# Evaluating virtual agents for e-commerce

**Fabio Abbattista[1], Verner Andersen[2], Hans H.K. Andersen[2], Pasquale Lops[1], Giovanni Semeraro[1]**

**[1] Dipartimento di Informatica, Universitá di Bari, Italy**
***e-mail: fabio@di.uniba.it; lops@ di.uniba.it; semeraro@ di.uniba.it***
**[2]Risoe National Laboratory, Denmark**
***e-mail: verner.andersen@risoe.dk; hans.andersen@risoe.dk***

## Introduction

The rapid evolution of interactive Internet services has led to both a constantly increasing number of modern Web sites, and to an increase in their functionality, which, in turn, makes them more complicated to use [1]. Thus, any attempt to enhance the consumer-supplier relationship in e-commerce has to meet the challenge of coping with two almost contradictory goals: A useful e-commerce application should not only mimic traditional catalogues, order forms and other printed material which used to be the basis of communication between consumers and suppliers. Instead, the inherent potential for interactive data processing and man-machine dialogue should be used by e-commerce applications to meet the user's need for immediate situation-specific response, instantly available problem-specific advice, and better ways to access and inspect the supplier's offer. However, the currently prevailing graphical user interfaces, which rely on menu selection and navigation, require a considerable cognitive overhead. This may be tolerable to frequent users, but will in many cases deter casual users, especially those who are not yet used to computers. Hence, we need to combine the usefulness of a value-added service with a high degree of usability, and dedicated measures to build up trust and confidence in inexperienced users [2].
To meet these conditions the interaction must be, at the same time, as natural as possible, thus enabling users to rely on their communicative skills, it must convey precise and relevant information, and address the personal background of the individual user. The interface must use best practice solutions to achieve a high degree of dialogue intelligence, and employ an appropriate graphical design.
The solution we propose for project COGITO[1] is based on "intelligent personalized agents" which represent virtual assistants or advisors (also visually) by modelling their ability to support customers. There are many possible applications for such virtual assistants. They could instruct customers in the use of a Web site, point out new offers, help sift through products, and other support. There have already been some efforts made in developing chat robots ("chatterbots") based on expert systems.

## Personalized intelligent agents for electronic commerce

In the e-commerce area, applications today offer a wide range of functions. Advanced Web sites use state-of-the-art technology in order to satisfy their customers. Considering the vast selection of offerings, it is of particular interest to give the customers personal advice which reflects their *individual needs* and *interests*. Therefore, advanced features are provided to tailor the presentation of the Web site to the customer's personal tastes, e.g., by re-arranging the shelves in order to match what the customer reveals about her likes and dislikes. Nevertheless, the current interface technology provides only limited solutions for specific e-commerce related requirements. For instance, the interface should support affective marketing communication, allow for interactive guided tours, and provide dynamic FAQs. In order to enhance personal assistance, e.g., in cross- and up-selling, we need to include *pro-active* features, and we have to build and exploit *user profiles*. Thus, personal recommendations, individual purchase incentives, and session-independent assistance are facilitated, establishing a "learning relationship" between customer and supplier.
Due to the inherent limitations of contemporary GUI technology, many of these goals are not sufficiently supported by traditional interface designs. On the other hand, *intelligent agents* can converse with the user in written natural language and can be regarded as an alternative interface metaphor which can be used to implement many of the advanced requirements of e-commerce in an efficient, elegant way. Natural language dialogues are easy to understand, the users may express their needs in an unrestricted form, and, in principle, it is possible to realize sophisticated system behaviour within the same way of interaction.
Such virtual assistants must be capable of flexible behaviour if they are to be acceptable to users on a long-term basis. Simple chatterbots (see http://bots.internet.com/search/s-chat.htm for an overview) only simulate

[1] *COGITO is an EU-funded project in the 5th Framework Programme, (IST-1999-13347)*

