

IntersectionExplorer, a Multi-Perspective Approach for Exploring Recommendations

Bruno Cardoso^{a,*}, Gayane Sedrakyan^a, Francisco Gutiérrez^a, Denis Parra^b, Peter Brusilovsky^c, Katrien Verbert^a

^a*Katholieke Universiteit Leuven, Celestijnenlaan 200A, 3001 Leuven, Belgium*

^b*Pontificia Universidad Católica de Chile, Departamento Ciencia de la Computación, Vicuña Mackenna 4860, Macul, Santiago, zip code 7820436, Chile*

^c*University of Pittsburgh, School of Computing and Information, 135 North Bellefield Avenue, Pittsburgh, PA 15260, USA*

Abstract

In this paper, we advent a novel approach to foster exploration of recommendations: IntersectionExplorer, a scalable visualization that interleaves the output of several recommender engines with human-generated data, such as user bookmarks and tags, as a basis to increase exploration and thereby enhance the potential to find relevant items. We evaluated the viability of IntersectionExplorer in the context of conference paper recommendation, through three user studies performed in different settings to understand the usefulness of the tool for diverse audiences and scenarios. We analyzed several dimensions of user experience and other, more objective, measures of performance. Results indicate that users found IntersectionExplorer to be a relatively fast and effortless tool to navigate through conference papers. Objective measures of performance linked to interaction showed that users were not only interested in exploring combinations of machine-produced recommendations with bookmarks of users and tags, but also that this “augmentation” actually resulted in increased likelihood of finding relevant papers in explorations. Overall, the findings suggest the viability of IntersectionExplorer as an effective tool, and indicate that its multi-perspective approach to exploring recommendations has great promise as a way of addressing the complex human-recommender system interaction problem.

Keywords: interactive visualization, exploration of recommendations, recommender systems, set visualization, scalability, user study

1. Introduction

Recommender systems play an important role in our daily lives, specially where decision making is involved. They are a common presence in the web, helping us find our way through

*I am the corresponding author.

Email addresses: b.m.pinto.cardoso@gmail.com (Bruno Cardoso), gsedrakyan@gmail.com (Gayane Sedrakyan), francisco.gutierrez@cs.kuleuven.be (Francisco Gutiérrez), dparras@uc.cl (Denis Parra), peterb@pitt.edu (Peter Brusilovsky), Katrien.Verbert@cs.kuleuven.be (Katrien Verbert)

Preprint submitted to International Journal of Human-Computer Studies

the ever-growing information jungle [1, 2]. We can find them in many different domains such as e-commerce websites, recommending items that we would likely be interested in, or in social networks suggesting new connections for us to expand our network of contacts.

Research has shown that improving recommendation diversity [2, 3, 4], providing users with means to control the recommendation process [5, 6] and facilitating exploration can have a positive impact on user acceptance of recommendations [7], the latter being of particular interest to our work. There are, of course, different ways to foster exploration. PeerChooser [8] and Smallworlds [9] are two prominent examples that enable end-users to explore a rich set of recommendations. Both approaches focus on the relationships between users and items as a basis to foster exploration and hence improve user trust and acceptance of recommendations [8, 9, 5].

Our work proposes a different approach to address the challenge of exploring recommendations in a visual manner: the exploration of multiple, intertwining *perspectives of relevance* in an integrated visual user interface. The concept of perspective of relevance is a broad term that describes the source and nature of recommendations. It was pioneered by social tagging systems, which first introduced the ability to explore different community relevance perspectives [10].

To the extent of our knowledge, existing social systems with recommendation features such as Conference Navigator 3 [11], which recommends conference papers to its users, allow users to explore only a single relevance perspective at a time: either the algorithmically-derived suggestions of recommendation engines (or agents), or one of two alternatives derived directly from human activity, specifically items marked by other users as relevant (e.g., bookmarked, bought, or viewed) or lists of items tagged with specific labels.

In our work, we use a synergistic approach for combining different perspectives of relevance, and identify three main types of perspective: (1) the *perspective of personal relevance*, (2) the *perspective of social relevance* and (3) the *perspective of content relevance*. More specifically, the suggestions of different personalized recommender engines produce ranked lists that can be considered as personal relevance perspectives, as they leverage previous knowledge about users to provide suggestions that would likely fit their interests and goals (e.g., the collection of items suggested by amazon.com based on a user’s recent purchases). Sets of items that are marked as relevant by known users offer a perspective of social relevance: if these users are perceived as like-minded, a collection of their items may be considered as a set worth exploring (e.g., a set of movies rated highly by a friend). Likewise, sets of items labelled by the community of users with specific tags offer a perspective of content relevance. Indeed, since tags are usually keywords drawn or derived from the contents of the tagged item or the experience of other users with it [12], they provide insightful glances about the items to people that find them relevant (e.g., a set of music pieces tagged by the community with the labels “relaxing” or “inspiring”).

Although these perspectives of relevance have undeniable potential by themselves, the ability to explore collections of items from multiple perspectives simultaneously can potentially increase user trust in the value of a given recommended item [10]. For instance, a user may feel more inclined to explore items that are both recommended by an automatic agent and marked as relevant by another, known and like-minded user. In other words, users may

be interested in “augmenting” the recommendations of automatic agents with items that interest other users or items that are also tagged with a specific topic of interest, such as “visualization”.

In prior work, we investigated this multi-perspective approach using TalkExplorer, an interactive recommender interface based on the cluster map visualization [5]. Although we were able to demonstrate the potential value of this approach, the user interface was found to be challenging for users in terms of usability. Further, scalability problems of the cluster map visualization made it hard to explore more than three sets simultaneously and, therefore, such combinations were never explored in field studies. Consequentially, the effects of the multi-perspective approach could not be thoroughly evaluated in our previous work.

To address these problems, we propose IntersectionExplorer: a novel, web-based platform that uses a scalable relevance-based visualization, UpSet [13], to allow users to simultaneously explore multiple sets of items from the three previously mentioned perspectives of relevance. As the user interface of IntersectionExplorer is based on a scalable visualization, we hoped that it could allow us to overcome problems revealed during our studies of the cluster map visualization [10].

A key feature of IntersectionExplorer is the seamless way to combine sets from the three perspectives of relevance, making no distinction between them in terms of interaction or representation, other than clearly stating their original perspective. In this way, IntersectionExplorer effectively treats recommendations derived directly from user activity (i.e., items tagged or bookmarked by other users) the same way it does for the suggestions of recommendation agents. The main questions that guided the research work we present in this paper are:

- *RQ1: How does IntersectionExplorer’s multi-perspective approach facilitate the exploration of collections of items?*
- *RQ2: How flexible and usable is IntersectionExplorer in responding to the requirements of users with different technological backgrounds?*

To provide a comprehensive response to these questions and assess the overall viability of IntersectionExplorer, we performed three user studies in different contexts. The user studies allowed us to evaluate the interface along several relevant dimensions of user experience with recommender systems and more objective metrics of performance, based on an analysis of user interaction data. Study 1 was designed as a comparison between IntersectionExplorer’s multi-perspective approach against the baseline of traditional, relevance-ordered list of recommendations. The study revealed that users found IntersectionExplorer to be a relatively fast and effortless tool to help them explore collections of items, while fostering the discovery of new relevant items with increased levels of overall satisfaction. Although this study provided interesting results, it also had some limitations. It was conducted in a controlled setting after the conference, in a one-hour scheduled session with the main researcher present to answer questions. To assess the viability of the tool in real settings we performed two additional studies live at conferences, more realistically capturing the behavior of participants trying to find relevant talks to attend and papers to bookmark.

While Study 2 was conducted with less technically-oriented participants, at the Digital Humanities conference in 2016, Study 3 was conducted with highly technical end-users at the 2017 edition of the IUI conference. These studies provide insight into the utility of the tool for users with different backgrounds, and the tool’s flexibility for accommodating users with different profiles and interests. Overall, the findings of our three studies evidence the viability of IntersectionExplorer as an effective tool, and indicate that its multi-perspective approach to recommendation exploration has great promise as a way of addressing the complex human-recommender system interaction problem, although technical end-users interact more seamlessly with the features of IntersectionExplorer, resulting in a better user experience than for less technically-oriented users.

The main contributions of this work are then: we present (1) IntersectionExplorer, a multi-perspective approach to the exploration of recommendations, (2) the results of three user studies designed to assess the tool’s usability and value in different contexts and user profiles, and (3) a comprehensive assessment of the value of exploring recommendations from multiple perspectives in parallel.

2. Related Work

2.1. Recommender Engines

Recommender systems have become an important research field since the mid-1990s. The goal of a recommender system is to help people deal with information overload by providing personalized content and services [1, 2, 14]. Recommender systems have been developed in a wide variety of application areas, including Web search [15, 16], multimedia [17], personalized advertisement [18], and e-learning [19]. Recent research also applies recommender systems in high-risk sectors, such as health care [20].

Recommender algorithms can be broadly categorized into three groups: *collaborative filtering*, *content-based filtering* and *hybrid recommenders*. The collaborative filtering approach is based on finding commonalities between users [21]. These commonalities are detected based on explicit relevance indicators, such as ratings [22] and tags [23], or implicit relevance indicators, such as the consumption of suggested media [24]. Recommendation lists are then produced based on the preferences of other users with similar interests. The content-based filtering approach proposes suggestions by matching descriptions of items to descriptions of users [25], based on information about individual users and items. Hybrid recommender systems propose items based on the combination of different recommendation techniques [26, 1].

While initial research on recommender systems was mainly focused on the challenge of improving algorithmic accuracy, an emergent research direction is to consider user experience as a core dimension in the evaluation of the performance of these systems. A survey of existing research on recommender systems from a user-centric perspective [27] revealed the importance of diversity [28], the impact of transparency on user trust [29] and the role of explainability [30] on improving the user experience. A common issue found in existing recommender systems is that they often present themselves to users as “black boxes”, not offering their users any insights into their inner logic or justifications for the suggestions made

[29]. This black box nature of recommender systems prevents users from comprehending recommended results and can lead to trust issues when recommendations fail [31]. Research has shown that making recommender systems more transparent, enabling users to control the recommendation process and facilitating exploration can lead to improved user trust. For instance, Herlocker et al. [31], Sinha and Swearingen [32] and Tintarev and Masthoff [30] concluded that by giving users information to personally justify recommendations through a good explanation interface could help inspire users’ trust and satisfaction, increasing user involvement and educating users on the internal logic of the system [31, 30].

One way to enhance user experience is to depart from the traditional relevance-ranked list of recommended items and to provide alternative visual interactive representations that leverage system explainability [33] and user control [34, 35, 36].

2.2. Visualization

An important line of research that is closely related to the work we propose in this paper concerns visualization, and we begin by mentioning an approach that is particularly relevant: relevance-based visualization. This visualization approach had its origin in the field of Information Retrieval [37, 38] and focuses on improving the display of search results to users. The main goal of relevance-based visualization is to emphasize which results were relevant to different parts of multi-term queries. The simpler example of this approach is direct relevance visualization, which indicates by means of color or color scales which terms are relevant to each retrieved item/document. We can find different implementations of direct relevance visualizations both in research literature and in industry. For instance, TileBars [39] is a technique that indicates relative document length and query term frequency and distribution among search results, simultaneously and in a compact way. The visualization can quickly be scanned and deciphered by users, thereby supporting them in judging the potential relevance of retrieved documents. HotMap [40], a meta-search system that retrieves documents from a web search API, is a recent example that fosters exploration of the retrieved results through inspection of visual document representations and through a nested sorting interaction feature. The ultimate goal is to allow users “to see the information without having to read the information” [40], therefore using color encoding to convey the frequency of query terms in the documents. Another example of highly-interactive relevance-based visualization in the information retrieval context is provided by uRank [41], a visual approach that combines lightweight text analytics and an augmented ranked list to assist in exploratory search of textual documents.

Set-based visualizations are another type of relevance-based visualizations that share important concepts with our work. This approach uses sets and spatial layouts to meaningfully organize search results. For instance, a three-termed query would be represented in as many set areas as there are different combinations of terms, to illustrate the relevance of each result to each term combination. In this example, there are seven set areas: one for each unique term, one for each of the three pairs of terms, and one more for all three terms combined. The classic example of this approach is InfoCrystal [42], that visualizes all possible relationships between a number of concepts and allows users to explore and filter information in a flexible and dynamic way by assigning relevance weights to concepts, and

selecting interesting relationships by means of thresholding. Another relevant set-based visualization is the Cluster Map [43], a visualization of lightweight ontologies, or taxonomies, that describe a domain through a set of classes and their hierarchical relationships. Classes are connected by direct edges that indicate hierarchies, class instances are connected to their most specific class and instances that share the same class are visually grouped in clusters.

Another stream of research that is relevant to our work concerns various visualization approaches that improve user experience with recommender systems by facilitating exploration, transparency and user control. For example, PeerChooser [8] and SmallWorlds [9] are two approaches that focus on explaining the inner logic of collaborative filtering recommender engines to users. Both systems allow users to interact with representations of relations between items and other users to support transparency and user control. PeerChooser uses a graph-based visualization to represent these relationships. SmallWorlds allows users to explore the relationships between recommended items and similar friends in multiple layers of similarity.

In addition, a number of visualizations have been developed to interact with hybrid recommender systems. TasteWeights [35] is a system that allows users to control the impact of different algorithms as well as different input data sources on the recommendation results, eliciting 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. [44] on meta-recommendation systems, where users are provided with personalized control over the generation of recommendations by altering the importance of specific factors on a scale from 1 to 5. Similarly, SetFusion [6] (see Figure 1) is a more recent example that allows users to fine-tune weights of a hybrid recommender system, using a set-based visualization - a Venn diagram [45] - to represent the relationships between recommendations. MoodPlay [46] is a hybrid music recommender system that integrates different techniques in an interactive interface supporting explanation and control of affective data. The system allows the user to explore a music collection through latent affective dimensions, thereby improving acceptance and understanding of recommendations. MyMovieMixer [47] is an interactive movie recommender that integrates different recommender techniques with interactive faceted filtering methods, called “blended recommending”. The approach allows users to interact with a set of filter facets representing criteria that can serve as input for different recommendation methods, including collaborative and content-based filtering.

