University of Waterloo

Technical Report

CS-2010-19

Understanding How Users Express Preferences: a User Study

Ingrid Oliveira de Nunes

Simone Diniz Junqueira Barbosa

Carlos José Pereira de Lucena

Faculty of Mathematics

DAVID R. CHERITON SCHOOL OF COMPUTER SCIENCE

UNIVERSITY OF WATERLOO

WATERLOO, ONTARIO, CANADA N2L 3G1

Technical Report CS-2010-19 October, 2010
David R. Cheriton School of Computer Science University of Waterloo

Understanding How Users Express Preferences: a User Study [1]

Ingrid Oliveira de Nunes, Simone Diniz Junqueira Barbosa, Carlos José Pereira de Lucena

{ionunes,simone,lucena}@inf.puc-rio.br

Abstract. With the exponential growth of information available on the Web and our 24/7 availability through social networks, instant messengers and e-mail, people are facing the challenge of processing huge amounts of data and playing multiple roles at the same time. Personal Assistance Software (PAS) aims at aiding users in these tasks from making recommendations to acting on their behalf. Even though extensive research has been done on improving such systems, little attention has been given to their acceptance by users. With the goal of building users' trust on PAS, we aim at developing an end-user language in order to empowering users to control and instruct their PAS to provide task delegation. However, such language would be meaningless if users are not able to adequately express their preferences. This paper presents a user study whose goal was to provide a deeper understanding of how users express their preferences. Seven research questions are investigated, including how the knowledge about a domain influences the expression of preferences and how users change them after being exposed to decision-making situations. This study allowed us to identify kinds of support users need to better express their preferences so that a system can be able to act on their behalf.

Keywords: Personal Assistance Software, User Study, User Preferences, Domain Specific Language.

[1] Trabalho patrocinado pelo Ministério de Ciência e Tecnologia da Presidência da República Federativa do Brasil e FINEP.

In charge of publications:
Helen Jardine
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
Tel: +1 519 888-4567 x33293 Fax: +1 519 885-1208
E-mail: hjardine@uwaterloo.ca
Web site: http://www.cs.uwaterloo.ca/research/tr/

Contents

1 Introduction

The growing popularity of the Web is turning interactivity and access to information two constants in people's lives. As a consequence, information overload and playing multiple roles at the same time turn to be challenges that, without appropriate support, are hard to manage. Agent-based approaches used in combination with artificial intelligence techniques have been explored in the context of Personal Assistance Software (PAS) in order to develop agents as assistants able to aid users in tasks of diverse natures, from making recommendations to acting on their behalf.

Most of current research work focused on increasing the accuracy of elicitation and learning methods. Only a few (Chen & Pu 2010, Glass, McGuinness & Wolverton 2008, Schiaffino & Amandi 2004) have involved an extensive interactivity with users and has investigated acceptance of PAS by them. This is essential since there is no sense in developing PAS for users if they are not willing to accept it. Our goal is to develop an approach that gives users a first level of autonomy, empowering them to control and instruct PAS in task delegations. We aim at developing an end-user language that can be instantiated across different domains. An initial version of an end-user Domain-specific Model (DSM) to be used in a software architecture for PAS has already been proposed in (Nunes, Barbosa & Lucena 2010).

Nevertheless, our goal relies on assumptions about users and their ability to express their preferences. An end-user language is meaningless if users are unable to appropriately instruct their agents. Therefore, we have performed a user study, presented in this paper, to validate these assumptions as well as to provide a deeper understanding of how users express their preferences. We focus on answering two main research questions: (i) are users able to express their preferences in such a way that a domain specialist is able to make an adequate choice in this domain on their behalf?; and (ii) do users need to be exposed to a concrete decision-making situation to able to express their preferences about a familiar domain? The study involves collecting preferences specifications expressed in natural language before and after experiencing a decision-making situation. Later a domain specialist uses the initial specification to make recommendations according to each specification. Other aspects are also analyzed such as which kind of preferences users typically forget to specify. Our study allowed us to identify some kinds of support users need to better express their preferences and relevant concepts that should be part of an end-user preferences language.

2 A Two-level Software Architecture for Agent-based Personal Assistance Software

One of the most challenging tasks in building PAS is to capture users' characteristics and preferences. If this task is not correctly performed or does not capture enough data about the user, the system tends to present an unexpected and often undesirable behavior, even causing users to reject the system and stop using it. User preferences can be gathered by means of explicit or implicit techniques. The former requires explicit interactions with the user either in the form of queries or feedback that is incorporated into the preferences capturing mechanism. Users can directly set their preferences, such as checking their preferred areas of interest, or indirectly, such as answering personalized-based quizzes. On the other hand, in implicit techniques, the system constantly monitors users' actions in order to learn about their preferences. Both approaches have pros and cons, which are discussed elsewhere, for instance in (Anand & Mobasher 2005).

Our research focuses on agent-based PAS in which systems do not only assist users on their tasks, but also are able to act on their behalf. In this kind of PAS, the accuracy level of user information is more critical, because the consequences of an agent action have a higher impact, as opposed to useless recommendations of products or interface changes that users may ignore or revert. Our research aims at taking a first step on the users' task automation: our goal is to empower users to control and instruct their agents by exposing user models at a high-level of abstraction, closer to the users' language. Even though ideally an agent that "guesses" what people want would be even better, freeing users from time-consuming tasks that they can specify and delegate for a computer system can also be helpful, while granting users more control over the agent's behavior. The group of users we are targeting are users that are willing to specify a task for an agent to execute on their behalf.

In our previous work (Nunes et al. 2010), we have proposed a first version of a domain-neutral high-level user metamodel in conjunction with a software architecture to implement agent-based PAS. The main idea is to decouple an agent-based implementation that supports variability and to provide a high-level view of user customizations. Dynamic model transformations keep these two models consistent. The advantages of providing this high-level model are: (i) user customizations are implementation-independent; (ii) the vocabulary used in the user model becomes a common language for users to specify configurations and preferences; (iii) the user model modularizes customizations, allowing a modular reasoning about them. Even though our main goal is to empower users to control their agents by being capable of managing their user models, we are aware that providing this kind of information is a time-consuming task. We do not exclude the possibility of using learning algorithms to aid the user in the user model management. Therefore, our approach can be used in mixed-initiative systems.

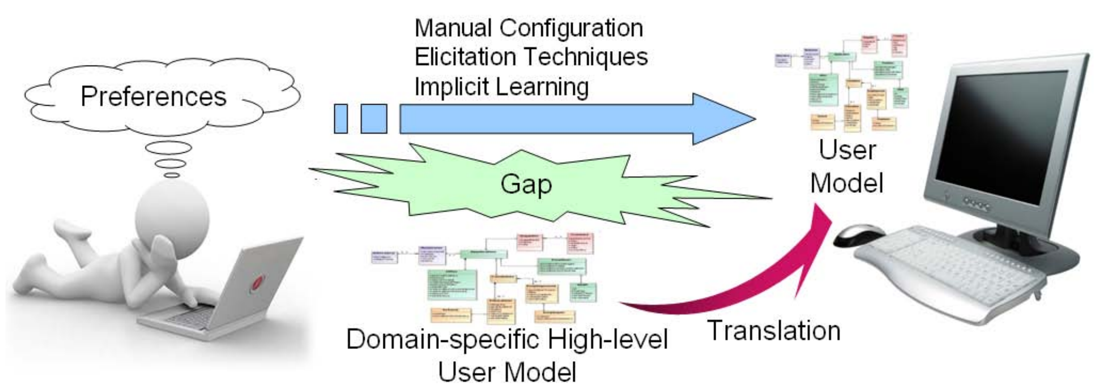


Figure 1: High-level User Model.

Figure 1 illustrates how we represent the process of modeling user customizations. At one side, there are users, who possibly already went through a certain kind of situation several times and have a mental model of their preferences, e.g., business travel. At the other side, there are computers with programs with low-level implementations of user concepts, for instance they may be based on a specific preferences model used in algorithms to reason about preferences. In addition, we have a potentially large gap between the preferences mental model of the user and that of the software. As a consequence, the main goal is to translate one model to another. We show in Figure 1 three common possibilities: (i) manual configuration – users themselves are responsible for setting the (restricted) preferences model of the software, thus potentially requiring a huge user effort to perform this task; (ii) elicitation techniques – users are also involved in this process, however software systems help them to express their preferences, typically with guiding questions that directly translate the users' answers to the preference model. This process may become very tedious, and there is the problem of knowing when the system should stop posing questions; and (iii) implicit learning – the system keeps track of users' actions and builds statistical models, which allows to infer users' preferences. However, this approach needs a significant amount of data, and even then errors may occur.

Our aim is to reduce the gap between the preferences mental model of the user and the preferences model of the software not by adopting techniques that automate this translation, but by providing users with a high-level preferences model that they can manually customize, and letting them to express themselves in a language similar to the one they use in human conversations and instructions. In this approach, a new need arises for translating these high-level models to lower-level ones, which computers can handle. For instance, a user may say: "My preference is not to buy a laptop of the brand X, but I don't care if the brand is A, B or C." Suppose now that the system is able to process preferences expressed as a partial order relation. We can represent that preference like this: $preference = \{< A, X >, < B, X >, < C, X >\}$.

In our approach we are assuming that (at least a specific group of) users have this preferences mental model and are able to describe it. This activity is similar to instruct an individual to perform a task on a person's behalf. This task might have already been executed by the person several times (a tedious repetitive task), and she is able to describe her preferences. Therefore, the user study presented in this paper aims at identifying this group of users, understanding the relationship of the knowledge about a domain and other aspects, such as gender and age, and how users are able to express their preferences. This study not only allows to validate our assumption, but it also helps to identify deficiencies (missing and wrong information) in users' preferences specifications and how we can support users to better express their preferences in such a way that they will be able to delegate a task for their PAS.

