What if Interactive Explanation in a Scientific Literature Recommender System

Mouadh **Guesmi**, Mohamed Amine **Chatti**, Jaleh **Ghorbani-Bavani**, Shoeb **Joarder**, Qurat Ul **Ain** and Rawaa **Alatrash**

Social Computing Group, Faculty of Engineering, University of Duisburg-Essen, 47048 Duisburg, Germany

Abstract

Despite the vast amount of research on interactive recommender systems (RS) and explainable recommendation, there is a lack of research on how to incorporate interactivity features in explainable RS. To address this research gap, a possible solution to achieve interactive explanation could be to provide *What if* explanations that can help users iteratively build better mental models of how the RS works. Through an iterative human-centered design (HCD) process, we designed *What if* interactive explanations in the transparent Recommendation and Interest Modeling Application (RIMA) and explored how this type of explanation could impact the understandability of and interaction with an explainable RS. Our investigation showed that providing *What if* explanations has a positive effect on different aspects of a RS such as transparency, trust, user control, satisfaction, and user experience.

1. Introduction

Similar to most artificial intelligence (AI) applications, recommender systems (RS) often act as black boxes for the end-users. To alleviate this problem, explainable recommendation has attracted much attention in the RS research community. Explanations are a necessary condition to help users build an accurate mental model of the RS [1, 2, 3]. Generally, explanations seek to show how a recommended item relates to a user's preferences [4]. Explanation is inherently a social process [5, 6]. The social nature of explanation implies that an explainable RS has to be interactive. Over the past years, a growing body of research has focused on visual interaction and control mechanisms for RS [7, 8]. It has been shown that interactive recommendation and user control can improve user experience and trust in the RS [9, 10, 11]. Several researchers, however, identified that there is a lack of research on interactive explanation in RS [12, 13, 14, 15]. While many recognize the necessity to providing user control and interaction mechanisms in the context of explanations, how to design, implement, and evaluate interactive explanation in RS remains an open question.

To address this research gap, in this paper, we focus and elaborate on the concept of *What if* explanation as a viable path to interactive explanation. Our aim is to follow a human-in-the-loop approach where users can steer the explanation process to build an accurate mental model of the RS. Taking scientific literature RS as our domain, we systematically designed *What if* explanations in the transparent Recommendation and Interest Modeling Application (RIMA). When interacting with RIMA, users can ask exploratory *What if* questions to keep closing the gap of understanding and scrutinize the RS (i.e., correct the system's assumptions) if necessary. Further, we conducted moderated think-aloud sessions and semi-structured interviews with students and researchers (N=12) to systematically study how users perceive *What if* interactive explanations in an explainable RS. The results of our study show that providing *What-if* explanations has a positive effect on different aspects of a RS such as transparency, trust, user control, satisfaction, and user experience.

The main contribution of this paper is twofold: first, we follow a human-centered design (HCD) approach to effectively design *What if* interactive explanations aiming at clarifying the background behavior of the RS by allowing users to adjust the set of inputs and monitor the possible results. Second, we provide evidence on the positive impact of *What if* explanation on the perception of explainable recommendation.

This paper is organized as follows. We first outline the background for this research and discuss related work. We then present the systematic design of the *What if* explanations in RIMA. Afterwards, we describe the user study and present its results. Finally, we summarize the work and outline future research plans.

2. Related Work

2.1. Interactive Explanation

In a broader view, an explanation aims to make the reasons behind a decision or recommendation comprehensible to humans. Thus, work on XAI in general and explainable RS in particular must take a human-centered approach. To this end, the HCI community has called for interdisciplinary collaboration and user-centered approaches to XAI [16, 17]. For instance, Wang et al. [17] proposed a conceptual framework to connect XAI techniques and cognitive patterns in human-decision making to guide the design of user-centric XAI systems. Liao et al. [18] provided an XAI question bank and discussed how it can be used for creating user-centered XAI. Miller [5] synthesized perspectives on human explanation from philosophy, social science, and cognitive science and identified a list of human-friendly characteristics of explanation, including that human explanations are contrastive (i.e., "sought in response to particular counterfactual cases"), selective (i.e., selected in a ''biased manner" from a "sometimes infinite number of causes"), and social (i.e., conversational process, where an "explainer transfer knowledge to an explainee"). Similarly, Hilton [6] stressed that explanation is inherently a social process. It involves the interaction between explainer and explainee engaging in information exchange through dialogue, visual representation, and other communication modalities. The social nature of explanation implies that an XAI has to be interactive or even conversational [18]. With the goal of bridging the gap between XAI and HCI, research on designing and studying user interactions with XAI has emerged over the past few years [19, 20, 21, 22]. However, little is

known about how interactive explanation should be designed and implemented in RS, so that explanation goals such as scrutability, transparency, trust, and user satisfaction are met [12, 13].

Although interactive and more recently conversational RS have been well studied [7, 23, 8, 24, 25], there has been little work on how to incorporate interactivity features in explainable RS. Both in the literature and in real-world systems, there are only a few examples of RS that provide interactive explanations, mainly to allow users to scrutinize the provided recommendations and correct the system's assumptions [12, 26, 27, 28, 29, 30], or have a conversation, i.e., an exchange of questions and answers between the user and the system, using GUI-navigation or natural language conversation [13].

A distinction is to be made here between interactive recommendation and interactive explanation. While both empower users to take control of the RS process, they differ in the goal of the control action. While the primary goal of interactive recommendation is to improve and personalize the recommendation results, the goal of interactive explanation is to help both the users for better understanding and the system designers for better model debugging. In this paper, we focus on *What if* explanation as a possible mechanism to achieve interactive explanation in RS.

