Scalable Exploration of Relevance Prospects to Support Decision Making

Katrien Verbert
Department of Computer Science
KU Leuven
Leuven, Belgium
katrien.verbert@cs.kuleuven.be

Karsten Seipp
Department of Computer Science
KU Leuven
Leuven, Belgium
karsten.seipp@cs.kuleuven.be

Chen He
Department of Computer Science
KU Leuven
Leuven, Belgium
chen.he@cs.kuleuven.be

Denis Parra
Dept. of Computer Science
Pontificia Universidad Católica de Chile
Santiago, Chile
dparras@uc.cl

Chirayu Wongchokprasitti
Department of Biomedical Informatics
University of Pittsburgh
Pittsburgh, PA, USA
chw20@pitt.edu

Peter Brusilovsky
School of Information Sciences
University of Pittsburgh
Pittsburgh, PA, USA
peterb@pitt.edu

ABSTRACT

Recent efforts in recommender systems research focus increasingly on human factors that affect acceptance of recommendations, such as user satisfaction, trust, transparency, and user control. In this paper, we present a scalable visualisation to interleave the output of several recommender engines with human-generated data, such as user bookmarks and tags. Such a visualisation enables users to explore which recommendations have been bookmarked by like-minded members of the community or marked with a specific relevant tag. Results of a preliminary user study (N=20) indicate that effectiveness and probability of item selection increase when users can explore relations between multiple recommendations and human feedback. In addition, perceived effectiveness and actual effectiveness of the recommendations as well as user trust into the recommendations are higher than a traditional list representation of recommendations.

CCS Concepts

•**Human-centered computing** → **Information visualisation; Empirical studies in visualisation;** *User interface design;*

Keywords

Interactive visualisation; recommender systems; set visualisation; scalability; user study

1. INTRODUCTION

When recommendations fail, a user's trust in a recommender system often decreases, particularly when the system acts as a "black box" [7]. One approach to deal with this issue is to support exploration of recommendations by exposing recommendation mechanisms and explaining why a certain item was selected [19]. For example, graph-based visualisations can explain collaborative filtering results by representing relationships among items and users [11, 3].

Our work has been motivated by the presence of *multiple relevance prospects* in modern social tagging systems. Items bookmarked by a specific user offer a *social relevance prospect*: if this user is known or appears to be like-minded, a collection of her bookmarks is perceived as an interesting set that is worth to explore. Similarly, items marked by a specific tag offer a *content relevance prospect*. In a social tagging system extended with a personalised recommender engine [12, 15, 4], top items recommended to a user offer a *personalised relevance prospect*.

Existing personalised social systems do not allow their users to explore and combine these different relevance prospects. Only one prospect can be explored at any given time: a list of items suggested by a recommender engine, a list of items bookmarked by a user, or a list of items marked with a specific tag. In our work, we focus on the use of visualisation techniques to support exploration of *multiple* relevance prospects, such as relationships between different recommendation methods, socially connected users, and tags, as a basis to increase acceptance of recommendations. In earlier work, we investigated how users explore these recommendations using a cluster map visualisation [20]. Although we were able to show the potential value of combining recommendations with tags and bookmarks of users, the user interface was found to be challenging. Further, the nature of the employed visualisation made our approach difficult to scale: in a field study, users only explored relations between a maximum of three entities. Due to these limitations, the effect of using multiple prospects could not be fully assessed.

In this paper, we present the use of a *scalable* visualisation that combines personalised recommendations with two additional prospects: (1) bookmarks of other users (a *social* relevance prospect), and (2) tags (*content* relevance prospect). Personalised recommendations are generated with four different recommendation techniques and embodied as *agents*

IntRS 2016, September 16, 2016, Boston, MA, USA.

to put them on the same ground as users (i.e., recommendations made by agents are treated in the same way as bookmarks left by users). We use the UpSet visualisation [9], which offers a scalable approach to combine multiple sets of relevance prospects, i.e. different recommender agents, bookmarks of users, and tags. We aim to assess whether the combination of multiple relevance prospects shown with this technique can be used to increase the effectiveness of recommendations while also addressing several issues related to the "black box" problem. In particular, we explore the following research questions:

- **RQ1** Under which condition may a scalable visualisation increase *user acceptance* of recommended items?
- **RQ2** Does a scalable set visualisation increase *perceived effectiveness* of recommendations?
- **RQ3** Does a scalable set visualisation increase *user trust* in recommendations?
- **RQ4** Does a scalable set visualisation improve *user satisfaction* with a recommender system?

The contribution of this research is threefold:

1. First, we present a novel interface that integrates a simplified version of the UpSet visualisation, allowing the user to flexibly combine multiple prospects to explore recommended items.
2. Second, we present a preliminary user study that assesses the effect of combining multiple relevance prospects on the decision-making process. We find that users explore combinations of recommendations with users and tags more frequently than recommendations only based on agents. Further, this combination is found to provide more relevant results, leading to an increase in user acceptance.
3. Third, we find indications of an increase in user trust, user satisfaction, and both perceived and actual effectiveness of recommendations compared to a baseline system. This shows the positive effects of combining multiple prospects on user experience.

