

# Establishing criteria for evaluating intelligent agents in E-commerce

Hans H. K. Andersen, Verner Andersen, Camilla B. Hansen

Risø National Laboratory  
System Analysis Department  
DK-4000 Roskilde, Denmark  
e-mails: [hans.andersen@risoe.dk](mailto:hans.andersen@risoe.dk),  
[verner.andersen@risoe.dk](mailto:verner.andersen@risoe.dk),  
[camilla.buhr.hansen@risoe.dk](mailto:camilla.buhr.hansen@risoe.dk)

**Abstract.** Based on a joint European project, COGITO<sup>1</sup>, concerning e-commerce, the objective of our research was to establish a set of evaluation criteria and investigate user acceptance of intelligent agents, supporting the user navigation on the web-site, and whether the human-computer interaction may be improved by an expressive visualisation of the virtual agent. We have used the Cognitive Systems Engineering means-ends analysis tool for creating the criteria in the form of a set of user requirements. We have used interview sessions followed by a focus group. for collecting data. In this paper we present a series of strategical evaluation criteria, a set of problems related to the use of existing web-agents and the results from a word association test on agent appearance

**Keywords:** Intelligent agents, evaluation, focus group, cognitive systems engineering

## Introduction

The rapid growth of interactive Internet services has led to a constantly increasing number of modern Web sites (Berners-Lee et. al, 1992). We have also experienced an increase in their functionality, which, in turn, makes them more complicated to use (Nielsen, 2000). Thus, any attempt to enhance the usability of the web applications has to meet the challenge of coping with two almost contradictory goals: A useful web application should be based on traditional graphical user interface design which rely on menu selection and navigation. But these interfaces 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. (Nielsen, 1993). Instead the COGITO project aim to pursue:

*"...the inherent potential for interactive data processing and human-machine dialogue should be used by web 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. "* (Thiel, et al 1999)

The solution we propose in COGITO is based on "intelligent personalised agents" which represent virtual assistants or advisors (also visually) by modelling their ability to support users in buying product through the use of the Internet. This is in line with Maes et al (1999), who argue that personalisable, continuously running, and autonomous agents are well-suited for *"mediating consumer behaviours involving information filtering and retrieval, personalised evaluations, complex coordination an time based interaction."* (Maes et al, 1999). Papazoglou (2001) proposes a typology of agents based on their functionality and competencies: Application agents (meta-level agents), personal agents (interface agents that work directly with users), general business activity agents (performing general commerce support activities, e.g., legal agents, search agents and negotiation agents), and system-level agents (takes care of system related tasks like, planning, scheduling, interoperation security, and business transaction. With respect to these categories the COGITO agent could be characterised as a personal agent. It is visual at the interface, it aims to support users in buying and help them through the process. It aim to be pro-active to suggest users new products based on learned user-profiles.

Hence, we need to combine the usefulness of a value-added service with a high degree of usability. In order to verify the assumptions underlying the design decisions outlined above, and to find out appropriate ways to adjust the system parameters, the technical development has been and will further be accompanied and influenced by user studies and evaluations of both the individual components as well as the system as whole.

## A Method for Establishing the Criteria for Evaluation of Agents

Several studies exist that determine demographics and some behavioural characteristics of web users via self-selection (Pitkow and Recker 1994a & 1994b). Though highly informative, such studies only provide high level trends in web use (e.g. frequency of web browser usage to access research reports, weather information, etc). Tauscher and Greenberg (1997) employed a broader method, combining a quantitative approach with post-study interviews. However, their interviews were aimed at evaluating the design of current history mechanisms and revolved around a narrow range of questions focused on page visitation and re-visitation

The methodological approach applied in our exploratory study has mainly followed the principles and concepts offered by the Cognitive Systems Engineering (CSE) approach developed at Risø National Laboratory (Rasmussen et al, 1995). Among many other tools it allows the evaluator to analyse a system in terms of means-ends relationships (see Table 1). The approach is not so much to focus on what the participants do when they use an agent but rather to try to get understanding as to why the agents was used the way they were. The methodics applied in study all belong to the qualitative area of research. We suggest that the CSE framework can be used to establish criteria for evaluation of web-based agents. We also suggest that these criteria can be formulated on the basis of an analysis of elicited user requirements. In addition we will argue that a combination of “interviewing by doing” and focus group can be used in collecting data in systematic way.