2.2. *What if* Explanation

An explanation can be seen as an answer to questions, also called intelligibility queries or types, such as *What, Why, How, Why not, How to*, and *What if* [31, 5, 18, 32]. Relevant to our work are *What if* explanations. These explanations allow users to speculate what the application would do given a set of user-set input values [33]. In XAI, they show how the prediction changes corresponding to changes of a feature (often in a visualization format) [18]. *What if* explanation differs from counterfactual (*How to*) explanation in that while the former asks about prospective future behavior (i.e., what if the factors were different, then what the effect would be?), the latter asks retrospectively (i.e., what needs to change for the alternative outcome to happen?) [34]. In an RS context, *What if* explanations deal with the manipulation of inputs to the RS. These explanations illustrate how the manipulation of inputs affects the output of the RS, i.e., recommendations. These explanations involve users' interaction with the system when they can change an input to the RS and want to know what will happen as a consequence. *What if* explanations aim to answer the question: "What if there is a change in my profile, what would happen?" [35].

The majority of previous research on *What if* scenarios in RS has focused on *What if* interactions with the RS, rather than with the explanation components of the RS. For example, Schaffer et al. [14] studied the effects of what they called hypothetical recommendations (i.e., recommendations generated by What if exploratory profile manipulations). In particular, the authors evaluated the effects of dynamic feedback from the RS on profile manipulations, the resulting recommendations, and the user's overall experience. Zürn et al. [36] discussed possible UI extensions to explicitly support *What if* interactions with RS, which allow users to explore, investigate and question algorithmic decision-making. Our work differs from previous approaches in that we focus on *What if* interactions with the explanation components of the RS and that we attempt to determine the impact of *What if* explanation on the perception of explainable recommendation.

3. *What if* Explanation for scientific literature recommendation

3.1. RIMA Application

We developed the transparent Recommendation and Interest Modeling Application (RIMA) with the goal of providing explainable interest models and recommendations. RIMA is a content-based RS that produces content-based explanations. It follows a user-driven personalized explanation approach by providing explanations with different levels of detail and empowering users to steer the explanation process the way they see fit [28, 26]. The application provides on-demand explanations, that is, the users can decide whether or not to see the explanation [15]. In this work, we focus on recommending scientific publications and leveraging explanatory visualizations to provide *What if* interactive explanations aiming at clarifying the background behavior of the RS by allowing users to adjust the set of inputs and monitor the possible results.

The user interest models in RIMA are automatically inferred from users' publications [37]. Based on these inferred interest models, the recommendation engine provides scientific publication recommendations. The top five interests (based on their weights) are initially used as input for the recommendation process. For obtaining the candidate publications, we use the semantic scholar API to fetch publications that contain, or are related to, one or more user interests that are used as input for the recommendation. We then apply an unsupervised keyphrase extraction algorithm on the fetched publications to extract keywords from the title and the abstract text. In order to compare the similarity between the user interests and the candidate publications, we use word embedding techniques to generate vector representations of the user interest model and the recommended publications. After getting the two embedding representations (i.e., interest model embedding and publication embedding), we calculate the cosine similarity between them in order to obtain a semantic similarity score. The top ten similar publications will be then recommended to the user. Initially, publications with a semantic similarity score above a threshold of 40% will be displayed to the user.

3.2. *What if* Explanation Design

An elegant translation of machine-generated explanations needs carefully designed human-understandable and satisfying explanations in the user interface [32]. The *What if* explanation presents an exploratory profile manipulation (i.e., addition, deletion, or re-weighting interests). Similar to counterfactual or contrastive explanations, *What if* explanations not only pinpoint the causes of a model decision but also provide users with actionable levers to change the recommendation [38]. According to Lim [39], the *What if* explanations would have to be interactive and dynamic, as they depend on example scenarios that users define themselves. Moreover, the results from Szymanski et al. [40] revealed that most users prefer visual explanations. Therefore, we aim at providing *What if* explanation using interactive visualizations.

The explanation scope or interpretation scale is another important dimension to be considered when designing explanations, which could be either local or global [32, 38]. Global explanation (or model explanation) is an explanation type that describes how the overall machine learning model works, while local explanations (or instance explanation) aim to explain the relationship between specific input-output pairs or the reasoning behind the results for an individual user query [32]. Moreover, local explanation is thought to be less overwhelming for novices, and it

can be suited for investigating edge cases for the model or debugging data [32]. Inspired by this distinction in terms of explanation scope, in this work, we aim at providing local and global *What if* explanations.

Building upon the insights from the literature outlined above, we followed the Human-Centered Design (HCD) approach [41] to systematically design interactive visualizations of the *What if* explanation. Designing with the HCD approach ensures that the needs and requirements of the user are taken into consideration as it is based on involving users from the very beginning and regularly consulting them for the evaluations of incremental prototypes. The HCD process consists of four consecutive activities, namely *Observation*, *Ideation*, *Prototyping*, and *Testing*. These four activities are iterated; that is, they are repeated over and over, with each cycle yielding more insights and getting closer to the desired solution [41]. The final design of these *What if* interactive explanations was the result of three HCD iterations. For evaluating the different explanation prototypes, a group of potential users was selected to participate in the design process. Our target group was researchers and students who are interested in scientific literature. For each design iteration, five different potential users were involved to test and give feedback on the provided prototypes, as recommended by Nielsen [42] in the case of qualitative user studies.

3.2.1. First iteration

Through this initial step, we aim at understanding users' needs and initiating the first low-fidelity prototypes for the *What if* explanation.

Observation. We conducted interviews with five potential users in order to gather the user's requirements in an explainable scientific literature RS. Through the interview, we investigated users' expectations from a *What if* explanation, as well as the expected level of interactivity and controllability over this explanation. Based on the interviews, we gained a better understanding of the end-user expectations and needs. The interviewed users agreed on four main scenarios for using *What if* explanations: (1) when they are not satisfied with the whole RS results, (2) when the recommended publications are not expected, (3) they want to interact with the RS to discover more recommendations, and (4) they are not satisfied with a specific recommended publication and interested to know why it was recommended.

Ideation. The ideation phase was focused on generating ideas about how to provide interactive *What if* explanations that address the four scenarios from the observation phase. A brainstorming session involving four authors and seven students from the local university having knowledge in RS and information visualization was carried out to collect as many ideas as possible for each scenario where we put quantity of ideas over quality. For each scenario, every idea was written down then discussed following a "pitch and critique" approach to gather both positive and negative feedback for each idea. The last step was the voting process to select the best ideas. In the end, we selected and categorized the top three voted ideas for each *What if* explanation scenario.