This paper is organized as follows: first, we present related work in the area of interactive recommender systems. We then introduce the design of IntersectionExplorer, an interactive visualisation that allows users to explore recommendations by combining multiple relevance prospects in a scalable way. We assess its impact on the decision-making process and finish with a discussion of the results.

2. RELATED WORK

In a recent study, we analyzed 24 interactive recommender systems that use a visualisation technique to support user interaction [6]. A large share of these systems focuses on transparency of the recommendation process to address the "black box" issue. Here, the overall objective is to explain the inner logic of a recommender system to the user in order to increase acceptance of recommendations. Good examples of this approach are PeerChooser [11] and SmallWorlds [5]. Both allow exploration of relationships between recommended items and friends with a similar profile using multiple aspects.

In addition, TasteWeights [3] allows users to control the impact of the profiles and behaviours of friends and peers on the recommendation results. Similar to our work, TasteWeights provides an interface for such hybrid recommendations. The system elicits preference data and relevance feedback from users at run-time in order to adapt recommendations. This idea can be traced back to the work of Schafer et al. [17] concerning meta-recommendation systems. These meta-recommenders provide users with personalised control over the generation of recommendations by allowing them to alter the importance of specific factors on a scale from 1 (not important) to 5 (must have). SetFusion [13] is a recent example that allows users to fine-tune weights of a hybrid recommender system. SetFusion uses a Venn diagram to visualise relationships between recommendations. Our work extends this concept by visualising relationships between different relevance prospects, including human-generated data, such as user bookmarks and tags in addition to outputs of recommenders, in order to incite the exploration of related items and to increase their relevance and importance in the eye of the user. To do so, we employ a set-based visualisation that allows users to quickly discern relations and commonalities between the items of recommenders, users, and tags for a richer and more relevant choice.

Relevance-based or set-based visualisation attempts to spatially organize recommendation results. This type of visualisation has its roots in the field of information retrieval and was used for the display of search results. For example: for a query that uses three terms, this type of visualisation would create seven set areas. Three sets will show the results separately for each term. Another set of three will show results for any combination of two of these terms. Finally, one set will show results that are relevant to all three terms together. The classic example of such a set-based relevance visualisation is InfoCrystal [18]. The Aduna clustermap visualisation [1] also belongs to this category, but offers a more complex visualisation paradigm and a higher degree of interactivity. The strongest point of both approaches, however, is the clear representation of the query terms and their relevant items, separately or in combination.

In the context of similar work, the novelty of the approach suggested in this paper is twofold: first, we use a set-based relevance approach that is not limited to keywords or tags, but which combines these with other relevance-bearing entities (users and recommendation agents). The major difference and innovation of our work is that we allow end-users to combine *multiple* relevance prospects to increase richness and relevance of recommendations. Second, we present and evaluate the use of a novel scalable visualisation technique (UpSet [9]) to perform this task and thereby demonstrate this approach's ability to increase recommendation effectiveness and user trust.

3. INTERSECTIONEXPLORER

IntersectionExplorer (IE) is an interactive visualisation tool that enables users to combine suggestions of recommender agents with user bookmarks and tags in order to find relevant items. We describe the visualisation and interaction design of the system, followed by its implementation.

3.1 Set Visualisation Design

We have adapted the UpSet [9] technique to visualise relations between users, tags, and recommendations. UpSet

represents set relations in a matrix view: while columns represent sets of different entities (such as recommender agents or other users' bookmarks), rows represent commonalities between these (Figure 1). The column header shows the name of the agent, user, or tag. The vertical bar chart below the column headers depicts the number of items belonging to each related set. Set relations are represented by the rows. In such a row, a filled cell indicates that the corresponding set contributes to the relation. An empty cell indicates that the corresponding set is not part of the relation. The horizontal bar chart next to each row shows the number of items that could be explored for this relation set. For example, the first row in Figure 1 indicates that there are three items that belong to both the set of recommendations suggested by the bookmark-based recommender agent, and the set of recommendations suggested by the tag-based agent. The second row shows suggestions of the bookmark-based agent only, whereas the third row only shows suggestions of the tag-based agent. For the convenience of the reader, we also depicted this relation in a traditional Venn diagram to support the understanding of the concept.