We have used a new developed version of the CSE means-ends abstraction model (see Table 1, right side) as an analytic tool in formulating the evaluation criteria / user requirements at three levels: Strategic, procedural and operational. The strategic requirements mirrors the goals, purposes and constraints governing the interaction between users and the agent application under consideration and their environment, e.g., trustworthiness and mediation quality. The procedural requirements characterise the general functions and activities reflects the actual use and functionality of the agent, e.g., - procedures for promoting the functionality of agent and system, criteria for agent appearance, rules for agent self-explanation, structures for supporting natural dialogue. The operational requirements represents the physical activities, such as use of tools and equipment. Furthermore, the operational requirements signifies the physical processes of equipment and the appearance and configuration of material objects, e.g., agents must stimulate positive user association by being confidence-inspiring, obliging, friendly, happy, relaxed, and charming, agents must match user language style in terms of young/old, sub-culture, education, provide an opportunity to communicate with a human being and avoid agent software plug-in's. (Andersen et al, 1999). The data collection was carried out in the following way:

*“Interviewing by doing” sessions* where we exposed the participants to agents on the Internet. In these sessions we focused on eliciting general e-commerce problems and agent use requirements. This procedure was preferred to the use of think-aloud protocols (Ericsson & Simon, 1984) as it imposes less of a cognitive burden on the participants and did not alter the activity that we wanted to study. It also allowed us to ask the participants leading questions, for example, about the things they did not do, as well as questions about what they had in fact done. The procedure was as follows: 8 participants took part in these sessions. We first gave them an introduction to the evaluation test and then we tried interviewing by doing ([www.dr.dk](http://www.dr.dk)) to let them get familiar on this approach. We then went on to the actual tests giving the participants several tasks to perform on 3 sites (Mr.Clean, one2one, Amazon). Finally we gave them a short questionnaire. The lengths of the different session varied from 1-2 hours length.

*Focus group discussion:* The same 8 participants took part in a focus group. Here we discussed general e-commerce problems, the agent appearance and in particular the agent use problems and requirements. We also carried through a word association tests related to the evaluation of agents visual appearance. The advantages and disadvantages of focus groups as a tool for establishing criteria for evaluation will be discussed in some detail below.

*Data analysis:* During the analysis we transcribed the video tapes and categorised user requirements in terms of means-ends relationships and categorised the evaluation of agent appearances (based on the word association tests in three categories - Positive, negative and neutral. We used the following agents:

Virtual Friend: <http://www.virtualfriend.com/default2.asp> (cited 30th Nov. 2000. The site has changed since the experiment and is currently under construction); Nicole: <http://nicole-sf1.nativeminds.com/default.htm> (cited 30th Nov. 2000); Lego: <http://shop.lego.com/gateway.asp> (cited 30th Nov. 2000); Quest: (cited 30th Nov. 2000); Tokiama: <http://www.vperson.com/sapphire2000/index.htm> (cited 30th Nov. 2000); E-cyas: [http://www.e-cyas.com/index\\_eng.php3](http://www.e-cyas.com/index_eng.php3) (cited 30th Nov. 2000); Eve: <http://www.egain.com/egainassistant.htm> (cited 30th Nov. 2000); Cogito: (Private communication)

*The background of participants* is as follows: Eight participants – six females and two males took part in the interviewing by doing sessions and the focus group interview. Their ages ranged from 24 to 38 with

an average of 28 years. Five of the participants had a background in library and information science and three had business school/university degrees in business and economics. Three persons were students at the librarian school, one of them at master's level. Further-more one participant worked as a lecturer at the librarian schools, one worked as a librarian in a large, broadcasting corporation, two worked with category management and marketing and one was currently unemployed. All participants had a high degree of experience using computers i.e. more than five years. They all used the Internet everyday either at work or at home and had done so for several years. All of the participants used the Internet for email, information seeking and home banking. A majority of seven out the eight participants had been shopping on the Internet, but only one did so on a regular basis. The other six seldom shopped online. The participants mostly bought books but a minority also bought movies and groceries.

## **Focus Group as a Basis for Evaluation of Agents**

In order to specify the user requirements for the best possible type of agent satisfying the subjective judgement of the end users, a number of agents taken from other applications on the Internet were used as test objects and background for a focus group discussion about the appearance of intelligent agents.

Gibbs (1997) lists a set definitions of a focus group based on an extensive literature study. Focus group incorporates features like organised discussion (Kitzinger 1994), collective activity (Powell et al 1996), social events (Goss & Leinbach 1996) and interaction (Kitzinger 1995).

