerna du använder för tillfället?
 - o Hur lång tid i genomsnitt använder du maskinen/maskinerna?
 - o Riskberäkning - Screeningformulär
 - Beräkningar utifrån medelvärden för de maskiner du använder nu visar att din vibrationsbelastning
- Risker med ditt arbete - Screeningformulär
 - o Vill du veta hur mycket dina maskiner vibrerar i jämförelse med andra maskiner?
 - o Är du fullt frisk?
 - o Allmän information - Screeningformulär
 - Var arbetar du?
 - Vilken är din ålder?
 - Är du?
 - o Vanor - Screeningformulär
 - Röker du?
 - Snusar du?

Du kan byta ordning på frågorna genom att dra dom i den ordning du vill ha dom.

Ta bort de valda

Allmän information

Testa överblick

Ingen vald Testa steg för steg

Fig. 2. Initial version of the assessment protocol, used in Phase 1.

Redesign of ACKTUS. The different evaluation methods yielded different design issues to be addressed. The results from the first phase of the user observations and activity analysis indicated that obtaining a full impression or overview of how objects relate and function in the system is problematic. The system's presentation of these aspects needed to be addressed, for instance in the labeling of controls and objects, using clear titles, explaining concepts, using a natural structure and suggesting the next step in a workflow, reducing information load to facilitate user decisions, etc. The results also showed that the rule editor requires specific attention (Figure 5). In addition, the cognitive walkthrough, performed concurrently with the first phase of user observations, identified a number of recurring issues, chiefly concerning visibility, recognizability and understandability of controls and objects in the interface.

In Figure 2 an assessment protocol is shown as an example of an object with the previous version of the interface. Other objects followed the same color theme and design. In the case of assessment protocol, the interface did give some hints as to the structure of the content, by displaying the contents of a sub-protocol with indentation. However, in general there were many types of controls available, unsorted by importance or suggested workflow. This was true for other objects in the system as well, and may have contributed to the lower scores in the AAIMA assessments in the first phase of evaluations.



Fig. 3. Assessment protocols in normal and edit view created by User 3 in Phase 2.

These issues were addressed in the re-design following two overarching themes, namely providing objects in the interface with unique, recognizable and recurring appearances, and hiding certain controls and information when they are not needed. Objects (such as rules, critical questions, assessment protocols etc.) were given unique appearances that recur throughout the interface wherever said objects are used or referenced (Figures 3, 4). The workflow in the system was improved by removing unnecessary steps and to provide more feedback at certain points in the interaction. The connection between controls and their affected components and between related objects were strengthened, and interface improvements were done to support a natural workflow, weighing content and information visually according to its relative importance to achieving the task at hand.

Some of the changes implemented for the second phase of evaluations are exemplified in Figure 3 and Figure 4. Two participants increased their performance in the managing of interaction protocols to Level 3, indicating a higher understanding of

the role of the objects in question in the system as a whole. Also, novice users reached high levels of performance and understanding even from only a short introductory session. Changes such as the ones displayed in Figure 3 and 4 may have contributed to this increased understanding. For one, each object now has a separate color theme, which should help the user to identify and separate objects more quickly.

Another change that may have contributed to the users' increased understanding is the edit mode (Figure 3). Hiding editing controls when first viewing an object allows the user to focus on getting an overview and understanding its content before deciding what to do with the object.

Consistent use of colors, along with icons for each object, should help the user get a sense of "what goes where" in the system. Additionally, the spine to the left in the assessment protocol is meant to further emphasize the structure of it, again to assist the user in recognizing the relations between different kinds of objects (Figure 4, 5).