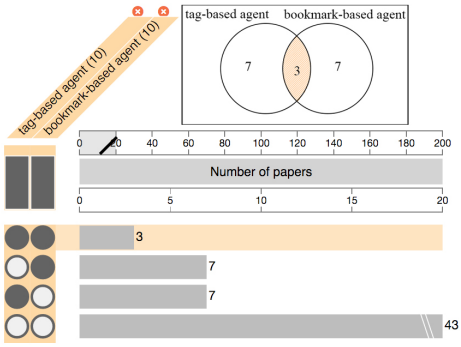

Figure 1: Set visualisation of IntersectionExplorer

One of the biggest advantages of a visual matrix is scalability. Whereas a Venn diagram can only display the intersections of a limited number of sets, the UpSet technique can present many sets in parallel, as only a single column has to be added to add another set to the visualisation. This greatly reduces space requirements while increasing the information density. The visual encoding of IE is identical for any number and constellation of sets. In practice, users may wish to first familiarise themselves with the display of a small number of sets, but due to the consistent and space-efficient design, they can seamlessly increase the set numbers without altering the view.

3.2 Interaction Design

An overview of the full IE interface is shown in Figure 2. The interface is separated into three connected parts. In the left part, the user can select different entities: agents, users and tags. If an agent is selected, the set of items suggested by this agent is added to the matrix visualisation in the canvas area. If a user is added, the set of bookmarks of this user is added. Similarly, if a tag is added, the set of papers marked with this tag is added to the view.

The canvas area represents user-selected sets as columns in a matrix view, allowing the user to explore overlaps between these sets. Each row represents relations between the different columns as explained in the previous section.

The user can explore the details of data items related to a certain row by clicking on the row. For example, after clicking the first row in Figure 2, the right part shows the title and authors of two papers that are bookmarked by "P Brusilovsky" and also suggested by three different agents.

The user can explore the items related to a specific set by clicking on the column header: all containing items of this set are then presented in the panel on the right. Meanwhile, the rows related to this set are also gathered at the top to facilitate exploration of relations with other sets.

At the top of the set view, the user can also sort the rows (set intersections) by *number of items* or *number of related sets* in ascending or descending order. The example of Figure 2 sorts the rows by the number of related sets in descending order. The first row represents items in the intersection of four sets. The second row represents items in the intersection of three sets and the next five rows represent items in the intersection of two sets. The other rows represent items related to a single set only.

3.3 Implementation

We have implemented IE on top of data from Conference Navigator 3 (CN3). CN3 is a social personalised system that supports attendees at academic conferences [14]. The main feature is its conference scheduling system where users can add talks of the conference to create a personal schedule. Social information collected by CN3 is extensively used to help users find interesting papers. For example, CN3 lists the most popular papers, the most active people, and the most popular tags assigned to the talks. When visiting the *talk page*, users can also see who scheduled each talk during the conference and which tags were assigned to this talk.

We use the list of conference talks as data items in IE. CN3 offers four different recommendation services that rely on different recommendation engines. The *tag-based* recommender engine matches user tags (tags provided by the user) with item tags (tags assigned to different talks by the community of users) using the Okapi BM25 algorithm [10]. The *bookmark-based* recommendation engine builds the user interest profile as a vector of terms with weights based on TF-IDF [2] using the content of the papers that the user has scheduled. It then recommends papers that match this profile of interests. Another two recommender engines, *external bookmark* and *bibliography*, are the augmented version of the *bookmark-based* engines [21]. The *external bookmark* recommender engine combines both the content of the scheduled papers and the research papers bookmarked by the user in academic social bookmarking systems such as Mendeley, CiteULike, or BibSonomy. Similarly, the *bibliography* recommender engine uses the content of papers published by the user in the past to augment the bookmarked papers.

The suggestions of these four recommender engines are represented as separate *agents* in IE. Users can explore which items are suggested by a single agent, for instance the tag-based recommender, but they can also explore which items are recommended by multiple agents to filter out the potentially more relevant recommendations. In addition, users can explore relations between agent suggestions and bookmarks of real users. As shown in Figure 2, the third row represents items suggested by the tag-based agent that have also been bookmarked by "P Brusilovsky", but that are not suggested by the two other agents and that have not been bookmarked by the active user ("K Verbert"). In this paper,