A focus group allow participants and researchers to become deeply absorbed in a subject and provides an opportunity for a rational as well as emotional consideration of the discussed theme. Using focus groups interviews makes it possible to collect information and user requirements from several persons at a time and the interaction going on between these people in form of exchanges of views and debates during the session provides more varied and detailed information than could have been obtained by interviewing the same persons individually. (Templeton, 1994)

The individual participants expressions serve as an input to the other participants thereby creating an effect of synergy in which different attitudes are probed and interpreted within a kind of collective hermeneutic. By means of the focus group interview requirements or certain angles on topics, which participants might have thought about but haven't been able to express a need for before, are generated and made explicit. The mechanism of group dynamics can then make the group discuss subjects that weren't obvious to the individual. (Gibbs, 1997)

## **The Advantages and Disadvantages of a Focus Group**

A focus group can be effective as means for establishing the criteria for evaluation - the user requirements. But it might not be that effective when it comes to the evaluation itself. That is, the focus group should not be used as the only usability data source in the evaluation of an agent. Like Nielsen (1997) phrase it: *"To assess whether users can operate an interactive system, the only proper methodology is to sit users down, one at a time, and have them use the system....the proper role of focus groups is not to assess interaction styles or design usability, but to discover what users want from the system.* In our study we used a interviewing by doing session as a secondary usability data channel.

There are other disadvantages. The moderator exhibits less control over the production of data (Morgan, 1988). The moderator cannot fully control interaction and discussions and the way they are heading. Interruptions can be misunderstood by the participants and even create a negative atmosphere among the participants. The only control that can be used is to try to guide and keep the participants close to the focus group topic.

Moreover, it can be difficult for the moderator to know if individuals express their own individual opinion or they speak with the "voice" of the group. That is they express their meaning within a certain context and a certain culture that might influence their individual perception and communication of problems or solutions discussed in the focus group. In addition, it can be difficult to gather a representative group of people. Some people are not used speak out in larger groups and may more or less refuse to take turns in the discussion while others can be very articulated and confident and therefore in a high degree influence the course of discussions. Finally, focus groups can produce imprecise data. As you will see below we tried to avoid this by exposing the participants to the most concrete examples of the technology being discussed as possible.

Our evaluation study was performed in two parts. One in which we had the participants as individuals using various sites including or not including intelligent agents, and one in which they all together took part in a common group discussion from which user requirements were elicited.

The first part of the experiment was an 'interviewing by doing' session, in which the participants were asked to perform a series of tasks on specific pre-selected sites. About half the sites had no intelligent agent, whereas the other half made use of an intelligent agent as an alternative to menu or icon-selecting procedures.

The subjects were advised to make use of the intelligent agent when available, partly to be acquainted with intelligent agents in general, which was important for the common focus group discussion following the individual sessions, and partly to be able to compare the use of agents with similar task solutions without having an agent available. During the session the subjects were asked to think aloud and furthermore interviewed by a moderator about their way of acting and their reactions to the agents. The complete sessions were logged on videotapes, and furthermore two observers taking notes surveyed the sessions. Each session took about one to two hours. Following the individual sessions the same group of people took part in a group discussion, discussing – with the experiences from the individual sessions in mind – once more user requirements for e-commerce in general and for intelligent agents in particular. This time the discussion or interviews were based on scenarios, in which the demands and wishes of the group of participants related to e-commerce purchase were expressed. This session lasted for about 3½ hours. During this session we also performed a word association test, in which the participants were presented to a number of various agents taken from the net. This was to unveil their spontaneous reaction to the agent in order to find the best appearance of an agent related to instilling a feeling of trust and confidence for discussing and exchanging information. Each agent was presented for 90 seconds and the participants were asked to give in writing the first words that came into their mind covering all kinds of aspects, like comments on appearance, indicate descriptive impressions or aspects, emotional feelings, etc. As mentioned the data from the focus group were analysed based on a new developed version of the means-end relation model. This new means-ends relation model is shown at the right in Table 1. The upper level, the strategical requirements, indicate the overall requirements of the agent, the middle level, the procedural requirements, indicate the procedures necessary for supporting the strategical level, and the bottom level, the operational requirements, specify the operations to be implemented for supporting the procedures fulfilling the user requirements. The procedure for the analysis is first to use the 5 level means-ends analysis, and then next to re-categorise into the three level means-ends hierarchy. The left side column of Table 1 shows the original 5 level means-ends hierarchy.