The ideation phase resulted in a global and a local *What if* explanation design. The global *What if* explanation aims at helping users interact with the explanation to answer the following questions: "What if I change the weights of my interests? " or "What if I change (add or delete) my interests?" in order to deal with the first three scenarios, namely when users are not satisfied

with or did not expect the RS results, or when they want to discover new recommendations. The local *What if* explanation addresses the fourth scenario when users might not be satisfied with a specific recommended publication and interested to know why it was recommended. In this case, users can interact with their interest profiles as well as with the set of keywords extracted from the publication for two reasons; either they want to understand the reason behind providing that specific publication as a recommendation through the provided similarity computations, or they are curious to know what would happen if their interests (or their weights) or the publication keywords changed. The local *What if* explanation can thus be used to question the recommendation of a specific publication by manipulating the interest model or the set of the extracted keywords from that publication. This explanation could answer the following questions: "What if I change my interests (add, delete, modify the weight)?" or "What if I change the keywords extracted from the publication (add or delete keywords)?".

Prototyping. The next step was to come up with possible visualizations for the global and local *What if* explanations. For each explanation scope, we discussed various visualizations and created low-fidelity prototypes as paper mock-ups. The goal of the global *What if* explanation is to provide users with an overview of the recommended publications, as well as reveal the relationship between these publications and the user's interests. We proposed three visualizations to show these relationships, namely bar chart, polar area chart, and stacked bar chart. Users can interact with the visualization by selecting publications for more details, or by changing the weight of an interest using a slider. The goal of the local *What if* explanation is to make users understand the relationship between their interests and a specific recommended publication. This explanation should allow users to manipulate (a) the interest model either by adding or deleting interests or changing the weight of specific interests or (b) the set of automatically extracted keywords from the publication by removing or adding new ones. We chose bar chart and polar area chart to visualize the impact of the interests' weights on the recommendation output, and heatmap to depict the similarity between all interests and all extracted keywords. Similar to the global *What if* explanation, users can interact with the visualization through selecting and changing actions.

Testing. The evaluation of the initial low-fidelity prototypes aims to receive feedback for optimization. This feedback was collected through a qualitative evaluation with five potential users following a think-aloud approach where we used open-ended questions to ask the users about their thoughts on each of the selected visualizations and their opinion towards the proposed *What if* explanations. The purpose is to understand to what extent each of the visualizations was able to convey the intended purposes of the global and local *What if* explanations to the user. Regarding the global *What if* explanation, users agreed that bar chart is the most suitable chart to present the similarity between their interests and the recommended publications. However, they mentioned that the new recommended publications obtained after interacting with the *What if* explanation are not specified or highlighted in the visualization. Also, they suggested avoiding polar area and stacked bar charts with the argument that if they have a considerable number of interests, the visualizations will be overwhelming and confusing to a certain degree. Similarly, they selected bar chart for the local *What if* explanation as polar chart, area chart, and heatmap should be avoided for the same previous reasons. Furthermore, they reported that they liked the feature of monitoring in real-time whether the publication will still be recommended or not after making the changes. Accordingly, the selected visualizations

for both global and local *What if* explanations were bar charts, as shown in Figure 1.
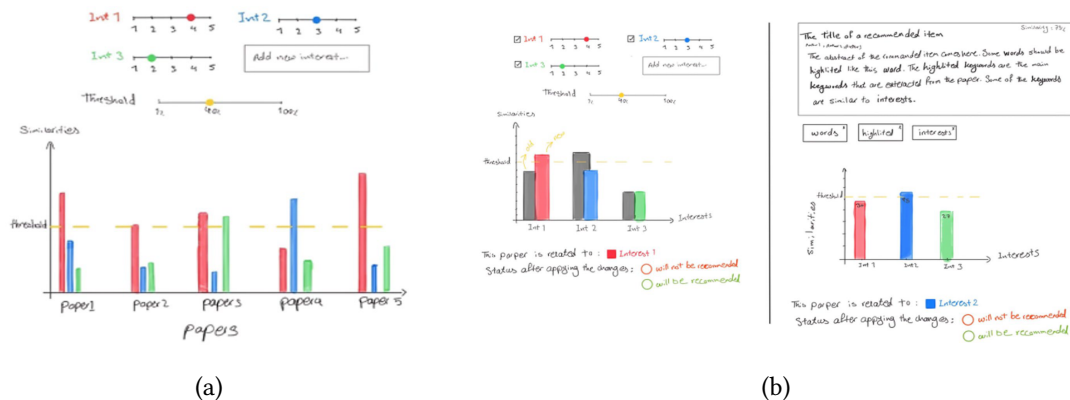
Figure 1: Low-fidelity prototypes: (a) global explanation, (b) local explanation.

3.2.2. Second iteration

This step aims to overcome the deficiencies of the previous designs by considering users' feedback collected from the previous testing phase. The prototypes in the second phase are designed using the Figma tool, but still considered as low-fidelity prototypes.

Prototyping. For the global *What if* explanation, the decided visualization was a bar chart that shows similarity scores between the user interest model and all the recommended publications (Figure 2). The control panel at the top of the visualization allows the user to change the weights of their interests. In addition, this visualization gives the user an overview of potential results (i.e., hypothetical recommendations) as a response to a user's change and asks for their decision to keep or cancel the changes (Figure 2a). In order to overcome the issue of specifying and highlighting the newly added publications, we used colors to indicate the publications' status and distinguish between the old recommendations (i.e., before interacting with the *what if* explanation), the omitted recommendations (i.e., those removed after interacting with the *What if* explanation), and the new recommended publications (i.e., those added after interacting with the *What if* explanation) (Figure 2b).