conversation without utilizing any knowledge about the individual users and their actual behaviour during online sessions. Such simple chatterbots are not powerful enough to serve as a medium for customer advice. Even those with a sophisticated repertoire of conversational skills will fail to be more than entertaining if they do not treat a user as an individual having specific needs, preferences, etc. This means that, in addition to some of the abilities already available (e.g., help question answering controlled by simple event-action rules), a further reaching dialogue management will be needed to help accomplish two major goals: First, in order to achieve an adequate, non-stereotypical repertoire of reactions, the individual dialogue situation must be interpreted; and second, dialogues that are more complex allow goal-directed strategies to be pursued (cooperative behaviour, convincing argumentation). This kind of dialogue intelligence will be based on elaborated dialogue rules allowing the system to interpret a wide variety of situations that may occur.
Whereas an increase in general dialogue intelligence can be achieved by elaborate rule sets, the naturalness of the dialogue depends on the degree of ability of the system to adapt itself to individual users, whether it is able to learn about their preferences and attitudes during the dialogue, and memorize them for later use. For this purpose, we will include learning mechanisms that extract permanent features of a given user from the dialogue (of course, the user must consent to this, and will be given an opportunity to inspect and change the data) and construct a corresponding user profile.
By user profile we mean all the information collected about a user that logs to a web site, in order to take into account her needs, wishes, and interests. A user profile, as intended within the COGITO project, is composed by two main frames: The frame of user data, that comprehends interaction data (number of searches or purchases within a category, number of connections,…) and the frame of user interests, that is the part of the profile built on the basis of supervised learning algorithms [3]. The application used for testing the COGITO agent is the virtual bookshop BOL (www.bol.de), and the preferences for the user automatically "learned" by the system concern the ten main book categories the BOL product database is subdivided into. User profiles are represented by XML files and are the key to personal recommendations because they enable the agent to customize its book recommendations to the individual user.

## The retrieval process in COGITO

On the ground of their profiles, the chatterbot offers a better support to customers during the interaction, providing personal recommendations, purchase incentives and helping users in problematic situations during the search.
This improves usability of the BOL web site, as shown by the following scenario.
A user is known by the COGITO system if she completes the BOL registration procedure. This step causes each customer to be assigned with a personal identification number. This is necessary to both recognize a user accessing to the online media shop and collect data about her preferences for generating/updating her profile.
A dialogue between the chatterbot, named Susanna, and an unknown user asking the chatterbot for a book by author "King" is undertaken:

| | |
|---|---|
| Susanna: | Good evening! Nice to see you found your way to BOL.DE. May I introduce myself? My name is Susanna and I am your personal assistant at Bertelsmann Online. What is your name? |
| User: | My name is Fabienne. |
| Susanna: | Nice to meet you, Fabienne. What can I do for you? |
| User: | I' m looking for a book. |
| Susanna: | Then this is the right place to be! We here at BOL have hundreds of thousands of deliverable books. Are you looking for a book by a specific author? |
| User: | Yes |
| Susanna: | What is the name of the author? |
| User: | King |
| Susanna: | I shall check whether we have a book by the author called King. One moment please… |

Susanna finds several books by the author "King" through a remote call (*deep linking*) to the search engine available on the BOL web site and displays them, as shown in Figure 1.


**Fig. 1. Susanna offers a long list of books belonging to several categories by authors whose last name is "King"**

It can be noticed that the books ranked first are by the author Stephen King. Books by other authors are found further down the list, which means that the user should scroll down a long list if she was not looking for a book by Stephen King. The customer not looking for a Stephen King book might now also choose to either refine the search by using an advanced search function or continue to chat with Susanna about different fields of interest.

If the user has already been chatting with Susanna, then a profile of this user is available to the system, which can exploit it to accomplish a more precise search in the product database.

Let us suppose that the profile of such a user contains the category *Computer_und_Internet* as one of the preferred categories with degree of preference 100%, and the category *Belletristik* (narrative) with a lower degree of preference (79%), and that the submitted query is again about books written by King.

The first book displayed in the page of the search results is a book about Windows 2000 written by Robert King (Figure 2).

This result is due to the fact that the original query "King" has been automatically expanded by the system in "King AND Computer & Internet" (highlighted by the circle in Figure 2), since *Computer_und_Internet* is the category with the highest degree of interest in the profile of the user.

This scenario highlights the dependence of the result set on the profile of the user that issued the query.

The following section describes the query expansion mechanism implemented in COGITO.