Table 1 Means-End relations in general, for user interests, and in new form utilised in COGITO (example from the on-line bookstore involved in COGITO)

<b>Means-End relations</b>	<b>User interest</b>	<b>Means-End relations</b>
Goals and Purposes, Constraints	Readers' ultimate goal: Education, Emotions, Profit, Power, Social career	Strategical requirements
Priority Measures: Flow of Information, Values, People and Money	Value Criteria Related to Reading Process and /or Product Knowledge, Data, Aesthetical, Psychological, Political experiences	
General Functions and Activities	General Topical Interest in Historical, Social, Geographical, Cultural Settings and Environments	Procedural requirements
Physical Processes in Work and Equipment	Topical Interest in Specific Kinds of Plots, Subject matter, facts, events	Operational requirements
Appearance, Location, and Configuration of Material Objects	Reading ability and Physical Characteristics of books (size, colour, pictures) and users (sight, age, sex)	

During this session we also performed a word association test, in which the participants were presented for a number of various agents taken from the net. This was to unveil their spontaneous reaction to the agent in order to find the best appearance of an agent related to instil a feeling of trust and confidence for discussing and exchanging information. Each agent was presented in 90 seconds and the participants were asked to give in writing the first words that came into their mind covering all kinds of aspects, like comments on appearance, indicate descriptive impressions or aspects, emotional feelings, etc. No list of words to choose among were given, so the participants were completely free to express their feelings. Based on these utterances each agent was evaluated.

## **Results**

The results of the study fall in three parts. We have established a set of criteria for evaluation based on a means ends relation analysis. The criteria were established on the basis of data from the interviewing by doing sessions and the focus group. That is, we have established a list of operational, procedural and strategical user requirements. We have also established the relations between the different categories of user requirements. Due to limited space we have chosen to present only the strategical requirements. We have generated use scenarios as a means for presenting these results. Another result stemming from the interviewing by doing evaluation is a list of usability problems with respect to the use of intelligent agents on e-commerce web-sites. A third result comes from the word association test where we evaluated different visual appearances of agents. The strategical requirements are presented in the following sections.

### **Trustworthiness**

The most important requirement for a customer/supplier relation - be it via a web-site or in any other form - is trust. Trust that information given by the customer will be treated in a confidential and decent way and not be misused in any way, trust about fulfilment of agreements and about the quality of purchased products, and trust concerning treatment of information related to credit cards and accounts.

Internet stores frequently offer several opportunities to the users to get tailored recommendations for books and other items, to create personal lists and to obtain news from the shops. In order to acquire these offers users must submit information about their personal preferences, email addresses, and in the case of buying items, a credit card number is imperative. Two problems arise from the need for submission of information. One concerns the creation of a user profile and another the handling and utilisation of the given information. Concerning the handling of information users often state a need for information about which of their data are saved and for what purpose.

They fear that information could be sold to other organisations and that they then would receive a massive amount of junk mail from both the specific store and other unknown companies. Users are generally cautious about submitting their email addresses and feel an even greater uncertainty about submitting their credit card number. This uncertainty is enhanced if it is not made clear who the owner of the site is or if detailed information about the product is not available. Trust in that a specific e-commerce site/organisation handles the user information properly and respects the privacy of the individual buyers is therefore paramount to the utilisation of the given offers.

It is especially important when collecting information for the user profile clearly to indicate the intentions for the query and not to ask for information, which to the customer seems unimportant for the purpose, e.g. married state, number of children or income. Furthermore, customers should not be rejected if they decline to expose the requested information, but be given – based on the information available – the best possible general profile instead of the specific personal one.

Other important issues for increasing trust is to specify clearly the rights of the customer concerning delivery of goods, rights concerning possible return, payment and information of account. If possible, a quality indication in form of guarantee from an authorised institution, like a bank, would be most helpful for obtaining trust in the supplier and for delivering banking information or credit card number.

Likewise, each purchase transaction should be followed by a detailed specification summing up the order including time of delivery and the agreed way of payment.

For the customer/agent relation it is extremely important that the customer feels he/she is in charge with the agent following the wishes of the customer regarding wanted information or support in navigation. The agent should not be the leading part unless the customer specifically has requested guidance.

Finally, a clearly indicated possibility of contact to a human person or a guaranteed answer to an e-mail within a specified limited time as well as the general corporate image of the company are of utmost importance.