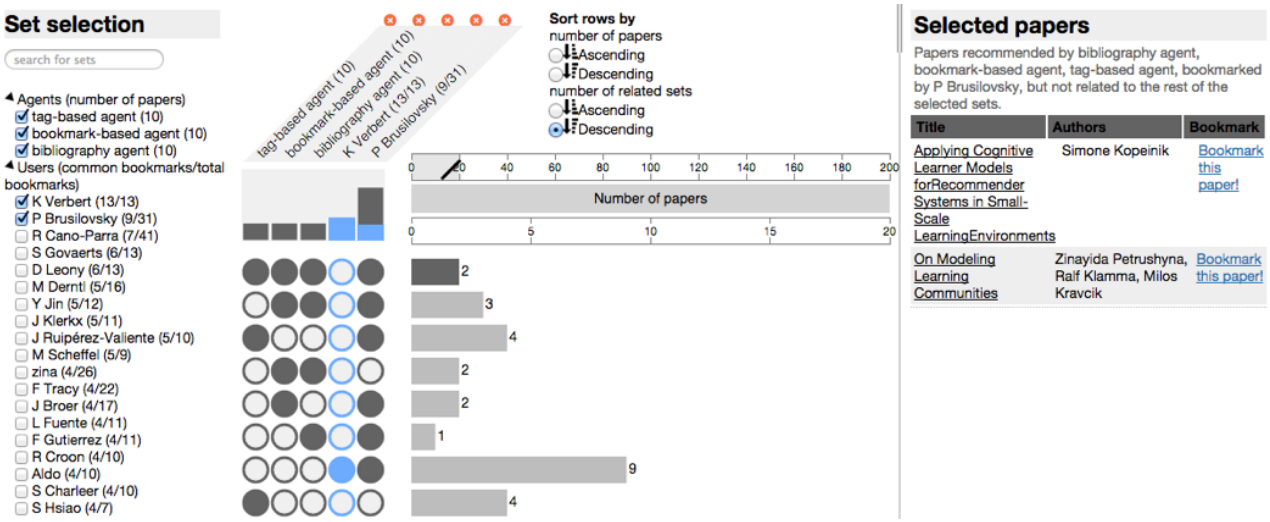

Figure 2: IntersectionExplorer visualises relationships between recommendations generated by multiple recommendation techniques (agents) and bookmarks of users and tags to increase relevance of recommendations.

we evaluate whether enabling users to explore relations between recommendations of different techniques, real users, and tags increases the acceptance of recommendations.

The set visualisation shows the relations of the selected sets as described in section 3.1. The column of the current user is displayed in blue while the other columns are represented in grey. As presented in Figure 2, the bar chart below the column headers of users overlays a blue bar that encodes the number of common bookmarks with the current user. The similarity between users is also represented next to the user name in the panel on the left: "P Brusilovsky (9/31)" indicates that the user "P Brusilovsky" has 31 bookmarks in total. Nine out of these 31 talks are also bookmarked by the active user ("K Verbert").

For the user study presented in this paper, we used the data from the EC-TEL conferences of 2014 and 2015. EC-TEL is a large conference on technology enhanced learning. We retrieved user bookmarks and tags of these conferences, and had access to the different recommender services for both the 2014 and 2015 edition of the conference. Attendees of the EC-TEL conference participated in the user study that is presented in the next section.

4. USER STUDY

To investigate to what extent the set visualisation may support users in finding relevant items, we conducted a within-subjects study with 20 users (mean age: 32.9 years; SD: 6.32; female: 3) in two conditions, both of which had to be completed by all participants.

In the first condition (baseline), users were tasked to explore recommendations presented to them using the CN3 "my recommendations" page with four ranked lists. In the second condition, users explored recommendations using IntersectionExplorer (IE). To avoid a learning effect, each condition used a separate data set from which to generate recommendations. The baseline condition (CN3) used the EC-TEL 2014 proceedings (172 items), the IE condition used the EC-TEL 2015 proceedings (112 items).

To prepare for the study, users bookmarked and tagged five items in each of the proceedings. In addition, users' publication history and academic social bookmark systems (CiteULike and Bibsonomy) were read. From the combined data, recommendations were generated in both conditions using the four different techniques described in Section 3. These were then presented as four individual agents: a tag-based agent, a bookmark-based agent, an external bookmark agent and a biography agent.

To explore the impact of the IE visualisation on the users' acceptance of items, users were tasked to explore the recommendations of the four agents freely and to bookmark five items. During this period we recorded the time and amount of steps taken to create a bookmark. In particular, we recorded the following actions: selection/deselection of agents, users and tags, sorting, hovering over a result row (if mouse position was held for more than two seconds), clicking onto a paper's title, and clicking the bookmark button. Further, we collected data using a think-aloud protocol, synchronizing screen recording and microphone input. Finally, users completed a questionnaire using a five-point Likert scale. The questions were based on ResQue [16] and the framework proposed by Knijnenburg et al. [8], both of which have been validated for the measurement of subjective aspects of user experience with recommender systems.

Before exploring the recommendations using IE, users were shown a three-minute video to explain the system's operation. In the IE view, users saw the intersections with the agents' recommendations. In the CN3 (baseline) view, users saw the full results of the bookmark agent and could navigate to the recommendations generated by the three other agents, as presented in Figure 3. The study was counterbalanced by mode of exploration (CN3/IntersectionExplorer). Five users completed the study with a researcher present in the same room, whereas 15 users completed the study via an on-line video call. To establish users' background-knowledge, we asked each participant a set of questions using a five-point Likert scale after the study. Mean results were as follows:

- Users were familiar with technology-enhanced learning

(mean: 4; SD: 1.1).

- Users were familiar with recommender systems (mean: 4; SD: 0.95).
- Users were familiar with visualisation techniques (mean: 4.05; SD: 0.86).
- Users occasionally followed the advice of recommender systems (mean: 4.25; SD: 0.77).
- Eight participants had never heard of CN3 before. Twelve had heard of it, but had no particular familiarity with the system (mean: 3.25; SD: 1.13).