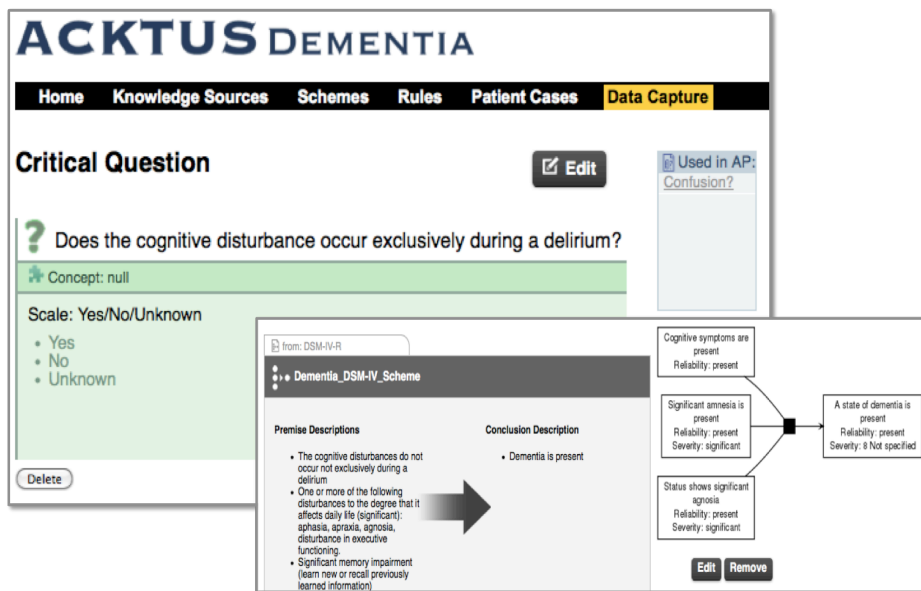


Fig. 4. Critical question (interaction protocol) view, and the scheme and rule view in Phase 2.

Phase 2. In the second phase of user observations, the participants did reach Level 3 of the activity, some needing occasional aid from the evaluators (i.e. the activity lay within their ZPD), since they were also able to create rules within the system (Table 2). However, their proper understanding of the rules and their application was limited, and the concepts, terms and workflow used in the rule creation process still caused breakdowns, though these were handled by the individual in most cases. Still, a progress was seen in, e.g., User 1. In Phase 2 this user was able actually to complete rules that govern question presentation and order, with some guidance from the evaluator. This was not accomplished in the first phase of evaluation, when he only managed to edit parts of existing rules but not create his own rules.

The possibility to apply and evaluate the results in user cases was not included in the first phase of the evaluation study. In the second phase the participants were able to evaluate the content they had entered into assessment protocols by viewing it in a prototype of the end user interface, and from there enter responses in the manner of specific types of users, in order to see whether these fictive users would be shown the dedicated tailored content (e.g., Figure 5). This task generally caused no problems for the participants; indeed, a few of them even chose to view the content as the end user would see it, even when editing basic content.

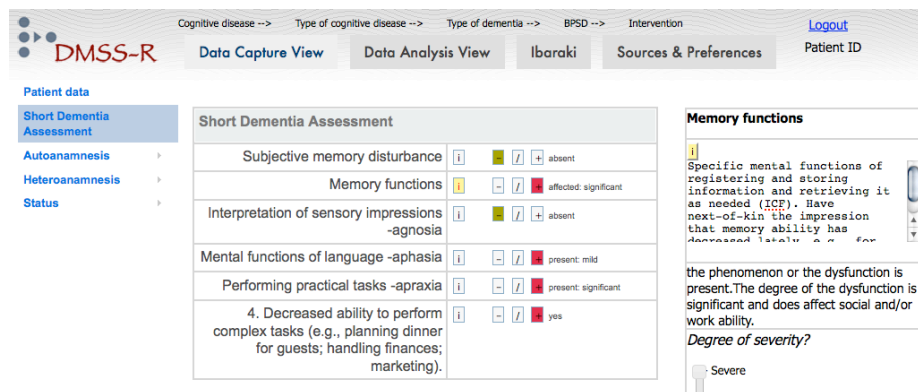


Fig. 5. DMSS end user prototype visualizing the short dementia assessment protocol.

3.2 Outcome of the Evaluation Sessions in Terms of Developed Applications

During the evaluation sessions, the participants continued with tasks related to their own work and were able to complete tasks yielding authentic results that will be usable in practice in the future.