3 Related Work

One of the biggest projects in the context of user agents as PAS is the Cognitive Assistant that Learns and Organizes (CALO) project[2] (Berry, Donneau-Golencer, Duong, Gervasio, Peintner & Yorke-Smith 2009), funded by Defense Advanced Research Projects Agency (DARPA), whose goal is to support a busy worker in dealing with information and task overload by creating cognitive software systems, which are able to reason, learn from experience, be told what to do, explain what they are doing, reflect on their experience, and respond robustly to surprise. Along the project, the research effort was mostly concentrated in PTIME (Berry, Peintner, Conley, Gervasio, Uribe & Yorke-Smith 2006), an agent that helps users to manage their schedules. The CALO project significantly advanced the research on user agents, however the solution is tightly coupled with the domain being

[2] http://caloproject.sri.com/

explored. This differs from our goal, which is to develop a domain-neutral language. Lessons learned from studies involving "real" users (Berry et al. 2009, Glass et al. 2008) provide us with a substantial basis to our approach. They are related to the acceptance of this kind of systems by users, showing that essential characteristics are expected from these systems, such as transparency.

Schiaffino & Amandi made solid contributions to the development of personalized user agents from a Human-computer Interaction (HCI) perspective. By means of an empirical study (Schiaffino & Amandi 2004), they showed what users really expect from user agents, such as the kind of interruptions they tolerate, when they are willing to delegate tasks to agents, and when agent mistakes are accepted. Nevertheless, their research focused on an issue that we are not directly addressing: when and how interrupt users. Their goal is to design agents that can provide context-aware assistance and make context-aware interruptions (Schiaffino & Amandi 2006).

User preferences have also been widely explored in the domain of recommender systems. In such systems, systems must anticipate user preferences in order to make a recommendation of a certain product in the future (which may not be of their interest). In addition, users typically have their preferences stored in several locations, such as different online stores. As a consequence, there is a lack of motivation of the user to provide their preferences, and for several locations. The domain of applications we are looking at is different in two main aspects: (i) the idea is to have a user personal computer system, i.e. users provide their preferences only once and for their particular purpose; and (ii) we are aiming at the automation of users' repetitive tasks, and therefore they have already been exposed to decision-making situations associated with that task. We aim at providing means for users to explicitly specify (or to make modifications on the specification of) this task to be done. The cost of performing this specification is amortized by reducing the cost of performing the task several times in the future. But still, there are user studies on recommender systems (Hu & Pu 2009) and evaluation frameworks (Chen & Pu 2010) that can be used in our context. The concepts of objective/perceived accuracy and objective/perceived effort can be also used to evaluate user agents.

User studies presented in this section are complementary to our approach. They explore different relevant aspects related to users and help in understanding how they interact with PAS. These studies motivated the proposal of our approach introduced in previous section. The user study presented in this paper explores a different angle, which is how users express their preferences before experiencing a decision-making situation, and helps to understand how users of different categories, mainly associated with different domain knowledge, express their preferences.

4 Experiment

In this section we detail the design of our user study, as well other relevant details, including the research questions we aim at answering and the participants involved in the study.

Following the experiment process presented in (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén 2000), we have elaborated the experiment definition. The purpose of the definition phase is to define the experiment goals, and for that, we have adopted the goal-question-metric (GQM) (Basili, Selby & Hutchens 1986) template. Following this template, the experiment goal is presented in Table 1. Both (Basili et al. 1986) and (Wohlin et al. 2000) provide guidance for performing experimental studies in the context of Software Engineering (SE), however they could be also adopted in our study.

Definition element	**Our experiment goal**
Motivation	To understand how users express their preferences,
Purpose	characterize and evaluate
Object	users' preferences specifications
Perspective	from a perspective of the researcher
Domain:user	as users with different knowledge about a domain express their preferences
Scope	in the context of the social network of the researcher.

Table 1: Goal Definition.

4.1 Research Questions

The main goal of this study was to evaluate how users would typically express their preferences about a domain without having just experienced a decision-making situation. This evaluation was performed in different directions, which are associated with seven research questions addressed in the study, as presented in the first column of Table 2.

Research Question	**Evaluation Approach**
Are users able to express their preferences about a familiar domain in such a way that a domain specialist is able to make an adequate choice in this domain on their behalf?	Comparison between laptops selected by users and the ones recommended by the domain specialist based on their specification.
Do users need to be exposed to a concrete decision-making situation to be able to express their preferences about a familiar domain?	Analysis of the differences between the initial preferences specification and the reviewed version of it.
Which type(s) of preferences users usually forget or incorrectly specify before being exposed to a concrete decision-making situation?	Analysis of the most common types of preferences that appeared only in the preferences review.
How different are specifications provided by users with a higher degree of knowledge about a domain from the ones provided by users with a lower degree of knowledge?	Comparison between the preferences specifications provided by users with higher knowledge about the domain and by users with lower knowledge about it.
Which user profiles take less time to express their preferences?	Comparison of how long users in different categories (domain knowledge, gender, ...) take to specify their preferences.
When users make a choice, which ones select fewer options from among the offered ones? In other words, which user profiles are more confident in which is the right choice for them?	Comparison of how many laptop options were specified by users in different categories (domain knowledge, gender, ...).
Which user profiles take less steps (filtering, comparing, analyzing, ...) in the process of decision-making (choosing among available options)?	Comparison of how many steps (filtering, looking details, comparing,...) users in different categories (domain knowledge, gender, ...) took to define their laptop options.

Table 2: Research questions and their evaluation approach.

With these research questions, we aim at: (i) verifying whether it makes sense to provide an end-user language for users expressing or adjusting their preferences so they can delegate tasks for PAS. If users are not able to specify their preferences in natural language in such a way a (human) expert in that domain is not able to make an appropriate choice on their behalf, it is unlikely that it will work with a restricted language and computer system as experts; (ii) understanding whether users change and which kinds of changes they make after being exposed to a decision-making situation; and (iii) investigating how the domain knowledge or other relevant aspects (age, gender, etc.) impact on the users' expression of their preferences. If we conclude that the initial preferences specification, i.e. prior to a decision-making situation, is not enough for making a decision on behalf of users – item (i) –, it is essential to identify what kind of support can be provided for each user category in order for users to better express their preferences, but still without having to go through a process of choice – items (ii) and (iii).

4.2 Procedure

The experimental study we planned to answer our research questions is based on a web-based questionnaire applied to a wide spectrum of users (see next section for details). The domain selected for performing our study is shopping for products, in particular we chose the laptop as the target product. This decision was made due to the availability of domain experts to collaborate with the study.

In a nutshell, the idea of the questionnaire is to first ask users to specify their preferences for someone who is going to buy a laptop for them. Later, they are asked to navigate on a laptop catalog and select from one to five laptops. Finally, we give users a chance to modify their preferences specification.

The applied questionnaire consisted of four parts:

- *User Information Data.* The questionnaire is anonymous, however we collect relevant information related to the study from the study participant: (i) age; (ii) location (city and country); (iii) working/studying field; (iv) how many laptops the participant have already had (current one included) (v) from these, how many were chosen by the participant herself; and (vi) how she rates her knowledge about the domain. These last three items are used to evaluate the participants' knowledge about the domain.

- *User Preferences.* The study participant is requested to imagine a situation in which she is going to ask someone to buy a laptop for her. Therefore, she is requested to specify all her preferences and restrictions, i.e. instructions. An example in the flight domain is provided. Besides storing the provided specification, we logged the current state of the specification every 15 seconds and the time the participant took in this part.

- *Choosing Product.* Next, the participant is requested to analyze a set of different computers and say which one she would have bought. We ask her to rank her favorite up to five laptops. We used the Best Buy[3] catalog, which had 144 laptops by the time we imported it (at the same day that the survey was released). We recorded each step (comparing, filtering, detailing, ...) the participant performed, as well as the time taken for choosing the laptops.

[3] `http://www.bestbuy.com/`

- *User Preferences (review).* Finally, after analyzing the available computers, the participant is given a chance to review her preferences and modify them, in case she realized that something was missing or wrong in her specification. We have notified participants in the third part that they would have this reviewing chance. We also asked the participant's opinion about what changed on her specification. The additional logs are the same as in the second part of the questionnaire.

After collecting all the data, a domain expert was involved in the experiment. The domain expert's responsibility was to analyze the first version of the preferences provided by the participants, and to rank up to five laptops he would have recommended for each participant.

Based on the questionnaires and the recommendations made by the domain expert, we analyzed this data according to two main aspects, related to the research questions 1 and 2: (i) were the participants able to express their preferences in such a way the domain expert could make the right indication for them?; and (ii) did the participants change their preferences specifications after experiencing the process of choosing a computer? Furthermore, we have also analyzed other relevant aspects in order to answer the additional research questions, from 3 to 7. In the second column of Table 2, we detail how we analyzed the survey data to answer each research question. The logs collected periodically from the preferences specification were not used in this study, but kept for future work.

The evaluation approach presented in Table 2 shows we have performed a mainly qualitative but also quantitative analysis of the data to answer all our research questions.

4.3 Participants

Our study involved a total of 192 participants, who answered our questionnaire, and one domain expert, who indicated laptops for each participant according to their initial preferences specification.