The aim of the local *What if* explanation is to explain to the user the factors influencing the recommendation process of a specific publication. This explanation should allow users to manipulate their interest models or the set of automatically extracted keywords from the publication and see the consequences of these actions. To achieve this, we used two bar chart visualizations, as suggested by users in the first iteration. In the first bar chart, each bar shows how similar are the individual interests to a specific publication (Figure 3a). The fixed gray bars in the background represent the initial similarity scores before interacting (i.e., adding/removing an interest or changing its weight) with the explanation. The displayed score on the bars represents the similarity score between a specific interest and the selected publication. In addition to manipulating their interest profiles, the users can also remove the extracted keywords from the publication or select new ones to be taken into consideration.
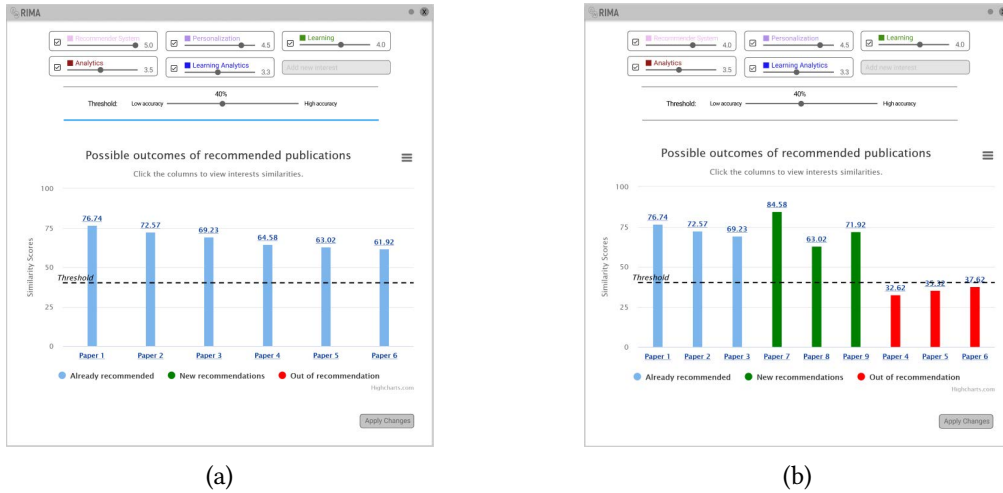
Figure 2: *What if* global explanation: (a) before interaction, (b) after interaction.

The second bar chart is used to show the similarities between the user-selected publication's keywords and the user interest model (Figures 3b and 3c).
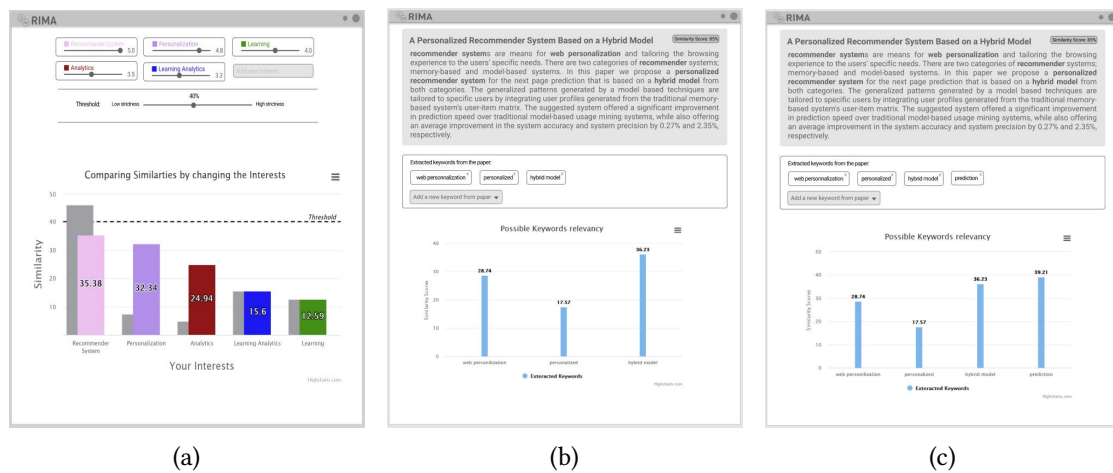

Figure 3: *What if* local explanation: (a) manipulating the interest model, (b) manipulating the set of the extracted keywords from the publication (before interaction), (c) manipulating the set of the extracted keywords from the publication (after interaction).

Testing. The second evaluation round was conducted with five other users. Related to the global *What if* explanation, users were satisfied with the provided bar chart visualization as an explanation and they reported that it helped them understand the reason behind getting such recommendation. Moreover, they liked that they can get real-time changes when they want to test different scenarios by manipulating their interests. Further, they mentioned that using colors to distinguish between the recommended publications' status was helpful for them to immediately see the impact of the changes they performed on their interest model. However,

they suggested making the bars clickable in the visualization to allow them to see more details about a specific publication. That means users sometimes want to get a local explanation from the global explanation interface.

As far as the local *What if* explanation is concerned, users mentioned that besides the similarity score between each interest and the recommended publication, they want to see the similarity score between their whole interest model and the selected publication. Also, they reported that they liked the background gray bars as they represent the initial similarity score before they make changes, so they can easily compare the results. However, one of the main critiques for this visualization was the lack of the final decision if the current publication will still be recommended or not after changing their interest models. Overall, the evaluation of the second iteration prototypes revealed that the users had mostly similar positive opinions regarding the proposed bar chart visualizations and the expected interactions with the global and local *What if* explanations.
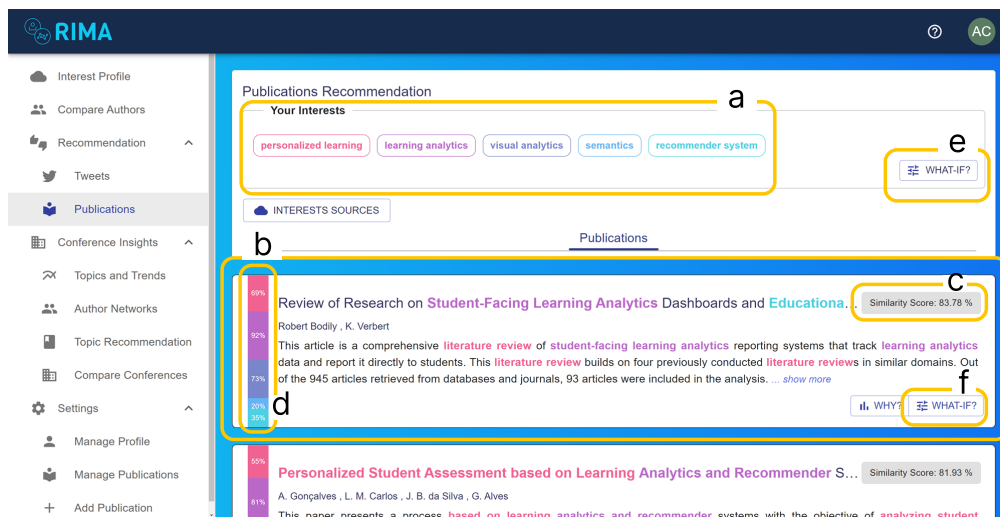
3.2.3. Third iteration