### **Mediation Quality**

A very important point related to mediation is the confidence the customer feels in the agent. This confidence reflects as well the direct appearance of the agent, how the agent conforms to the context in which he/she is placed, as the functionality of the agent concerning a natural conversation and the professional knowledge. If the customer lacks confidence in the agent he/she may not make use at all of the agent, but try to navigate on his/her own using available menus or by clicking on icons missing the additional support the agent is meant to offer. Preferences concerning the appearance of the agent are presented below in the section 'evaluation of agent type'. The age and sex of the agent were discussed by the participants of the group, but this does not seem to be of great importance; however, it was mentioned that more agents, among which the customer could choose him/herself, could be available. More important was the size and location of the agent. It should not take up space showing valuable information, so it would be advisable to offer an agent that could be resized and moved around on the screen by the customer, alternatively be closed down completely. A technical issue addressed by the participants was that when the site is opened the agent should be downloaded and stored locally avoiding additional downloads each time the agent is called.

A good introduction to a site with an agent could be to let the agent appear automatically – with the potential of closing down the agent for more experienced customers - giving an introduction to the site and the agent itself, explaining its abilities for conversation and mediation. The agent should provide more levels of support to the customer, explain these levels, and leave it open to the customer to select the appropriate level to be redefined at any time. For less experienced customers – low level of the agent – the agent could join the session on own initiative if this seems appropriate due to low level of activity of the customer, indicating that support is needed. For more experienced customers – selecting a higher level of the agent – the agent should not intervene unless explicitly asked for by the customer.

The agent must be kind and polite and - besides assisting the customer in general - try to achieve information about the customer to build up a customer profile. This may partly be based on the actions of the customer and partly on personal questions to the customer without being persistent, if the customer don't want to respond to these types of questions. From each interaction the customer profile may be improved and thereby the support to the customer may be more and more personalised. This may consequently add to the inspiration of the customer and ease his positive feelings of the purchase.

Experience from the individual sessions showed that it was difficult or new to the participants to communicate with the agents using full sentences. Most of them made use of single keywords as known from search engines, and most of the agents did not react reasonably to this way of communication. Therefore, it was stated as a wish that the agent should accept full sentences as in a conversion as well as keywords normally used for a search engine. Furthermore, a dialogue history should be available allowing going back to refresh ones memory about previous phases of the communication.

Based on the individual profile compared to a group of similar profiles the agent should be able to give constructive and innovative feed-back to the customer concerning items related to the products searched by the customer. All participants was very goal oriented; therefore general chat with the agent had very limited interest, whereas context related chat was accepted.

### **Seriousness / Reliability**

The participants expressed very explicitly that the agent should be serious and reliable if he should prove useful to them. The seriousness applies to both the form of the agent represented by his appearance and the contents of the agent represented by the agent's abilities and knowledge. The appearance must be serious and sober and inspire the user with confidence in that the agent can solve problems effectively and that he possesses professional skills and is not just an amusing toy feature at the web-site. The agent must provide focused, serious, correct and reliable answers to asked questions. Furthermore he must react to the user's input in a manner that corresponds with the question. This means that he should leave out irrelevant information and answer in a stable way possibly by providing the user with a predefined set of categories to select from. The participants also expect the agent to learn from questions that he isn't able to answer correctly and then improve his performance. These elements are all important because they contribute to a high quality of the provided information. In addition the participants mention the quality of performed searches as being of high priority and demand fast, relevant answers and a high response

rate. Finally in relation to the need for a serious agent a thorough domain expertise of the agent is claimed as being of great significance. The expertise not only implies that the agent should answer current questions effectively but it should also make it possible for him to be proactive in the elicitation of users' unrecognised or future needs. Possessing domain expertise should also make the agent able to sense whether there is need for help or alternative offers or whether he should remain passive.

### **Flexibility / Tailorability**

The participants stressed the need for flexibility of the abilities of the agent. This means that the agent must be able to support various search strategies allowing users to seek and retrieve information in several ways. The agent should also provide conventional navigation and search alternatives for instance by providing access to menus or by assuming the function of a search machine allowing the users to search in single words or phrases, use truncation, similarity search etc. according to their preferences. The participants wanted to be able to ask directly for help or just type in a specific word dependent on whether they were able to identify the problem or needed to explain the problematic situation. Concerning the use of menus, there must be accordance between the menus and the knowledge of the agent so that identical information can be found using both methods. The agent should lead users to the site where appropriate information is given instead of just linking to the site just as he should go directly to a specific chunk of information instead of providing superior categories with non-transparent content. The participants state a need for a flexible support of different kinds of information needs and purchase situations. In relation to this the agent must handle goal-oriented behaviour such as verification of titles or other specific needs and give guidance to a speedy check out. Moreover he must deal with muddled needs requiring exploration by browsing and adjust the proposing of related offers to whether the user is in a hurry or wants to spend time at the site.