The questionnaire was available online from May 20 to July 13, 2010 (almost two months). For selecting the participants, we used a convenience sampling, based on the social network of the researchers involved in this study. First, invitations for participating of the survey were sent by e-mail and people were asked to forward the invitation for other people. In addition, the survey was published in different Orkut[4] communities.

As result, we collected a database with 451 surveys that were at least initiated, from which 192 were completed (42.6%) – incomplete surveys were discarded. As the researchers that are performing this experiment are Brazilians, most of the participants are from this country (86.98%), and the remaining 13.02% are from four other countries. The same situation happens with the working area (63.54% participants work with informatics-related areas). In our analysis, we did not detail the other working areas because our focus was to identify participants with a higher knowledge about our study domain (laptops); however our database contains the working area of each participant. The description of the demographic characteristics of our study participants is detailed in Table 3.

The domain expert that was involved in our experiment has an MSc degree in Computer Science. Moreover, his work involves giving technical support to the Software Engineering Laboratory of PUC-Rio as well as specifying and recommending new computers and laptops

[4] `http://www.orkut.com`

Age	16-25 years 60 (31.25%)	26-35 years 83 (43.23%)	36-45 years 21 (10.94%)	>45 years) 28 (14.58%)	
Country	Brazil 167 (86.98%)	Germany 10 (5.21%)	Canada 10 (5.21%)	United Stated 4 (2.08%)	Peru 1 (0.52%)
Gender	Male 134 (69.79%)	Female 58 (30.21%)			
Working Area	Informatics 122 (63.54%)	Non-informatics 70 (36.41%)			
Domain Knowledge	No Knowledge 5 (2.60%)	Beginner 16 (8.33%)	Intermediate 40 (20.83%)	Advanced 83 (43.23%)	Expert 48 (25.00%)

Table 3: Demographic Characteristics of Participants

for the laboratory and its individual members. Therefore this expert is used to listening to clients specifying their preferences and to recommending computers and laptops for them.

5 Results and Analysis

In this section we provide the results we collected from the execution of our experiment as well as interpretations for those results. We have made a qualitative analysis of the preferences specifications (initial and reviewed versions) given by the survey participants and a quantitative analysis of part of the collected data, such as time taken to accomplish the parts of the survey. Due to space restrictions, the paper contains only the charts and "raw" data we consider most relevant to report.

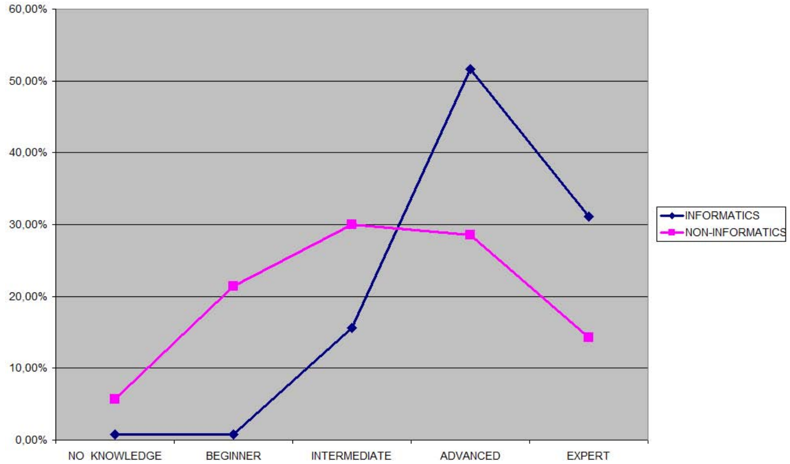
The first analysis that we made was how to measure the participants knowledge about laptops. The fields (iii) to (vi) from the *User Information Data* part of the questionnaire were used for that. Based on this data, we make the following observations (see Figure 2):

- Participants that work in the computer science area have at least an INTERMEDIATE[5] domain knowledge. In other fields, participants are mostly INTERMEDIATE.
- Most of the participants who have had several laptops have at least an ADVANCED knowledge, the more laptops participants have, the higher their knowledge.
- Almost all of the participants chose their laptops, only the ones who had several laptops chose only some of them (possibly because they get laptops from work companies).
- Not having had a laptop does not indicate a low knowledge about the domain – participants chose not to have a laptop.

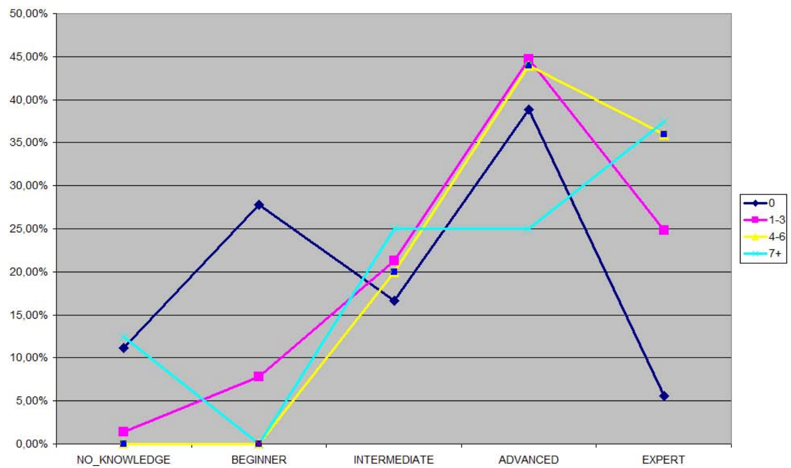
The relationship between the other fields and the domain knowledge provided by the participants presented the expected behavior. Therefore, when analyzing the collected data, we adopted the domain knowledge that the participants themselves provided as the criteria to determine their knowledge about the domain.

Research Question 1. Based on the initial preferences specifications of all participants, and the domain specialist recommendation for each of them, we have compared the

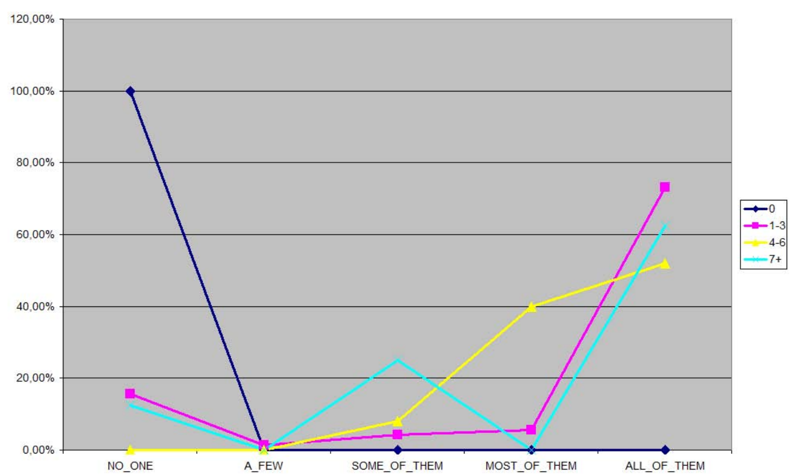
[5]Values for evaluating the domain knowledge: NO_KNOWLEDGE, BEGINNER, INTERMEDIATE, ADVANCED, EXPERT.



(a) Domain Knowledge x Working Area (Percentage).



(b) Domain Knowledge x Owned Laptops (Percentage).



(c) Owned x Chosen Laptops (Percentage).

Figure 2: Domain Knowledge Analysis.

laptops recommended and the ones the participants chose. The goal was to investigate the ability of users to express their preferences and how their knowledge about the domain influences this ability. From all 183 surveys[6], 53 (28.96%) had at least one of the specialist recommendations that matched at least one of the participants' choices. Furthermore, we have calculated a similarity score SS between the recommendations and choices according to the function below, which takes into account the positions matched to calculate a weighted average.

$$SS = \frac{\sum_{i=0}^{size(CL)} (5-i) * \left(\frac{\sum_{j=0}^{size(SR)} (5-|i-j|)*(sim(CL[i],SR[j]))}{\sum_{j=0}^{size(SR)} 5-|i-j|} \right)}{\sum_{i=0}^{size(CL)} 5-i}$$

where CL is the chosen laptops (by the participant), SL is the specialist recommendation (for the participant), $size(v)$ returns the size of a vector v and $sim(x,y)$ is the function that calculates the similarity between two laptops. If they are equal, its value is 100, otherwise it is the sum of each feature compared (1 for equal, 0.5 for an unspecified value in that feature in x, and 0 for different). The latter is then normalized for 100. The only feature treated differently was the price, which is 0 for a difference bigger then $100, or $1 - |price1 - price2|$, otherwise, also normalized for 100. Table 4 presents the values found for our study. The column *matches* is the number of surveys in which at least one of the laptops matched, and the columns *SS(AVG)* and *SS(STDEV)* are the average and standard deviation of the similarity score, respectively.

	Matches	SS**(AVG)**	SS**(Median)**	SS**(STDEV)**
Domain Knowledge				
NO_KNOWLEDGE	3 (60.00%)	60.05	54.38	23.50
BEGINNER	2 (13.33%)	46.50	44.92	5.13
INTERMEDIATE	9 (23.08%)	48.76	47.85	8.49
ADVANCED	27 (33.33%)	51.58	49.47	11.37
EXPERT	12 (27.91%)	51.25	47.41	12.55
Gender				
FEMALE	17 (29.82%)	50.26	48.02	10.20
MALE	36 (28.57%)	50.93	47.94	11.78
Age				
16-25 years	14 (24.14%)	49.55	46.30	11.67
26-35 years	27 (35.06%)	52.16	48.36	12.10
36-45 years	8 (38.10%)	51.91	49.74	10.98
>45 years	4 (14.81%)	48.19	47.62	7.49
Total	**53 (28.96%)**	**50.72**	**48.02**	**11.29**

Table 4: Domain Specialist Recommendation - Matches per Group.