There are also a number of approaches that focus primarily on the tags used by social recommenders. SFViz (Social Friends Visualization) [48] represents relationships between users and user interests in order to suggest potential new friends with similar interests, using the Radial Space-Filling technique [49]. Tagsplanations [50] are recommendation explanations based on community tags with two core concepts: tag relevance, the degree to which a tag describes an item, and tag preference, which conveys a user’s sentiment toward a tag. FaceTag [51] is a collaborative tagging tool for bookmarking domain resources that mixes a faceted classification scheme with the keywords space of user-generated tags in order to improve the information architecture of social tagging systems. MrTaggy [52] is a tag-based exploratory system that fosters rapid browsing of topic spaces using relevance feedback on tags to indicate users’ interest in the topics. Finally, Pharos [53] is a social recommender

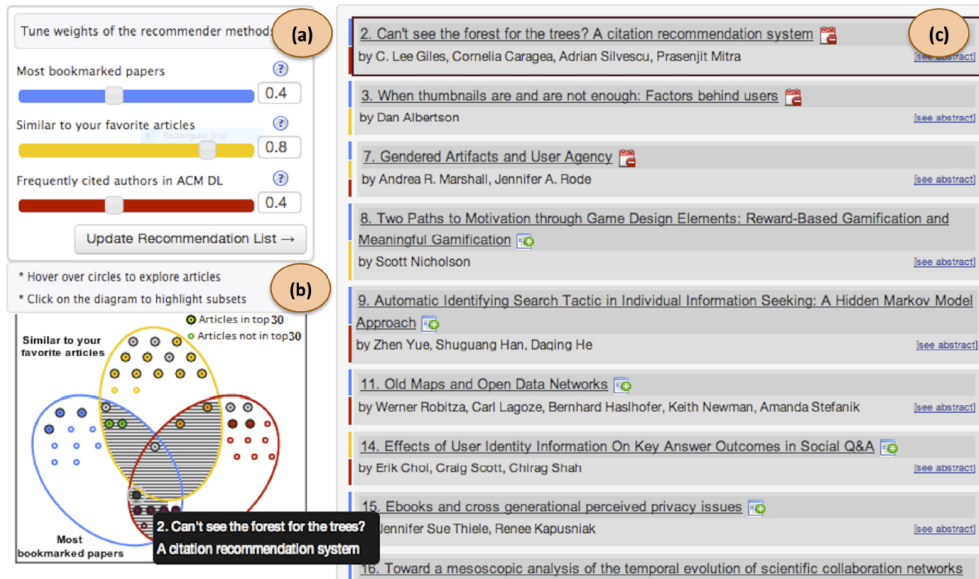


Figure 1: The SetFusion [6] user interface, featuring (a) sliders to tune the recommender algorithm weights, (b) interactive Venn diagram to select papers for exploration and (c) list of recommended papers [used with permission]. Like IntersectionExplorer, SetFusion is also a set-based, multi-perspective approach.

system that visualizes the social network activity of different communities in a map-based user interface. One of its most distinctive features is the way it addresses the cold-start issue by using topic modeling [54] to offer a summary of available content to new users.

2.3. Prior Work

Our previous work [7] is focused on various factors that affect acceptance of recommendations, such as user satisfaction, trust and sense of control. Specifically, based on the analysis of research on interactive recommender systems, we derived a framework proposing four important attributes for trust-aware and interactive recommender systems, namely: transparency, justification, diversity and context. We also investigated how information visualization can improve user understanding of the rationale behind recommendations in order to increase their perceived relevance and meaning and to support exploration and user involvement in the recommendation process. To this end, we performed two user studies using TalkExplorer [5] (see Figure 2), an interactive visualization tool developed for attendees of academic conferences based on the previously mentioned Cluster Map [43]. We combined different user-generated data sources in the study, but rather than automatically merging these data as it is done in hybrid recommender systems, end-users were allowed to select which users or tags should be considered. In addition, users could select different recommendation techniques that are represented as agents, similar to Ekstrand et al.’s [55] idea of enabling users to switch between recommenders.

The two user studies were conducted at two separate conferences, which allowed us to obtain interesting insights to enhance user interfaces that integrate recommendation technology. We observed that the effectiveness of recommendations and the probability of rec-

ommended item selection are both increased when users are able to explore and interrelate multiple perspectives of relevance, i.e., items bookmarked and tagged by other users and recommendations of automatic agents [5]. However, the user interface was found to be challenging for users in terms of usability. Furthermore, scalability problems of the cluster map visualization hindered the number of sets that users were able to explore simultaneously, and consequentially, the effects of the multi-perspective approach could not be thoroughly evaluated. In addition, we assessed the approach with a technical audience only.

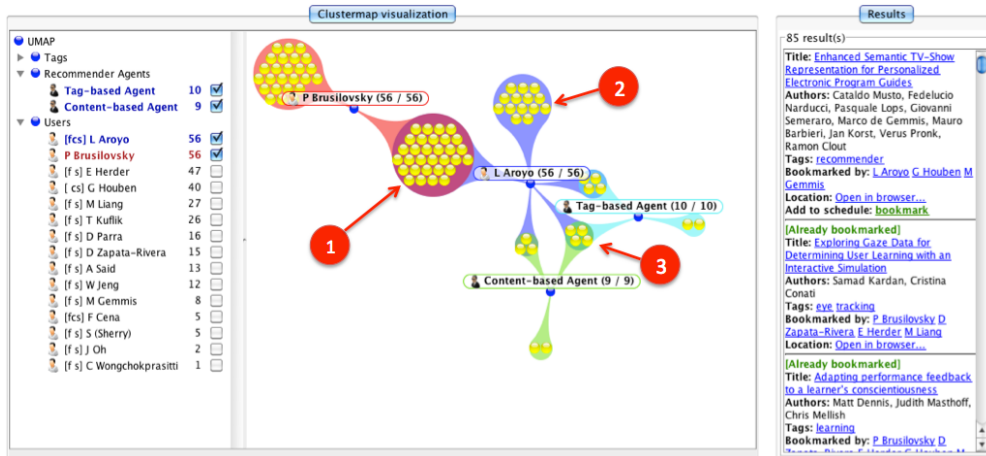


Figure 2: The TalkExplorer [5] user interface. The current user (P Brusilovsky) can explore which of his bookmarked papers are also bookmarked by user L Aroyo (1), which additional talks are bookmarked by L Aroyo but not recommended by any agent (2) and which talks are recommended to him by the two selected recommender agents and also bookmarked by L Aroyo (3) [used with permission]. Like IntersectionExplorer, TalkExplorer is also a set-based, multi-perspective approach.

2.4. Current Work

The system we propose in this work, IntersectionExplorer, shares many commonalities with the systems mentioned before in this section. Indeed, IntersectionExplorer is also a visualization-based approach for exploring recommendations, aimed at improving the user experience. Specifically, IntersectionExplorer is built on UpSet [13] (more details in Section 3), a set-based visualization technique itself, dedicated to the analysis of sets, their intersections, and aggregates of intersections. Similar to TalkExplorer, TasteWeights and SetFusion, IntersectionExplorer is also a visual hybrid recommender system, allowing its users to mix the suggestions of different recommender engines and from other sources seamlessly (more details can be found in section 4). Also, because IntersectionExplorer allows users to explore items through community tags, among others, IntersectionExplorer bears some commonalities with systems like FaceTag and MrTaggy. However, while all of these systems are generally oriented to letting users explore a single perspective at a time, IntersectionExplorer aims to let users explore multiple perspectives simultaneously. Finally, our work shares its core concept with TalkExplorer and SetFusion, being focused on the exploration of multiple perspectives of relevance. Nevertheless, its novelty lies in the use

of UpSet, whereas TalkExplorer used the cluster map visualization and SetFusion used a classical Venn diagram. Combinations of more than three sets were never explored in previous field studies due to the scalability limitations of these two visualizations [37]. Derived from the increased scalability potential of UpSet’s intersection-based visualization, IntersectionExplorer has less limitations than these previous proposals and, therefore, an increased potential for leveraging the multi-perspective approach and the realization of its promise.

3. UpSet - Scalable Set Visualization

UpSet [13] is a visualization technique dedicated to the analysis of sets, their intersections, and aggregates of intersections. UpSet divides a dataset into all possible intersections, as illustrated in Figure 3. These intersections correspond to the atomic areas of a Venn diagram, and are defined as “exclusive intersections”. According to Lex et al. [13], the use of exclusive intersections has two benefits: “it enables users to focus on the sets relevant to their analysis, and it addresses scalability” [13].

UpSet visualizes set intersections in a matrix layout. The columns represent the different sets and the rows represent commonalities between them. The vertical bar chart below the column headers depicts the number of items of each set, allowing to compare sets by size. Set relations are represented by the rows, where intersecting sets are identified by filled circles in the matrix. The horizontal bar charts at the right of the matrix display the number of items in each row/intersection.

For example, the first row in Figure 3 indicates that there is an intersection between the sets named “Set 1” and “Set 2”, as the circles corresponding to their respective columns are filled, and this intersection contains three elements. Likewise, the second row reveals that there are seven elements exclusive to “Set 2”. The third row represents seven elements that are exclusive to “Set 1”. For clarity purposes, this relation is also depicted in the equivalent Venn diagram of Figure 4. Finally, the fourth row conveys the existence of 43 elements that belong to neither “Set 1” nor “Set 2”.

Notably, one of the most prominent advantages of the UpSet matrix-based visualization is its scalability: comparing to a Venn diagram [45], arguably one of the most common visual representations of set intersections, the UpSet technique can present many sets in parallel, having the same visual encoding for any number and constellation of sets. This makes the UpSet visualization a better approach to use in the design of user interfaces when scalability is important, as it has more information density with less space requirements [13]. The application of UpSet in domains that deal with large amounts of complex data, such as overlaps in genomic data [56] and medical data models [57], provides further evidence of UpSet’s scalability.

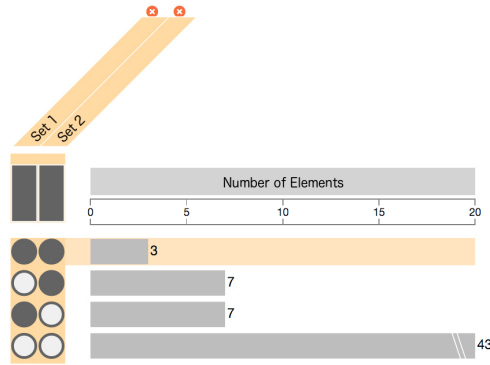


Figure 3: The UpSet matrix visualization applied to two sets of 10 items each, respectively named “Set 1” (left column) and “Set 2” (right column). The rows represent the intersections between these two sets: filled circles represent intersecting sets and empty circles indicate non-intersecting sets. Respectively from top to bottom, the four rows mean that (1st row) there are 3 elements in common between “Set 1” and “Set 2”, (2nd row) there are 7 elements exclusive to “Set 2”, (3rd row) 7 elements exclusive to “Set 1” and (4th row) 43 items in the “universe” of elements that do not belong to either set. The vertical bar chart below the column headers indicates the relative size of each set and the horizontal bar chart at the right of the circle matrix displays the absolute number of items in each intersection.

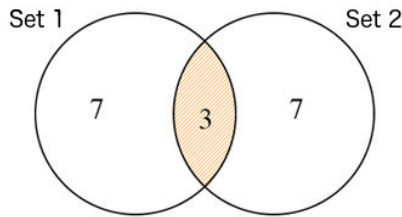


Figure 4: Venn diagram representation of the relation between two sets, “Set 1” and “Set 2”, each with 10 items, 3 shared elements and with 7 exclusive elements (equivalent to the relation depicted in Figure 3).

4. IntersectionExplorer

4.1. Information Support

While the principles that guided the development of IntersectionExplorer are general enough to make it potentially applicable to other domains, the tool was developed as a web-based application that aims to foster the exploration and finding of relevant conference papers. Therefore, we have deployed IntersectionExplorer on top of Conference Navigator 3 (CN3), a social personalized system that supports attendees at academic conferences [11]. The main feature of the CN3 platform is a scheduling system where users can add conference papers to create a personal schedule. Users can also add tags to papers and connect with other users by following them (unidirectional relationship) or connecting with them (bidirectional relationship). Social information collected by CN3 is used to help users find interesting papers. For example, CN3 lists the most popular papers, the most active attendees, and the most popular tags assigned to papers. When visiting a *talk page*, users can

also see who scheduled each paper during the conference and which tags were assigned to this paper. This social information is also used to provide recommendations to other papers (“*People who scheduled this presentation, also scheduled:*”), mimicking the well-known Amazon.com’s suggestions. Similarly, when visiting *user pages*, users can see which talks other users are planning to attend. Finally, tags can also be explored to find other papers that have been labelled with particular tags.

Additionally, the CN3 system offers the recommendations of five different recommendation engines, or agents, four of which offer personalized suggestions. The simplest, non-personalized one is the *top-N* engine, which suggests the N papers that have been bookmarked the most in any given conference. There are four recommender engines that offer personal suggestions:

- the *tag-based* recommender engine matches the tags assigned to papers by the current user to those assigned to other papers by the community of users, using the Okapi BM25 algorithm [58], a bag-of-words retrieval function that ranks a set of documents based on the query terms appearing in each document;
- the *bookmark-based* recommender engine builds the user interest profile as a vector of terms with weights based on the TF-IDF statistic [59], a measure of the importance of words in document collections, using the contents of the papers bookmarked by the user;
- the *external bookmark* recommender engine is based on the bag-of-words model augmented with information from external bookmarks. Stemmed terms from bookmarks in CN3 are combined with terms extracted from the bookmarks in social bookmarking systems, such as Mendeley, CiteUlike, and BibSonomy, to create a bag-of-words vector space. Items are ranked by cosine similarity [60].
- the *bibliography* recommender engine is based on a bag-of-words model augmented with bibliographical information. Stemmed textual terms from the papers bookmarked are combined with terms extracted from the user’s own bibliography to create a bag-of-words vector space and items are ranked through cosine similarity [60].

On a brief note, we mention that the CN3 made for an ideal platform for deploying IntersectionExplorer given its considerable community of users, extensive database of conference papers, different recommender engines, and a comprehensive API that facilitated the integration of our tool. Additionally, CN3’s API allowed for some of the actions performed by a user on IntersectionExplorer, like the bookmarking of papers, to immediately reflect in the user’s own CN3 account and therefore be persistent. To an extent, this may have been beneficial since, because participant actions had consequences, our collected data is likely more reliable, better describing the use of IntersectionExplorer out of experimental settings (versus, for example, a hypothetical situation in which participants, not committing to the study’s protocol, created bookmarks at random).

4.2. User Interface

An overview of the user interface of IntersectionExplorer can be seen in Figure 5. The interface is composed of three connected parts, or views, arranged from left to right (see the

top green callouts of Figure 5): the *Set Selection View*, the *Set Exploration View* and the *Intersection Exploration View*, further detailed in the following sections.



Figure 5: The IntersectionExplorer user interface. The interface is composed of three views, identified by the top green callouts: the *Set Selection View* lists (a) the suggestions of recommender agents (*Perspective of Personal Relevance*), (b) the bookmarks of other users (*Perspective of Social Relevance*) and (c) papers tagged by the community (*Perspective of Content Relevance*); the *Set Exploration View* allows users to explore the intersections between the selected sets of papers as rows (d) (currently explored intersections/rows are colored in darker gray and the intersections with the current user’s bookmarked papers are highlighted in blue); and, finally, the *Intersection Exploration View* displays the items (e) of the intersections selected in the Set Exploration View (fourth row in the example), thereby allowing users to explore and bookmark the suggested papers contained in this set. The particular example (row 4) contains five papers that are suggested by the bookmark-based agent and that are also bookmarked by User 1, but that are not suggested by the top-10 agent and the tag-based agent, not tagged with Tag 1 and also not yet bookmarked by the Current User.

Set Selection View

The Set Selection View allows users to search, select and deselect the sets of papers they would like to explore. It groups these sets in three separate sections, according to their source (see Figure 5 left): the *Perspective of Personal Relevance* section contains the sets of paper suggestions of different recommendation engines (see Figure 5, label a); the *Perspective of Social Relevance* section groups the paper sets that have been bookmarked by other conference attendees (see Figure 5, label b); and the *Perspective of Content Relevance* section lists the sets of papers that have been tagged with specific keywords by other conference attendees (see Figure 5, label c). The names of every CN3 user that is registered as an attendee of the conference being explored are alphabetically listed in the perspective of social relevance and, likewise, all tags assigned to papers of said conference are alphabetically listed in the perspective of content relevance.