One user had no publications, four had two to four publications, fifteen had five publications or more. Within the last group, 93.3% had published on an EC-TEL conference in the past.

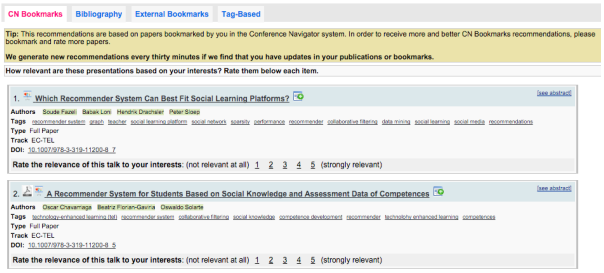

Figure 3: CN3 baseline interface with four ranked lists provided by four recommender engines.

4.1 Results and Evaluation

4.1.1 *Quantitative results*

The main focus of this study was to investigate under which condition the visualisation may increase user acceptance of recommended items. To answer our question, we need to analyse the in-depth behaviour of users exploring the recommendations using various combinations of recommender agents and the bookmarks and tags of other users.

In order to be able to determine the impact of visualising relations between agents, users, and tags, we defined two measures: effectiveness and yield.

Effectiveness measured how frequently the exploration of a specific set providing a number of intersections (henceforth called 'size') eventually led to the user bookmarking another paper (from the recommended set of papers). By the exploration of a set we mean clicking on a row of intersections in the visualisation (Figure 1, Figure 2) to show the items belonging to the intersection of the selected sets.

Effectiveness was calculated as the number of cases where the exploration of an intersection of a specific type and size resulted in a new user bookmark, divided by the number of times this intersection type and size was explored. Intersection types could be a single agent, a combination of agents, or a combination of agents with another entity (user or tag). The size represented the number of sets in the intersection. For instance, users explored suggestions of a single agent 26 times in total (one agent, Figure 4, first row). Exploration of these sets resulted in the creation of five bookmarks. Thus, the single agent's effectiveness is 5/26 = 19%.

Yield measured the fraction of items of an explored set that were actually bookmarked by all users in total. For instance, if the results of the intersection with one agent listed a total of 93 items for all users combined, but only five bookmarks were created from this type and size of intersection across the whole study, its overall yield was 5/93=0.05 (Figure 5, first row).

Figure 4 and Figure 5 reveal an interesting effect: sets which included the recommendations of agents and other entities, such as other users' bookmarks and tags, appeared to have a higher yield and effectiveness than sets based on agent recommendation alone, even if the number of intersections were the same. To further explore this aspect, we divided the results for effectiveness and yield into two groups: those obtained for interaction with one to four agents, and those obtained from interaction with the recommendations of different numbers of agents and another entity (user or tag). A Friedman test indicated a significant effect of recommendation source on effectiveness, $c^2(1) = 4$, $p = .046$ revealing that users who explored the recommendations of agents combined with another entity in the recommendation matrix of IE (median: .43), tended to find more than twice as many relevant items as when only using the agents for the recommendation (median: 0.21) (Figure 4). These results correspond to our findings that the richer the set (the more "perspectives" contribute the recommendation), the higher the yield (Figure 6). In general, Figure 7 shows that the larger the amount of intersections with a specific type, the higher the yield. Pearson's correlation showed a positive correlation between the number of intersections and yield (r = .839, n = 6, p = .037).

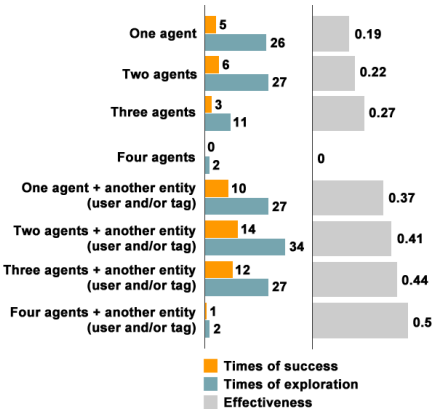

Figure 4: Effectiveness of the combinations of various amounts of agents and the combinations of various amounts of agents and other entities, such as users or tags. Effectiveness was higher when agents were combined with another entity.

Overall, these results suggests that enriching automated recommendations based on tags, previous bookmarks, publication history and academic social bookmarks with *socially collected* relevance evidence, such as the bookmarks made by other users of the same conference or a tag, greatly increases the relevance of recommendations, resulting in a higher acceptance rate.