Table 4 shows that the number of matches was higher for participants with a higher knowledge about the domain. This can also be seen in the similarity score. Nevertheless, the highest number of matches (in percentage) were in the group of NO_KNOWLEDGE participants. We observed that these specifications, even though they do not contain specific details of the laptop, provided key information about for what the laptop will

[6] Due to failures in the web application to save the domain specialist recommendation, 9 recommendations were lost, and their specifications were discarded for this research question.

be used. But it is important to highlight that we are not aware with which criteria the participants chose the laptops, as they do not have knowledge about the domain.

Table 5 presents the number of matches according to each rank position of the laptops chosen by participants. For some participants, more than one position matched. It can be seen that the number of matches is higher in the first positions. It means that when the specification provides good details of what users want, it is more likely that the exact laptop they want is matched.

Position Matched	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}
#Matches	30	17	11	8	2

Table 5: Domain Specialist Recommendation - Position Matched.

In order to test if the difference among the matches for the groups with different domain knowledge is statistically significant, we used the one-way ANOVA. The recommendations did not differ significantly across the five levels of domain knowledge, $F(4, 178) = 1.84816$, $p = 0.12163$.

Research Question 2. From the 192 participants, only 62 (32.29%) modified their preferences specification after experiencing choosing laptops and navigating through the catalog. This result differs from the claim that users usually construct their preferences as they see the available options (Pu & Chen 2008). Table 6 shows the participants that changed their preferences according to the domain knowledge and age. In addition, it presents the average number of changes (we explain how we counted it in the next research question).

Domain Knowledge	Who Changed (%)	#Changes (AVG)
NO_KNOWLEDGE	0 of 5 (0.00%)	0.00
BEGINNER	4 of 16 (25.00%)	2.50
INTERMEDIATE	14 of 40 (35.00%)	2.29
ADVANCED	28 of 83 (33.73%)	3.04
EXPERT	16 of 48 (33.33%)	2.13
Gender	Who Changed (%)	#Changes (Avg)
MALE	47 of 134 (35.07%)	2.53
FEMALE	15 of 58 (25.86%)	2.80
Age	Who Changed (%)	#Changes (Avg)
16-25 years	27 of 60 (45.00%)	2.07
26-35 years	25 of 83 (30.12%)	3.12
36-45 years	6 of 21 (28.57%)	2.33
>45 years	4 of 28 (14.29%)	3.25

Table 6: Preference Changes.

From these 62 participants, no one had NO_KNOWLEDGE about the domain. Even after searching laptops and seeing their features, these users were unable to describe preferences in terms of the laptop features – they did not know (and maybe did not want to know) what these features mean. A few BEGINNERS changed their specification, but they added high-level features such as “modern design” and “installed software”, and particular features learned from the catalog did not influence their specification. Approximately one third of the three remainder categories changed their specification. They understand the domain (some of them better), but not necessarily know the latest news (this was the main reason for changes made by EXPERTS). When they see new and updated features, or features they forgot to mention, they provide further details on they specification.

Analyzing changes and ages, the older the participant is, less changes she made. Older people made less detailed specifications (see research question 4), but still did not change them after going through the process of decision-making. However, when they changed their specification, they made more changes (the average number of changes grows as the age increases).

Research Question 3. We have analyzed all specifications that changed when they were reviewed and we have classified each change in the following way. Each change has a target and a type. There are three kinds of types: *add*, *remove*, or *change*. Also, there are three kinds of targets: (i) *Feature*: it describes a characteristic of the laptop, e.g. "HDMI"; (ii) *Feature value*: it describes the value of a feature, e.g. "Processor i5" changed to "Processor i5 or i7"; (iii) *Value*: it describes a high-level characteristic of the laptop, e.g. "Mobility". When a participant added a feature and its value in the preferences review, it was considered as a feature, because the feature would not make sense without a value. But if the participant only added a value to an existing feature, it was considered as add feature value. Figure 3 shows the occurrence of preference changes according to their nature (target and type).

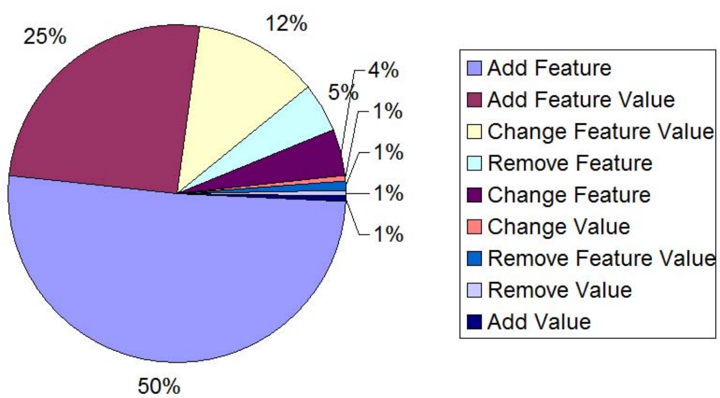


Figure 3: Nature of Preference Changes.

As it can be clearly seen, the three most common types of preference change that participants made in the preferences review are: (a) Add Feature (50%); (b) Add Feature Value (25%); and (c) Change Feature Value (12%). What happened was that users forgot to specify some characteristics that are important for them, or there is a new characteristic that they did not know about. At the moment they saw them in the laptop catalog, they remembered to specify them.

Moreover, some of the users were not aware of the current average or top values (price, processor, etc.) and as they know this by searching an up-to-date catalog, they realize that the value is different that they thought (it is mainly related to feature values). However, some participants specified feature values in terms of relative values ("second best value"), instead of absolute ones ("4GB"). Using this kind of specification makes the preferences specification reusable in different occasions.

Figure 4 presents how preference changes occurred distributed among the different domain knowledge categories. Even though the distribution of the three most recurrent preference changes are different among the different categories, they are still the most frequent categories. The only exception is in the BEGINNER category, in which 60% of the changes are of the type remove feature. However, it happened because a single participant changed the way he provided his specification, and therefore he removed the previously provided features and added a different kind of information (provided a specific model).

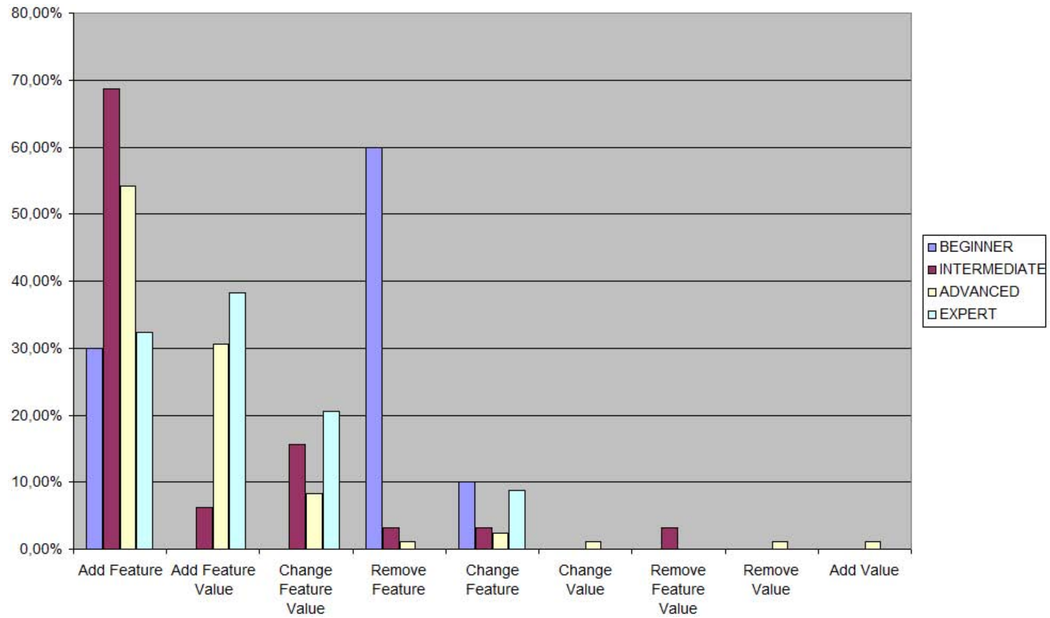


Figure 4: Preference Changes x Domain Knowledge (Percentage).

Research Question 4. The goal of this research question is to investigate how users with different knowledge about the domain express their preferences. As users that do not know too much about laptops are not aware of their features, they tend to use an alternative vocabulary in comparison with domain experts (higher domain knowledge). In other words, our research questions can be stated as: do users with higher knowledge about the domain express themselves with fine-grained features (e.g. laptop specific features) and users with lower knowledge with high-level features?

We analyzed each preferences specification and classified it in four different types, which take into account only the laptop specific features. In addition, we have identified particular characteristics and common patterns. Figures 5 to 8 present the charts that show the data collected (percentage) from the preferences specifications from our study. Figures show two perspectives from the specifications: (i) their type – Figures 5, 6(a), 7(a) and 8(a); and (ii) their characteristics – Figures 5, 6(b), 7(b) and 8(b). Figure 5 shows the results related to the whole group of participants, and Figures 6, 7 and 8 present the results related to the different groups of domain knowledge, gender and age, respectively. The four types of preferences specification are:

- *Basic* specifications mention characteristics for features that are part of every laptop (processor, RAM memory, hard drive, screen size). Characteristics can be specific values, or adjectives, such as “good” and “big”;

- *Brief* specifications do not cover laptop basic features (they mention none or few of them). Usually other kind of specification is provided, such as for what the participant will use the laptop;

- *Detailed* specifications give more details about laptops than the basic features, i.e. they are more specific, tending to narrow the laptops search space. We added to this category brief and basic specifications of *Apple* laptops, namely *Brief but Enough*, because participants who want laptops of this manufacturer, by describing only a few features, already indicate a unique laptop; and

- *No Description.* Some participants did not provide a specification but informed the specific model they wanted.