For instance, in Phase 2 User 1 worked with two assessment protocols that will be used for assessing environmental health risk factors, concerning vibration specifically, for end users of the EKF subsystem. User 1 was able to review an extensive amount of questions, correcting mistakes such as spelling errors, reviewing the order the questions were presented in when answering a specific way, and editing these factors when necessary. For presentation purposes, it was also necessary for the user to create sub-assessment protocols with which to group the questions. All in all, an assessment protocol that had been developed was fine-tuned and made presentable during the participant's evaluation session. Another assessment protocol was reviewed and adjusted, giving information about what remains to be done. What caused the user the most pause was managing the rules used for deciding question order and presentation conditions.

In the case of User 3, this physician had been involved in developing a knowledge base for dementia integrated in a stand-alone prototype, subjected to evaluations with medical professionals in Asia. Following their request of a quick assessment protocol that would fit their workload and resources (preferably to be finished within three

minutes, which is, however, in practice impossible), he modeled during a single session a minimal assessment protocol for diagnosing dementia by reusing components integrated in the full assessment protocol. He was able to redesign the assessment flow with some assistance from a knowledge engineer to follow the associated clinical guideline using rules, and modeled critical questions to appear in the user interface for guiding the end user. He also combined the outcome with an additional short protocol for assessing Alzheimer's disease. This was accomplished within 30 minutes. This 3-minute protocol was seen as a guided entrance into the full protocol in the complex patient cases, and as such it was seen as a major contribution to the development and tailoring of the dementia support system to different user needs (see Figures 3-5).

In the case of User 4 and User 5, the setting was an initiation of a collaborative project where User 5 modeled content she will have use of in an end user situation involving older adults. User 4 functions as the leader of their common project and had a larger perspective on the use of ACKTUS in their setting, leading to her to envision a diagnostic tool that integrates research-based knowledge sources as well. Following her idea, she modeled general knowledge for this purpose, supplementary to the knowledge modeled by User 4. The modeled content was anticipated to be used by older adults in a dialogue system, where parts could be re-used and possibly adapted by other professionals in the domain.

An example of the interaction designs produced by the participants during the evaluation period is shown in Figure 5. To summarize, the results of the modeling done by the participants are accustomed, personalized interaction flows adapted to given end users' needs and requests. For instance, in the case of DMSS (Figure 5), the minimal assessment protocol for diagnosing dementia gives the end user the option of using this protocol to quickly assess a patient's mental state and only continue to the larger, more detailed protocol if they decide that this is necessary. In the case of EKF, the questions and information shown to the end user is decided by their previous responses to questions, according to the rules defined by the participants. At the end of the dialogue, the end-user receives personalized advice based on their answers.

4 Discussion

Defining assessment protocols generally caused little problems for the users in Phase 2. While the users did reach Level 3 in both phases, User 2 and User 4 faced breakdown situations during phase 1 and the activity lay within their ZPD. This was not the case in the second phase. One reason for this increase in skills could be the major revisions in the interface that underwent between the two evaluation phases (see Figures 3-5). This is also the case for the task of creating interaction protocols, which had received improved functionality and presentation since the first phase, most likely reflected in the increased performance for the users 1, 2 and 4.

One notable change is the increased performance in the task of creating rules. User 1 and User 2 increased their score from barely being able to edit rules or not wanting to participate, to being able to create new rules with some guidance in phase 2. The beginner users (User 4 and User 5) also scored high in the second phase. However,

since only parts of the intended interface changes for managing rules had been implemented at the time of the second phase of evaluation due to time constraints, this change likely has other causes as well, such as an increase in motivation or learning. The EKF users may for instance have gained insight into potential applications of rules and therefore were more motivated to learn about them before and during the second phase. These two users' initial perception of the modeling tasks was that they would describe the content and a programmer then would implement the content. Their attitude changed when presented with the system and when starting to model the content, though with some delay. Another motivational factor for the two EKF users was probably the approval of, and encouragement for, the inclusion of personal advice for the end users, coming from a project reference group, which had not been approved at the time for Phase 1. They put large interest in their last sessions on modeling personal advice to appear for end users, which were to be demonstrated for the reference group. User 4, who scored 3 when creating rules, immediately saw potential applications for the rules and the system in their line of work and as such was highly motivated to test and learn.

