An Experiment to Evaluate how to Better Present User Models to the Users

Fabiana Vernero, Alessandra Petromilli, Federica Cena and Cristina Gena

Dipartimento di Informatica, Università di Torino
Corso Svizzera 185, Torino, Italy
{vernero,petromilli,cena,gena}@di.unito.it

Abstract. The externalisation of user models may allow users to understand how a user-adapted Web application makes its adaptation decisions, enabling them to inspect and modify the values stored in their user model. When externalising a user model both the underlying representation of the user model and the visualization used to present it have to be taken into account. In this paper we present a study aimed at evaluating different ways to represent and visualize user models. We use a social recommender system in a cultural events domain (iCITY) as case study. To our purposes, we conducted two experiments: i) a large between-subjects on-line evaluation aimed at confronting different representation and visualization modalities; ii) a within-subjects experiment aimed at confronting the same experimental condition.

1 Introduction and Related Work

Usually, user models tend to be hidden and out of the user access and control [9]. However, many systems have started to involve users in the maintenance of their model, especially in educational context, for example by enabling them to edit it [8], or to negotiate the contents of the learner model with the system [7, 4].
Open user models are models of the users that are available for viewing, and sometimes maintaining them by the users themselves (and sometimes also by other users, such as peers and teachers in educational context) [1]. A further step is the *scrutable user model* [9, 10], an open user model containing not only the user model data but also the evidence about how such data have been derived by the adaptive system. A *transparent system* [6] allows the user to understand the way it works and explains system choices and behaviour. Understanding, accepting and trusting a personalisation system may additionally improve the user-system interaction.
Bull and Kay [1] sustain that a model should be available in a form similar or identical to its underlying representation for greater accuracy. However in case of complex representation the similarity is not mandatory. What is important is that the user might understand the model. Thus, in case of complex underlying representation, a simpler representation, and consequently visualization, could be preferred.
Concerning the representation of the user models, and the visualization used to present them, several solutions have been proposed in the past. The visualization of user models can take a simple textual form such as in ELM-ART [17], Personis [11], SIV - Scrutable Inference Viewer [12]. Other systems visualize the user model content in

graphical ways. Some systems use very simple and intuitive visual representations, such as *sliders* (LOZ [13]), *emoticons* (Subtraction master) [3], *stars* (UMPTEEN [2]), *colors* (The Fractionator [2]), *bar charts* (PSAT/NMSQT [19]. In other cases, the information presented can be more complex, such as a graphical externalisation in the form of a Bayesian network [18], a hierarchical tree structure (Viewer[9]), a conceptual graph [7], multiple views (Flexi-OLM [2]). Finally, other systems exploit special metaphors such as *magic wands* (Wandies [2]) or *cups* (INSPIRE [15]).
In this paper we present a set of evaluations aimed at identifying the best modality of representation and visualization of an open user model in an existing social recommender system in a cultural events domain, iCITY[1]. The paper is structured as follows. In Section 2 we introduce our motivations and background. In Section 3 we present our experiments, describing in detail how we conducted the evalations, and which results we obtained. Section 4 presents the conclusions we draw from the experiments.

2 Motivations and Background

As highlighted by Norman [14] it is important to explain *why* and *on what basis* an application shows an adaptive behaviour. Knowledge about the inner working of an application helps users in interpreting the answers it provides, especially when personal data are manipulated. In order to reach these goals, we decide to externalize the user model of iCITY [5], a social recommender system in a cultural events domain which integrates adaptivity principles with Web 2.0 social features. In iCITY users are allowed to publish and share their own events, as well as rating, commenting, bookmarking and tagging other content; moreover, part of the events are provided via RSS-feed by the Turin Municipality. Great emphasis is also put on social networking. As regards adaptation, events are recommended according to their estimated interest for a certain user, balanced with their average rating and also considering the event date and location.
In iCITY, the user model maintains different types of information, such as user level of participation, user skills, and user preferences for the classes of the domain taxonomy[2]. A probability distribution of user interests is associated with each class of the taxonomy. Notice that the values in a probability distribution always sum up to 1. This means that if the value expressing user interest in a class increases, the values representing her interest in the other classes at the same level of the taxonomy proportionally decrease, in such a way that the sum remains equal to 1. For example, considering only two classes, if the user level of interest is 0.2 in “Music” and 0.8 in “Cinema”, and the user changes the first value to 0.3, then her level of interest in “Cinema” should be changed to 0.7.
We have conceptualized this representation of the user model interests overlaying the domain as *“relative representation”*. Given that, our aim is to find the better way to represent this conceptualization to the user, and to allow the user to modify her model maintaining such probability distribution[3]. As far as we know, all the systems in the lit-