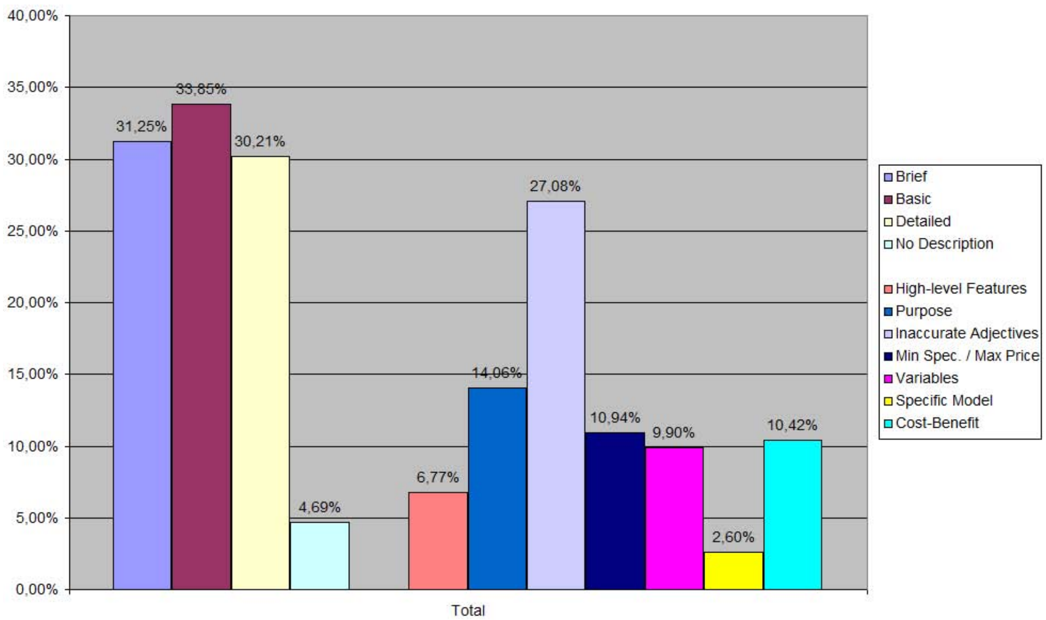
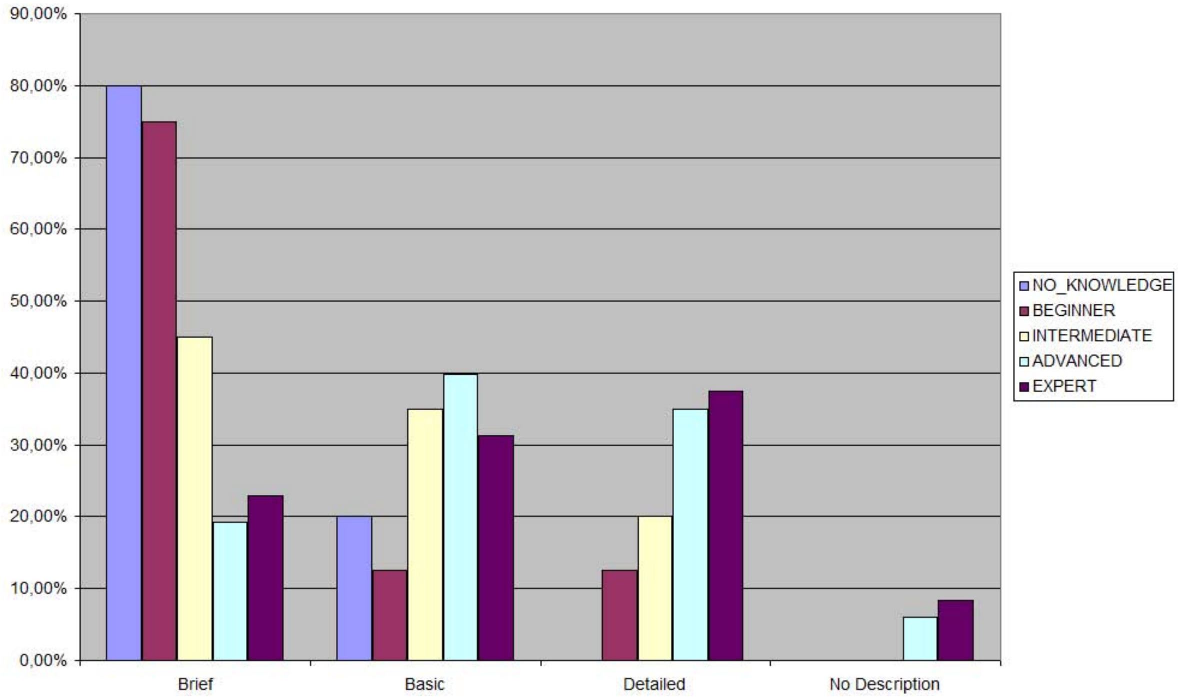
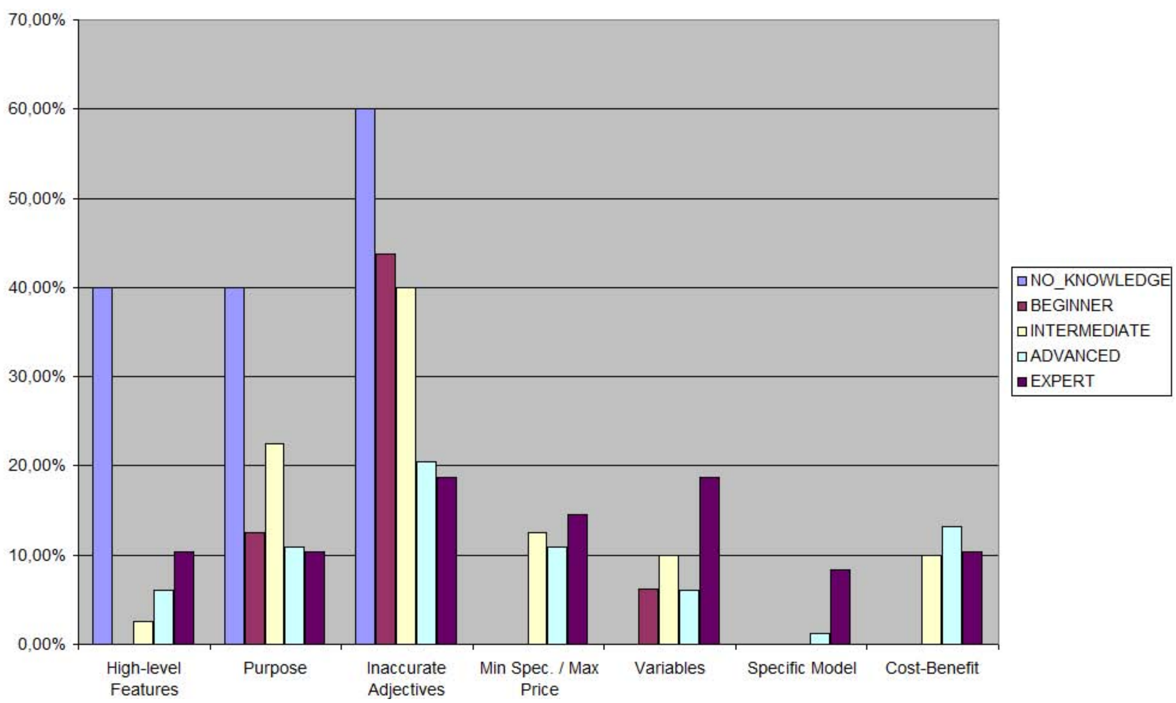


Figure 5: Preferences Specification Analysis.

As stated previously, this categorization has taken into account only how laptop specific features were described. We also observed other characteristics of the specifications, which are: (i) presence of *High-level Features*, which describe the consequences of having a value for a (set of) specific laptop feature, e.g. mobility, readability, performance; (ii) description of *Purpose* – specifications that contain for what the participant wants the laptop, for instance playing games; (iii) presence of *Inaccurate Adjectives*, which are adjectives subjective to the point of view of the participant, e.g. "good", "modern design", "beautiful"; (iv) *Minimum Specification/ Maximum Price* – pattern of specification that specifies a minimum specification for the laptop features and establishes a maximum price that the participant is willing to pay; (v) presence of *Variables*, which is when the participants used variables for feature values on their specification; (vi) *Specific Model* – specifications that do not describe laptops but indicate the specific model the participant wishes; (vii) *Cost-benefit* – participants that mentioned this characteristic on their specifications.