Another aspect of flexibility lies within the tailorability of the agent. Many participants uttered that tailorability of both functionality and appearance was preferable. They want it to be easy to learn how the agent functions and not be forced to learn all functionality at once. In accordance with this the agent should give an introduction showing his abilities and knowledge to first time users and enabling other users to skip the intro and proceed directly to the information-seeking phase. The agent must at all times be able to explain the user's action possibilities in an understandable way. The agent must facilitate and handle the constant evolution of the user's preferences in relation to the user profile and make it easy for the user to regret and redo former choices. Finally the participants clearly stated that they wanted a possibility of disabling the agent during an interaction and either continue on their own or ask for other kinds of assistance from the agent. Concerning tailorability of the agent's appearance the participants mentioned a need to be able to select some sub-functions of the agent and avoid others in order to enhance the transparency and usability of the agent.

### **Value Added Services / Surprise Me**

The participants all expressed a wish for value added services on the web-site. Value added means services or products that provide an extra, possibly not foreseen value or experience to the user, something that surprise the user in a positive way and thereby enhance the quality of the site. One of the basic elements that the participants need is product amendment, which consists of a large selection of products and products related to the main product. In the case of selling books related products could be pens and paper articles. Participants also wished for a lower price level than in physical stores and a delivery service that could be matched to the current user's needs for instance by offering wrapping in plastic, delivery at user-specified hours of the day or delivery at the user's work place. The necessity of being independent of opening hours of the local post offices was stressed. Among the services that the participants really felt would enhance the site's value were: a chat room focused at topics related to the site, wish lists with clear indication of the purpose of the lists, recommendations from experts and other users containing short presentations of the product and options for further exploration of the recommendations, reviews from papers and consumer magazines, results from customer satisfaction analyses, and graphical presentations of the products that can be tailored by the user and that can transfer the experiences provided by traditional catalogues and also a sense of actually touching the product. Furthermore the participants would like some kind of receipt when closing the deal indicating which products have been purchased and at what prices. Finally some participants found that context relevant commercials from the web-organisation could vitalise and touch up the site.

Another kind of value added services are incitements that the web-organisation utilises specifically with the purpose of getting users of the site to use the agent or carry through a purchase. The participants

mentioned the need for some kind of welcoming offer or discount the first time a user interacted with the agent. Some participants felt that the agent should be able to give the user an impression of having eased his shopping experience for example by guiding the user through the site and by that serve as a reminder of things the user should do at the site. The agent could offer products related to specific user-defined events for instance a picnic thereby relieving the user's cognitive burden of remembering which things to buy for the event. This feature makes the agent proactive. The participants expressed a need for marketing campaigns through traditional media that could make first time purchasers recognise the value of buying online assisted by the intelligent agent.

The participants saw great potentials in user profiles providing them with value added services if they work efficiently and in a non-obtrusive manner. However they also showed mistrust in the web-organisation's correct handling of personal information required for the development of the profile, and therefore the wish of being able to build user profiles is somehow ambivalent.

### **Entertainment**

Entertainment by the agent was discussed among the participants of the focus group and this topic seems to be a very delicate problem. On one side entertainment could be valuable for increasing the interest in the site, intensify the communication, and even increase impulsive purchase. On the other hand the risk of tiring or offending people may be very high. It was agreed that a high knowledge of the customer through the specific customer profile is very important. Traits such as culture, gender, age, education, etc. are crucial for the interpretation or misinterpretation of jokes or witticism. Furthermore, and stressing the difficulty in amusement, is the fact that the actual mood of the customer plays an important role for the perception of witticism.

Likewise, the use of jokes will be context dependent. A joke presented to a customer looking for books about cancer, e.g., may be very inappropriate; whereas a person who has previous been buying a lot of book related to, e.g., stand-up comedians, and is not looking for a new one, may very well be open to and pleased by jokes related to this specific topic.

### **Problems Concerning the Use of Intelligent Agents on E-commerce Web-sites.**

When using the agent the participants of the investigation encountered several problems. Some problems were so serious that they inhibited the participants from using the agent effectively and some just lead to a mild frustration. Consequently both types of problems made the participants feel they wouldn't have used the agent if they were allowed to choose for themselves and not being part of a test.