[1] http://www.icity.di.unito.it/dsa-en/
[2] Regarding preferences, the iCITY user model overlays the domain model
[3] So far, the section of iCITY user model open to the users regards the user preferences, visualized in a plain textual way. The user is not allowed to modify her preferences (see for details [5])

erature (see Section 1) make use of *"absolute representation"*, wherein the user model values are presented in a scale, and each value is independent from the others. Even if an order can be derived and distances can be measured, such relations are not explicit to the user, who assessed each element separately. None of the reviewed past systems presents values by means of either an *"ordered representation"* (wherein the user model values are ranked in a list) or a relative representation. In an ordered representation, relations such as "superior to" and "inferior to" can be established among the various items. However, it is impossible to measure the "distance" or "difference" between two elements, as well as to assign the same value to different elements, unless co-winners are allowed. Finally, with a relative representation, users have to explicitly assess both order relations and distances among elements. The relative representation provides therefore more information: in fact, not only the relative values express how much a user is interested in a certain topic (or she knows about it) but they also highlight the relation among different elements. For example, the statement: "I assign 20% of my interest to literature and 80% to sport" is more informative than both "I like sport more than literature" (ordered values) and "I rate sport 100 out of 100 and I rate literature 25 out of 100"(absolute values).

However we hypothesized that, even if more informative, the relative representation conveys a more elaborated conceptual model of the inner representation of the system. This could increase the cognitive load of the users. On the contrary the other two representations seem to convey and represent concepts in a way easier to be comprehended by the final users.

Moreover also the visualization modalities chosen to present the user model can have an impact on the comprehensibility of the user model itself. We hypothesized that visualizations which are more often used in social web sites will be the most appreciated since users are already familiar with them. Thus, we decided to perform a set of experiments with the aim of verifying such hypotheses. More in particular, we wanted i) to investigate which graphic visualization, for each of the three representations, is the most understandable and usable. In particular we want to discover which metaphors could be used to better convey the ideas underlying the three representations; ii) to verify whether the "relative representation" is more informative for the user but also more elaborate to manage and to understand in comparison to the absolute and the ordered representation. Furthermore, we wanted to verify if user features (and in particular age, gender, education) influence in some way her preferences in user model externalisation, both regarding the modalities of value representation (ordered, absolute, relative) and the modalities of visualization (e.g., sliders, bar charts, etc). Finally, we wanted to verify whether users really feel useful to inspect and modify their models.

3 The evaluations

With the goal of investigating user preferences both for the user model representation and visualization modalities, as well as user opinions about user model externalisation, we performed in parallel two experiments: i) a large between-subjects on-line evaluation, where we compared different visualizations, given a certain representation modality, and ii) a within-subjects experiment, which allowed an in depth-evaluation and a

comparison of the three representation modalities and their corresponding visualizations. In designing the user interfaces used to convey the different visualizations, we took inspiration from the user model of our case-study application. Therefore, we represented user preferences with respect to different categories of events, which correspond to the classes in the taxonomy of iCITY (Appointments, Cinema, Art, Music, Books, Theatre).
The two experiments involved 9 visualizations, implemented as dynamic web pages by means of JavaScript, Ajax, and PHP scripting. Notice that some of them were selected thanks to a previous pilot study based on the paper-prototyping technique. Visualizations were divided into three homogeneous groups, based on the user model representation (see Figure1):