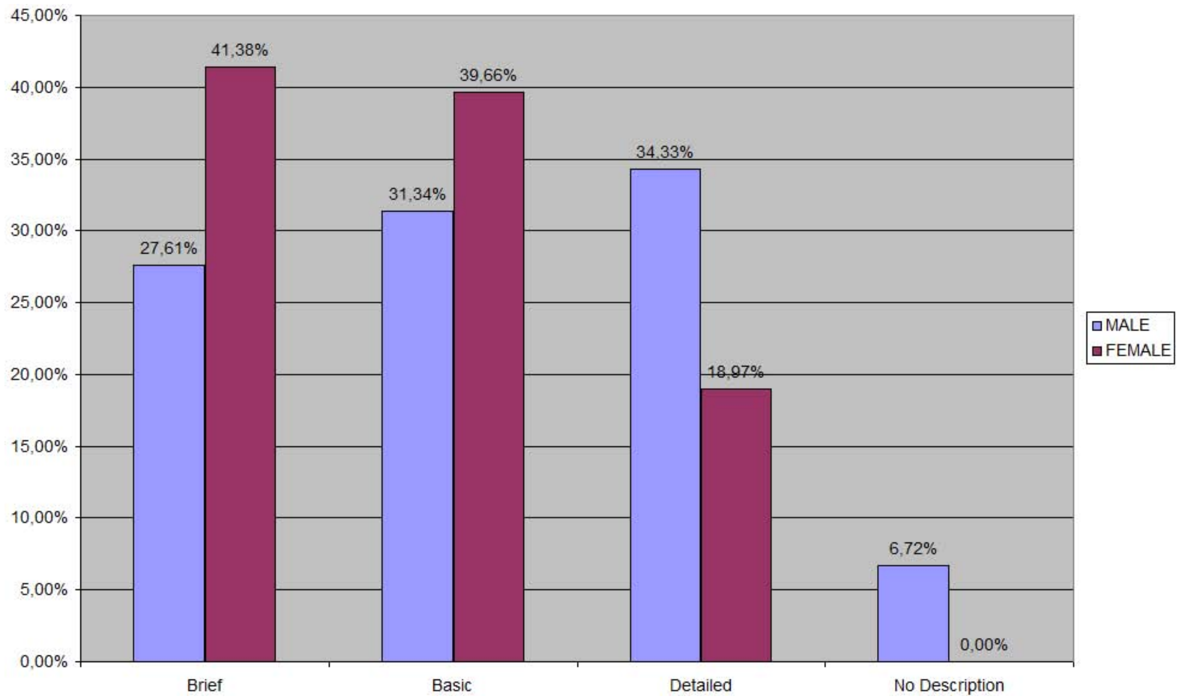


(a) Specification Types (percentage).

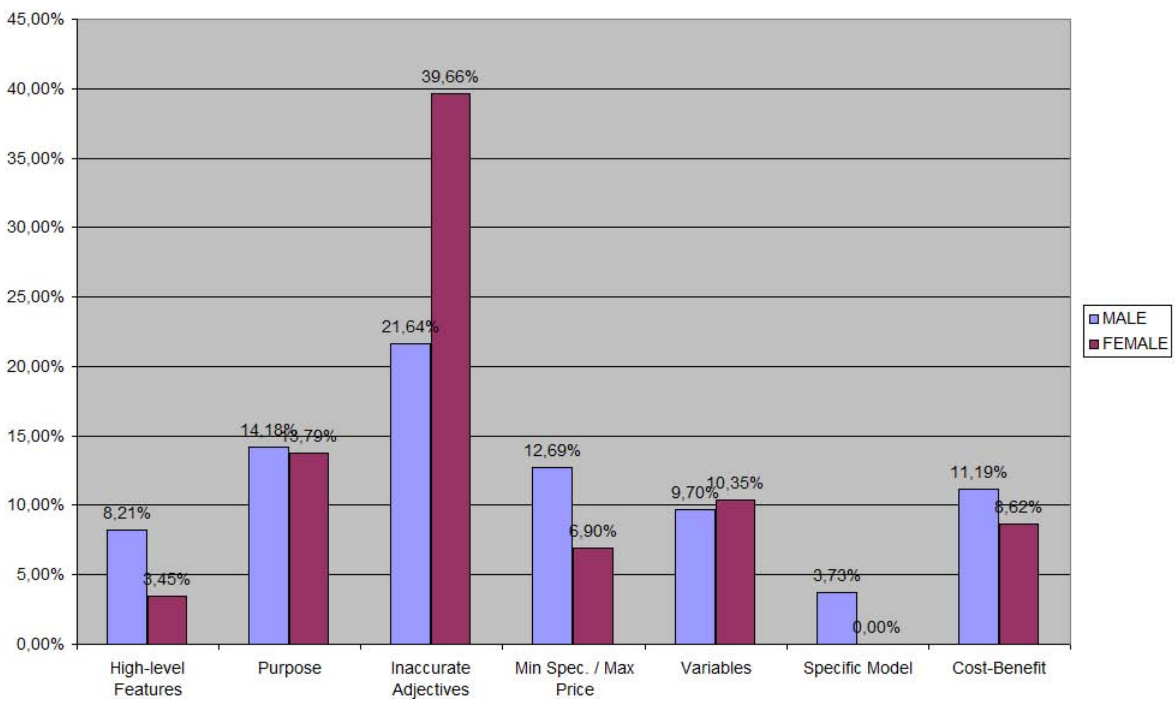


(b) Specification Characteristics (percentage).

Figure 6: Preferences Specification Analysis – Domain Knowledge.

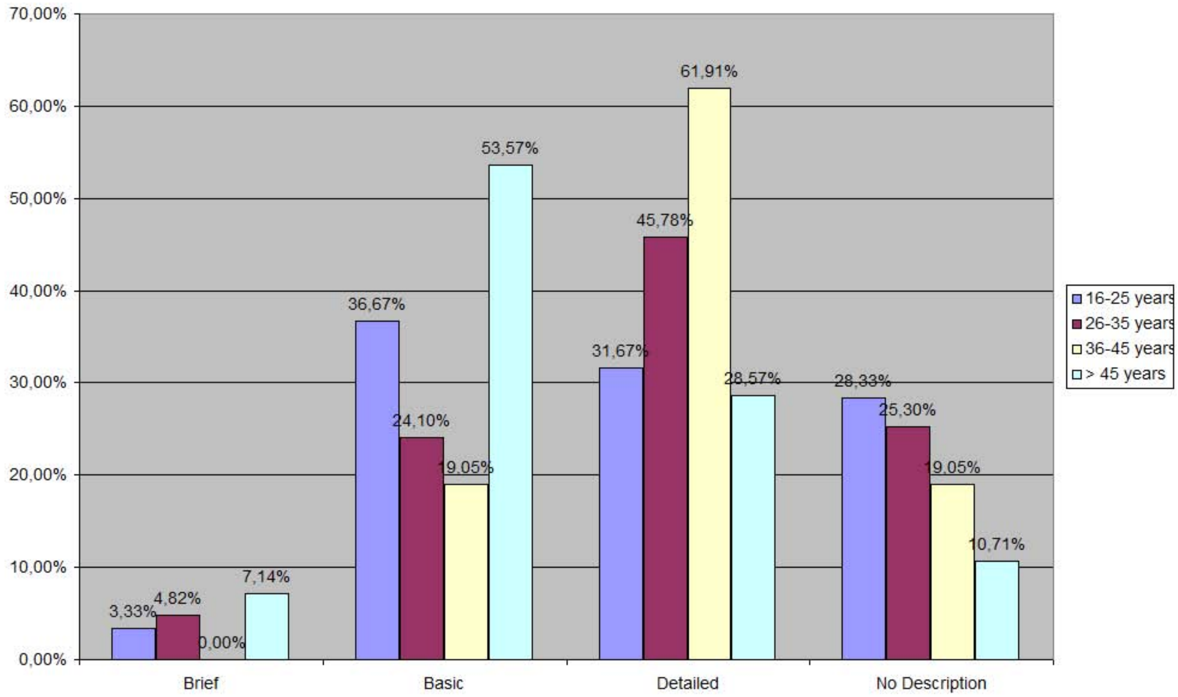


(a) Specification Types (percentage).

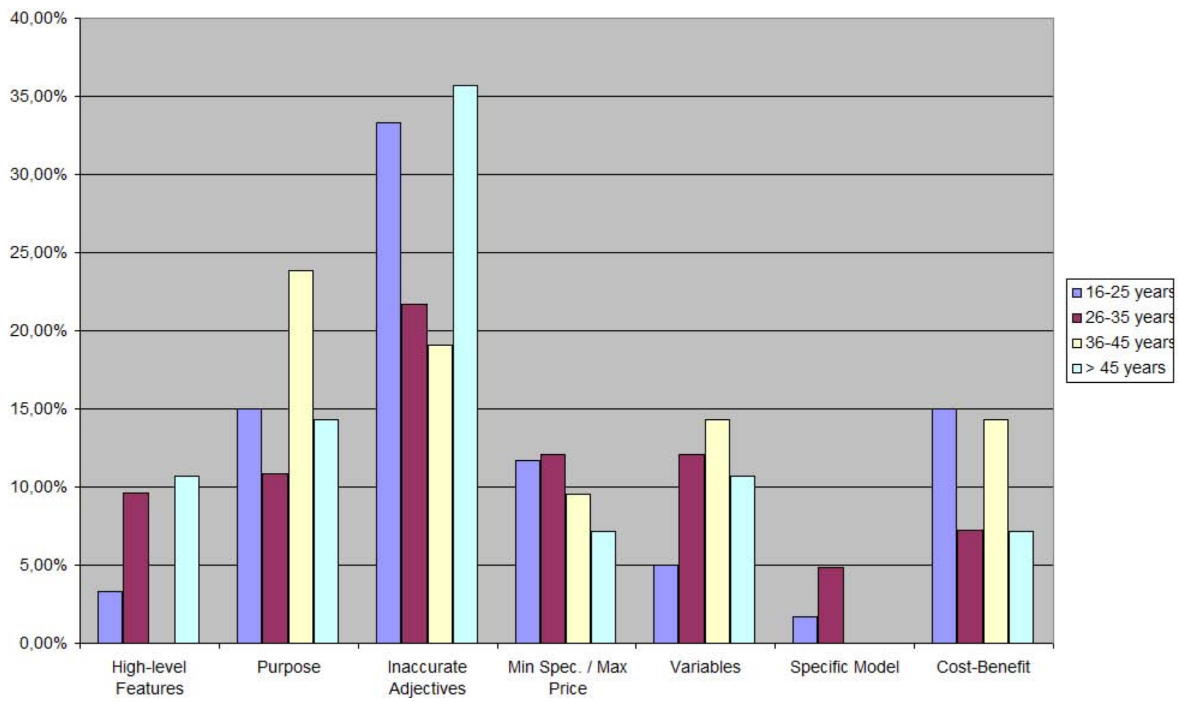


(b) Specification Characteristics (percentage).