Figure 4: RIMA application main user interface.

Based on the previous iterations and following the feedback from users, we designed and implemented the final prototypes of the global and local *What if* explanations in the RIMA application (Figure 4). The main user interface consists of the list of the top five user's interests (Figure 4a), as automatically generated by the system. In order to easily identify the interests and their impact on the recommendation, we used unique color for each interest. This user interface displays the list of the recommended publications in form of separated boxes (Figure 4b), containing a relevance score for each publication (Figure 4c). We provide a color band next to each publication. The colors are the same ones used for the user interests on the top. The height of each color bar indicates how relevant is this publication to the related interest (Figure 4d). Users can access the global *What if* explanation through a "WHAT-IF?" button provided

on the upper box where they can see their interests (Figure 4e). The local *What if* explanation is provided through another "WHAT-IF?" button on the bottom-right side of the box for each recommended publication (Figure 4f).

For the global *What if* explanation (Figure 5), we provided a bar chart visualization that shows the similarity scores between the recommended publications and the interest model (Figure 5a). A slider on top of the bar chart visualization allows users to control the similarity threshold used by the RS. Only publications above the user-specified threshold value will be displayed. We implemented the color feature regarding the indication of the publication's status (old: blue, new: green, or omitted: red) (Figure 5b). As requested by users in the previous iteration, we made the bars clickable so they can see a more detailed explanation for each publication. This visualization shows similarity scores of the selected publication to each interest after a user clicks on a specific publication in the initial bar chart (Figure 5c).

(a)

(b)

(c)
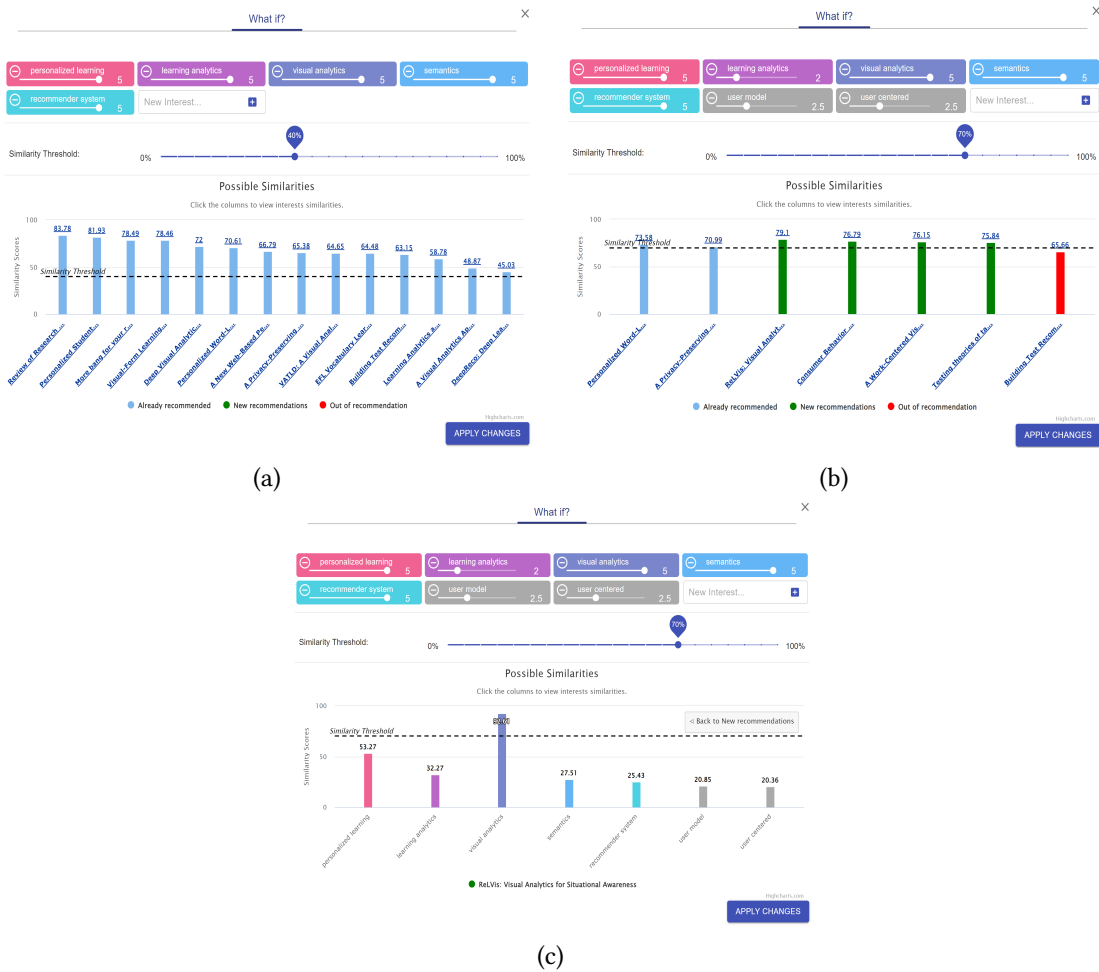
Figure 5: *What if* global explanation: (a) similarities between publications and interest model (before interaction), (b) similarities between publications and interest model (after interaction), (c) similarities between a selected publication and each interest.