User 4 expected a different procedure for relating the assessment protocols to the rules, making her somewhat unwillingly adjust to the procedure implemented in ACKTUS. It will be investigated whether this alternative procedure should be supported by the system in the future.

Since a convenience sample was used for selecting the participants (i.e., all professionals involved in modeling tasks during the evaluation period) and due to the fact that the sample contains only five participants, the results are only representative for the group. However, the results indicate that it is possible to learn and use the re-designed system after only a short introduction with the correct motivation grounded in a need for tailoring a knowledge system to specific users or user groups.

Adapting and applying the AAIMA protocol yielded a structured baseline for evaluation and for a continuing evaluation during the ongoing development of ACKTUS and the different end user applications. It also gave insight into a user's comprehension and capability in using the system by incorporating elements of activity theory, such as ZPD. The results from using the protocol to score user performance provided a basis for a structured comparison both between different users and between different sessions with the same user. From this, the development and learning of a participant using the system can be extrapolated, providing insight into areas of concern in the interface and how well the system supports learning. The protocol did not eliminate the subjectivity in determining whether, for instance, a user understands the relationship between different kinds of objects, but as long as the activity levels are interpreted and applied consistently according to the same criteria, the subjectivity is reduced. A way to further investigate its inter-rater reliability would be to have several evaluators apply it to the same set of data and compare their results. The protocol was originally used to assess the skills in computer-mediated activity in elderly users [5], and has in this study been used with participants from the medical professional domain. In this study it was shown that the protocol is useful for a significantly different domain, provided the protocol is adapted based on an activity analysis of the application domain. This adaptation need may prevent the application of the protocol in cases where there are limited resources for assessing the use context. Furthermore, the protocol may show less applicability also in the cases where

an actual use situation with dedicated end users cannot be arranged for evaluation. The major contribution with the AAIMA protocol is that it supplements other methods in that it supports a structured theory-based interpretation of the data that can be collected e.g., in a study of focus shifts in activity execution [17]. In addition, it allows for detecting changes in knowledge and behavior over a longer period of time.

5 Conclusions

A pilot evaluation study was conducted of ACKTUS, an early prototype for modeling knowledge and interaction design of knowledge-based support systems to be used in the health domain. The purpose was to investigate how well the system supports knowledge modeling and interaction design for domain experts in authentic use situations, as well as their development of a conceptual understanding of the system. Consequently, we focused on users with minor or no experience from using the application. Five health-care professionals with expertise in four different domains were observed using the system in 1-3 sessions each, and were interviewed. All participants were working on particular applications relevant for their domain as part of three different research projects. The concept of “zone of proximal development” (ZPD) was used to assess the development of skills, following the AAIMA assessment protocol [5] adapted to the result from an activity analysis done of the activity in focus for evaluation. The protocol proved to be a valuable instrument for detecting and describing the development of the users’ skills and understanding of the system. This evaluation also demonstrates the applicability of the protocol to a new user group and a new type of system than it was originally designed for.

Between the phases the ACKTUS user interface was re-designed based on the results from the initial phase. The results show that the users who participated in both phases increased their understanding and skills, and that the two new users who entered in Phase 2 were able to reach the same level of understanding as the three who had used the system in Phase 1. In Phase 2 they were able to model components and interaction design and had gained a sufficient overall understanding of the system. The re-design of the user interface was interpreted as the major reason, and the fact that a certain amount of learning took place between sessions was interpreted as a minor factor. A more important factor was the motivational level in the users. These results illustrate the development of understanding, and that the ACKTUS application does not require extensive training and learning to use, in contrast to other approaches for integrating medical knowledge presented in the literature.

The results are fed into further re-design of the user interface, with the aim at providing novice users of the ACKTUS modeling environment a transparent and intuitive interaction design. When this is accomplished, also end users may contribute with knowledge as part of a community of users sharing a common living knowledge repository. The AAIMA protocol will be used as baseline in future evaluation studies, enabling the detection of development of skills and understanding. The protocol is also being integrated into ACKTUS for developing adaptive functionality using automated methods for detection of activity and skills.

Acknowledgments. The authors would like to thank the professionals participating in studies and development work. The project is partly funded by VINNOVA (Sweden's innovation agency) and Emil and Wera Cornell foundation.