Figure 7: Preferences Specification Analysis – Gender.



(a) Specification Types (percentage).



(b) Specification Characteristics (percentage).

Figure 8: Preferences Specification Analysis – Age.

It can be seen in Figure 6 that the behavior of the preferences specifications inverts as the knowledge about the domain grows. Participants with a lower level of knowledge specify their preferences without detailing too much about the specific features of the laptop. They use high-level features to describe what they want and for what they need a laptop. Some of them mention that they would ask a friend who understands about the domain for receiving a recommendation – these are not even interested in learning about the domain. On the other hand, participants with a higher level of knowledge were much more specific, stating the exact values (or a range) for most of the laptop specific features. The level of precision in defining feature values decreases as the domain knowledge increases, however even EXPERTS use inaccurate adjectives (18.75%).

Even though this specification is supposed to instruct an individual to execute a task for the participant, there is a certain degree of autonomy – choosing the laptop. Some participants (6.02% ADVANCED and 8.33% EXPERT) did not provide a specification but gave the exact model they want. One of the participants stated "I would never delegate such a decision for someone else." This shows a category of users that do not trust other parties to decide on their behalf (at least for certain tasks). However, there is still other kinds of support that could be provided, such as checking prices in different stores or making recommendations from among which users could choose and make the final decision.

We have attributed numbers for each type of preferences specification, in a scale ranging from 1 to 5 (Brief, Basic, Brief but Enough, Detailed and No description). Then, a one-way ANOVA was used to test for specification differences among five levels of domain knowledge. Preferences specifications differed significantly across the five levels of domain knowledge, $F(4, 187) = 7.12467$, $p = 2.32319e^{-5}$.

In this research question, we aimed at looking at how the domain knowledge influences the preferences specification, however we add an observation about the differences found according to the gender. Figure 7 shows that specifications provided by FEMALE participants are less detailed than the ones provided by MALE participants, in addition, they used inaccurate adjectives much more then MALE participants. However, this difference appears to occur due to their domain knowledge – 39.66% of the FEMALE participants has an ADVANCED or EXPERT domain knowledge and 80.60% of the MALE participants has an ADVANCED or EXPERT domain knowledge. Further investigations about it are outside the scope of this paper.

Research Question 5. Table 7 shows how much time (average, median and standard deviation) participants took for providing their initial preferences specification, according to their domain knowledge, gender and age. We have observed difference among the time taken by participants with different domain knowledge.

Participants with NO_KNOWLEDGE or BEGINNER domain knowledge took less time for giving their specifications. One reason is that their specifications are smaller than others. Second, their specifications contain details about for what they need the laptop or high-level specifications, which are details that may be easier to remember. The participants who took more time specifying what they wanted were the ones with INTERMEDIATE or ADVANCED knowledge. Their specifications are more detailed, but they did not promptly remember what they wanted (we observed that in the specification logs). Sometimes they went backward and changed or added details to their specifications. Finally, EXPERTS participants also provided detailed specifications, but as they are more familiar with the domain, their preferences have come easier to their mind.

		Specification Time		
		AVG	Median	STDEV
Domain Knowledge	NO_KNOWLEDGE	0:04:30	0:02:42	0:05:38
	BEGINNER	0:04:57	0:03:40	0:04:05
	INTERMEDIATE	0:06:35	0:05:12	0:05:11
	ADVANCED	0:06:45	0:05:49	0:05:28
	EXPERT	0:06:14	0:04:44	0:07:57
Gender	MALE	0:05:33	0:04:12	0:04:47
	FEMALE	0:06:44	0:05:21	0:06:27
Age	16-25	0:06:09	0:05:14	0:04:37
	26-35	0:06:32	0:05:17	0:05:53
	36-45	0:05:49	0:05:11	0:03:03
	>45	0:06:49	0:04:03	0:09:48

Table 7: Time Taken for Specifying Initial Preferences.

Research Question 6. Regarding the number of laptops chosen by participants, we can observe that no group has an average or a median less than three (see Table 8). It indicates that even when an individual knows very well the domain, there are different options that satisfy her needs. In addition, BEGINNER and INTERMEDIATE participants have a slightly higher average and median than the other categories of domain knowledge. Possibly, they do not care about minor details of the laptops, as ADVANCED and EXPERT participants.

		Options		
		AVG	Median	STDEV
Domain Knowledge	NO_KNOWLEDGE	3.40	3.00	1.67
	BEGINNER	3.88	5.00	1.54
	INTERMEDIATE	4.10	5.00	1.30
	ADVANCED	3.67	4.00	1.44
	EXPERT	3.21	3.00	1.53
Gender	MALE	3.74	4.50	1.49
	FEMALE	3.62	4.00	1.46
Age	16-25	3.60	4.00	1.55
	26-35	3.75	4.00	1.41
	36-45	3.76	4.00	1.37
	>45	3.43	3.50	1.55

Table 8: Number of Chosen Laptops.

The framework proposed in (Chen & Pu 2010) considers the objective accuracy as one of the criteria for evaluating recommender systems, which compares what the system recommended to the user best option. However, as our participants did not chose only one laptop, it might lead to the conclusion that such best option does not exist. In the field of marketing, it is more common to talk about "client satisfaction", which is more related to the perceived accuracy criteria of (Chen & Pu 2010).

Research Question 7. Besides storing the laptops chosen by participants, we have also logged their actions each time they executed one of these actions in order to analyze the steps participants take in the decision-process. Table 9 shows the data we have collected.

The catalog we presented for participants initially presented all laptops, with a short description of it and a small picture. Additionally, the following actions could be performed

		Steps		
		AVG	Median	STDEV
Domain Knowledge	NO_KNOWLEDGE	7.40	1.00	13.35
	BEGINNER	5.56	2.00	9.67
	INTERMEDIATE	3.15	1.00	4.91
	ADVANCED	3.75	2.00	4.46
	EXPERT	4.04	1.00	7.20
Gender	MALE	3.86	1.00	6.41
	FEMALE	3.98	2.00	6.08
Age	16-25	3.57	1.00	6.38
	26-35	4.05	1.00	5.97
	36-45	3.62	3.00	3.71
	>45	4.68	2.00	7.77

Table 9: Number of Steps Taken to Choose Laptops.

in the catalog: (i) Sort: laptops could be ordered according to the value selected (price, name, etc.); (ii) Filter: different filters (price range, brand, ...) could be added or removed, when the filter links are clicked; (iii) Show laptop details: by clicking on the laptop name, a new window was opened with the specification of the selected laptop; and (iv) Compare laptops: from two to three selected laptops could be compared (a table was displayed with laptop features side-by-side).

Table 4.3 shows that the standard deviation of each group is high. In means that within a group, there are participants that took much more steps to choose laptops than others. Observing the mean value, we see that the participants with less level of knowledge took more actions to choose their options. When users have a low knowledge about the domain, they need to search the catalog to learn about it.

Participants with NO_KNOWLEDGE executed random actions in the catalog, indicating that they had little idea about how to choose the laptop. BEGINNER and INTERMEDIATE participants asked much more to detail laptops, showing their exploration of the domain. And ADVANCED and EXPERT participants made an extensive use of filters. As they have a more precise idea of what they wanted, they reduced the search space in order to look only at the laptops they were interested. In case of applications that aid users on the decision-process, it is essential to give a personalized assistance that considers their domain knowledge.

We have used the one-way ANOVA to analyze the variance of data related to these last three research questions within the different domain knowledge groups. The test showed no significant difference in any of them: (i) Specification time – $F(4,187) = 0.44132$, $p = 0.77863$; (ii) Chosen Laptops – $F(4,187) = 2.21783$, $p = 0.06868$; and (iii) Number of Steps – $F(4,187) = 0.85659$, $p = 0.49116$. It means that even though we have identified small differences among groups of our study, users with different knowledge: (i) take about the same time to specify preferences, even thatey are specified using different criteria; (ii) they are equally precise in choosing different options; and (iii) they take about the same number of steps to make a decision.

6 Discussion

Even though the domain specialist was able to make a decision on behalf of our study participants for some of them, others did not provide enough details for the specialist accomplish that task. This result implies that: (i) there is a group of users that are able to express their preferences in such a way someone can make an appropriate decision on their behalf; and (ii) other users need help to specify their preferences. Based on these two groups, we identify different kinds of support for each of them: (a) a language that is expressive enough for users of group (i) to state their preferences; and (b) help for users of group (ii) to better express their preferences. In this section, we present a discussion related these two points we stated above.