1. ***Ordered representation***
   - *The list*: preferences are represented as an ordered list, sortable at will;
   - *The podium*: each category is represented by a sphere, positioned on a certain step of a podium, according to the level of interest. Preferences can be modified by moving the spheres;
   - *The medals*: preferences are on list where the order is indicated by means of gold, silver and bronze cups and medals; the order can be modified by sorting the names of the categories;
2. ***Absolute representation***
   - *The stars*: each category can be awarded from a minimum of zero to a maximum of 5 stars;
   - *The sliders*: preferences can be adjusted by means of sliders;
   - *The tag cloud*: preferences are represented as tags in a tag cloud: the bigger a tag, the higher the level of interest; preferences can be modified by increasing or reducing the size of the tags;
3. ***Relative representation***
   - *The coins*: each category is associated with a box containing some coins. Preferences are represented by the number of coins; there is a fixed number of coins. Preferences can be modified by moving coins;
   - *The bricks*: user interest in a category is represented by a pile of bricks - the higher the pile, the higher the level of interest. Preferences can be modified by moving the bricks from one pile to another;
   - *The pie chart*: each category is represented by a slice in a pie chart. Preferences are the size of the slice and they can be modified adjusting it.

3.1 On-line Evaluation

The first evaluation was carried out as an on-line test aimed at evaluating the proposed user model visualizations with a large number of users. We wanted to discover i) which visualization is the most appreciated, given a particular user model representation, ii) whether users actually appreciate the possibility to inspect and modify their user models and iii) if significant correlations exist between demographic features and user preferences in visualizations.

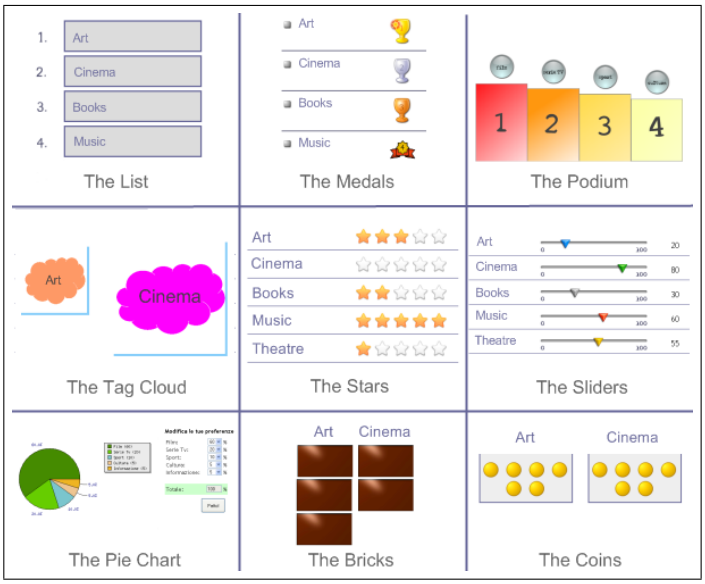

Fig. 1. The figure shows the main features of the visualizations used in the two experiments

Hypothesis. We hypothesized that visualizations which are more often used in social web sites will be the most appreciated since users are already familiar with them. Moreover, we thought that users would prefer prototypical interfaces which give prominence to visual aspects and allow direct manipulation. For such reason, we hypothesized that the preferred visualizations will be the "list" for the ordered representation, and "the stars" for the absolute representation. Since there are no examples of relative representations in existing systems, in this case we hypothesized that only the input modality (textual input vs direct manipulation) will impact on user preferences; consequently, the "coins" could result as the preferred visualization. Moreover, we thought that users appreciate open user models and that demographic variables had some influence in determining preferences.

Experimental design. Multiple factors (user model representations; visualizations) between-subjects design.

Subjects. Subjects were users of Facebook[4] and were therefore familiar with social media, as iCITY. They were recruited among the contacts of the authors and randomly assigned to one of the three groups: in this way, we obtained 100 subjects for the "ordered representation", 96 subjects for the "absolute representation" and 103 subjects for the "relative representation" group (299 subjects in total, 16-65 years old, 133 females and 166 males).

Measures and material. User preferences for the different visualizations were collected through an online questionnaire, personalized according to the group. The 3

[4] Facebook (http://www.facebook.com/) is one of the most popular social networking web sites.

groups of user model representation, and their corresponding visualizations, were made available online.

Experimental tasks. Subjects first accessed a page displaying a short thank-you message and the instructions. They were explained that they would access a series of visualizations of their preferences with respect to different categories of events, as they could have been automatically inferred by the system. They were invited to examine the visualizations and to try to "reply" by modifying/correcting the values. Subjects were also informed that they would be asked to fill in an anonymous questionnaire. After that, users accessed 3 different pages, each one containing an interactive user model visualization. The presentation order of the visualization was randomly changed for each user.