One of the major problems that all participants were faced with was that they stated their needs or questions in a single word instead of sentences. The participants were all accustomed to seek information by the means of search engines using only single words and they automatically transferred their knowledge and habits from that kind of information retrieval during their communication with the agent. Some participants eventually found out how to phrase the questions by browsing in the menus and others were guided by the experimenters, but even when the participants phrased whole sentences, problems could arise. The agent sometimes still didn't react and gave no hints to whether it was because of a malfunction in the system or inaccurate formulations of the sentence. This again caused frustration of the users. Another type of problem concerned the appearance of the agent.

### **Evaluation of Agent Type**

All of the participants felt that the agent's appearance was very important since it expressed the seriousness and quality of the agent and the web-site and by that also the trustworthiness of the agent and the information he provided and finally trust in the correct handling of customer data.

The participants all expressed a clear need for context dependent agents because these seemed more trustworthy and knowledgeable. The agents that received the most negative response were judged to be very weird, artificial, and silly. These were based on computer generated visualisations.

As an interesting point we had included a site without any agent, due to the fact that some participants during the individual session claimed that no agent was needed, and they could just as well communicate with the 'text field'. However, presented to this situation during the word association test, this situation was evaluated as impersonal, dull, too much text, and trivial, resulting in a very high negative score. The discussion that followed up upon the word association test showed that the participants didn't agree on

the most suitable appearance of an intelligent agent and that the majority wanted an opportunity to tailor the agent according to personal preferences. Though, if we look at the number of word and positive value of the words attached to each agent there was a preference for a of a human being

## **Plan for Final Evaluation**

The plan for the final evaluation consists of set of objectives for the evaluation, a description of the work needed for meeting the objectives and a concrete design of the final evaluation study. The objectives are:

1. Whether the interface succeeds in getting the users into a dialog where they reveal enough information about the contents as well as the context of their searches to provide the tailoring techniques with a data material sufficiently rich to make probable inferences.

2. Whether the techniques for tailoring the dialog toward the individual user succeeds in providing users with an interface they consider more useful and book suggestions they consider more relevant than those provided by a system without the tailoring techniques.

3. Whether the users notice the tailored parts of the dialog and book suggestions and whether and how they focus their interaction on these tailored parts, as compared to the standard parts of the interface.

4. Whether users notice and accept the Visual Persona, and whether the human-computer interaction is improved by an expressive visualisation of the virtual agent.

Objective 1 and 3 can be measured (quasi) quantitatively by, for example, counting the number of words the user type, classifying the user input and counting the number of categories covered, counting the interactions directed at tailored parts of the interface compared with those directed at the rest of the interface, and keeping track of the user's line of gaze.

Objective 2 involves the elicitation and analysis of the users' subjective assessment of using the system and of the outcome of their searches. To fulfil the objectives we will base our evaluation experiments on analysis of the users actual interaction with the system, verbal protocols, eye tracking, and thorough debriefing interviews.

Objective 3, 4, and involves evaluation of different system versions, e.g., comparing versions which employ no visualization of the virtual agent with versions that employ different instances of the Persona's character and appearance. Ad-hoc designed questionnaires and interviews will also be used.

The concrete evaluation activities will be split into two parts. The first part is a verification part related to a check of indicated user requirements as related to what has really been implemented in terms of operational, procedural and strategical requirements.

The second part is an evaluation session based on user groups appointed by the German application user in accordance with their customer profile database. We will have a size of groups of 6 - 10 persons depending on the number of groups to be involved in the evaluation. The suggested number of groups is 4, but this is dependent on the time available for the test, which again depends on the availability of a prototype finalised for testing. The four groups will be:

1. Novices without agent
2. Novices with agent
3. Experienced users without agent
4. Experienced users with agent

All sessions are performed on individuals and logged partly by video in order to have the feedback and mimics from the participants, and partly by the system itself concerning actions and written dialogue. Pre-planned tasks should be solved and indicators for the performance will be discussed based on the actual prototype, but also the time spent or the number of steps in the process for solving a number of pre-planned tasks will be taken into account. The data collection methods are listed above. The results will indicate

- User satisfaction based on
  - Agent appearance
  - Agent functionality
  - System functionality
  - User interface
- List of missing implementation of user requirements
- Suggestions for improvements

## Conclusion

We have proposed to use “intelligent personalised agents” to support users in buying products through the Internet. Our study shows that such agents need to be carefully evaluated before released to users. That is, “bad” exemplars provoke negative and neglecting attitudes among users. These attitudes may be rather enduring and as such they might impact the introduction and future use of well designed agents in a negative way.