Supporting the Preferences Expression. Research work on preferences elicitation has been reported different techniques for it. The kind of support we are looking at is not to elicit preferences from scratch, but to identify issues in preferences specified by users and help them to be more specific. According to the preferences changes of our study, we identified users do not provide wrong information, but incomplete or not updated in case of values that change over time. In such situations, information about the domain should be provided, such as features not mentioned, new features and updated values. However, this must take into account the domain knowledge of users so as not to annoy them with things they are aware about. Moreover, some of our participants provided templates of how they specify preferences about laptops, with variables for features that change over time. This can be really helpful for users having a starting point for their specification.

Our study also showed that users typically adopt inaccurate adjectives in their preferences statements, even when they are domain experts. A *good* video card has a different meaning for a user who plays games and another who watches movies. Therefore, these adjectives should be identified and scales be shown to users so they can rate what is "good" or "fast". But the point is to let users express themselves for getting better specifications later. The same situation happened frequently with the term "cost-benefit". Only one of the participants provided an accurate specification for that. A common issue is also dealing with subjective characteristics, e.g. "modern design", "beautiful'. In these cases, samples of groups of items could be shown to understand what the user means. Naturally, our approach does not exclude the help of learning algorithms as a complementary approach.

Providing Different Forms of Expressing Preferences. This second point is the one we have been working on (Nunes et al. 2010), which focuses on a domain-neutral user DSM, to represent user preferences as well as configurations, which are user settings on a system. The model can be instantiated for different applications. By analyzing the preferences specifications of our study, we have concluded that they are significantly different when provided by users with different degrees of domain knowledge. Yet, we could not reject the null hypothesis of research question 1: even specifications are indeed different, there is no significant difference in the domain specialist matches among the groups with different knowledge. Therefore, different forms of preferences expression must be provided to users, and they are equally important. We have already taken one step in this direction by providing users the concept of *value*, which refers to high-level features of a domain. Inaccurate adjectives must be part of the users' vocabulary, so as not to restrict their expression, and further mechanisms can be used to solve these issues, as previously discussed. In addition, specifications provided by less experienced users let the domain specialist make a good decision on their behalf, and such specifications contains preferences that could be reused over time, e.g. "I always look for the second best model." This is an interesting way to capture preferences about domains that evolve over time, as it is the case of laptops. Even though new features of laptops appear constantly, the process of looking for them and stating features relative to a reference point can be used as a pattern for future executions of the task.

7 Conclusion

This paper presented a user study, whose focus is to provide a deeper understanding about user preferences specification. We have investigated users' ability in expressing their preferences about domains that they might be familiar with or not, without having experienced a prior decision-making process in such domain. We have targeted the identification of the characteristics of preferences specifications provided by users with different degrees of domain knowledge, and how effective they were in order for a domain specialist to use those specifications to make decisions on the users' behalf.

Seven research questions were analyzed individually. Our main findings were that users with different knowledge about our study domain, laptops, provide different types of specifications – they are significantly different. Users with lower degree of knowledge mainly give high-level preferences and personal information, such as for what the laptop will be used. On the other hand, expert users provide information about fine-grained features. Despite these differences, domain specialists are able to provide recommendations of the same quality for all groups. Moreover, we observed users typically provide the right information about their preferences, but they might be incomplete or outdated for preferences whose values evolve over time.

Therefore, it is essential to provide a rich vocabulary for users expressing their preferences, including course and fine-grained preferences. In addition, mechanisms to help to eliminate subjectivity in specifications must be adopted. As future work, we are evolving our user DSM to contemplate the vocabulary of the preferences we collected in our study as well as to create a language based on this model.

Acknowledgments

We would like to thank our domain expert, Luis Fernando Barbosa, for taking the time to collaborate with this study. This work has been partially supported by CNPq 557.128/2009-9 and FAPERJ E-26/170028/2008. It is related to the following topics: Software technologies for web applications - A Multi-Agent Systems Approach for Developing Autonomic Web Applications - G1. Design techniques to improve the development of autonomic Web applications. Ingrid Nunes #141278/2009-9, Simone Barbosa #313031/2009-6, Carlos Lucena #304810/2009-6 also thank CNPq for respective research grants.

References

Anand, S. & Mobasher, B. (2005), Intelligent techniques for web personalization, *in* 'Intelligent Techniques for Web Personalization', Vol. 3169 of *LNCS*, Springer, pp. 1–36.

Basili, V. R., Selby, R. W. & Hutchens, D. H. (1986), 'Experimentation in software engineering', *IEEE Trans. Softw. Eng.* **12**(7), 733–743.

Berry, P. M., Donneau-Golencer, T., Duong, K., Gervasio, M., Peintner, B. & Yorke-Smith, N. (2009), Evaluating user-adaptive systems: Lessons from experiences with a personalized meeting scheduling assistant, *in* 'IAAI'09', USA, pp. 40–46.

Berry, P., Peintner, B., Conley, K., Gervasio, M., Uribe, T. & Yorke-Smith, N. (2006), Deploying a personalized time management agent, *in* 'AAMAS '06', ACM, Japan, pp. 1564–1571.

Chen, L. & Pu, P. (2010), User evaluation framework of recommender systems, *in* 'Workshop on Social Recommender Systems (SRS'10) at IUI'10'.

Glass, A., McGuinness, D. L. & Wolverton, M. (2008), Toward establishing trust in adaptive agents, *in* 'IUI '08', ACM, pp. 227–236.

Hu, R. & Pu, P. (2009), A comparative user study on rating vs. personality quiz based preference elicitation methods, *in* 'IUI '09', ACM, pp. 367–372.

Nunes, I., Barbosa, S. & Lucena, C. (2010), An end-user domain-specific model to drive dynamic user agents adaptations, *in* 'SEKE 2010', KSI, pp. 509–514.

Pu, P. & Chen, L. (2008), 'User-involved preference elicitation for product search and recommender systems', *AI Magazine* **29**(4), 93–103.

Schiaffino, S. & Amandi, A. (2004), 'User - interface agent interaction: personalization issues', *Int. J. Hum.-Comput. Stud.* **60**(1), 129–148.

Schiaffino, S. & Amandi, A. (2006), 'Polite personal agents', *IEEE Intelligent Systems* **21**(1), 12–19.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B. & Wesslén, A. (2000), *Experimentation in software engineering: an introduction*, Kluwer Academic Publishers, Norwell, MA, USA.

A Questionnaire

This appendix presents the questionnaire used in our study. The participant has the option of doing it either in English or in Portuguese.

A.1 Introduction: Survey about User Preferences

The purpose of this survey is to collect data that helps on understanding the user preferences expression. The survey is completely anonymous and all information collected will be used solely for statistical analysis within the context of this study. The survey has four steps and the estimated time for implementation of the survey is around 20 minutes.

Please, click on the image below to start the survey in English:

A.2 Part I: User Data

- Age: an a positive integer;
- Gender: a value from {Male, Female};
- Country: a value from a provided list of countries;
- City: a string;
- Working/Studying Field: a string;
- How many laptops have you already had (including current ones)? a value from {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10+};
- If you had(have) at least one laptop, did you yourself choose it(them)? a value from {All of them, Most of Them, Some of them, A few, No one};
- How do you rate your knowledge about which computer features to consider when buying a laptop? a value from {Expert, Advanced, Intermediate, Beginner, No knowledge};

A.3 Part II: Preferences Specification

Suppose that you want to **buy a new laptop** and somebody is going to buy it for you. You are going to specify all preferences and restrictions to this person, who will buy the laptop for you with no further communication after the initial specification.

We present below a simple example of a preferences specification on the flight domain.

Example

1. I like to minimize the price, I always pay promotional fares.
2. I don't like making connections.
3. I prefer the shortest flying time as possible, as long as I have at least one hour to make connections.
4. Flying time and number of connections are more important to me than the price.

Please, write down below the specification that you would provide to this person so he/she can buy the **laptop** for you.

[text area in which participants write their specification]

A.4 Part III: Options Selection

Now, let's assume that all laptops available for you are the ones listed below. Please, indicate which laptop would you choose. You can rank up to five options (at least one is required).

If you notice, during this selection process, that your previously specified preferences are incomplete, please do not go back. You will have the chance to review your preferences in the next (and last) step.

1. Option 1: {laptop list};
2. Option 2: {laptop list};
3. Option 3: {laptop list};
4. Option 4: {laptop list};
5. Option 5: {laptop list}.

Laptop Catalog

Use the catalog below to choose your laptop options. In the selection boxes above, laptops are identified by their SKU number. In addition, laptop names are also displayed. Instead of selecting a laptop manually, you may also click on the select button of the chosen laptop, which will be selected in the first empty select box.

The following actions can be performed in the catalog:

- *Sort*: laptops can be ordered according to the value selected in the box "Sort by";
- *Filter*: different filters (price range, brand, ...) can be added or removed, when the filter links are clicked;
- *Show laptop details*: by clicking on the laptop name, a new window is opened with the specification of the selected laptop; and
- *Compare laptops*: you can selected 2 or 3 laptops to be compared. After selecting the laptops, click on the "Compare" button, and a new window will be opened with a comparison table.

Obs. All prices are in American dollars.

[the laptop catalog]

A.5 Part IV: Preferences Specification Review

After choosing the laptops from the previous page, would you have specified your preferences and restrictions in a different way? If so, please make the necessary modifications in your specification. Please, note that you do not know about the available laptop options while making this specification.

Initial Preferences Specification

[initial preferences specification provided by the participant]

Chosen Laptops

You may click on the name of the laptops to see their details.

[the up to five laptops chosen by the participant]

Reviewed Preferences Specification

[text area in which participants write their reviews specification – it is initialized with the initial specification]