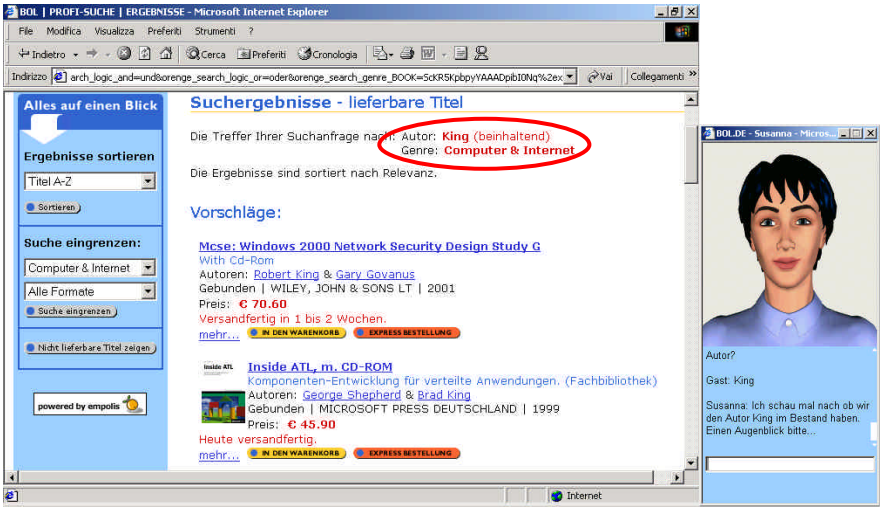


**Fig. 2. List of books by authors whose last name is "King" belonging to the book category *Computer & Internet* (in the circle).**

## The query expansion process

When a user asks the chatterbot for a book by author "King", the system dynamically builds an XML file containing the string "King" as the value of the proper tag – in this case <author> – and sends it to a module called Prompter [4], that performs the expansion of the original query by using the favorite book categories stored in the profile of the user.
The query expansion process consists of an improvement of the criteria used for the specification of a query. This is usually achieved by adding search terms to an already defined query. The additional terms may be taken from different sources:

- Product thesaurus. Products like books or other kinds of media are usually characterized by a textual description. The most relevant words contained in these descriptions are clustered according to their relation to the most frequently appearing ones, thus generating a "thesaurus" of terms.
- User profiles. They are accessed for identifying the book categories preferred by a user, which can be enclosed in a query for a more specific result identification, as already described in the previous section.
- Usage patterns. The application of association rules to a specific user can lead to infer a possible interest of this user in a product or service [5]. In this way, the chatterbot can decide which dialogue context to use when the dialogue comes to a dead end, i.e. when the user does not want to neither take the initiative nor mention a specific topic of discourse.

The information retrieved by the interface to the information sources (i.e. the expanded keywords, the preferred book category or the dialogue context coming out from an applied usage pattern) is used for the generation of a deep linking, which is directly forwarded to the chatterbot by the Prompter.
The Prompter is responsible for determining a suitable query expansion method to be used, according to the information available in the input file. The decision process is thoroughly described in [6] and summarized in Table 1.

**Table 1. Query expansion decision process**

| Input Information | Profile Available? | Expansion? | Information to use for expansion |
|---|---|---|---|
| Any combination of author, title and publisher_name | No | No | - |
| Publisher_name | Yes | No | - |
| **Author** | **Yes** | **Yes** | **Preferred book category** |
| **Title** | **Yes** | **Yes** | **Preferred book category** |
| Keywords | No | Yes | Product Thesaurus |
| **Keywords** | **Yes** | **Yes** | **Product Thesaurus and Preferred book category** |

## Evaluation Methods

The framework used for evaluating the performance of the agent is based on the means-end hierarchy already utilised during the phase of requirements specification.

### User requirements specification method

In order to facilitate the evaluation of the COGITO outcome, the user requirements were presented categorised in terms of means-end relations (see Andersen et al., [7]). For further information about the means-end hierarchy see Rasmussen et al. [8]. The framework for this categorisation is shown in Figure 3, presenting by the middle column the hierarchy in general and by the left-hand column the means-end hierarchy in a condensed form utilising fewer levels in the hierarchy. This presentation has proved sufficient and successful for specification of user requirements, and was utilised in the COGITO project, indicating the strategic goals on the highest level, procedures supporting these goals at the next lower level, and – at the lowest level – the operations from which these procedures are created. In this representation, each level will be specified by the next upper level concerning the reason or background for an action, and by the next lower level concerning how this action may be supported (see the right-hand column).