Regarding the overall operability of IE, an ANOVA of task completion time showed an effect of task number $F(4, 44) = 20.5$, $p < .001$ on interaction time. However, a post-hoc

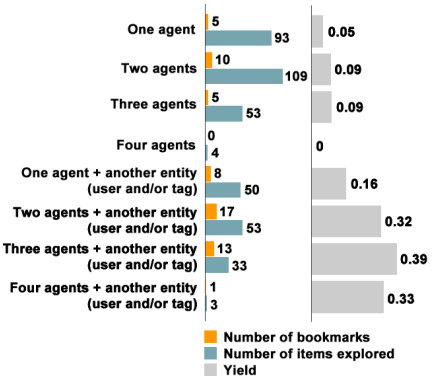

Figure 5: Yield of the combinations of various amounts of agents and the combinations of various amounts of agents and other entities, such as users or tags. Yield was higher when agents were combined with another entity.

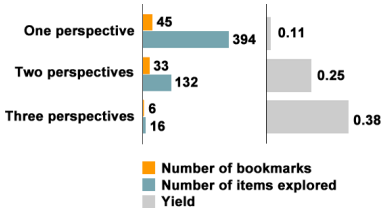

Figure 6: Yield of different numbers of perspectives in an exploration. Pearson's correlation showed a positive correlation between number of perspectives in an exploration and yield (r = 1.0, n = 3, p = .015).

Figure 7: Yield of different numbers of entities in the intersection. Pearson's correlation showed a positive correlation between number of entities in an intersection and yield (r = .839, n = 6, p = .037).

Bonferroni-Holm-corrected Wilcoxon signed-rank test indicated that differences were not statistically significant.

A Greenhouse-Geisser corrected ANOVA of the amount of steps needed to complete the bookmarking tasks showed an effect of condition, $F(1, 11) = 7.86$, $p = .017$, and an effect of task order, $F(2.09, 23) = 168.82$, $p = .002$. A Wilcoxon signed-rank test showed a trend for task one taking more steps when using IE (median: 11) than when using CN3 (median: 4), $Z = 2.5$, $p = .012$, but after applying a Bonferroni-Holm correction, differences were not statistically significant. This suggests that while IE may have a higher learning curve than CN3, no statistically significant differences exist in terms of efficiency of operation after acquaintance with the system (Figure 8).

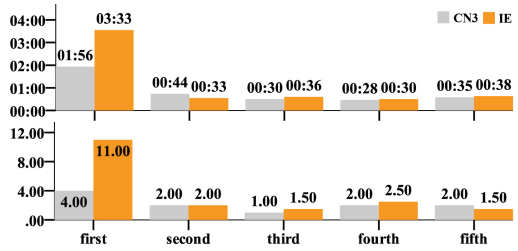

Figure 8: Median time (mm:ss) and steps of each task with IntersectionExplorer (IE) and CN3.

4.1.2 *Questionnaire results*

Results are reported in Figure 9, 10 and 11. Running a set of Bonferroni-Holm-corrected Wilcoxon signed-rank tests on the questionnaire results revealed the following:

- Papers explored with IE were perceived to be of a higher quality than with CN3 ($Z = 3.54$, $p < .001$).
- IE was perceived to be more effective than CN3 ($Z = 4.24$, $p < .001$).
- User satisfaction was higher with IE than with CN3 ($Z = 3.22$, $p = .001$).
- Users would me more willing to use IE frequently than CN3 ($Z = 3.42$, $p = .001$).
- Users perceived the recommendations shown in IE to be more trustworthy than CN3 ($Z = 2.55$, $p = .011$).

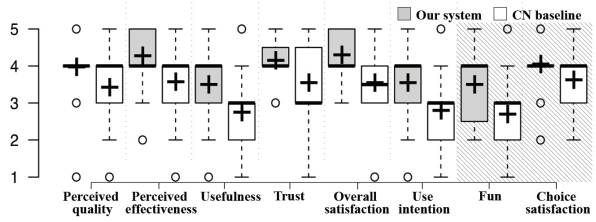

Figure 9: Questionnaire results with statistical significance. Differences between the aspects "Fun" and "Choice satisfaction" were not significant after the Bonferroni-Holm correction.

In addition, a trend was observed that users experienced IE to be more fun than CN3 ($Z = 2.28$, $p = .023$) and to provide a higher choice satisfaction ($Z = 2.1$, $p = .039$). However, after applying a Bonferroni-Holm correction, differences were not statistically significant.

Similarly, the results for the novelty of items (median: 4), effort to use the systems (median: 2), usefulness (median: 4), and ease of use (median: 4) were the same for both systems. Users tended to perceive the creation of bookmarks as more difficult in IE (median: 3) than in CN3 (median:

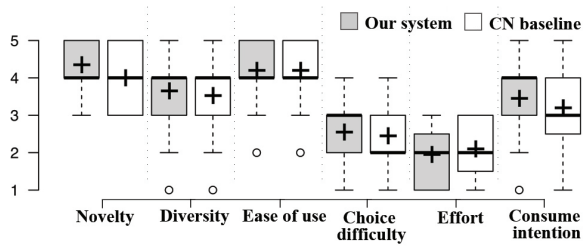

Figure 10: Questionnaire results without statistic significance.

2), but tended to read the bookmarked papers afterwards more frequently when using IE (median: 4) that when using CN3 (median: 3).