The names of the sets themselves are different according to their sources: if the set is a list of recommendations of an automatic agent, the set is named after the recommending agent; if the set is a list of papers bookmarked by some other conference attendee, the set’s name is

the name of the attendee; and finally, if the set is a group of tagged papers, the displayed set’s name is the tag itself. In the example of Figure 5, the user has selected the recommendations of three automatic recommendation agents (metaphors for recommendation algorithms), the top-10, the tag-based and the bookmark-based agent, the bookmarks of two conference attendees, including his/her own bookmarks, and those of another attendee named “User₁”, and the set of papers tagged with the “Tag₁” label. When users select a set in any of the aforementioned sections, the selected set is added to the Set Exploration View and made available for exploration therein.

An interesting aspect of IntersectionExplorer to highlight is that the sets in the perspectives of social- and content-relevance can, at various levels, be considered as “human recommenders”, as their recommendations are not sets of suggestions ranked by relevance by automatic algorithms but instead groups of papers resulting from direct user action (bookmarking or tagging). Our tool does not make any distinction between sets of papers in its user interface, other than clearly stating their origin. Thereby, IntersectionExplorer offers a unique opportunity for users to freely explore both types of recommenders, human and automatic, in a seamless manner, all carrying the same weight and potential relevance, thus preventing biases induced by the user interface.

Set Exploration View

The Set Exploration View enables users to explore all of the possible combinations between the previously selected sets. The design of this view is based on an adaptation of the UpSet [13] technique, a visualization dedicated to the analysis of sets, their intersections, and aggregates of intersections (more details are presented in Section 3). The column headers display the names of the paper sets that were previously selected in the Set Selection View, and the matrix of circles below the headers represent the relations between those sets. Filled circles represent an intersection, while empty circles indicate that the sets are disjoint. At the top of the Set Exploration View, the users can also sort the rows - i.e., the intersections - by number of items or the amount of intersecting sets. In the example depicted in Figure 5, the user has sorted the rows by the number of papers. To explore a specific intersection, users click anywhere on the corresponding row, and the intersection items will be displayed in the Intersection Exploration View. In the example of Figure 5, the user has selected the fourth row (see Figure 5, label d) which became darker to indicate this selection. This row’s intersection contains five papers, in common between the suggestions of the bookmark-based agent and the bookmarks of the user named “User₁”, but not suggested by the two other agents, not tagged with “Tag₁”, and also not yet bookmarked by the current user.

Intersection Exploration View

The Intersection Exploration View allows users to browse through the papers of an interesting intersection, viewing important details like the paper authors and titles (see Figure 5, label e). Additionally, it is also possible to bookmark papers in this view by clicking on the corresponding “*Bookmark this paper*” links. Papers bookmarked in this way will be automatically added to the user schedule in the CN3 system. In the example of Figure 5, the user is currently exploring the five papers of the intersection represented in the

fourth row of the Set Exploration View. This row contains papers that have been suggested by the bookmark-based agent and bookmarked by the user named “User₁”. In this paper, we research whether exploring such relations, such as recommendations of an agent that are also bookmarked by a known user in this example, helps to filter out the more relevant recommendations. If the user is interested in scheduling a specific paper in his/her CN3 account (say, to attend the talk at the conference or to keep its reference for the future), he/she may do so by clicking on the “*Bookmark this paper*” link. Additionally, by clicking on the paper’s title, the user will be redirected to the paper’s CN3 page where further details, such as the abstract, are available for exploration.

5. User Studies

5.1. Common Methodology and Study Design

We have conducted three user studies in order to achieve a comprehensive understanding of the main potentials and limitations of IntersectionExplorer. To that end, we conducted each study in different settings, although we followed the same general guidelines when designing the three studies to minimize the impact of uncontrolled variability. Study 1 was an assessment of IntersectionExplorer’s general viability against the traditional recommender system user interface, relevance-ordered lists items. In turn, studies 2 and 3 were performed to understand how users with different backgrounds interact with and find value in our tool.

Since IntersectionExplorer targets the domain of academic conferences, the recommended items are conference papers (see Section 4.1). Therefore, in order to have a set of meaningful items to recommend and to have users interested in the recommendations, we conducted all of the three studies in the context of distinct academic conferences. Participants were recruited from the list of conference attendees through direct invitation, and participation took place live (also remotely via video call, in Study 1). Participation was voluntary and not compensated and all participants were asked to perform the same task: to freely explore the conference proceedings through IntersectionExplorer, and bookmark five relevant papers.

Before starting the tests, participants received the same introduction to IntersectionExplorer, which explained the functionality and covered essential concepts. To address the well-known cold start problem and to place participants at the same level regarding the relevance of the recommendations, we requested participants to bookmark and tag a minimum of five relevant papers from the conference proceedings using the CN3 proceedings page. Other than emphasizing that the bookmarked papers should be relevant, we did not impose any restrictions on this operation and allowed participants to take how much time they needed. As previously stated, CN3 also processes publications of participants and their activities in other academic social bookmarking systems like CiteULike and Bibsonomy. From this data, the CN3 recommender engines are able to offer personalized recommendations, metaphorized as “agents” in IntersectionExplorer (see Figure 5, label a).

During the studies, we collected the following IntersectionExplorer use data while participants interacted with the tool: the time taken to perform the task, selection and deselection of sets from different perspectives, sorting of rows, selection of rows and columns, clicking

in paper titles and also the bookmarking of papers. In order to enrich our data collection procedure, we applied the think-aloud protocol [61], asking participants to voice and describe their current thoughts and feelings about the task at hand while performing the paper-bookmarking task. We collected this data by recording both the screen and microphone input. For analysis, we replayed these recordings later, taking manual annotations as needed.

In order to have a general metric of performance to compare across studies, we define *precision* as the fraction of papers that were visualized and actually bookmarked, across all users. It is calculated as the ratio between the papers of a given combination of sets that were bookmarked and the total number of papers contained in that combination (that were listed in the “Intersection Exploration View” - see Figure 5). For instance, imagine that users have explored the 10 papers that were both recommended by the “bookmark-based agent” (perspective of personal relevance) and also bookmarked by “User₁” (perspective of social relevance), having bookmarked 1 paper. Then, they have explored the 20 papers that were both recommended by the “tag-based agent” and bookmarked by some other user, and bookmarked 2 of these papers. This scenario would result in a precision of 3/30, or 0.1, for the combination of the “personal” and “social” perspectives of relevance.

Since we are particularly interested in the way that IntersectionExplorer facilitates user interaction with automatic recommendation agents, we segment and analyze agent-user interaction data in the following three categories: (1) *single agent* means interactions with the recommendations of a single agent, (2) *multiple agents* are combinations of recommendations of two or more automatic agents and (3) *augmented agent* stands for combinations of agent recommendations and sets of papers from other perspectives (i.e., bookmarks of other users or sets of tagged papers, also defined as “human recommendations” in Section 4.2).

Finally, after completing their participation in the studies, we asked participants to respond to a questionnaire designed to assess some demographic information, as well as their user experience along several relevant dimensions. The questionnaire was answered online and participants were directed to it by the researcher after the experiment, using the same browser instance that was used in the experiment. In other words, the questionnaire’s web page was loaded and presented to users once they completed the paper-bookmarking task. Participants were free to answer the questionnaire as they pleased, without time or other restrictions imposed by the researchers. In order to assess the previously mentioned dimensions in a reliable way, the questionnaire was derived from two validated tools for the assessment of subjective aspects of user experience with recommender systems: the ResQue [62] questionnaire and the framework proposed by Knijnenburg et al. [63]. The questionnaire assessed the following 18 constructs using 5-point Likert items:

- *perceived quality* - how much users like the recommended items;
- *perceived effectiveness* - how effective users think the system is;
- *usefulness* - the extent to which users find that using the system would improve their performance compared with their previous experiences without the help of a recommender;
- *trust* - whether or not users find the whole system trustworthy;

- *overall satisfaction* - what users think and feel while using a recommender system;
- *use intention* - if users intend to use the chosen recommendations;
- *fun* - if users enjoy using the system;
- *choice satisfaction* - how much users like their chosen items;
- *novelty* - the extent to which users receive new and interesting recommendations;
- *diversity* - the diversity level of items in the recommendation list;
- *ease of use* - measures users' ability to quickly and correctly accomplish tasks with ease and without frustration;
- *choice difficulty* - how easy it is to choose the most relevant items;
- *effort* - how much effort is needed to use the system;
- *consume intention* - whether or not the system is able to influence users' decision to use the system and consume some of the recommended results;
- *interaction adequacy* - generally refers to the system's ability to present recommendations, allow for user feedback and to explain the reasons behind recommendations;
- *interface adequacy* - users' subjective evaluations of the system's interface in terms of displayed information, labels and layout adequacy and clarity;
- *information sufficiency* - the recommender's user interface provides sufficient information;
- *control* - whether users felt in control while interacting with the system.

5.2. Study 1: European Conference on Technology Enhanced Learning 2015

5.2.1. Methodology

We conducted this study at the 2015 edition of the European Conference on Technology Enhanced Learning (EC-TEL 2015). EC-TEL is a yearly, technology-oriented venue dedicated to “engaging researchers, practitioners, educational developers, entrepreneurs and policy makers to address current challenges and advances in the field” [64]. EC-TEL has a focus on applied technology and is not a very large event. Indeed, the number of papers published at EC-TEL's 2014 and 2015 editions were, respectively 172 and 112, and both editions had around 200 attendees [65, 66].

In this first study, in order to assess the overall viability of IntersectionExplorer (i.e., to what extent does the tool facilitate users to find relevant papers in conference proceedings), we compared it against another relevant approach that served as a baseline, the CN3 “*My Recommendations*” page. The reasons for this choice are twofold: first, since IntersectionExplorer uses the recommendation engines of CN3 (albeit having them metaphorized as “agents”), the strategies/algorithms used for recommending items to participants are the same for both tools; and second, CN3 presents its engines' recommendations following the classical relevance-ranked-list approach. This makes the IntersectionExplorer-CN3 comparison inherently unbiased, since the potential relevance of the suggestions to the user is the same and the results of this comparison enable contrasting between one of the most ubiquitous approaches to recommendation visualization, single-perspective ranked lists, and the novel, multi-perspective approach that we propose in this work.

We followed the protocol discussed in the previous section regarding data collection, recruitment and introduction of participants to the experiment (see Section 5.1). This study followed the within-subjects design, meaning that all participants were asked to perform the five-paper-bookmarking task on both tools. To minimize the impact of carryover effects and result biases, the study was counterbalanced by mode of exploration (CN3/IntersectionExplorer) (more details about Study 1 can be found in prior work [67]).

5.2.2. Results

Demographic Data

Our recruitment procedure yielded 20 participants, 15 of which completed the study via an on-line video call after the conference. The responses to the demographic items of the final questionnaire were as follows: participants averaged 32.9 years old (SD: 6.32) and three were female; users were *familiar with technology-enhanced learning* (mean: 4; SD: 1.1), *familiar with recommender systems* (mean: 4; SD: 0.95), *familiar with visualization techniques* (mean: 4.05; SD: 0.86) and *often followed the advice of recommender systems* (mean: 4.25; SD: 0.77); eight participants had *never heard of CN3 before*, whereas the remaining 12 had *previously heard of it, but had no particular familiarity with the system* (mean: 3.25; SD: 1.13); one participant had *no publications*, four of them had *between two and four publications*, 15 participants had *five or more publications* and, out of these, 14 had *publications in previous editions of EC-TEL*.

Interaction Data

In order to understand how participants interacted with IntersectionExplorer’s multi-perspective approach, we measured the amount of times that participants explored intersections combining different perspectives. Table 1 displays the results of the particular case of participant interaction with agents.

Table 1: Precision scores for explorations of agent recommendations “augmented” with other perspectives, i.e., sets of papers that were recommended by some agent while also being in some other user’s bookmarks or tagged with some keyword, across all participants of Study 1 (more details about the precision metric can be found in Section 5.1).

	Bookmarks	Papers Viewed	Precision	Explorations
Single Agent	5	93	0.05	26
Multiple Agents	15	166	0.09	40
Augmented Agent	8	50	0.16	27

A chi-square test of independence was conducted between the type of agent and the number of papers viewed and bookmarked. No significant association was found between the type of agent and the number of papers viewed and bookmarked, $\chi^2(2) = 4.454$, $p = .097$. In turn, results of the overall precision scores indicate that users are interested and find value in augmenting the recommendations of automatic agents with perspectives other than those of other agents. The precision of a single agent is 5%. Combining multiple agents results in a precision value of 9%. Augmented agents, that is agents combined with bookmarks of other

users or tags, have a precision of 16%. The higher precision values for augmented agents when comparing with single- and multiple-agent explorations indicate that users tend to find value in augmenting traditional recommendations offered by agents with other perspectives.

Table 2: Precision scores for single-perspective explorations, i.e. explorations of paper sets from the same perspective, across all participants of Study 1 (more details in section 5.1). The top row reports values for the perspective of personal relevance, the middle row shows the values for the perspective of social relevance and, finally, the bottom row contains the precision values for the perspective of content relevance. A significantly higher proportion of papers have been bookmarked when users explored sets of tagged papers exclusively.

Sets	Bookmarks	Papers Viewed	Precision	Explorations
1 Agent	5	93	0.05	26
≥1 Agents	20	259	0.08	66
1 Attendee	10	75	0.13	14
≥1 Attendees	14	107	0.13	30
1 Tag	10	27	0.37	18
≥1 Tags	11	28	0.39	19

An analysis of the precision scores for each individual perspective can be seen in Table 2, which reports data from explorations of combinations of one or more sets sharing the same perspective. As the data shows, the precision scores for within-perspective, single- and multi-set explorations are relatively similar (compare the precision values of both lines of each of the three rows of Table 2). Between perspectives, higher overall precision scores were obtained for the explorations of the perspective of content-relevance (bottom row) whereas explorations of the two other perspectives registered lower precision values (top and middle rows). A chi-square test of independence conducted between the perspective explored (regardless of number agents combined) and the number of papers viewed and bookmarked indicated a statistically significant association between the explored perspective and the number of papers viewed and bookmarked, $\chi^2(2) = 25.285$, $p < .001$. A set of Bonferroni-corrected chi-square post-hoc pairwise comparisons indicate that significantly more papers are viewed and bookmarked for single-perspective explorations of the perspective of content-relevance than for the other two perspectives (agents vs. attendees: $p = .108$; agents vs. tags: $p < .001$; tags vs. attendees: $p = .001$).

The precision scores regarding perspective involvement (i.e., present or not) in explorations may be found in Table 3. For example, an intersection combining the recommendations of an automatic agent with other sets involves the perspective of personal relevance, since agent recommendation sets are what composes that perspective.