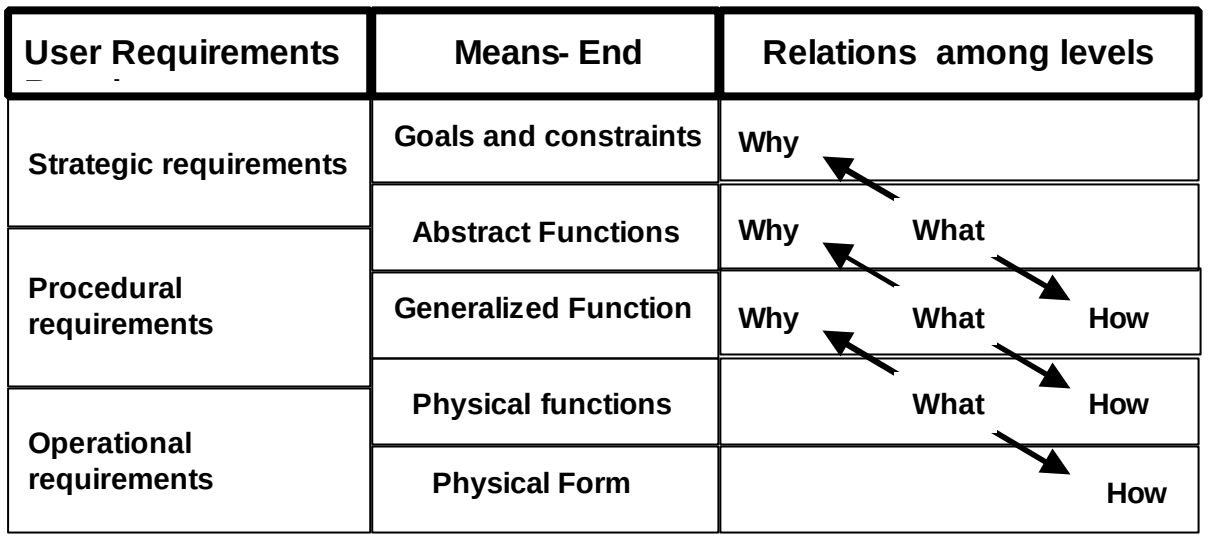


**Fig. 3. Means-end relations presented in general form in the middle column, in squeezed form utilised for specifying user requirements for COGITO at the left hand column, and with indication of the relations among the various levels at the right hand column.**

**System evaluation**

System evaluation normally constitutes three levels of evaluation procedures:

- verification, which is a check of implementation of operations specified in the user requirements, and therefore directly related to the lower level of the hierarchical representation of user requirement, the operational requirements;
- evaluation, which is a check of the presence of the functionality specified in the user requirements, i.e. is the system capable of executing all the sequences of operations needed for fulfilling the goals specified in the requirements. This part of the test procedure is directly related to the middle part of the hierarchical presentation of the requirements, the procedural requirements;
- last, but not least, the validation takes care of – based on user satisfaction – testing whether the system is of any value to the end users, i.e. do they perform better, more efficient and with a higher success rate than without having the system available. The question here is the difference between developing the system right, i.e. following carefully all the elicited requirement specifications, or developing the right system, i.e. a system that really is of benefit to the end users.

The logical way of evaluating a system is to take the top-down approach in which the evaluation and validation is tested by user interaction with the system. In case this test does not end up satisfactorily, the next step would be to take the bottom-up approach starting with the verification phase checking the implementation of operational features and continuing with check of the functional features.

That is, the COGITO system evaluation is based on the top-down approach focusing on the system/user interaction. In more detail the ‘ evaluation and validation’ of the COGITO agent has been performed by letting groups of test persons solve various tasks related to searching general information or specific products utilising the agent and its linking with specific BOL sites related to the questions and wishes of the users. The evaluation is partly based on quantitative measures, such as, e.g., the length of sentences of the users indicating a real conversation in contrast to using just a search-engine technique, the stereotypical use of the sentences utilised by the agent, and the number of fall back sentences indicating a missing interpretation of the request from the user. Likewise, objective and quantitative measure of eye-tracking specifying the time the user spent looking at the agent, the answers given by the agent, or the BOL site itself. Furthermore, the evaluation is qualitatively based on the users’ subjective assessment of using the system and of the outcome of their search. This is done through direct interviews and through fulfilment of detailed questionnaires enlightening the general impression and understanding of the agent when it reacted to requests from the user and presented the suggestions through effective links to the BOL site.

The test persons were requested to complete the questionnaire revealing their satisfaction with the system and the agent concerning various aspects, such as impression, control, effectiveness, navigability, learnability, aidability, and comprehension of the agent.

In order to have a reference for evaluating the COGITO proactive agent, a baseline session was performed using the BOL site equipped with a ‘ BOL state-of-the-art agent’ for comparing the two agents (see Figure 4).