Regarding the local *What if* explanation, we provided two visualizations for the user. The first visualization allows users to understand the reason behind getting a specific recommendation by manipulating their interest models. As requested by users in the previous iteration, we provided another bar chart next to the initial visualization to show the similarity score between the whole interest model and that specific publication, which is changed in real-time based on the user's actions (Figure 6). The second visualization enables users to change the list of keywords that were automatically generated by the RS, and see how this affects the computed similarities between the user-selected publication's keywords and the user interest model, and whether in that case the publication will be recommended or not (Figure 7).
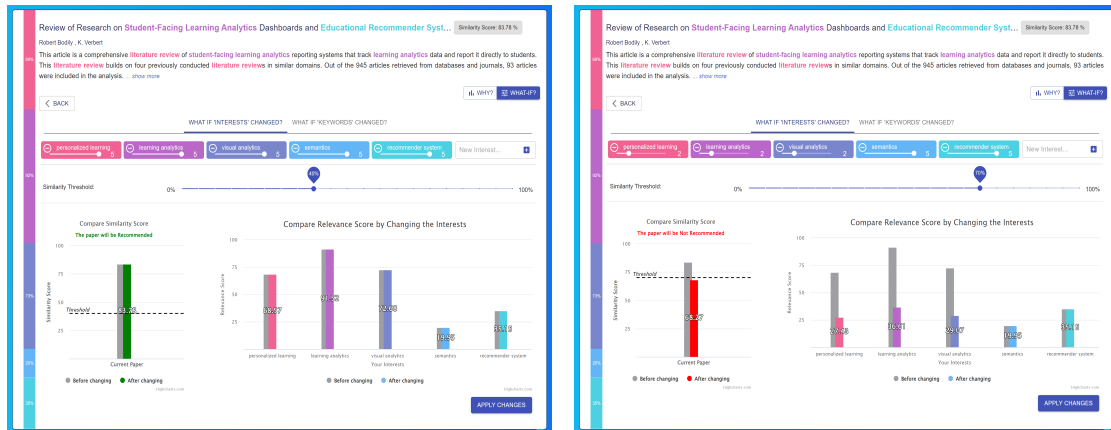
(a)   (b)

Figure 6: *What if* local explanation – similarities between a publication and each interest: (a) before interaction (b) after interaction.

(a)   (b)

Figure 7: *What if* local explanation – similarities between publication's keywords and interest model: (a) before interaction (b) after interaction.

4. Evaluation

4.1. Study Design

After systematically designing the *What if* explanations and implementing them in the RIMA application, we conducted a quantitative and qualitative user study to explore the usage and attitudes towards our scientific literature RS, considering the *What if* explanations. Researchers and students interested in scientific literature were invited to participate. 12 participants took part in this study. Participants were between 20 and 39 years old, where half of them were master's graduates or higher, and the other half were master's students. All participants gave informed consent to study participation.

Participants were initially given a short introductory video about the RIMA application in general, and another short demo video about the *What if* explanation feature in the application. Next, they answered a questionnaire in SoSci Survey which included questions about their demographics and familiarity with RS and visualization. Afterwards, we conducted moderated think-aloud sessions where participants were asked to (1) create an account using their Semantic Scholar ID (users who do not have Semantic Scholar IDs can generate their interest models manually) in order to create their interest models, (2) interact with the application based on given scenarios, and (3) take a closer look at the *What if* explanations provided by the system. Following a think-aloud approach, the participants were also asked to say anything that comes to their mind during each interaction. After that, we conducted semi-structured interviews to gather in-depth feedback. The interviews took place online and were recorded with the consent of the participants. They lasted 10 to 15 minutes with the following questions: *(1) What do you like the most about the provided What if explanations? (2) What do you like the least about the provided What if explanations? (3) Why / When (in which situation) / How often would you like to use each of the provided explanations? (4) How much has the controllability of the What if explanations influenced your satisfaction with the recommended publications? Do you have any suggestions to improve the controllability of the What if explanations? (5) Which What if explanation gives you a better sense of transparency of the recommender system? Why? (6) Which What if explanation gives you a better sense of trust in the recommender system? Why? (7) Do you have any suggestions to improve the system?*

After the semi-structured interviews, participants were also invited to fill out a questionnaire containing questions regarding usability aspects and attitudes towards the RS, based on the ResQue evaluation framework [43].

4.2. Analysis and Results

The results of the ResQue questionnaires are summarized in Figure 8. Besides our quantitative analysis, we also conducted a qualitative analysis of the moderated think-aloud sessions and the semi-structured interviews to gain further insights into the reasons behind the individual differences in the perception of the RS in terms of the *What if* explanations. We followed the instruction proposed by Braun and Clarke [44] to code the data and identify patterns to organize the codes into meaningful groups. Notes and transcripts of the interview recordings were made for the analysis. The analysis was rather deductive as we aimed to find additional explanations for the findings of our quantitative analysis. We present the results of the evaluation organized

by four themes, which were adapted from the ResQue framework, namely *Transparency and Trust*, *User Control*, *Satisfaction*, and *Overall User Experience*.

Figure 8: Results from the ResQue questionnaire.

4.2.1. Transparency and trust

This theme concerns the perception of the *What if* explanations and visualizations in terms of transparency and trust. In this regard, participants stated that the visualizations had a good effect on transparency in the system and they pointed to the possibility of making changes and observing the effect on the results. For instance, **P6** mentioned that *"Through what if explanations, I could see how the results will change"*. However, some of them could not identify some aspects of the system behavior clearly. When we asked the participants in the think-aloud session to explain how the system works and name the factors influencing the recommendation process, all of them could successfully mention that interests and weights play the main role in the recommendation process. They also stated that the similarity between their interests and the publication is the criteria to select which publication to recommend. However, three participants pointed out that the role of keywords was not clear to them.