We find a higher difference between the involvement of the content-relevance perspective, as precision is notably higher when sets of tagged papers are involved vs. not involved (see the bottom row of Table 3). In turn, precision differences are not so outstanding regarding the involvement of the perspectives of personal and social relevance (see top and middle rows). These results are further supported by statistics as we conducted three chi-square

Table 3: Precision scores for perspective involvement, i.e. for explorations involving one or more sets of papers from the same perspective, across all participants of Study 1 (more details in Section 5.1). The top row reports values for the perspective of personal relevance, the middle row shows the values for the perspective of social relevance and, finally, the bottom row contains the precision values for the perspective of content relevance. There is a statistically significant association between the number of papers viewed and bookmarked and the involvement of tags.

Perspective Involvement	Bookmarks	Papers Viewed	Precision	Explorations
Agent(s) Involved	59	398	0.15	156
Agent(s) Not Involved	25	145	0.17	59
Attendee(s) Involved	45	239	0.19	119
Attendee(s) Not Involved	39	304	0.13	96
Tag(s) Involved	25	71	0.35	56
Tag(s) Not Involved	59	472	0.13	159

tests of independence between perspective involvement (one per perspective) and the number of papers viewed and bookmarked. We did not find statistically significant associations between the number of papers bookmarked and viewed for neither the involvement of agents, $\chi^2(1) = 0.475$, $p = .491$, nor attendee bookmarks, $\chi^2(1) = 3.683$, $p = .055$. However, there was a significant association for the involvement of tags, $\chi^2(1) = 24.344$, $p < .001$, indicating that a significantly higher number of papers are bookmarked when sets of tagged papers are involved than when they are not.

Another interesting result concerns the number of multi-set explorations, since this is an indication of the way participants leveraged the scalability potential of IntersectionExplorer. Participants explored 58 times single sets, 57 times combinations of two sets, 43 times combinations of three sets, 29 times four sets, 12 times five sets and 16 times 6 sets. These results indicate that our participants were interested in exploring different-sized combinations of multiple sets.

In order to further understand the way participants interacted with our tool, we measured the time and number of steps (clicks on the user interface) taken by participants to complete the five-paper-bookmarking task. The median values for these metrics can be respectively seen in Figures 6 and 7, for both IntersectionExplorer and the CN3 system.

As can be seen in Figure 6, IntersectionExplorer and CN3 are comparable in terms of interaction time and execution steps for all tasks except the very first one. This indicates there is a learning effect at play that is more marked in IntersectionExplorer than in CN3, but that this is diluted upon continued contact with the tool. Together, the results indicate that IntersectionExplorer and CN3 can be considered comparable in terms of time and execution steps metrics. Further supporting these results, statistical analysis confirmed no statistically significant differences between the two tools (initial ANOVA and Wilcoxon significant differences not confirmed after Bonferroni-Holm correction - details of these analyses can be found in prior work [67]).

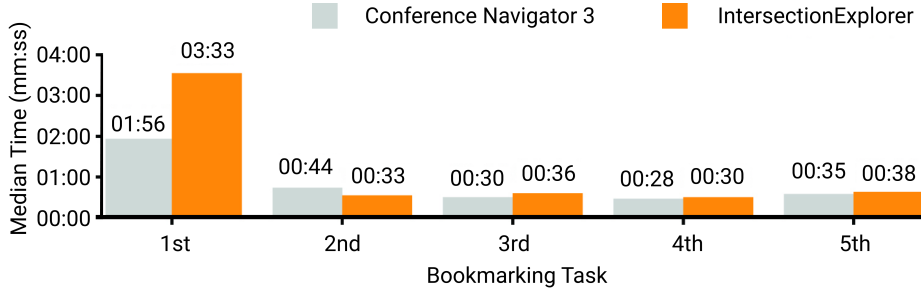


Figure 6: Median time (mm:ss) taken by participants of Study 1 to complete the five-paper-bookmark task, for both the IntersectionExplorer (in orange) and the CN3 (in gray). The horizontal axis indicates the number of the paper-bookmarking task.

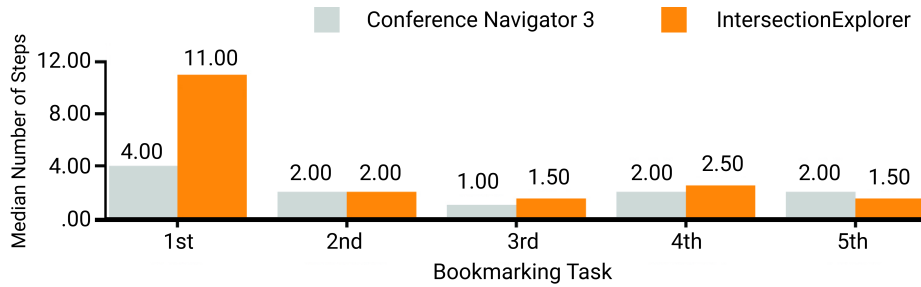


Figure 7: Median number of steps (navigational clicks in the interface) taken by participants of Study 1 to complete the five-paper-bookmark task, for both the IntersectionExplorer (in orange) and the CN3 (in gray). The horizontal axis indicates the number of the paper-bookmarking task.

Questionnaire Data

The results of the questionnaire items can be seen in Figure 8, comparatively for IntersectionExplorer and CN3. For a detailed definition of all these dimensions of recommender system usability, we refer the reader to Section 5.1. A set of Bonferroni-Holm-corrected Wilcoxon signed-rank tests ran on the questionnaire responses yielded the following statistically significant results:

- *papers explored with IntersectionExplorer were perceived to be of a higher quality than those explored with CN3's ranked list ($Z = 3.54, p < .001$);*
- *IntersectionExplorer was perceived to be more effective than CN3's list ($Z = 4.24, p < .001$);*
- *satisfaction was higher with IntersectionExplorer than with CN3's list ($Z = 3.22, p = .001$);*
- *participant willingness to use the tool was higher for IntersectionExplorer than for CN3 ($Z = 3.42, p = .001$);*
- *IntersectionExplorer recommendations were deemed more trustworthy than those in CN3's list ($Z = 2.55, p = .011$).*

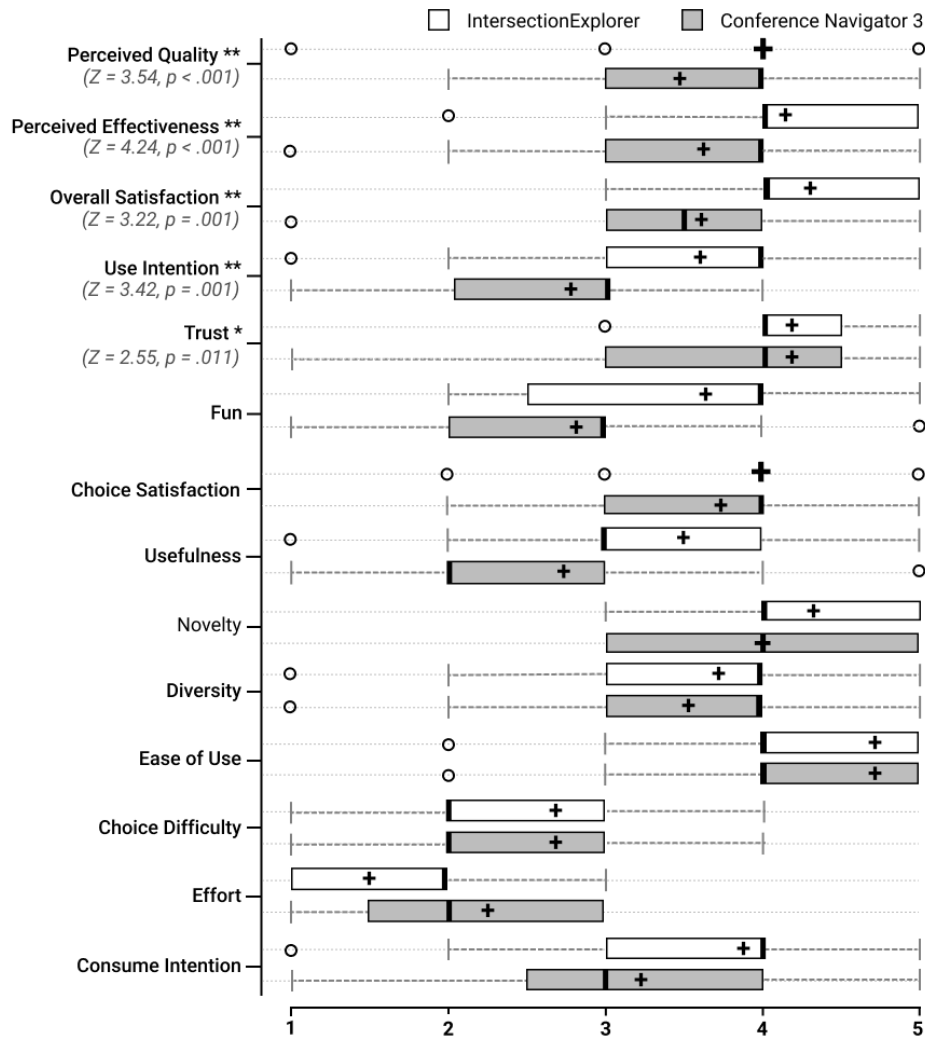


Figure 8: Questionnaire results for Study 1, comparing the questionnaire scores for IntersectionExplorer and CN3 (■ depict the medians, ○ represent the outliers, + stand for averages, the boxes represent the middle 50% of the data, and the whiskers stretch from the minimum to the maximum values excluding the outliers). Dimensions with statistically significant differences between tools are annotated with the Wilcoxon signed-rank test scores in parenthesis and asterisks (** significant at $p < .01$, * significant at $p < .05$).

Even though participants also reported *IntersectionExplorer* to be more fun than *CN3* ($Z = 2.28$, $p = .023$) and providing higher choice satisfaction ($Z = 2.1$, $p = .039$), these results were not statistically significant after a Bonferroni-Holm correction. The Wilcoxon signed-rank test did not indicate statistically significant differences between the tools in the remaining dimensions (see Figure 8).

Regarding the questionnaire items specific to IntersectionExplorer, participants judged highly (median of 4) IntersectionExplorer’s *interface adequacy*, *information sufficiency* and *control*. The only dimension with a somewhat lower score (median of 3.5) was *interaction adequacy* (see Figure 9 for more details).

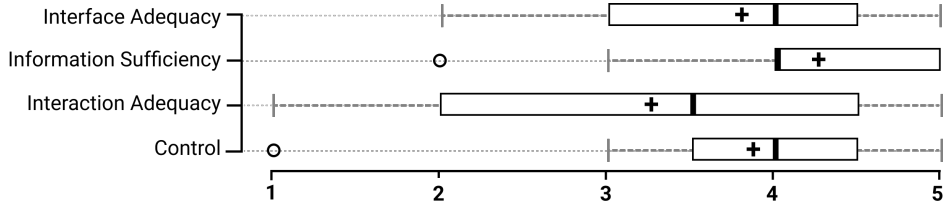


Figure 9: Questionnaire results for IntersectionExplorer-specific questionnaire items, regarding the tool’s overall interaction and information sufficiency (■ depict the medians, ○ represent the outliers, + stand for averages, the boxes represent the middle 50% of the data, and the whiskers stretch from the minimum to the maximum values excluding the outliers).

Think-aloud Protocol Data

An analysis of the think-aloud protocol data revealed that three participants misinterpreted the empty circles to convey a match of bookmarks or recommendations. Three participants initially did not properly understand the meaning of the circles and another three participants stated that they did not know that a tag-based agent was available, and that the list of sets in the Set Selection View (see Figure 5) was too long. Additionally, two participants reported difficulties in understanding the meaning of “sets”, “related sets” and the numbers representing the amount of papers in a set. The low number and disparity of the reported issues, however, suggests that they are likely incidental and not indicative of a fundamental flaw in IntersectionExplorer’s design and concept.

5.3. Study 2: Digital Humanities 2016

5.3.1. Methodology

We conducted Study 2 at the Digital Humanities 2016 Conference (DH 2016), a large event in the field of Digital Humanities, bringing together academics mainly from the areas of social sciences and humanities. In the 2016 edition of the DH conference, the event registered over 1000 [68] attendees and had 417 works accepted for publication [69], with submission topics ranging from “Data Mining/Text Mining” to “Music” and “Theology” [70].

The study took place over the five days of the DH 2016 conference, and we followed the task-based protocol explained in Section 5.1. We conducted Study 2 at this venue to explore IntersectionExplorer’s applicability with users with more diverse, less technology-focused backgrounds. The overall objective was to obtain data in the diverse landscape of DH 2016 to better understand how well IntersectionExplorer responds to the needs of a more heterogeneous audience, likely less biased by technology-oriented backgrounds.

5.3.2. Results

Demographic Data

We recruited 36 participants out of the DH 2016 attendees, averaging 38 years old (SD: 10). Eleven participants were female. An analysis of the results of the 5-point demographic items of the final questionnaire describes our participant sample as: *moderately familiar with recommender systems* (mean: 3.42; SD: 0.96), *familiar with visualization techniques* (mean:

4.05; SD: 0.71), *occasionally followed the advice of recommender systems* (mean: 3.53; SD: 1.08) and *one participant had heard of Conference Navigator 3 before*.

Interaction Data

Table 4 displays data resulting from participant interaction with agents. Like in Study 1, the highest precision was achieved using augmented agents. Moreover, in this study it emerged as the most popular exploration approach that involves agents. It suggests that participants of Study 2 found value in augmenting the recommendations of automatic agents with other perspectives, such as bookmarks of known users or tags. In contrast to study 1, combining multiple agents, such as the tag-based and bookmark-based agents, did not yield a higher precision value than the precision of a single agent, however this data might not be reliable since DH participants almost never used such combinations. A chi-square test of independence was conducted between the type of agent and the number of papers viewed and bookmarked, finding no statistically significant association $\chi^2(2) = 0.455$, $p = .797$.

Table 4: Precision scores for explorations of agent recommendations “augmented” with other perspectives, i.e., sets of papers that were recommended by some agent while also being in some other user’s bookmarks or tagged with some keyword, across all participants of Study 2 (more details about the precision metric can be found in Section 5.1).

	Bookmarks	Papers Viewed	Precision	Explorations
Single Agent	41	196	0.21	31
Multiple Agents	1	7	0.14	4
Augmented Agent	15	63	0.24	37

An analysis of the precision scores for each individual perspective can be seen in Table 5, which reports data from explorations of combinations of one or more sets sharing the same perspective. This data reveals that the precision score results are very similar between single- and multi-set explorations of the same perspective (compare the precision values of both lines of each row). Across perspectives, it is interesting to observe that at a large and diverse conference where the attendees are less familiar with each other, attendee-based exploration was not as productive as agent-based exploration. Yet, as in Study 1, tag-based exploration was the most productive: both single- and multiple-set explorations of the perspective of content-relevance have a higher precision (bottom row) than those of the other two perspectives (top and middle rows). We conducted a chi-square test of independence between the perspective explored and the number of papers viewed and bookmarked. There was a statistically significant association between perspective and the number of papers bookmarked, $\chi^2(2) = 45.336$, $p < .001$. Like in Study 1, a set of post-hoc Bonferroni-corrected pairwise comparisons revealed that significantly more papers are viewed and bookmarked for single-perspective explorations of the perspective of content-relevance than for the other two perspectives (agents vs. attendees: $p = .069$; agents vs. tags: $p < .001$; tags vs. attendees: $p < .001$).