Another result is that it proved useful to apply the CSE mean-ends analysis in the study. It provided us a valuable means for establishing a set of evaluation criteria in the form of a series of user requirement categorized and related to each other in three levels - strategical, procedural, and operational. We have found that a combination of interview sessions, where users try existing agents and web-sites while they are being interviewed, together with a focus group follow up proved useful as tools for data-collection. The focus group should not be used on its own, users have to use the actual systems beforehand to get a good feeling and enough background to go into detailed discussions about the subject matter. The result from the word association test showed that our participants preferred photos of agent compared to 2D / 3D animations. The results from test should be interpreted with care. On the other hand it is easy to perform and it is possible to get results very quickly. In addition, the test gave a good background for initializing a discussions on agent appearance. Finally, on the basis of this initial study we have generated a set objectives and concrete plans for the final evaluation study.

## References

- Andersen, H.K. and Verner Andersen, User Requirement for Converging Agreement by Networking Telematics, Risø National Laboratory, Roskilde Denmark, Risø-I-1517, (1999)
- Andersen, V, Andersen H.H.K., Hansen, C:B.: Study of end-user needs and behaviour, COGITO Focus group experiment, Risø-R-1264, (2001)
- Gibbs, A.: Focus Groups, Social Research Update, Edited by Nigel Gilbert, Department of Sociology, University of Surrey, England, Issue 19, (1997)
- Berners-Lee, T., R. Cailliau, J.F. Groff and B. Pollermann: World-Wide Web: The Information Universe, Electronic Networking: Research, Applications and Policy, (1992)
- Catledge, L. & Pitkow. J.: Characterizing Browsing Strategies in the World-Wide Web, Proceedings of the third Conference on the World Wide Web, Darmstadt, (1995)
- Ericsson, K. A. & Simon, H. A. (1984) Protocol Analysis: Verbal reports as data, MIT Press, Cambridge, MA
- Goss J.D., Leinbach T.R. Focus groups as alternative research practice, *Area* 28 (2), (1996) 115-23.
- Kitzinger J. The methodology of focus groups: the importance of interaction between research participants, *Sociology of Health* 16 (1), (1994) 103-21
- Kitzinger J. Introducing focus groups, *British Medical Journal* 311 (1995) 299-302.
- Maes, P., Guttman R.H., Alexandros, G.M: Agents that buy and sell, In *Communication of the ACM*, March, Vol 42, No 3, (1999). 81-91
- Nielsen, J.: The use and misuse of focus groups, *IEEE Software* 14, 1 (January), (1997) 94-95.
- Nielsen, J.: *Designing Web Usability: The Practice of Simplicity*. New Riders Publishing, Indianapolis, (2000)
- Nielsen, J.: *Usability engineering*. Boston, AP Professional, (1993)
- Papazoglou, M.P.: Agent oriented technology in support of e-business: Enabling the development of "intelligent" business software for adaptive, reusable software, In *Communications of the ACM*, April, Vol. 44, No4, (2001), 71-77,
- Pitkow, J. and Recker, M. Results from the First World-Wide Web Survey. Special issue of *Journal of Computer Networks and ISDN systems*, Vol. 27, no. 2, (1994)
- Pitkow, J.E., Recker, M.M.: Integrating Bottom-Up and Top-Down Analysis for Intelligent Hypertext, Conference on Intelligent Knowledge Management. Intelligent Hypertext Workshop, Dec. 12, National Institute of Standard Technology, (1994)
- Powell R.A. and Single H.M.: Focus groups, *International Journal of Quality in Health Care* 8 (5) (1996), 499-504
- Rasmussen, J., Pejtersen, A.M., Goodstein, L.P.: *Cognitive Systems Engineering*. Wiley Series in Systems Engineering, (1994)
- Templeton, J.F.: *The Focus Group: A Strategic Guide to Organizing, Conducting and Analyzing the Focus Group Interview*, Probus Publishing, (1994)
- Thiel, U.: *E-Commerce with Guiding Agents Based on Personalized Interaction Tools*, EU IST Technical Annex, IST-1999-13347, GMD-IPSI, Darmstadt, Germany, (1999)

---

<sup>1</sup> E-Commerce with Guiding Agents Based on Personalised Interaction Tools (COGITO IST-1999-13347)  
The project is partly funded by CEC