At the end they filled in the questionnaire. The first 4 questions were aimed at collecting basic demographic data (gender, age, education and job); then, users had to indicate the best and the worst visualization, and to give reasons for their choices. In the following 3 questions, subjects had to select one or more adjectives to describe each visualization, choosing from this list: "easy to use", "difficult to use", "pleasant", "unpleasant", "comprehensible", "incomprehensible", "amusing", "boring". The last 3 questions investigated: i) the subjects' opinion about the possibility to correct the values in their user model (answers were collected by means of a 4-point Likert scalewhere the different steps -in ascending order- corresponded to "very negatively", "negatively", "positively", "very positively"; ii) whether subjects would bother to correct their preferences in everyday usage and iii) whether subjects preferred a system where they could modify their preferences or a "traditional" one.

Results. As far as the "*ordered representation*" group is concerned, 63% of the users chose "the list" as their favourite visualization, followed by "the medals" (20%) and "the podium" (17%); this distribution of values is significant ($\chi^2(2) = 28.42$; $p < 0.001$). More than half of the subjects indicated "the podium", as the least favourite one; the distribution of values for this variable is also significant ($\chi^2(2) = 38.36$; $p < 0.001$).

User opinions. The list results "easy to use" (86% of subjects), "comprehensible" (65%) and "pleasant" (25%). Notice that the adjectives "difficult to use" and "incomprehensible" were used only once, while almost the same small number of users (14 and 11, respectively) described "the list" with the two opposite adjectives "boring" and "amusing", suggesting that the corresponding underlying dimension is not relevant. The distribution of values for the description of the list is statistically significant ($\chi^2(7) = 282.31$; $p < 0.001$).

On the contrary, the podium is described with opposite adjectives by almost the same number of subjects: it is "difficult to use" for 23% of subjects, "easy to use" for 28%; "unpleasant" for 18%, "pleasant" for 16%; "boring" for 21%, "amusing" for 15%. The only exception is the "comprehensible-incomprehensible" dimension, where 25% of subjects chose the first adjective and only 5% the second one. The observed values for the description of the podium are significant ($\chi^2(7) = 19.01$; $p < 0.01$).

Finally, "the medals" visualization is strongly and positively characterized on both the "pleasant-unpleasant" and the "amusing-boring" dimensions, with 36% of subjects de-

scribing it as "pleasant" and 29% as "amusing"; the distribution of values for the description of this visualization is also significant ($\chi^2(7) = 57.78$; $p < 0.001$).

As regards the "*absolute representation*" group, there is no clearly defined favourite visualization, since 39% of the subjects expressed their preference for "the sliders", while 36% chose "the stars" and the remaining 25% the "tag cloud" ($\chi^2(2) = 5.1$; this value is not significant ($p > 0.05$). Notice, however, that users of social media, as our subjects are, are quite accustomed to expressing their preferences by means of stars and sliders, which in fact received most votes. Coherently, the most innovative interface in this group, the tag cloud, was indicated as the worst visualization by almost three quarters of the users (72%) and the distribution of values for the "least favourite" visualization is significant ($\chi^2(3) = 56$; $p < 0.001$).

User opinions. Both the sliders and the stars, similarly to "the list", have very high values for the adjectives "easy to use" (chosen by 66.7% of users for the sliders and 77% for the stars) and "comprehensible" (63.5% for the sliders, 57.3% for the stars) and a significant agreement around the adjective "pleasant" - notice that the preferences collected by the sliders and the stars, 36.5% and 49% respectively, are higher than those of the list. Moreover, a few users also described these visualizations as "amusing". The chi square values for the descriptions of both "the stars" ($\chi^2(7) = 194.54$; $p < 0.001$) and "the sliders" ($\chi^2(7) = 166.3$; $p < 0.001$) are significant.