The precision scores regarding perspective involvement in intersections of sets (i.e., present or not) are presented in Table 6. For all perspectives, precision results are consis-

Table 5: Precision scores for single-perspective explorations, i.e. explorations of paper sets from the same perspective, across all participants of Study 2 (more details in Section 5.1). The top row reports values for the perspective of personal relevance, the middle row shows the values for the perspective of social relevance and, finally, the bottom row contains the precision values for the perspective of content relevance. A significantly higher proportion of papers have been bookmarked when users explored sets of tagged papers exclusively.

Sets	Bookmarks	Papers Viewed	Precision	Explorations
1 Agent	41	196	0.21	31
≥ 1 Agents	42	203	0.21	35
1 Attendee	45	274	0.16	25
≥ 1 Attendees	49	335	0.15	41
1 Tag	44	91	0.48	78
≥ 1 Tags	44	94	0.47	80

tently higher when perspectives are involved than when they are not (compare the precision values of both lines of all three individual rows). Differences in precision are more clear, however, for the perspective of content-relevance, whose presence and absence seems to have the most effect (note the larger difference between the values of both lines of the bottom row). These results are further corroborated by chi-square tests of independence on the association between perspective involvement and the number of papers viewed and bookmarked, for each of the three perspectives. The associations found were statistically significant for all perspectives: involvement of agents, $\chi^2(1) = 55.213$, $p < .001$; involvement of attendees, $\chi^2(1) = 30.708$, $p < .001$; and finally, involvement of tags, $\chi^2(1) = 165.428$, $p < .001$.

Table 6: Precision scores for perspective involvement, i.e. for explorations involving one or more sets of papers from the same perspective, across all participants of Study 2 (more details in Section 5.1). The top row reports values for the perspective of personal relevance, the middle row shows the values for the perspective of social relevance and, finally, the bottom row contains the effectiveness and precision values for the perspective of content relevance. The associations between perspective involvement and the number of papers viewed and bookmarked are significant for the involvement of all individual perspectives.

Perspective Involvement	Bookmarks	Papers Viewed	Precision	Explorations
Agent(s) Involved	57	267	0.21	73
Agent(s) Not Involved	96	1383	0.07	133
Attendee(s) Involved	66	408	0.16	86
Attendee(s) Not Involved	87	1242	0.07	120
Tag(s) Involved	48	110	0.44	94
Tag(s) Not Involved	105	1540	0.07	112

Regarding the number of multi-set explorations (regardless of perspectives), our participants explored 134 times single sets, 46 times intersections of two sets, 18 times three sets,

5 times four sets, and combined five sets only once. These results suggest that, even though participants of Study 2 did explore combinations of more than two sets, they favoured single sets and intersections between two sets.

The median amount of time and steps taken by participants of Study 2 to complete the five-paper-bookmarking task can be seen in Figures 10 and 11. The notable difference that exists between the median times taken to perform the first and second bookmarks suggests the existence of a learning curve that, once overcome, will no longer have an impact on the time required to explore and create new bookmarks. The same conclusion is also indicated by the chart illustrating the median number of steps (Figure 11), which indicates 2 steps or less with continued use of IntersectionExplorer. These results are corroborated by two Friedman tests¹, with pairwise Bonferroni correction for multiple comparisons, that we ran to determine if the differences in number of steps and time taken were statistically significant between tasks. The differences were statistically significant between tasks, both for the number of steps, $\chi^2(4) = 53.409$, $p < .001$, as well as for time, $\chi^2(4) = 61.966$, $p < .001$. Post-hoc analyses revealed statistically significant differences between the first bookmark and the other bookmarks regarding time taken, $p < .001$, and number of steps, $p < .001$. Bookmarks 2, 3, 4, and 5 had no statistically significant differences between one another, either in time or number of steps.

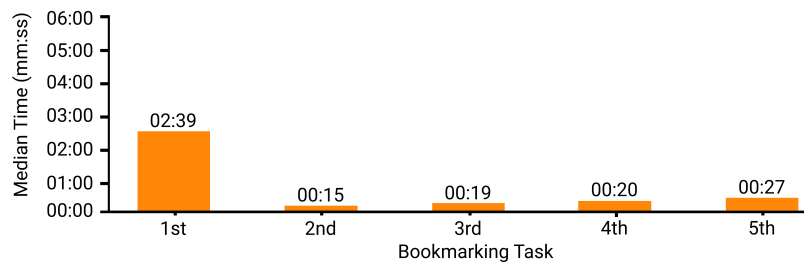


Figure 10: Median time taken by participants of Study 2 to complete the five-paper-bookmarking task (mm:ss).

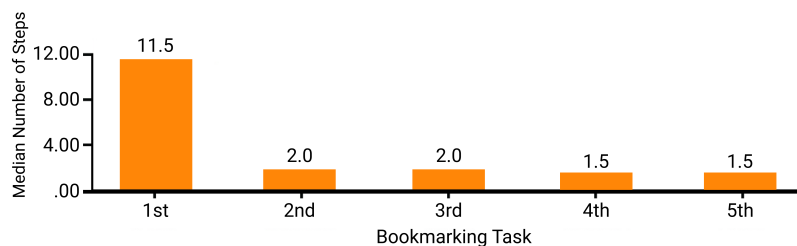


Figure 11: Median steps required by participants of Study 2 to complete the five-paper-bookmarking task.

¹Time and steps data deviate significantly from the normal distribution, as found by Shapiro-Wilk tests.

Questionnaire Data

The questionnaire results were generally positive, with participants evaluating IntersectionExplorer highly (median of 4) on some very important dimensions like *quality of recommendations*, *trust*, *overall satisfaction* and *consume intention*, having a positive to moderate assessment (median of 3.5) of the tool's *interaction adequacy* and *ability to modify preferences to find relevant papers*. In turn, the lowest scores (median of 2.5) were obtained for the dimensions of *fun to use* and *effort needed to find a relevant paper*. Overall, these results indicate that participants of this study were moderately satisfied with IntersectionExplorer (see Figure 13 for a complete overview of the questionnaire results and Section 5.1 for more details on the questionnaire constructs).

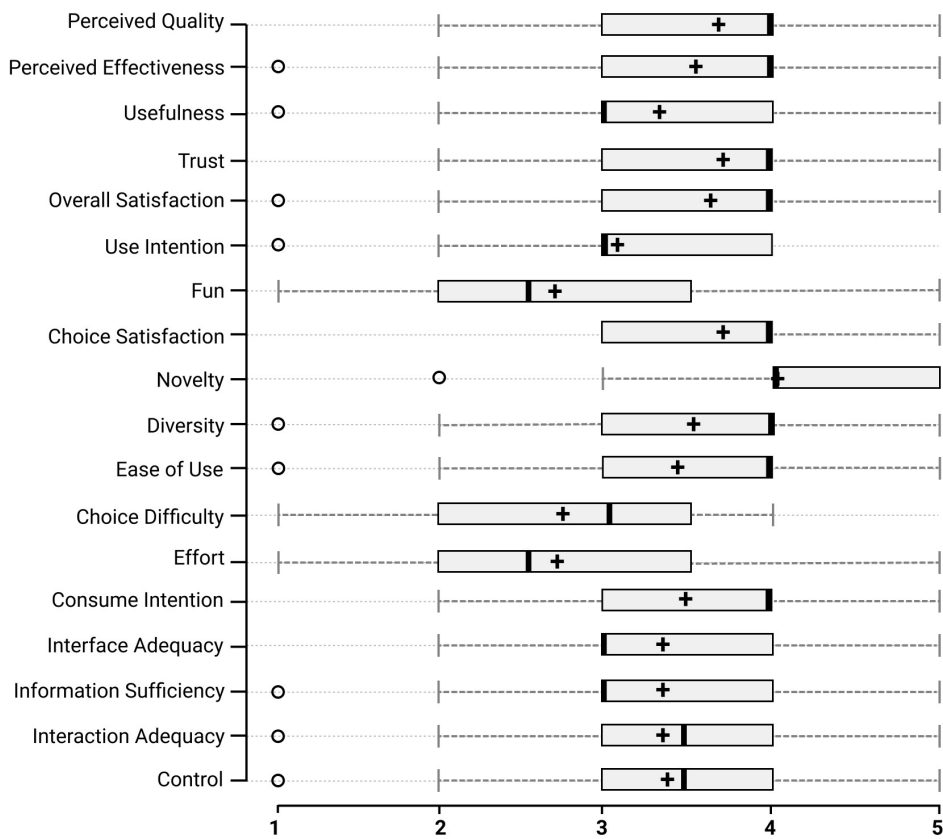


Figure 13: Questionnaire results for Study 2 (■ represents the medians, ○ the outliers, + averages, the boxes represent the middle 50% of the data, and the whiskers stretch from the minimum to the maximum values excluding outliers).

5.4. Study 3: Intelligent User Interfaces 2017

5.4.1. Methodology

This study was conducted at the Intelligent User Interfaces 2017 (IUI 2017) conference. IUI is an annual, small-to-medium size conference, focused on the interaction between

machine and human intelligence, attracting over 200 attendees from the community of intelligent interfaces [73]. In the 2017 IUI edition, 111 papers were accepted for publication, including full papers, demos, posters and doctoral consortium works [74].

We ran the study over the four days of the conference, once again following the protocol explained in Section 5.1 regarding data collection, and participant recruiting and briefing. The main motivation for this study was repeatability and comparison, as the likely technology-focused backgrounds of the IUI 2017 attendees makes them an interesting sample to compare against the more generic, social-sciences and humanity oriented attendees of the DH 2016-sampled participants of Study 2.

5.4.2. Results

Demographic Data

We recruited 42 participants for this study out of the population of IUI 2017’s attendees, averaging 32.4 years old (SD: 9.69) and 17 were female. The results of the demographic items of the questionnaire revealed that: participants were *familiar with recommendation systems* (mean: 3.79; SD: 1.19); *familiar with visualization techniques* (mean: 3.51; SD: 1.08); and *occasionally followed the advice of recommendation systems* (mean: 3.56; SD: 1.1).

Interaction Data

Table 7 displays data resulting from participant interactions with agents. As observed in the previous studies, augmented agents offered the best precision. Moreover, just like in Study 2, augmented agents emerged as the most popular among explored combinations: the users interacted with augmented agents 110 times. Interactions with single or multiple agents were lower, respectively 79 and 63 interactions. Both the higher interaction numbers, as well as the tendency of higher precision supports the hypothesis that users prefer augmenting the recommendations of automatic agents with perspectives other than other agents, such as bookmarks of known users and tags, and that they find such combinations useful. Like in previous studies, a chi-square test of independence was conducted between the type of agent and the number of papers viewed and bookmarked. A statistically significant association was found between the type of agent and the number of papers viewed and bookmarked $\chi^2(2) = 16.603$, $p < .001$. A set of post-hoc pairwise Bonferroni corrected comparisons indicated that there is a significantly higher proportion of papers bookmarked for explorations of augmented vs. single and multiple agents. This suggests that participants are more likely to accept the suggestions of agents that have been augmented with sets of tagged or bookmarked papers.

An analysis of the precision scores for each individual perspective can be seen in Table 8, which reports data from explorations of combinations of one or more sets sharing the same perspective. The precision scores for within-perspective, single- and multi-set explorations are notoriously similar (compare the values of both lines of each of the three rows of Table 8). Comparing between perspectives, higher overall precision scores were obtained for explorations of the perspective of content-relevance (bottom row), whereas explorations of the two other perspectives registered lower precision values (top and middle rows). A chi-square test of independence between the perspective explored (regardless of number

Table 7: Precision scores for explorations of agent recommendations “augmented” with other perspectives, i.e., sets of papers that were recommended by some agent while also being in some other user’s bookmarks or tagged with some keyword, across all participants of Study 3 (more details about the precision metric can be found in Section 5.1). A significantly higher proportion of papers have been bookmarked when users explored the suggestions of augmented agents.

	Bookmarks	Papers Viewed	Precision	Explorations
Single Agent	50	805	0.06	79
Multiple Agents	28	595	0.05	63
Augmented Agent	41	360	0.11	110

Table 8: Precision scores for single-perspective explorations, i.e. explorations of paper sets from the same perspective, across all participants of Study 3 (more details in Section 5.1). The top row reports values for the perspective of personal relevance, the middle row shows the values for the perspective of social relevance and, finally, the bottom row contains the precision values for the perspective of content relevance. A significantly higher proportion of papers have been bookmarked when users explored sets of tagged papers exclusively.

Sets	Bookmarks	Papers Viewed	Precision	Explorations
1 Agent	50	805	0.06	79
≥ 1 Agents	78	1400	0.06	142
1 Attendee	6	122	0.05	20
≥ 1 Attendees	12	151	0.08	25
1 Tag	28	79	0.35	50
≥ 1 Tags	31	84	0.37	54

agents) and the number of papers viewed and bookmarked revealed a statistically significant association between the perspective explored and the number of papers bookmarked, $\chi^2(2) = 328.682$, $p < .001$. Like in studies 1 and 2, a set of post-hoc Bonferroni-corrected pairwise comparisons revealed that significantly more papers are viewed and bookmarked for single-perspective explorations of the perspective of content-relevance than for the other two perspectives (agents vs. attendees: $p = .236$; agents vs. tags: $p < .001$; tags vs. attendees: $p < .001$).

The precision scores regarding perspective involvement (i.e., present or not) in explorations are presented in Table 9. The results for precision are diverse (see all three rows of Table 9). We find a higher difference between the involvement of the personal and content-relevance perspectives (respectively, the top and bottom rows): precision is lower when agents are involved vs. not involved, and higher when tags are involved vs. not involved. The difference in precision is less perceptible regarding the involvement of the perspective of social relevance (middle row). This indicates that a larger proportion of papers were bookmarked when users included sets of tagged papers in their explorations (last column, bottom row), the lower when sets of agent recommendations (last column, top row) were involved and did not make such a big difference whether sets of attendee

Table 9: Precision scores for perspective involvement, i.e. for explorations involving one or more sets of papers from the same perspective, across all participants of Study 3 (more details in section 5.1). The top row reports values for the perspective of personal relevance, the middle row shows the values for the perspective of social relevance and, finally, the bottom row contains the precision values for the perspective of content relevance. The associations between perspective involvement and the number of papers viewed and bookmarked are significant for all individual perspectives.

Perspective Involvement	Bookmarks	Papers Viewed	Precision	Explorations
Agent(s) Involved	134	2800	0.05	329
Agent(s) Not Involved	49	257	0.19	97
Attendee(s) Involved	64	1393	0.05	212
Attendee(s) Not Involved	119	1664	0.07	214
Tag(s) Involved	50	315	0.16	114
Tag(s) Not Involved	133	2742	0.05	312

bookmarks were used or not (last column, middle row). These observations are supported by chi-square tests of independence between perspective involvement and the number of papers viewed and bookmarked, for each of the three perspectives. The found associations were statistically significant for the involvement of all perspectives: involvement of agents, $\chi^2(1) = 85.297$, $p < .001$; involvement of attendee bookmarks, $\chi^2(1) = 8.809$, $p = .003$; and involvement of tags, $\chi^2(1) = 60.996$, $p < .001$.