References

1. Kaplan, B.: Evaluating informatics application – clinical decision support systems literature review. *Int. J. Med. Inform.* 64, 15--37 (2001)
2. Lindgren, H., Winnberg, P.: A Model for Interaction Design of Personalised Knowledge Systems in the Health Domain. In: Grasso, F., Paris, C. (Eds) *proc. 5th International workshop on Personalisation for e-Health 2010*. Casablanca, Morocco (2010)
3. Lindgren, H., Winnberg, P.: Evaluation of a Semantic Web Application for Collaborative Knowledge Building in the Dementia Domain. In *proc. E-Health 2010*. Casablanca, Morocco (2010)
4. Seyfang, A., Miksch, S., Marcos, M., Wittenberg, J., Polo-Conde, C., Rosenbrand K.: Bridging the Gap between Informal and Formal Guideline Representations. In: Brewka, G., Coradeschi, S., Perini, A., Traverso, P. (eds.) *ECAI 2006. Frontiers in Artificial Intelligence and Applications*, vol. 141, pp. 447--451. IOS Press (2006)
5. Lindgren, H.: Personalisation of Internet-Mediated Activity Support Systems in the Rehabilitation of Older Adults – A Pilot Study. In *Proc. Personalisation for e-Health 2009*. pp. 20--27 (2009)
6. Peleg, M., Tu, S., Bury, J., Ciccarese, P., Fox, J., Greenes, R.A., Hall, R., Johnson, P.D., Jones, N., Kumar, A., Miksch, S., Quaglini, S., Seyfang, A., Shortliffe, E.H., Stefanelli, M.: Comparing computer-interpretable guideline models: a case-study approach. *J Am Med Inform Assoc.* 10, 52--68 (2003) doi: 10.1197/jamia.M1135.
7. Isern, D., Moreno, A.: Computer-based execution of clinical guidelines: A review. *Int. J. Med. Inform.* 77(12), 787--808 (2008)
8. Shalom, E., Shahar, Y., Maimon, M.: Can physicians structure clinical guidelines? Experiments with a mark-up-process methodology. In: *Knowledge Management for Health Care Procedures*, LNAI, vol. 5626, pp. 67--80. Springer, Heidelberg (2009)
9. Sutton, D.R., Taylor, P., Earle, K.: Evaluation of PROforma as a language for implementing medical guidelines in a practical context. *BMC Med. Inform. Decis. Mak.* 6, 20 (2006)
10. Eccher, C., Seyfang, A., Ferro, A., Stankevich, S., Miksch, S.: Bridging an Asbru Protocol to an Existing Electronic Patient Record. In: Riano, D. et al (Eds.) *K4CARE 2009*. LNAI vol. 5943, pp. 14--25, Springer, Heidelberg (2010)
11. Lindgren, H.: Towards personalised decision support in the dementia domain based on clinical practice guidelines. *UMUAI*, DOI: 10.1007/s11257-010-9090-4.
12. Patel, V., Kaufman, D., Arocha, J.: Emerging paradigms of cognition in medical decision-making. *Journal of Biomedical Informatics* 35, 52--75 (2002).
13. Lindgren, H., Nilsson, I. Designing Systems for Health Promotion and Autonomy in Older Adults. In: *Interact 2009*. LNCS vol. 5727, pp. 700--703. Springer, Heidelberg (2009)
14. Davidson, E. and Heslinga, D.: Bridging the IT Adoption Gap for Small Physician Practices: An Action Research Study on Electronic Health Records. *Information Systems Management*, 24 (1), 15--28 (2007)
15. Kaptelinin, V., Nardi, B.A.: *Acting With Technology: Activity Theory and Interaction Design*. MIT Press, Cambridge, MA (2006)
16. Vygotsky, L.: *Mind in Society*. Harvard University Press, London (1978)
17. Bodker, S.: Applying activity theory to video analysis: How to make sense of video data in HCI. In: B. Nardi (Ed.) *Context and Consciousness. Activity theory and human computer interaction*, pp. 147--174. MIT Press, Cambridge, MA (1996)