On the contrary, for the "tag cloud", opposite adjectives obtained almost the same number of preferences (in fact, chi square test is not significant), with the negative adjective prevail for all dimensions. The most unbalanced dimension is "easy to use-difficult to use", with 38.5% of subjects choosing the negative adjective and only 19.8% the positive one; in contrast, the "amusing-boring" dimension is very balanced, with 28.1% of subjects choosing "boring" and 24% "amusing", suggesting that this visualization, although considered difficult by most subjects, can prove engaging to some users.

Quite surprisingly, the favourite visualization in the "*relative representation*" group is the "pie chart", with 47% of the preferences (the value distribution for the "favourite visualization" variable is significant with $\chi^2(2) = 17.67$; $p < 0.001$), while no clear winner can be identified, as far as the least favourite visualization is concerned: 44% of the users chose "the bricks", 30% the "pie chart" and 26% the "coins"; chi square value for the "least favourite visualization" variable is equal to 6.07, while the critical value of the chi square distribution is 5,99; therefore, even if significant, it is too close to the critical value to be definitely considered ($\chi^2(2) = 6.07$; $p < 0.05$).

User opinions. The "pie chart" is described as "easy to use" (44.7% of subjects) rather than "difficult to use" (23.3% of subjects); pleasant (21.4%), rather than "unpleasant" (10.7%) and "comprehensible" (49.5%) rather than "incomprehensible" (3.9%). However, it is considered "boring" (30.1%), suggesting that the input modality, which forces users to correctly define the different percentages so that they sum up to 100, is too demanding. The observed values for the description of the pie chart are significant ($\chi^2(7) = 79.84$; $p < 0.001$).

The coins are positively assessed on all dimensions, even if they are judged a little less "comprehensible" (41.7% of subjects) in comparison with the "pie chart"; the observed values for the description of the "coins" are significant ($\chi^2(7) = 118.73$; $p < 0.001$).

"The bricks" are described as "difficult to use", "unpleasant" and "boring", the distri-

bution of values for this description is also significant ($\chi^2(7) = 70.8; p < 0.001$).
Finally, almost all users declared to prefer a system where they can access their user model compared to a "traditional" one ($\chi^2(1) = 143.36; p < 0.001$) and that they would like to inspect and modify their preferences also in their everyday usage ($\chi^2(1) = 152.87; p < 0.001$).
A correlational analysis was also performed in order to discover correlations between demographic features and user preferences in visualization. However, no significant correlations were found, disconfirming our hypothesis of a relation.

3.2 Empirical evaluation

The second evaluation aimed at gaining a deeper insight about i) user preferences in specific visualizations; and ii) their opinion about the possibility to inspect and modify their models. With respect to the first experiment, we also have the goal to investigate iii) which type of user model representation (ordered, absolute or relative) is the most meaningful and user-friendly.

Hypothesis. In comparing different user model representations, we hypothesized that "relative" representations would be considered more difficult, but also more informative. The easiest-to-use visualizations should be those based on the "absolute representation", which is normally used for the externalization of user models. As far as goals i) and ii) are concerned, we expect to confirm the results of the previous experiment.

Experimental design. Multiple factors (user model representations, visualizations) within-subjects design.

Subjects. We selected a group of 28 subjects, 16-45 years old, 12 females and 16 males, among colleagues and students at the Computer Science Department, University of Turin, according to an availability sampling strategy[5]. All subjects were frequent Internet users, familiar with social media.

Measures and material. We measured user opinions by means of an on-line questionnaire. Oral comments were elicited through *thinking aloud* technique. Both the subjects' comments and their performance were recorded by means of a screen capture software, as a support for thinking aloud. The nine visualizations were made available online and shown to the subjects by means of a laptop computer.

Experimental tasks. The experiment, which took approximately twenty minutes to each subject, was carried out in a laboratory at the University, one subject at a time. After being welcomed, subjects were invited to sit in front of the computer, where they could read a short thank-you message and a text with the same instructions of the first experiment. Specifically, users were invited to read and modify their preferences with the proposed interfaces, "thinking aloud" if they felt comfortable with it. Also in this case, they were informed that they would be asked to fill in an anonymous questionnaire.

After that, subjects could autonomously access all the nine visualizations. These were clearly divided into the three groups (ordered, absolute, relative representation) and each visualization was displayed in a separate page. Notice that the experimenters

[5] Notice that, even though non-random samples are not statistically representative, they are often used in much psychology researches, as well as in usability testing, especially in early evaluation phases [16]

carefully observed the users, while they were interacting, without providing any explanations or suggestions, unless they were explicitly questioned.