The way that participants of Study 3 combined and explored sets of papers were more similar to Study 1 than Study 2. Indeed, single sets were explored 149 times, two sets 144 times, three sets 88 times, four sets 33 times and five sets were explored 9 times. Although the majority of explorations were of single sets and 2-set intersections, a considerable number of combinations of three, four and fives were also explored. These results indicate that, similar to Study 1, participants of Study 3 were also interested in combining and exploring combinations of multiple sets.

In order to further understand how participants interacted with our tool, we again measured the time and number of steps (clicks on the user interface) taken by participants for the five-paper-bookmarking task. The corresponding median time and number of steps taken by Study 3 participants to complete this task can be seen in Figures 14 and 15.

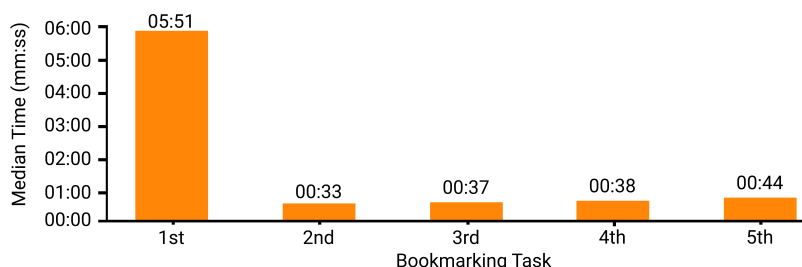


Figure 14: Median time taken by participants of Study 3 to complete the five-paper-bookmarking task (mm:ss).

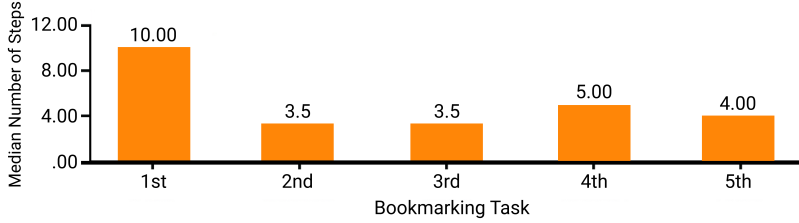


Figure 15: Median number of steps performed by participants of Study 3 to complete the five-paper-bookmarking task.

Observing Figure 14, it is straightforward to note that users took more time to create the first bookmark than they did for achieving the four remaining bookmarks. Although time seems to increase after the creation of the first bookmark, it is a very slight growth and remains below the 45-second threshold, suggesting that this threshold may be asymptotic. A Friedman test² with pairwise Bonferroni correction for multiple comparisons supports this observation, indicating a statistically significant difference between the times taken in the tasks, $\chi^2(4) = 72.778$, $p < .001$. Post-hoc analysis revealed statistically significant differences between bookmark 1 and the other bookmarks regarding time taken, $p < .001$, but not between the times taken for bookmarks 2, 3, 4 and 5.

In turn, an analysis of Figure 15 indicates that participants took a much higher number of steps to create the first bookmark than they did for creating the other. Contrasting to what happens with the median time, there is not a monotonic increase in the number of steps suggesting that, with continuous use, IntersectionExplorer users will create new bookmarks with a low number of steps. These findings are corroborated by a Friedman test² with pairwise Bonferroni correction for multiple comparisons, which found a statistically difference between the number of steps taken in the tasks, $\chi^2(4) = 42.378$, $p < .001$. As before, a post-hoc analysis indicated a difference between the number of steps taken for bookmark 1 vs. all other bookmarks, $p < .001$, but no statistically significant differences between bookmarks 2, 3, 4 and 5.

Interaction Process Discovery

Similar to what we did in Study 2, we also present the process-discovery maps extracted from the interaction data of Study 3 (see Figure 16 and, for a detailed explanation of these visualizations, we refer the reader to subsection *Interaction Process Discovery*, in Section 5.3.2).

This map indicates that the exploration of augmented agent recommendations was the most frequent action for participants of Study 3 (frequency of 110) and that these explorations led to a higher number of bookmarks (frequency of 41). Single- and multi-agent explorations were, respectively, the second (frequency of 79) and third (frequency of 63) most frequent actions. Like in Study 2, there was a high number of successive explorations

²As indicated by Shapiro-Wilk tests, time and steps data deviate significantly from the normal distribution.

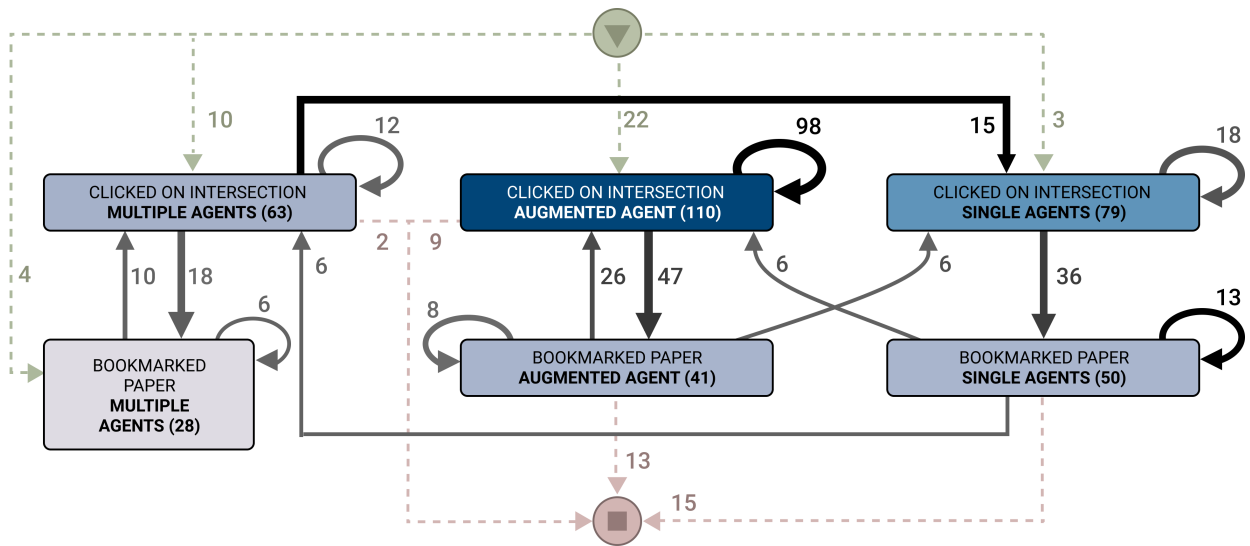


Figure 16: Process discovery map of interaction logs of Study 3, presenting cross-user interaction patterns for the exploration of different perspectives. Interaction is represented in terms of paths leading from intersection exploration events (CLICKED ON INTERSECTION) to subsequent bookmarking events (BOOKMARKED PAPER). Dashed arrows point to activities that occurred at the very beginning or at the very end of the process.

of augmented agents (frequency of 98) supporting the notion that allowing users to interleave the recommendations of automatic agents with other sources does foster exploration. The number of successive explorations of augmented agents, however, was more expressive in Study 3 than in Study 2, where the majority (98 for Study 3 vs. 15 for Study 2) of the explorations of augmented agents were preceded by explorations of the same kind. This discrepancy may be explained by the technology orientation of the backgrounds of participants of Study 3, which may have resulted in ease of use and thus increased exploration of intersections between agent recommendations and user tags and bookmarks.

Questionnaire Data

The results of the questionnaire reveal that participants were, overall, positive about IntersectionExplorer. The only assessed dimensions of usability that did not receive a high rating (i.e., a median of 4) were *fun*, *effort* and *interface adequacy* (all with a median of 3) (see Figure 17 for a complete overview of the questionnaire results and Section 5.1 for more details on the questionnaire constructs).

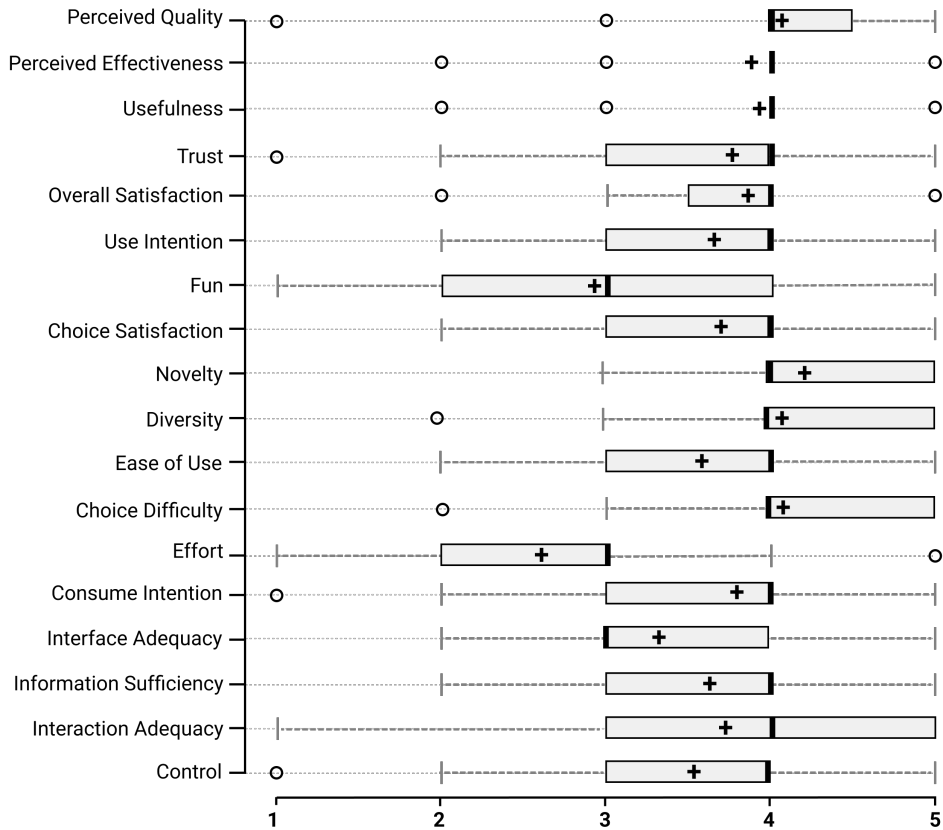


Figure 17: Questionnaire results for Study 3 (I depicts the medians, O the outliers, + the averages, the boxes represent the middle 50% of the data, and the whiskers stretch from the minimum to the maximum values excluding the outliers).

6. Comparison with Prior Work

As previously stated (see Section 2.4), IntersectionExplorer shares its core concept with TalkExplorer and SetFusion: both are set-based visualization tools, dedicated to the exploration of collections of items through multiple perspectives of relevance. However, the results of our previous work [37] suggest that scalability limitations in TalkExplorer’s and SetFusion’s user interfaces, respectively based on cluster maps and Venn diagram visualizations, prevented participants from exploring combinations of more than three sets.

Addressing this issue, IntersectionExplorer’s user interface is based on UpSet, another set-based visualization, that was developed to allow exploration of large amounts of complex data (more details in Section 3). Therefore, in order to understand the impact of this approach in IntersectionExplorer’s potential for exploration, we briefly contrast some of the results of Study 1 (presented in Section 5.2) with findings from a prior user study that asked 18 participants to freely use TalkExplorer to explore conference papers [37]. Although these two studies were not originally designed to draw a comparison between these tools, both were conducted with participants sampled from the attendees of two editions of the EC-TEL conference (2012 and 2015, respectively for TalkExplorer and IntersectionExplorer),

and both have no limitations on the number of sets that can be combined and explored by users. Results of the two studies are presented in Table 10.

Table 10: Number of explorations and precision scores (more information in Section 5.1) of different-sized combinations of sets, for IntersectionExplorer Study 1 (see Section 5.2) and a prior user study [37] that evaluated TalkExplorer with 18 participants. For Study 1, there is a statistically significant positive correlation between the amount of papers combined and the precision scores (respectively, the 1st and 5th columns).

Sets	TalkExplorer Study		IntersectionExplorer (Study 1)	
	Explorations	Precision	Explorations	Precision
1	234	0.15	25	0.13
2	28	0.35	16	0.09
3	8	0.17	21	0.19
4	0	0	13	0.35
5	0	0	3	0.23
6	0	0	6	0.38

The most straightforward observation concerns the number of explorations made with IntersectionExplorer and TalkExplorer. While participants of Study 1 of the current work explored combinations of up to 6 sets with IntersectionExplorer, the largest combination explored with TalkExplorer involved only 3 sets [37]. This suggests that the UpSet-based visualization of IntersectionExplorer’s user interface is indeed more scalable than our previous approaches, thereby fostering the exploration of larger combinations of sets.

A Pearson product-moment correlation coefficient indicates a statistically significant positive, strong correlation between the amount of sets combined and the precision scores for multi-set explorations made in IntersectionExplorer ($r = 0.8383, p < .05, n = 6$). In turn, such a trend could not be identified for TalkExplorer, as a maximum of 3 sets combined were explored at one time, resulting in a positive yet non-statistically significant correlation coefficient between the amount of sets combined and the precision scores. It is noteworthy that the correlation between these two variables is also not statistically significant for IntersectionExplorer, if we only consider the data from explorations of combinations involving up to 3 sets (the first 3 rows of Table 10). This suggests that the value of multi-set explorations becomes more prominent when users explore larger combinations of sets, further illustrating the value of IntersectionExplorer’s scalable visualization for increasing precision and helping users find useful items.

7. Discussion

7.1. Study Design

The first thing that should be discussed to position our contributions concerns the design of our studies. We designed the three studies to provide a comprehensive understanding of IntersectionExplorer in terms of the way that users interact with and find value in it. Study 1 was designed as an assessment of IntersectionExplorer’s general viability against more

traditional user interfaces for recommender systems such as the one implemented in the CN3 platform. In turn, studies 2 and 3 were performed to understand how participants with different backgrounds interact with and find value in IntersectionExplorer. Since an assessment of IntersectionExplorer’s overall viability was already established in Study 1, and given the considerable amount of time it would require participants to complete the five-paper-bookmarking task of our studies on both tools, we did not repeat the IntersectionExplorer vs. CN3 comparison in the following studies. Another important distinction between Study 1 and studies 2 and 3 concerns the settings in which they were conducted. While Study 1 was conducted in a controlled setting, studies 2 and 3 took place live and in more realistic conditions. To provide some details, as mentioned before in Section 5.2, the sessions of Study 1 took place either live or via video call after the conference, and in the presence of an observer during the whole session. In turn, the sessions of studies 2 and 3 took place during the conference and, besides the introductory briefing given by a researcher, participants were left to perform the tasks freely by themselves.

As explained in Section 5.1, our experiments required participants to find and bookmark five papers of interest in the CN3 platform before starting the experiments, so as to allow the production of relevant recommendations and thereby address the “cold start” problem. We did not impose any limits on the time that our participants had to make this preliminary paper-bookmarking setup step or in some other way attempted to control the exploration, other than emphasizing that the bookmarked papers should be of relevance. This approach, we believe, placed participants at the same level regarding the relevance of the recommendations made by the CN3 recommenders and, by extension, IntersectionExplorer’s agents.

Because we deployed IntersectionExplorer on top of the CN3 platform, bookmarks created through it were actually created in the users’ CN3 accounts and were persistent. This may explain the amount of bookmarks created by participants in our studies, as some may have taken the chance during the experiment to bookmark more than the requested number of papers in order to keep them for future reference in their CN3 accounts.