Although six participants did not speak very confidently about trust in the system, they described that the consideration of interests in selecting publications to recommend, showing the similarity score between interests and publications, and how they are displayed with colors and numbers had caused more trust in the RS. However, on the other hand, they mentioned that there are uncertainties in extracting keywords from the publications and calculating the similarities in the *What if* explanations, which justified negative opinions about the system

trustworthiness: *"P8: It's difficult to say, some words are not extracted but they are close to one of my interests and this can make me disappointed"*. *"P2: It comes to my mind that there are some relevant publications that maybe the system is not recommending and I might have lost them"*.

4.2.2. User control

Most of the participants indicated a favorable opinion towards the control of the recommendations (see Figure 8): *"P1: When I change my interests or the keywords in a way that I wanted it has a good influence on the item which will be recommended and I like that it's exactly based on my interest"*. Their answers to the questions in the semi-structured interview about the controllability of the recommendations also indicate their satisfaction with the system. Also, we found that the provided controllability over the system and the explanations is sufficient and users were satisfied with it as they declared that it was enough and they didn't suggest any improvements: *"P3: I think the controllability is enough because there is nothing out of my control."*

4.2.3. Satisfaction

Most of the participants were satisfied with the provided *What if* explanations in the RS (see Figure 8). Participants indicated that they tend to use the *What if* explanations when the results were not favorable to them and that the global *What if* explanation to adjust their interests in order to get more accurate recommendations is the most appropriate explanation in this case. Regarding the local *What if* explanation, the participants mentioned that they tend to use it when they feel dissatisfaction with the recommendation in a particular case and that either the similarity calculations or keywords extracted from the publication might not be incorrect, so they have the ability to correct the system and improve the recommendation quality: *"P1: When I find unexpected and irrelevant items I would like to use the local What if explanations"*. Moreover, they mentioned that the local *What if* explanation helps them discover which one of their interests is most similar to the publication. However, most of the participants stated that they would use the global *What if* explanation more often than the local one.

4.2.4. Overall user experience

For this theme, we gathered feedback concerning the situations where each explanation is used. Controlling interests, real-time visual explanations of the made changes, and changing the publication keywords were the features that participants liked most in the *What if* explanations. Overall, the users liked that whenever the recommended items are not aligned with their expectations, they can easily interact with the system through the *What if* explanation interface to report errors and understand why a publication is recommended. All the answers indicated that changing the input and observing its effect on the results, especially in the global *What if* explanation, was a favorite feature for all participants: *"P10: I liked that I could change the keywords and interests when the recommended items were not desired"*. On the other hand, two participants did not like the similarity threshold and stated that it was unnecessary for them. Also, three participants found that the visualization is complicated and reported that *"P3: It may take some time to figure out what information the charts is providing"*.

When participants were asked about how to improve the system, most of the answers were about improving the application user interface. For example, **P3** mentioned that she didn't like the amount of color used in the application. According to the participants' opinions about the user interface, it can be concluded that the interface should be improved from a usability perspective.

While interacting with the system, we observed and examined user actions. The result was that a few users were challenged to find the *What if* keywords explanation, and in other cases, they could handle the defined situations well. In addition, they quickly realized the functionality of control panels, visual explanations, and how to interact with visualizations. Participants perceived the system as easy to use and showed an overall positive attitude towards the user experience. We could observe that they were able to use the explanations in the right situation and with a true reason. In addition, participants showed a high perception of the system's usefulness as they pointed out several times that they were able to find relevant publications by controlling their interests and weights. They also said that by limiting the domain of search, which is applicable by changing the interests and weights, one can get close to the right recommendation.

5. Conclusion and Future Work

Explanations are important to help users build an accurate mental model of how a recommender system (RS) works. Despite the wide agreement on the benefits of interaction and user control in RS, explanations in RS have so far been presented mostly in a static and non-interactive manner. To address this research gap, in this paper, we focused and elaborated on the concept of *What if* explanation as an effective mechanism to achieve interactive explanation in a scientific literature RS. Our qualitative study showed that providing *What if* explanations has a positive effect on different aspects of a RS such as transparency, trust, user control, satisfaction, and user experience. This paper contributes to a richer understanding of why and how to design *What if* interactive explanation in RS. Quantitative studies can build forth on our work to investigate the effect of *What if* explanation on the perception of and interaction with explainable recommendation, with different user groups and in different contexts.

References

[1] M. A. Chatti, M. Guesmi, L. Vorgerd, T. Ngo, S. Joarder, Q. U. Ain, A. Muslim, Is more always better? the effects of personal characteristics and level of detail on the perception of explanations in a recommender system, in: Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, 2022, pp. 254–264.

[2] I. Nunes, D. Jannach, A systematic review and taxonomy of explanations in decision support and recommender systems, User Modeling and User-Adapted Interaction 27 (2017) 393–444.

[3] Y. Zhang, X. Chen, et al., Explainable recommendation: A survey and new perspectives, Foundations and Trends® in Information Retrieval 14 (2020) 1–101.

[4] J. Vig, S. Sen, J. Riedl, Tagsplanations: explaining recommendations using tags, in: Proceedings of the 14th international conference on Intelligent user interfaces, 2009, pp. 47–56.

[5] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, Artificial Intelligence 267 (2019) 1–38.

[6] D. J. Hilton, Conversational processes and causal explanation., Psychological Bulletin 107 (1990) 65.

[7] C. He, 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] M. Jugovac, D. Jannach, Interacting with recommenders—overview and research directions, ACM Transactions on Interactive Intelligent Systems (TiiS) 7 (2017) 1–46.

[9] 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, 2012, pp. 43–50.

[10] 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, 2013, pp. 351–362.

[11] C.-H. Tsai, P. Brusilovsky, The effects of controllability and explainability in a social recommender system, User Modeling and User-Adapted Interaction 31 (2021) 591–627.

[12] D. Jannach, M. Jugovac, I. Nunes, Explanations and user control in recommender systems, in: Proceedings of the 23rd International Workshop on Personalization and Recommendation on the Web and Beyond, 2019, pp. 31–31.

[13] D. C. Hernandez-Bocanegra, J. Ziegler, Effects of interactivity and presentation on review-based explanations for recommendations, in: IFIP Conference on Human-Computer Interaction, Springer, 2021, pp. 597–618.