As for the IE-specific aspects shown in Figure 11, users perceived the visualisation to be adequate (median: 4) and the amount of information provided by the system to be sufficient to make a bookmark decision (median: 4). Users tended to be undecided regarding the interaction adequacy of IE (median: 3.5, see [16] for a definition), but found it easy to modify their preference to find relevant papers (user control, median: 4).

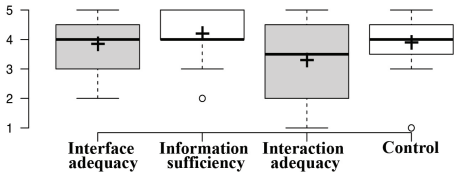

Figure 11: Interaction and visualisation sufficiency.

4.1.3 Observation

The think-aloud protocol revealed the following:

Interface: Some users misinterpreted empty circles to be a match of bookmarks or recommendations (three users) or initially failed to understand the meaning of the circles (three users). Others stated that they did not know that a tag-based agent was available and that the list of entities on the left was too long (three users).

Terminology: Two users had problems understanding the meaning of "sets", "related sets" or the numbers representing the amount of papers in a set.

It was further observed that some users only explored sets recommended by the agents, the majority explored sets recommended by agents and related to other users or tags.

4.1.4 Answering the research questions

RQ1 *Under which condition may a scalable visualisation increase user acceptance of recommended items?*

Our research showed that user acceptance of recommended items increased with the amount of sources used. However, the most important finding is that the addition of human-generated data – such as bookmarks of other users or tags – to the agent-generated recommendations resulted in a significant increase of effectiveness and yield. Our data suggests that providing users with insight into relations of recommendations with bookmarks and tags of community members increases user acceptance. We thus recommend to combine automated sources and personal sources whenever possible.

RQ2 *Does a scalable set visualisation increase perceived effectiveness of recommendations?*

Perceived effectiveness (expressed in the questionnaire) and actual effectiveness (how frequently users bookmarked a recommended paper) were increased by this type of visualisation.

RQ3 *Does a scalable set visualisation increase user trust in recommendations?*

The evaluation of the subjective data showed that user trust into the recommended items was increased with set-based visualisation of recommendation sources.

RQ4 *Does a scalable set visualisation improve user satisfaction with a recommender system?*

Overall, user satisfaction was higher when using the visualisation, suggesting this to be a key feature of the approach.

4.2 Discussion

4.2.1 Simplicity vs. effectiveness

The analysis of task completion time and amount of steps needed to complete the bookmarking tasks has shown that users require more time and interactions to set their first bookmark in IE, but that after this 'training phase', the operational efficiency between IE and CN3 does not differ. This corresponds to the observations made during the analysis of the think-aloud study, where it was found that some users initially struggled to understand the meaning of the different circle types or what a 'set' was.

However, the analysis of the subjective data has shown that users perceived IE to be more effective and its recommendations more trustworthy than those given by CN3. Especially the last point may be the result of removing the frequently lamented "black box" problem of recommenders by simply visualising how and why certain items are selected. In addition, users perceived items resulting from their use of IE to be of higher quality and found the overall experience more satisfying. This positive user experience may compensate for the initial conceptual problems encountered in the first exploration of the application and suggests that IE may be a helpful addition to the conference explorer service.

4.2.2 Comparison to previous work

In our previous work we presented the idea of combining recommendations embodied as agents with bookmarks of users and tags as a basis to increase effectiveness of recommendations [20]. A cluster map technique was used to enable users to explore these relations. Whereas the approach seemed promising, the cluster map was challenging for users to understand. In a first controlled user study, we asked users explicitly to explore recommendations of agents, bookmarks of users, tags and their combinations to try to find relevant items. Results of this user study indicate that there is an increase in effectiveness. In a follow-up uncontrolled field study users did not explore many intersections between different relevance prospects. As a result, the effect of combining relevance prospects could not be confirmed when users were not pushed to do so.

IE employs the novel UpSet visualisation technique that was presented at IEEE VIS in 2014. We simplified the interface and deployed it on top of data collected by Conference Navigator. The approach addresses the previous limitations

regarding ease of use and scalability: in this study users did explore many intersections, enabling us to investigate the effect of the approach on acceptance of recommendations.

4.2.3 Limitations

One limitation is the low number of participants. Further, the study was conducted with researchers from the field of technology enhanced learning with a high degree of visualisation expertise (mean: 4.05, SD: 0.86). Such users may be biased due to their graph literacy. In addition, our data was limited to that provided by the EC-TEL conferences.

5. CONCLUSION AND FUTURE WORK

We presented a study that used the UpSet visualisation technique to combine agent-based recommendations with human-generated recommendations in the form of bookmarks and tags. Despite the initial learning curve (when compared to the baseline system CN3), we found that this combination resulted in a higher degree of item exploration and acceptance of recommendations, than when using agent-only results. This way, user trust, usefulness, quality, and effectiveness were increased. We could thereby demonstrate the positive effects of the combination of multiple prospects on user experience and relevance of recommendations.