7.2. Use Effort and Experience

Regarding the results of our studies, we begin by discussing the baseline comparison between IntersectionExplorer and the ranked list representation of CN3 in Study 1. An analysis of median task-completion time and amount of steps needed to finish the five-paper-bookmarking task reveals that participants of Study 1 required both more time and interactions to create their first bookmark in IntersectionExplorer than they did in the CN3. However, after this “training phase”, the operational efficiency of both tools tends to converge. In other words, IntersectionExplorer has a steeper learning curve than CN3, but the difference tends to disappear as users get acquainted with the former, rendering both tools comparable in terms of time and interaction effort.

A likely explanation for this initial discrepancy lies in the standard interface and interaction design of CN3’s recommendations page, which is composed of single-perspective, relevance-ranked lists, versus the novel, more information-dense visualization that comprises the user interface of IntersectionExplorer. Indeed, lists of items ranked by relevance are traditional ways of presenting recommendations to users and we can find them in popular web

sites like YouTube [75] and the results page of both Google’s [76] and Yahoo’s [77] web search engines. Their widespread use makes them representative of the current standard in recommender systems’ user interfaces and, in that capacity, have previously been used as baselines in research (e.g., [78, 41, 40]). On the other hand, UpSet, the intersection-based visualization implemented in IntersectionExplorer, is rather novel and, to the extent of our knowledge, has been used mostly in research (e.g., [56, 57]). This novelty requires users to try and explore the tool, likely accounting for the initial median extra time and steps taken by participants when interacting with IntersectionExplorer versus the CN3. The tendentially decreasing time and number of interaction steps to create bookmarks as users get familiar with the tool that we found in Study 1 has been confirmed by the results of both studies 2 and 3. At this point, an interesting observation to be made concerns the time measurements in our studies and the impact of the think-aloud protocol, which may have accounted for an overhead in our time measurements. However, because we used this protocol in all our experiments, the potential overhead is the same for all measurements and, consequentially, can be considered negligible for the purposes of comparison. However, the times we present in this work may be assumed to include this overhead and, consequentially, to be an upper bound for IntersectionExplorer’s real operation time.

The difference between both tools is more accentuated in the questionnaire results, though, as participants tended to favor IntersectionExplorer multi-perspective visualization over CN3’s traditional, single-perspective, ranked-list recommendation layout. Even if the dimensions assessed by the questionnaire items mainly pertain to subjective evaluations of usability, it is revealing, and encouraging, that they generally seem to support the objective results of task-taken time and steps. Participants tended to consider IntersectionExplorer to be more effective than the ranked-lists of CN3, and the recommendations of the former to be more trustworthy. This last result in particular may be the result of IntersectionExplorer’s addressing of the “black box” issue through the clear statement of the recommendations’ sources and rationales. By augmenting agents with other perspectives such as bookmarks of similar users or tags, users can understand the rationale of several recommender engines that rely on these information sources. Although for the constructs of *ease of use* and *choice difficulty* responses were comparable for both tools, participants favored IntersectionExplorer over CN3 for all other questionnaire dimensions, notably *overall satisfaction* and *perceived effectiveness*. Additionally, participants reported that they perceived the papers recommended through IntersectionExplorer to be of higher quality, and also found the overall use experience more satisfying. Altogether, these results suggest that IntersectionExplorer is an interesting alternative to more traditional, single-perspective approaches for recommendation-based exploration of large item collections.

Across studies (see Figure 8, 13 and 17), median questionnaire results are very comparable and almost all dimensions were evaluated with a median score of 4. Exceptions are *usefulness* (3 vs. 3 vs. 4, respectively for studies 1, 2 and 3), *use intention* (4 vs. 3 vs. 4), *fun* (4 vs. 2.5 vs. 3), *choice difficulty* (2 vs. 3 vs. 4), *effort* (2 vs. 2.5 vs. 3), *interface adequacy* (4 vs. 3 vs. 3), *information sufficiency* (4 vs. 3 vs. 4), *interaction adequacy* (3.5 vs. 3.5 vs. 4) and *control* (4 vs. 3.5 vs. 4). It is revealing that Study 3 yielded higher median results for the dimensions of *usefulness* and *interaction adequacy* than studies 1 and 2, and

also that participants of both studies 1 and 3 evaluated the dimensions of *use intention*, *fun*, *information sufficiency* and *control* higher than participants of Study 2. A possible explanation for these observed disparities may be found in the different general backgrounds of the attendees of the conferences at which the studies took place. While Study 2 took place at the venue of a large event gathering mainly academics from the area of social sciences and humanities (DH 2016), studies 1 and 3 were conducted at technology focused events (respectively EC-TEL 2015 and IUI 2017; see demographic data in Sections 5.2.2, and 5.3.2 and 5.4.2). In light of these different backgrounds, it is generally expectable for participants of Studies 1 and 3 to be more familiar with novel, technology-based approaches to complex interaction problems than those of Study 2, and thereby to have a more positive appraisal of their challenges, value and promise. This assumption finds support in the research of Knijnenburg et al. [79] which concluded that the best interaction mechanisms are associated to user characteristics, as users with more domain knowledge generally report higher user satisfaction with interfaces that provide more control while less experienced users prefer simpler interfaces without many controllable variables.

7.3. Exploration of Agents

From the prospect of log analysis, the most notable result discovered in our studies is the value of “augmenting” agent recommendations with other perspectives (Table 1, 4 and 7). In all three studies the highest precision scores obtained through the use of recommendations (i.e., agents) were achieved by exploring combination of agents with tags or bookmarks of attendees. These results support the notion that combining recommendations with other information sources such as bookmarks of known users or tags increase the probability of finding a relevant paper. Moreover, these augmented recommendations also emerged as the most popular way of using agents in all three studies. While the value and popularity of using single and multiple agents varied across conferences (in particular, the least technologically-oriented attendees in Study 2 almost never combined two or more agents), the “augmented agents” were popular and efficient for all audiences.

The results of the process discovery maps of studies 2 and 3 (see Figure 12 and Figure 16, respectively) provide further support on the value of augmented agents: the explorations of augmented agent recommendations were not only the most frequent action, but also the actions that most frequently were followed by explorations of the same type. This effect was most pronounced in Study 3 where we also observed that a significantly higher proportion of papers was bookmarked through explorations of augmented agents. Altogether, our data suggests that participants were not only interested in exploring combinations of machine-produced recommendations with bookmarks of users and tags, but also that this “augmentation” actually resulted in increased likelihood of finding relevant papers in explorations.

7.3.1. Exploration of Items

As previously stated (see Section 1), we expected that IntersectionExplorer’s scalable visualization, the UpSet [13], would lend our tool more scalability potential than our previous approaches (see Section 3). The results of our three studies support this hypothesis, as

participants of three studies did explore a considerable number of multi-set intersections involving more than two sets (see Sections 5.2.2, 5.3.2 and 5.4.2). This contrasts with our previous work [37], where participants never explored combinations of more than three sets in a user interface based on cluster maps [80] in field studies. As such, it can be inferred that IntersectionExplorer is more scalable than our prior approaches, allowing successful exploration of intersections of multiple sets in parallel. However, in our three studies, the amount of explorations seems to decrease as the number of sets combined increases. A likely explanation for this lies in IntersectionExplorer’s interaction design, as users have to find and select the sets they want to combine in the Set Selection View (see Figure 5). Also, as soon as a user selects a set, IntersectionExplorer promptly adds it to the Set Exploration View and displays its combinations with the previously selected sets. To provide an example, before exploring combinations of say, three sets, users have already been presented with the combinations of the two sets selected before and also with the single set that they had selected initially. Because interesting papers may be potentially found while exploring combinations of any number of interesting sets (or a single set) it is likely that IntersectionExplorer users do not really need to combine multiple sets before finding relevant papers.

The analysis of single-perspective explorations and precision scores (Table 2, 5 and 8) reveals how participants interacted and found value in the different individual perspectives offered by IntersectionExplorer. An interesting aspect is that the within-perspective precision scores for single- and multi-set explorations do not differ greatly in any of the three studies. This indicates that participants found the same relative amount of relevant papers in explorations, regardless of the amount of same-perspective sets involved. This suggests that, when exploring a single perspective, the actual nature of the perspective itself exerts more influence in the percentage of relevant papers per exploration, than the amount of sets simultaneously explored. Nonetheless, although the precision is comparable between single- and multi-set explorations of the same perspective, it is worth to mention that the increased amount of explorations and the higher absolute number of bookmarked papers when multiple sets are explored highlight the value of allowing for multi-set explorations of single perspectives.

Comparing between perspectives (see Table 2, 5 and 8), the precision scores of the three studies tended to be higher for explorations of the perspective of content-relevance, indicating that the likelihood of a paper being relevant is higher for explorations involving only this perspective. This observation may be explained by the nature of this perspective: since its sets are composed of papers tagged with single keywords generally derived from the papers’ contents or the experience of users with it [12], they are likely composed of more focused and content-similar sets of papers that are potentially more relevant to users interested in the tag. The number of explorations, on the other hand, reveals that participants in different studies tended to focus their single-perspective explorations on different perspectives. Indeed, studies 1 and 3 had a higher number of explorations of the perspective of personal relevance (respectively, 66 and 142 explorations), whereas those of Study 2 tended to focus on the perspective of content relevance (80 explorations). Once again, the technology focus (or lack thereof) of our participants’ backgrounds may shed some light on this discrepancy. As studies 1 and 3 were conducted in events significantly more oriented towards technology,

it can be expected that due to their participants’ affinity with technology, their interest in the recommendations of automatic agents will be increased. In contrast, Study 2’s focus on humanities may explain why participants preferred to explore the perspective of content relevance by combining sets of tagged papers - an indeed more social way of getting recommendations. But perhaps the most important conclusion to be drawn from this discussion is that these results suggest that the IntersectionExplorer’s approach is flexible enough to let its users with different backgrounds adopt the system to their preferred ways of exploration.

8. Conclusions

We finalize this work addressing the research questions we posed initially (see Section 1) in light of the results obtained in the course of our three user studies. To recall, we asked “*how does IntersectionExplorer’s multi-perspective approach facilitate the exploration of collections of items?*” and “*how flexible and usable is IntersectionExplorer in responding to the requirements of users with different technological backgrounds?*”.

Addressing the first question, both in terms of subjective, self-reporting evaluations of usability as well as those of the more objective metrics (i.e., precision, time and steps taken), the results of our studies suggest that the multi-perspective approach to recommendation exploration is a promising way of addressing a very complex interaction challenge. IntersectionExplorer was found to be a relatively fast and effortless tool to navigate through conference papers. Objective measures of performance linked to interaction indicate that users were not only interested in exploring combinations of machine-produced recommendations with bookmarks of users and tags, but also that this “augmentation” actually resulted in increased likelihood of finding relevant papers in explorations. Overall, the findings point towards the viability of IntersectionExplorer as an effective tool, and indicate that its multi-perspective approach to exploring recommendations has great promise as a way of addressing the complex human-recommender system interaction problem.

To be sure, the comparison between IntersectionExplorer’s multi-perspective approach and CN3’s baseline (the traditional, relevance-ordered list of recommendations) allows us to conclude that the former is an interesting alternative to more traditional, single-perspective approaches, that potentially brings additional value to recommendation-based exploration of collections of items. In spite of the added conceptual and interaction complexity eventually brought by exploring item sets through mixed-perspective intersections, participants of our three studies were interested in this functionality and used it to find relevant items. This conclusion also finds support in the results of the usability questionnaires of the three studies which, in Study 1, favored IntersectionExplorer over the ranked list representation of CN3, and were generally positive in the other two studies. In terms of usability, the median time and number of interaction steps taken by participants of the three studies imply the existence of a process learning curve in IntersectionExplorer that, once surpassed, reveals a tool that is both relatively fast and effortless to use.

Additionally, after the definition of precision that we articulate in this work (see Section 5.1), our results suggest that allowing users to have different perspectives over the recommendations of automatic agents is an interesting and added-value approach. This ob-

ervation seems straightforward if we bear in mind that, in spite of the undeniable advances in recommender systems science and user interfaces, our social nature dictates that we are (and likely always will be) interested in the suggestions, judgments and interests of others in our circle of peers, family and friends. Thereby, based on the results of our studies, we believe that integrating multiple perspectives is indeed a valuable feature to consider when designing new recommender systems and their user interfaces.

Regarding our second research question, another interesting conclusion to draw from the results of our three studies concerns IntersectionExplorer’s flexibility, as they indicate that users with different professional backgrounds use IntersectionExplorer in different ways to explore and find items of interest. Indeed, the differences we found in our studies in terms of precision scores, explored perspectives and the questionnaire results suggest that IntersectionExplorer’s approach is flexible enough to allow its users to select and combine perspectives the way they judge the most productive, thereby coming to customized approaches to exploration that best fit their preferences.

Although we have explored IntersectionExplorer in different settings and derived important conclusions from our experiments, more studies are necessary to assess the tool’s viability in other domains. Therefore, as guidelines for future research, we would propose evaluating IntersectionExplorer in diverse scenarios of application to understand how multiple perspectives contribute to a more effective and fulfilling exploration experience in different settings and requirements. For instance, in the recommendation of emotion-eliciting items (e.g. music, movies, poetry) or in e-commerce, where the number of recommendable items likely scales beyond that of papers in conference proceedings. Moreover, studying IntersectionExplorer in other domains would allow to understand if the domains themselves change the way that users interact and find value in the perspectives. For example, items marked as relevant by work colleagues may be more interesting while searching for work-related items than they would be for selecting movies.

Additionally, even though UpSet was developed with scalability in mind, allowing to present the relations between many sets simultaneously using the same visual encoding [13], the visualization may still become crowded when a considerable number of sets are explored simultaneously. Therefore, we believe that other scalable visualizations should be investigated, and assessments of other dimensions of usability should be conducted (e.g., cognitive load) to further understand how the multi-perspective approach can be brought closer to its promise.

Acknowledgments

We thank the anonymous reviewers for the many useful comments that helped improve this work to a great extent. Part of this work has been supported by the KU Leuven Research Council (grant agreement C24/16/017 and StG/14/019) and the Research Foundation Flanders (FWO, grant agreement no. G0C9515N).