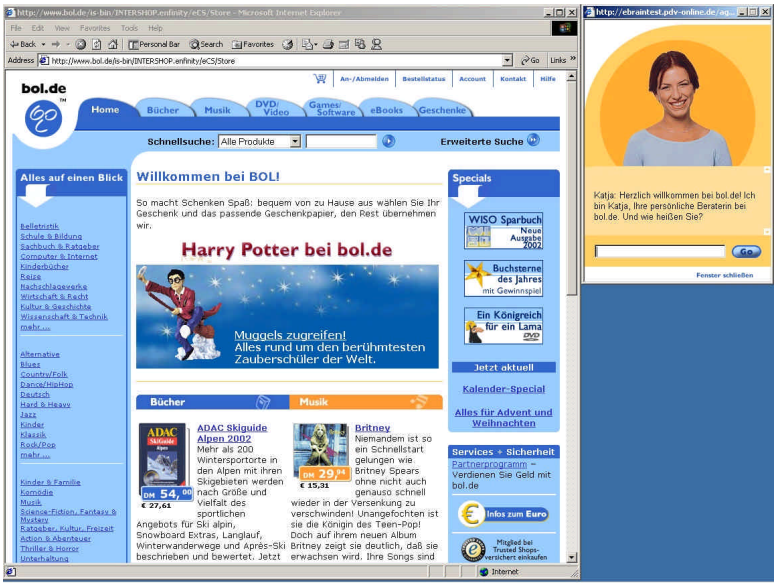


**Fig. 4. A snapshot of the BOL site equipped with the BOL state-of-the-art agent**

This agent had a level of chatting performance in line with existing agents of today and was integrated to the BOL site by having simple links to products being requested by the customer, i.e. this agent had no proactive features.

In order to test - for various levels of users - the benefit from the agent experienced by the users related to their previous experience from using the net, the test group was divided into two groups, one holding novices in using the net and one holding experienced users.

After being introduced to the system the test person was asked to deal with a number of tasks prepared for the test session. The tasks included problems like getting an overview about BOL, may be a guided tour of the site, find a specific book or books about a specific topic, or it could be a search for information about how to order and pay, or possibly about the security in using credit cards.

## Evaluation test sessions

Four groups of eight persons each were recruited for the test sessions, two groups of novices and two groups of experienced users in order to test the validity of the COGITO prototype for each of these types of end-users. These groups may be treated individually or – if preferred for improving statistical considerations – added two by two as groups of 16 persons for each agent to be compared.

As an example of the outcome concerning the criteria mentioned, the result of the effectiveness has been presented in Figure 5 (for a full presentation of the evaluation see Andersen et al. [9]).

In this scale, the questions were related to the fulfilment of task, i.e. to find the requested information. Furthermore, the easiness or difficulty in this process is included, exemplified by the number of steps needed to get the information.

For this scale, it seems obvious that the novices do not benefit fully from the features of the agent. The novices testing the COGITO agent seem even more unhappy that the novices testing the state of the art agent. Even though the functionality of the agent and its ‘cooperation’ or linking to the BOL site was explained to the test persons before each session, this may very well be forgotten during the session for persons not acquainted with the searching procedure using the Internet. So, the novices may have been too much focused on the communication with the agent and, therefore, not been noticing if the requested information was presented through a relevant link to the BOL. For the experienced users, however, this feature was observed and utilised, and these users rewarded the more effective linking of the COGITO agent as compared with the state of the art agent by increasing the overall satisfaction from 38% (the sum of the percentages concerning "Exp. 1" for the columns "Satisfied" and "Very satisfied") to 63% (the sum of the percentages concerning "Exp. 2" for the columns "Satisfied" and "Very satisfied").