Future work will explore the applicability of our findings to a more diverse dataset and audience, as well as different types of visualisations. We have currently deployed IntersectionExplorer for attendees of the ACM IUI 2016 conference and will evaluate whether the visualisation can be used in an open setting, without the presence of a researcher. In addition, we plan to deploy the visualisation on top of data from large conferences, including the Digital Humanities conference series. Follow-up studies will assess the added value of our visualisation on top of larger data collections and with a less technical audience. With these studies, we intent to reach a wider range of users to further evaluate the effect of the approach on the effectiveness of recommendations.

6. ACKNOWLEDGMENTS

We thank all participants for their participation and useful comments. Part of this work has been supported by the Research Foundation Flanders (FWO), grant agreement no. G0C9515N, and the KU Leuven Research Council, grant agreement no. STG/14/019. The author Denis Parra was supported by CONICYT, project FONDECYT 11150783.

7. REFERENCES

[1] Aduna clustermap. www.aduna-software.com/technology/clustermap. Retrieved on-line 20 Augustus 2014.

[2] R. Baeza-Yates, B. Ribeiro-Neto, et al. *Modern information retrieval*, volume 463. ACM press New York, 1999.

[3] S. Bostandjiev, J. O'Donovan, and T. Höllerer. Tasteweights: a visual interactive hybrid recommender system. In *Proc. RecSys'12*, pages 35–42. ACM, 2012.

[4] J. Gemmell, T. Schimoler, B. Mobasher, and R. Burke. Recommendation by example in social annotation systems. In *E-Commerce and Web Technologies*, pages 209–220. Springer, 2011.

[5] B. Gretarsson, J. O'Donovan, S. Bostandjiev, C. Hall, and T. Höllerer. Smallworlds: visualizing social recommendations. In *Computer Graphics Forum*, volume 29, pages 833–842. Wiley Online Library, 2010.

[6] C. He, D. Parra, and K. Verbert. Interactive recommender systems: a survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications*, 2016.

[7] J. L. Herlocker, J. A. Konstan, and J. Riedl. Explaining collaborative filtering recommendations. In *Proc. CSCW '00*, pages 241–250. ACM, 2000.

[8] B. P. Knijnenburg, M. C. Willemsen, Z. Gantner, H. Soncu, and C. Newell. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, 22(4-5):441–504, 2012.

[9] A. Lex, N. Gehlenborg, H. Strobelt, R. Vuillemot, and H. Pfister. Upset: visualization of intersecting sets. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):1983–1992, 2014.

[10] C. D. Manning, P. Raghavan, H. Schütze, et al. *Introduction to information retrieval*, volume 1. Cambridge university press Cambridge, 2008.

[11] J. O'Donovan, B. Smyth, B. Gretarsson, S. Bostandjiev, and T. Höllerer. Peerchooser: visual interactive recommendation. In *Proc CHI '08*, pages 1085–1088. ACM, 2008.

[12] D. Parra and P. Brusilovsky. Collaborative filtering for social tagging systems: an experiment with citeulike. In *Proc. RecSys '09*, pages 237–240. ACM, 2009.

[13] D. Parra and P. Brusilovsky. User-controllable personalization: A case study with setfusion. *International Journal of Human-Computer Studies*, 78:43–67, 2015.

[14] D. Parra, W. Jeng, P. Brusilovsky, C. López, and S. Sahebi. Conference navigator 3: An online social conference support system. In *UMAP Workshops*, pages 1–4, 2012.

[15] J. Peng, D. D. Zeng, H. Zhao, and F.-y. Wang. Collaborative filtering in social tagging systems based on joint item-tag recommendations. In *Proc CICM '10*, pages 809–818. ACM, 2010.

[16] P. Pu, L. Chen, and R. Hu. A user-centric evaluation framework for recommender systems. In *Proc. RecSys'11*, pages 157–164. ACM, 2011.

[17] J. B. Schafer, J. A. Konstan, and J. Riedl. Meta-recommendation systems: User-controlled integration of diverse recommendations. In *Proc. CIKM '02*, pages 43–51, NY, USA, 2002. ACM.

[18] A. Spoerri. Infocrystal: A visual tool for information retrieval & management. In *Proc. CIKM '93*, pages 11–20. ACM, 1993.

[19] N. Tintarev and J. Masthoff. Designing and evaluating explanations for recommender systems. In *Recommender Systems Handbook*, pages 479–510. Springer, 2011.

[20] K. Verbert, D. Parra, P. Brusilovsky, and E. Duval. Visualizing recommendations to support exploration, transparency and controllability. In *Proc. IUI'13*, pages 351–362. ACM, 2013.

[21] C. Wongchokprasitti. *Using external sources to improve research talk recommendation in small communities*. PhD thesis, University of Pittsburgh, 2015.