References

- [1] G. Adomavicius, A. Tuzhilin, Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, *IEEE transactions on knowledge and data engineering* 17 (6) (2005) 734–749.
- [2] M. Kunaver, T. Porl, Diversity in recommender systems a survey, *Know.-Based Syst.* 123 (C) (2017) 154–162. doi:10.1016/j.knosys.2017.02.009.
URL <https://doi.org/10.1016/j.knosys.2017.02.009>
- [3] L. Chen, W. Wu, L. He, Personality and recommendation diversity, in: *Emotions and Personality in Personalized Services*, Springer, 2016, pp. 201–225.
- [4] L. Chen, G. Zhang, H. Zhou, Improving the diversity of top-n recommendation via determinantal point process, *arXiv preprint arXiv:1709.05135* (2017) 8.
- [5] K. Verbert, D. Parra, P. Brusilovsky, E. Duval, Visualizing recommendations to support exploration, transparency and controllability, in: *Proceedings of the 2013 international conference on Intelligent user interfaces*, ACM, 2013, pp. 351–362.
- [6] D. Parra, P. Brusilovsky, User-controllable personalization: A case study with setfusion, *International Journal of Human-Computer Studies* 78 (2015) 43–67.
- [7] H. Chen, D. Parra, K. Verbert, Interactive recommender systems: a survey of the state of the art and future research challenges and opportunities, *Expert Systems with Applications* 56 (2016) 9–27.
- [8] J. O’Donovan, B. Smyth, B. Gretarsson, S. Bostandjiev, T. Höllerer, Peerchooser: visual interactive recommendation, in: *Proc CHI ’08*, ACM, 2008, pp. 1085–1088.
- [9] B. Gretarsson, J. O’Donovan, S. Bostandjiev, C. Hall, T. Höllerer, Smallworlds: visualizing social recommendations, in: *Computer Graphics Forum*, Vol. 29, Wiley Online Library, 2010, pp. 833–842.
- [10] K. Verbert, D. Parra, P. Brusilovsky, Agents vs. users: Visual recommendation of research talks with multiple dimension of relevance, *ACM Transactions on Interactive Intelligent Systems (TiiS)* 6 (2) (2016) 11.
- [11] P. Brusilovsky, J. S. Oh, C. López, D. Parra, W. Jeng, Linking information and people in a social system for academic conferences, *New Review of Hypermedia and Multimedia* 23 (2) (2017) 81–111.
- [12] I. Cantador, A. Bellogín, D. Vallet, Content-based recommendation in social tagging systems, in: *Proceedings of the fourth ACM conference on Recommender systems*, ACM, 2010, pp. 237–240.
- [13] A. Lex, N. Gehlenborg, H. Strobel, R. Vuillemot, H. Pfister, Upset: visualization of intersecting sets, *Visualization and Computer Graphics*, *IEEE Transactions on* 20 (12) (2014) 1983–1992.
- [14] S. M. McNee, J. Riedl, J. A. Konstan, Being accurate is not enough: how accuracy metrics have hurt recommender systems, in: *CHI’06 extended abstracts on Human factors in computing systems*, ACM, 2006, pp. 1097–1101.
- [15] D. J. Hand, H. Mannila, P. Smyth, *Principles of data mining*, MIT press, 2001.
- [16] I. Barry Crabtree, S. J. Soltysiak, Identifying and tracking changing interests, *International Journal on Digital Libraries* 2 (1) (1998) 38–53.
- [17] M. Albanese, A. d’Acierno, V. Moscato, F. Persia, A. Picariello, A multimedia recommender system, *ACM Transactions on Internet Technology (TOIT)* 13 (1) (2013) 3.
- [18] W.-P. Lee, C.-H. Liu, C.-C. Lu, Intelligent agent-based systems for personalized recommendations in internet commerce, *Expert Systems with Applications* 22 (4) (2002) 275–284.
- [19] J. Lu, Personalized e-learning material recommender system, in: *International conference on information technology for application*, Citeseer, 2004, pp. 374–379.
- [20] A. C. Valdez, M. Zieffle, K. Verbert, A. Felfernig, A. Holzinger, Recommender systems for health informatics: State-of-the-art and future perspectives, in: *Machine Learning for Health Informatics*, Springer, 2016, pp. 391–414.
- [21] J. L. Herlocker, J. A. Konstan, L. G. Terveen, J. T. Riedl, Evaluating collaborative filtering recommender systems, *ACM Transactions on Information Systems (TOIS)* 22 (1) (2004) 5–53.
- [22] J. Bennett, S. Lanning, et al., The netflix prize, in: *Proceedings of KDD cup and workshop*, Vol. 2007, New York, NY, USA, 2007, p. 35.

- [23] D. Parra, P. Brusilovsky, Collaborative filtering for social tagging systems: an experiment with citeulike, in: Proc. RecSys '09, ACM, 2009, pp. 237–240.
- [24] Y. Hu, Y. Koren, C. Volinsky, Collaborative filtering for implicit feedback datasets, in: Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on, Ieee, 2008, pp. 263–272.
- [25] M. J. Pazzani, D. Billsus, Content-based recommendation systems, in: The adaptive web, Springer, 2007, pp. 325–341.
- [26] R. Burke, Hybrid recommender systems: Survey and experiments, User modeling and user-adapted interaction 12 (4) (2002) 331–370.
- [27] J. A. Konstan, J. Riedl, Recommender systems: from algorithms to user experience, User modeling and user-adapted interaction 22 (1) (2012) 101–123.
- [28] C.-N. Ziegler, S. M. McNee, J. A. Konstan, G. Lausen, Improving recommendation lists through topic diversification, in: Proceedings of the 14th international conference on World Wide Web, ACM, 2005, pp. 22–32.
- [29] R. Sinha, K. Swearingen, The role of transparency in recommender systems, in: CHI'02 extended abstracts on Human factors in computing systems, ACM, 2002, pp. 830–831.
- [30] N. Tintarev, J. Masthoff, A survey of explanations in recommender systems, in: Data Engineering Workshop, 2007 IEEE 23rd International Conference on, IEEE, 2007, pp. 801–810.
- [31] J. L. Herlocker, J. A. Konstan, J. Riedl, Explaining collaborative filtering recommendations, in: Proc. CSCW '00, ACM, 2000, pp. 241–250.
- [32] K. Swearingen, R. Sinha, Interaction design for recommender systems, in: Designing Interactive Systems, Vol. 6, 2002, pp. 312–334.
- [33] N. Tintarev, J. Masthoff, Designing and evaluating explanations for recommender systems, in: Recommender Systems Handbook, Springer, 2011, pp. 479–510.
- [34] B. P. Knijnenburg, S. Bostandjiev, J. O'Donovan, A. Kobsa, Inspectability and control in social recommenders, in: Proceedings of the Sixth ACM Conference on Recommender Systems, RecSys '12, ACM, New York, NY, USA, 2012, pp. 43–50. doi:10.1145/2365952.2365966. URL <http://doi.acm.org/10.1145/2365952.2365966>
- [35] S. Bostandjiev, J. O'Donovan, T. Höllerer, Tasteweights: a visual interactive hybrid recommender system, in: Proc. RecSys'12, ACM, 2012, pp. 35–42.
- [36] D. Parra, P. Brusilovsky, C. Trattner, See what you want to see: visual user-driven approach for hybrid recommendation, in: Proceedings of the 19th international conference on Intelligent User Interfaces, ACM, 2014, pp. 235–240.
- [37] K. Verbert, D. Parra, P. Brusilovsky, The effect of different set-based visualizations on user exploration of recommendations, in: Proceedings of the Joint Workshop on Interfaces and Human Decision Making in Recommender Systems, Joint Workshop on Interfaces and Human Decision Making in Recommender Systems, 6 October 2014, 2014, pp. 37–44. URL <https://lirias.kuleuven.be/handle/123456789/476492>
- [38] P. Brusilovsky, Personalization in the context of relevance-based visualization, in: Proceedings of the 2017 ACM Workshop on Exploratory Search and Interactive Data Analytics, ESIDA '17, ACM, New York, NY, USA, 2017, pp. 49–50. doi:10.1145/3038462.3038474. URL <http://doi.acm.org/10.1145/3038462.3038474>
- [39] M. A. Hearst, Tilebars: visualization of term distribution information in full text information access, in: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press/Addison-Wesley Publishing Co., 1995, pp. 59–66.
- [40] O. Hoerber, X. D. Yang, The visual exploration of web search results using hotmap, in: Information Visualization, 2006. IV 2006. Tenth International Conference on, IEEE, 2006, pp. 157–165.
- [41] C. di Sciascio, V. Sabol, E. E. Veas, Rank as you go: User-driven exploration of search results, in: Proceedings of the 21st International Conference on Intelligent User Interfaces, IUI '16, ACM, New York, NY, USA, 2016, pp. 118–129. doi:10.1145/2856767.2856797. URL <http://doi.acm.org/10.1145/2856767.2856797>
- [42] A. Spoerri, Infocrystal: A visual tool for information retrieval & management, in: Proc. CIKM '93,

- ACM, 1993, pp. 11–20.
- [43] V. Geroimenko, C. Chen, *Visualizing the semantic web: XML-based internet and information visualization*, Springer, London, 2006.
- [44] J. Schaffer, T. Höllerer, J. O’Donovan, Hypothetical recommendation: A study of interactive profile manipulation behavior for recommender systems., in: *FLAIRS Conference*, 2015, pp. 507–512.
- [45] J. Venn, I. on the diagrammatic and mechanical representation of propositions and reasonings, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 10 (59) (1880) 1–18.
- [46] I. Andjelkovic, D. Parra, J. O’Donovan, Moodplay: Interactive mood-based music discovery and recommendation, in: *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, ACM, 2016, pp. 275–279.
- [47] B. Loepp, K. Herrmann, J. Ziegler, Blended recommending: Integrating interactive information filtering and algorithmic recommender techniques, in: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI ’15*, ACM, New York, NY, USA, 2015, pp. 975–984. doi:10.1145/2702123.2702496.
URL <http://doi.acm.org/10.1145/2702123.2702496>
- [48] L. Gou, F. You, J. Guo, L. Wu, X. L. Zhang, Sfviz: interest-based friends exploration and recommendation in social networks, in: *Proc. VINCI ’11*, ACM, 2011, p. 15.
- [49] M. C. Chuah, Dynamic aggregation with circular visual designs, in: *Information Visualization, 1998. Proceedings. IEEE Symposium on*, IEEE, 1998, pp. 35–43.
- [50] J. Vig, S. Sen, J. Riedl, Tagsplanations: Explaining recommendations using tags, in: *Proceedings of the 14th International Conference on Intelligent User Interfaces, IUI ’09*, ACM, New York, NY, USA, 2009, pp. 47–56. doi:10.1145/1502650.1502661.
URL <http://doi.acm.org/10.1145/1502650.1502661>
- [51] E. Tonkin, E. M. Corrado, H. L. Moulaison, M. E. Kipp, A. Resmini, H. D. Pfeiffer, Q. Zhang, Collaborative and social tagging networks, *Ariadne* 54 (Jan).
- [52] Y. Kammerer, R. Nairn, P. Pirolli, E. H. Chi, Signpost from the masses: Learning effects in an exploratory social tag search browser, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’09*, ACM, New York, NY, USA, 2009, pp. 625–634. doi:10.1145/1518701.1518797.
URL <http://doi.acm.org/10.1145/1518701.1518797>
- [53] S. Zhao, M. X. Zhou, Q. Yuan, X. Zhang, W. Zheng, R. Fu, Who is talking about what: social map-based recommendation for content-centric social websites, in: *Proceedings of the fourth ACM conference on Recommender systems*, ACM, 2010, pp. 143–150.
- [54] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent dirichlet allocation, *Journal of machine Learning research* 3 (Jan) (2003) 993–1022.
- [55] M. D. Ekstrand, D. Kluver, F. M. Harper, J. A. Konstan, Letting users choose recommender algorithms: An experimental study, in: *Proceedings of the 9th ACM Conference on Recommender Systems, RecSys ’15*, ACM, New York, NY, USA, 2015, pp. 11–18. doi:10.1145/2792838.2800195.
URL <http://doi.acm.org/10.1145/2792838.2800195>
- [56] L. T. Gray, Z. Yao, T. N. Nguyen, T. K. Kim, H. Zeng, B. Tasic, Layer-specific chromatin accessibility landscapes reveal regulatory networks in adult mouse visual cortex, *eLife* 6 (2017) e21883.
- [57] M. Dugas, P. Neuhaus, A. Meidt, J. Doods, M. Storck, P. Bruland, J. Varghese, Portal of medical data models: information infrastructure for medical research and healthcare, *Database* 2016.
- [58] C. D. Manning, P. Raghavan, H. Schütze, et al., *Introduction to information retrieval*, Vol. 1, Cambridge university press Cambridge, 2008.
- [59] R. Baeza-Yates, B. Ribeiro-Neto, et al., *Modern information retrieval*, Vol. 463, ACM press New York, 1999.
- [60] C. Wongchokprasitti, Using external sources to improve research talk recommendation in small communities, Ph.D. thesis, University of Pittsburgh (2015).
- [61] C. Lewis, Using the “thinking-aloud” method in cognitive interface design, *IBM TJ Watson Research Center*, 1982.

- [62] P. Pu, L. Chen, R. Hu, A user-centric evaluation framework for recommender systems, in: Proc. RecSys'11, ACM, 2011, pp. 157–164.
- [63] B. P. Knijnenburg, M. C. Willemsen, Z. Gantner, H. Soncu, C. Newell, Explaining the user experience of recommender systems, *User Modeling and User-Adapted Interaction* 22 (4-5) (2012) 441–504.
- [64] Ec-tel 2017, twelfth european conference on technology enhanced learning, accessed April 2017.
URL <http://www.ec-tel.eu/index.php?id=777>
- [65] Ec-tel 2014, ninth european conference on technology enhanced learning, accessed April 2017.
URL <http://ectel2014.httc.de/index.php?id=681>
- [66] Ec-tel 2015, ninth european conference on technology enhanced learning, accessed April 2017.
URL <http://ectel2015.httc.de/index.php?id=704>
- [67] K. Verbert, K. Seipp, C. He, D. Parra, C. Wongchokprasitti, P. Brusilovsky, Scalable exploration of relevance prospects to support decision making, in: *Proceedings of the Joint Workshop on Interfaces and Human Decision Making for Recommender Systems co-located with ACM Conference on Recommender Systems (RecSys 2016)*, CEUR-WS, 2016, pp. 28–35.
- [68] Clarin at digital humanities 2016, accessed April 2017.
URL <https://www.clarin.eu/blog/clarin-digital-humanities-2016>
- [69] Acceptances to dh2016 (pt.1), accessed April 2017.
URL <http://scottbot.net/acceptances-to-dh2016-pt-1/>
- [70] Submissions to dh2016 (pt.1), accessed September 2017.
URL <http://scottbot.net/submissions-to-dh2016-pt-1/>
- [71] W. M. van der Aalst, *Process mining: data science in action*, Springer, 2016.
- [72] W. M. Van Der Aalst, A. H. Ter Hofstede, Yawl: yet another workflow language, *Information systems* 30 (4) (2005) 245–275.
- [73] Acm iui 2017, accessed April 2017.
URL <http://iui.acm.org/2017>
- [74] Proceedings of the 22nd international conference on intelligent user interfaces, accessed April 2017.
URL <http://dl.acm.org/citation.cfm?id=3025171>
- [75] Youtube, accessed September 2017.
URL <https://www.youtube.com/>
- [76] Google web search, accessed September 2017.
URL <https://www.google.com>
- [77] Yahoo web search, accessed September 2017.
URL <https://www.yahoo.com/>
- [78] D. Parra, P. Brusilovsky, User-controllable personalization, *Int. J. Hum.-Comput. Stud.* 78 (C) (2015) 43–67. doi:10.1016/j.ijhcs.2015.01.007.
URL <http://dx.doi.org/10.1016/j.ijhcs.2015.01.007>
- [79] B. P. Knijnenburg, N. J. Reijmer, M. C. Willemsen, Each to his own: How different users call for different interaction methods in recommender systems, in: *Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys '11*, ACM, New York, NY, USA, 2011, pp. 141–148. doi:10.1145/2043932.2043960.
URL <http://doi.acm.org/10.1145/2043932.2043960>
- [80] Aduna clustermap, www.aduna-software.com/technology/clustermap, retrieved on-line 20 August 2014.