Finally, subjects accessed an extended version of the previous questionnaire. In particular, 9 further questions were added, aimed at assessing the *task* of "reading and modifying one's preferences" by means of each visualization: they were based on 4-point Likert scales ("very easy", "easy", "difficult", "very difficult"). Notice that no intermediate, neutral option was provided, in order to force the subjects to express a precise opinion. Users were also asked to choose which type of user model representation (ordered, absolute, relative values) was the most meaningful to them.

Results. As far as the best visualization is concerned, users indicated "the stars" (25%), which were never mentioned as the least favourite one, either. "The list" (18%), "the podium" and "the medals" (both 14%) follow (See Table1). Although these data seem to suggest appreciation for well-known, commonly used visualizations and confirm the evidence collected in the first experiment, the chi square test relative to the distribution of values for the favourite visualization is not significant.
The least appreciated visualizations were "the pie chart" and "the tag cloud", both with 28.6% of votes. However, they were still indicated as the best visualization by 10.7% and 7.1% of users, respectively.

	list	**medals**	**podium**	**cloud**	**stars**	**sliders**	**pie chart**	**bricks**	**coins**
Favourite	5	4	4	3	7	2	2	0	1
Least favourite	2	5	0	8	0	2	8	2	1

Table 1. Distribution of values for the favourite and least favourite visualizations

	list	**medals**	**podium**	**cloud**	**stars**	**sliders**	**pie chart**	**bricks**	**coins**
Very difficult	1	1	0	0	0	2	1	10	0
Difficult	0	8	7	9	0	14	15	11	4
Easy	11	14	14	10	9	7	10	5	18
Very easy	16	5	7	9	19	5	2	2	6

Table 2. Distribution of values for the task evaluation

An analysis of the comments collected by thinking aloud highlighted that directly manipulating shapes in order to change their size, as in "the tag cloud" visualization, is considered intuitive by some subjects, but not precise enough, according to others (in particular, it seemed difficult to correctly perceive and manage the possible small differences among similar-sized objects). On the other hand, the "pie chart" allowed a very fine-grained control, which was appreciated by some users, but also seemed too cumbersome to others. In addition, remember that "the pie chart" emerged as the favourite

visualization in the "relative values" group, when it was evaluated with more users, in the first experiment. Finally, it is interesting to notice that "the bricks" visualizations, which only 7.1% of users indicated as the worst, was never mentioned as the favourite. The observed data about the "least favourite visualization" are statistically significant ($X^2(8) = 25,36; p < 0,01$).

Task evaluation. All users judged the task of reading and modifying their preferences with "the stars" as either "easy" or "very easy" and the observed data for the corresponding variable are significant ($\chi^2(3) = 31.14; p < 0.001$) - see Table 2. On the contrary, both "the bricks" and "the tag cloud" were judged as "difficult" or "very difficult" to use by more than a half of the subjects ($\chi^2(3) = 19.14; p < 0.001$ and $\chi^2(3) = 11.14; p < 0.05$, respectively). These two more"innovative" visualizations, which had received negative feedback also in the first experiment, may actually have appeared as more difficult to use in comparison with a standard, familiar interface such as that provided by "the stars". Moreover, thanks to direct observation and thinking aloud, we can notice that an additional difficulty may have been caused by the some drag-and-drop mechanisms we used to implement these visualizations, which resulted unfamiliar to some subjects (direct manipulation was introduced on the web with AJAX is not yet a standard). Such hypothesis is confirmed by the fact that also the podium, which makes use of the same drag-and-drop mechanism, was considered "difficult" by a quarter of the subjects - the distribution of values for the task evaluation are significant also for this visualization ($\chi^2(3) = 14; p < 0.01$)

User opinions. An analysis of the adjectives used to describe the visualizations confirms our idea that simplicity, ease of use and familiarity are fundamental in determining the subjects' preferences. Both "the stars" and "the list", the two most appreciated , are in fact described as "easy to use" and "comprehensible" by most users, as in the first experiment. It is interesting to notice that they are also considered "pleasant", but by far less subjects (14% of the subjects for the list, 18% for the stars). On the contrary, only 5% of subjects describe "the stars" as amusing, while this adjective is never used for "the list": apparently, this feature is less relevant than ease of use. The value distributions relative to the user opinions, both for "the list" ($\chi^2(7) = 101.45; p < 0.001$) and for "the stars" ($\chi^2(7) = 67.9; p < 0.001$) are significant.