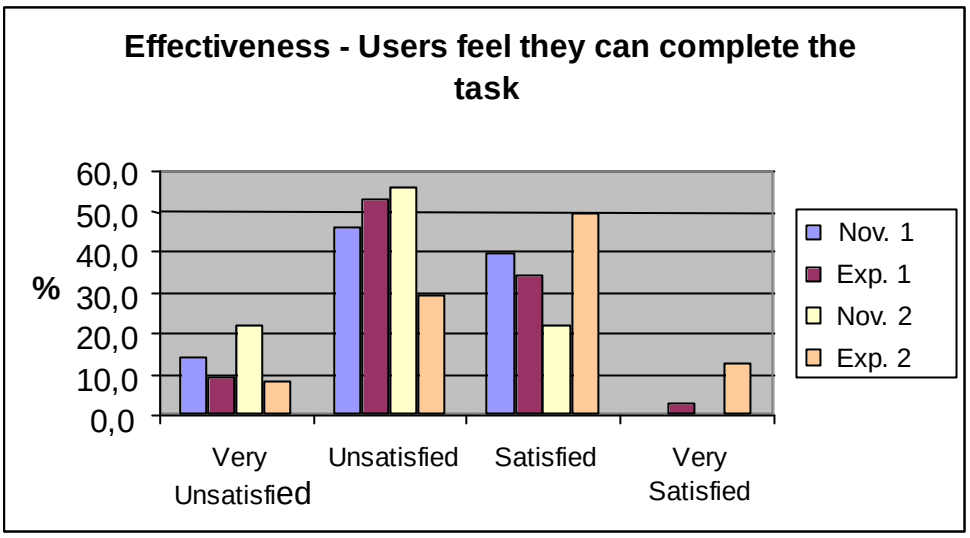


**Fig. 5. The degree to which users feel they can complete the task. Numbers are shown in percentage of the total number of user ratings in each category. ‘ Nov. 1’ is novices from the baseline evaluation of the BOL state-of-the-art agent, ‘ Nov. 2’ is novices from the evaluation of the COGITO proactive agent, ‘ Exp. 1’ and ‘ Exp. 2’ are similarly for the experienced users.**

## Conclusions

The COGITO agent has been evaluated for two groups of test persons, novices and experienced Internet users, to check if the agent is able to facilitate the interaction between the user and an e-commerce site exemplified in the COGITO project by the BOL site offering books, music, and gifts via the Internet.
The experienced users seem to appreciate more the developed improvements than the novices. Being more specific, for the topics selected for evaluating the satisfaction related to effectiveness this increased from 38% to 63%. Likewise, for results not presented in this paper the satisfaction related to the impression of the agent increased for the experienced users from 44% for the state of the art agent to 61% for the COGITO agent. For the novices, however, the satisfaction related to learnability increased from 37% to 55%.

## Acknowledgement

The COGITO project, ‘ IST-1999-13347, E-Commerce with Guiding Agents based on Personalised Interaction Tools’ was partly funded by EU. Furthermore, we would like to thank Osnabrück University for its hospitality in hosting the evaluation sessions, and especially Frank Ollermann for his help in preparing and being the moderator during the sessions.

## References

[1] Nielsen, J., *Designing Web Usability: The Practice of Simplicity*, New Riders Publishing, Indianapolis, 2000.
[2] Nielsen, J., *Usability engineering*, Boston, AP Professional, 1993.
[3] Abbattista, F., F. De Florio, N. Fanizzi and G.Semeraro. *Profile Extraction*. Deliverable 4.2, COGITO IST-1999-13347, University of Bari, IT, 2001.
[4] L’ Abbate, M. *Prompter*. Deliverable 4.5, COGITO IST-1999-13347, Fraunhofer Institute, D, 2002.
[5] Abbattista, F., O. Licchelli, P. Lops and G.Semeraro. *Usage Patterns Extractor*. Deliverable 4.3, COGITO IST-1999-13347, University of Bari, IT, 2001.
[6] L'Abbate, M., and U. Thiel (2001): *Intelligent Product Information Search in E-Commerce: Retrieval Strategies for Virtual Shop Assistants* in Proceedings of E-work and E-Commerce Conference 2001, Venice, Italy. B. Stanford-Smith & E. Chiozza (Eds.), IOS Press 2001. ISBN 1 58603 205 4
[7] Andersen, H.H.K., V. Andersen, C.B. Hansen, F. Abbattista, N. Fanizzi, F. Lioce, G. Semeraro, A. Thun. *Study of the end-user needs and behaviour*. Deliverable 7.1, COGITO IST-1999-13347, Risoe National Laboratory, DK, 2001.
[8] Rasmussen, J., A.M. Pejtersen and L.P. Goodstein. *Cognitive systems engineering*. New York: John Wiley & Sons, 1994. (Wiley series in systems engineering)
[9] Andersen, V. and H.H.K. Andersen. *Evaluation of the COGITO system*. Deliverable 7.2, COGITO IST-1999-13347, Risoe National Laboratory, DK, 2002.