Notice that "the tag cloud", although considered "difficult to use" by half the subjects, "boring" by 11.8% and "comprehensible" only by 7.8%, is also described as "amusing" by 13.7% people (one of the highest scores for this dimension): the corresponding value distribution is quite uniform and the χ^2 is not significant.

The "pie chart" is considered very "difficult to use" (33.3% subjects), "unpleasant"' (14%) and "boring" (21%); however, it scores well (12 preferences) as far as comprehensibility is concerned, suggesting that most problems are related to the cumbersome input modality, as previously hypothesized. In this case, the observed values for the user opinions are significant ($\chi^2(7) = 45.03; p < 0.001$).

Preferred user model representation. Subjects favoured "ordered representation" (46.4%) - the only group which contained no strongly disliked visualizations and a very successful one, i.e. "the list" - followed by "absolute representation" (32.2%)- the group containing "the stars", the most appreciated visualization, but also "the tag cloud", which was much criticized. The "relative representation" only obtained 21.4% of preferences,

since this group contained two visualizations which were much criticized, "the bricks" and, in particular, the "pie chart". Notice that this result partially disconfirms our hypothesis, since we supposed that users would favour the "absolute representation" as for ease of use and the "relative representation" as for its capacity to express rich information. "Ordered values" are the simplest user model representation, so the visuations belonging to this group probably require the least effort and time to users. However, the observed values for the preferred user model representation are not statistically significant ($\chi^2(3) = 2.64$). Finally, we notice that almost all users evaluated the possibility of inspecting and modifying their preferences in a positive way ($\chi^2(3) = 35.14$; $p < 0.001$), also declaring that they would prefer a system which offers such functionalities to a "traditional" one ($\chi^2(1) = 11.57$; $p < 0.001$) and that they would examine and correct their preferences also in their everyday usage ($\chi^2(1) = 17.28$; $p < 0.001$).

4 Conclusion

In this paper we described two experiments, and their results, aimed at verifying i) which visual metaphors used to present user models is more comprehensible for final users, given a specific user model representation, and ii) whether the "relative representation" is more informative than "absolute" and "ordered representation", even if more cumbersome. Regarding the *first experiment*, coherently with our hypothesis, the preferred visualizations are those which are commonly used in social websites, such as the stars and the sliders, for the absolute representation, and the list for the ordered representation. However, for the relative representation, the favourite visualization is the pie chart, disconfirming our *hypothesis* that users would prefer an easy-to-use, direct manipulation-based visualization, such as the coins. The pie chart is more demanding, but also more complete. Especially it allows more precise comparison between the values. Regarding the *second experiment*, our findings about the preferred user model representation are not significant, thus we will replicate this experiment with a larger sample, and only exploiting the visualizations that have obtained more success in the first experiment. This experiment was performed with a small user sample in order to collect a deeper insight in user opinions, which can be better reached through face-to-face interaction and with methods such as *direct observation* and *thinking aloud*. Comments collected through thinking aloud were particularly useful in order to confirm the idea, emerged in the first experiment, that the absolute representation, to which users are quite accustomed, is easy to understand and to use. However, the ordered representation is considered even easier. On the other hand, some users appreciated the visualizations based on the relative representation (the pie chart in particular), because they were more precise and allowed them to explicitly indicate relations among different categories. Noticed that in this experiment the pie chart was also indicated by a lot of users as least favourite visualization, so probably this visualization cause contradictory opinions. Therefore, our idea that the relative representation is more informative has been partially confirmed. We plan to conduct a further experiment with a larger sample, with the goal of statistically confirming these results. Notice that some of our findings may have been influenced by the specific interaction techniques we proposed. Some visualization allows direct manilation, which has been introduced only recently

in the web, and can be therefore unfamiliar to some users. However the list, which in the second experiment obtained a lot of preferences, allow direct manipulation. Probably in this case the kind of interaction proposed, even if it is not a standard interaction, is more intuitive than others. Thus this results suggest to carefully implement drag-and-drop mechanisms on the web.

To conclude, we must remember that what is important is that the user might understand the model. Consequently, if a system makes use of a complex internal representation, such as a relative one and if an effective and easy-to-use visualization cannot be designed based on such representation, the choice of a simple absolute or ordered representation, as far as externalization is concerned, can be the most appropriate.

References

1. S. Bull and Kay J. A framework for designing and analysing open learner modelling. In *Workshop on Learner Modelling for Reflection*, 2005.
2. S. Bull, M. Mangat, A. Mabbott, A.S. Abu Issa, and J. Marsh. Reactions to inspectable learner models: Seven year olds to university students. In *Proceedings of Workshop on Learner Modelling for Reflection, AIED*, pages 1–10, 2005.
3. S. Bull and M. Mckay. An open learner model for children and teachers: Inspecting knowledge. In *Level of Individuals and Peers, Intelligent Tutoring Systems: 7th Int. Conf*, pages 646–655. Springer-Verlag, 2004.
4. S. Bull and H. Pain. Did I say what I think I said, and do you agree with me? Inspecting and questioning the student model. In *Proceedings of AI in Education, AACE,*, pages 501–508, 1995.
5. F. Carmagnola, F. Cena, L. Console, O. Cortassa, C. Gena, A. Goy, I. Torre, A. Toso, and F. Vernero. Tag-based user modeling for social multi-device adaptive guides. *User Model. User-Adapt. Interact.*, 18(5):497–538, 2008.
6. H.S.M. Cramer, V. Evers, S. Ramlal, M. van Someren, L. Rutledge, N. Stash, L. Aroyo, and B.J. Wielinga. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Model. User-Adapt. Interact.*, 18(5):455–496, 2008.
7. V. Dimitrova. Style-olm: Interactive open learner modelling. *International Journal of AI in Education*, 13(1):35–78, 2003.
8. Kay J. Learner know thyself: Student models to give learner control and responsibility. In *AACE, International Conference on Computers in Education*, pages 17–24, 1997.
9. J. Kay. *A scrutable user modelling shell for user-adapted interaction*. PhD thesis, Basser Department of Computer Science, University of Sydney, Sydney, Australia, 1999.
10. J. Kay. Scrutable adaptation: Because we can and must. In *AH*, pages 11–19, 2006.
11. J. Kay, R. J. Kummerfeld, and P. Lauder. Personis: a server for user models. In *AH'2002, Adaptive Hypertext 2002*. AH'2002, Adaptive Hypertext 2002, 2002.
12. L. Li and J. Kay. Assess: promoting learner reflection in student self-assessment. In *Proceedings of Workshop on Learner Modeling for Reflection, to Support Learner Control. Metacognition and Improved Communication*, pages 32–41, 2005.
13. S. Mohanarajah, R. Kemp, and E. Kemp. Opening a fuzzy learner model. In *Proceedings of Workshop on Learner Modelling for Reflection, AIED*, page 6271, 2005.
14. D. Norman. *The Design of Future Things*. New York: Basic Books, 2007.
15. K.A. Papanikolaou, M. Grigoriadou, H. Kornilakis, and G.D. Magoulas. Inspire: an intelligent system for personalized instruction in a remote environment. In *In Proceedings of 3 rd Workshop on Adaptive Hypertext and Hypermedia*, pages 13–24, 2001.

16. Singleton A. Royce and Bruce C. Straits. *Approaches to Social Research (3rd Edition)*. New York: Oxford University Press., 1999.
17. G. Weber and P. Brusilovsky. Elm-art: An adaptive versatile system for web-based instruction. *International Journal of AI in Education*, 12(4):351–384, 2001.
18. J.D. Zapata-Rivera and J. Greer. Construction and inspection of learner models. In *Computer Support For Collaborative Learning CSCL 2002*, pages 495–497, 2002.
19. J.D. Zapata-Rivera, J.S. Underwood, and M. Bauer. Advanced reporting systems in assessment environments. In *Proceedings of Workshop on Learner Modelling for Reflection, AIED*, pages 23–32, 2005.