[14] J. Schaffer, T. Hollerer, J. O'Donovan, Hypothetical recommendation: A study of interactive profile manipulation behavior for recommender systems, in: The Twenty-Eighth International Flairs Conference, 2015.

[15] M. Guesmi, M. A. Chatti, L. Vorgerd, S. Joarder, S. Zumor, Y. Sun, F. Ji, A. Muslim, On-demand personalized explanation for transparent recommendation, in: Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, 2021, pp. 246–252.

[16] A. Abdul, J. Vermeulen, D. Wang, B. Y. Lim, M. Kankanhalli, Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda, in: Proceedings of the 2018 CHI conference on human factors in computing systems, 2018, pp. 1–18.

[17] D. Wang, Q. Yang, A. Abdul, B. Y. Lim, Designing theory-driven user-centric explainable ai, in: Proceedings of the 2019 CHI conference on human factors in computing systems, 2019, pp. 1–15.

[18] Q. V. Liao, D. Gruen, S. Miller, Questioning the ai: informing design practices for explainable ai user experiences, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, pp. 1–15.

[19] H.-F. Cheng, R. Wang, Z. Zhang, F. O'Connell, T. Gray, F. M. Harper, H. Zhu, Explaining decision-making algorithms through ui: Strategies to help non-expert stakeholders, in:

Proceedings of the 2019 chi conference on human factors in computing systems, 2019, pp. 1–12.

[20] K. Sokol, P. Flach, One explanation does not fit all: The promise of interactive explanations for machine learning transparency, KI-Künstliche Intelligenz 34 (2020) 235–250.

[21] J. Krause, A. Perer, K. Ng, Interacting with predictions: Visual inspection of black-box machine learning models, in: Proceedings of the 2016 CHI conference on human factors in computing systems, 2016, pp. 5686–5697.

[22] T. Kulesza, M. Burnett, W.-K. Wong, S. Stumpf, Principles of explanatory debugging to personalize interactive machine learning, in: Proceedings of the 20th international conference on intelligent user interfaces, 2015, pp. 126–137.

[23] Y. Jin, N. Tintarev, K. Verbert, Effects of personal characteristics on music recommender systems with different levels of controllability, in: Proceedings of the 12th ACM Conference on Recommender Systems, 2018, pp. 13–21.

[24] J. Harambam, D. Bountouridis, M. Makhortykh, J. Van Hoboken, Designing for the better by taking users into account: A qualitative evaluation of user control mechanisms in (news) recommender systems, in: Proceedings of the 13th ACM Conference on Recommender Systems, 2019, pp. 69–77.

[25] D. Jannach, A. Manzoor, W. Cai, L. Chen, A survey on conversational recommender systems, ACM Computing Surveys (CSUR) 54 (2021) 1–36.

[26] M. Guesmi, M. A. Chatti, Y. Sun, S. Zumor, F. Ji, A. Muslim, L. Vorgerd, S. A. Joarder, Open, scrutable and explainable interest models for transparent recommendation., in: IUI Workshops, 2021.

[27] M. Guesmi, M. A. Chatti, A. Tayyar, Q. U. Ain, S. Joarder, Interactive visualizations of transparent user models for self-actualization: A human-centered design approach, Multimodal Technologies and Interaction 6 (2022) 42.

[28] M. Guesmi, M. A. Chatti, L. Vorgerd, S. A. Joarder, Q. U. Ain, T. Ngo, S. Zumor, Y. Sun, F. Ji, A. Muslim, Input or output: Effects of explanation focus on the perception of explainable recommendation with varying level of details., in: IntRS@ RecSys, 2021, pp. 55–72.

[29] M. Guesmi, M. A. Chatti, L. Vorgerd, T. Ngo, S. Joarder, Q. U. Ain, A. Muslim, Explaining user models with different levels of detail for transparent recommendation: A user study, in: Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, 2022, pp. 175–183.

[30] K. Balog, F. Radlinski, S. Arakelyan, Transparent, scrutable and explainable user models for personalized recommendation, in: Proceedings of the 42nd international acm sigir conference on research and development in information retrieval, 2019, pp. 265–274.

[31] B. Y. Lim, A. K. Dey, Assessing demand for intelligibility in context-aware applications, in: Proceedings of the 11th international conference on Ubiquitous computing, 2009, pp. 195–204.

[32] S. Mohseni, N. Zarei, E. D. Ragan, A multidisciplinary survey and framework for design and evaluation of explainable ai systems, ACM Transactions on Interactive Intelligent Systems (TiiS) 11 (2021) 1–45.

[33] B. Y. Lim, A. K. Dey, Toolkit to support intelligibility in context-aware applications, in: Proceedings of the 12th ACM international conference on Ubiquitous computing, 2010, pp. 13–22.

[34] B. Y. Lim, Q. Yang, A. M. Abdul, D. Wang, Why these explanations? selecting intelligibility types for explanation goals., in: IUI Workshops, 2019.

[35] Q. U. Ain, M. A. Chatti, M. Guesmi, S. Joarder, A multi-dimensional conceptualization framework for personalized explanations in recommender systems, in: Proceedings of the Joint 27th International Conference on Intelligent User Interfaces, Helsinki, Finland, 2022, pp. 22–25.

[36] M. Zürn, M. Eiband, D. Buschek, What if? interaction with recommendations., in: ExSS-ATEC@ IUI, 2020.

[37] M. A. Chatti, F. Ji, M. Guesmi, A. Muslim, R. K. Singh, S. A. Joarder, Simt: A semantic interest modeling toolkit, in: Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, 2021, pp. 75–78.

[38] D. Afchar, A. B. Melchiorre, M. Schedl, R. Hennequin, E. V. Epure, M. Moussallam, Explainability in music recommender systems, arXiv preprint arXiv:2201.10528 (2022).

[39] B. Y. Lim, Improving understanding and trust with intelligibility in context-aware applications, Ph.D. thesis, Carnegie Mellon University, 2012.

[40] M. Szymanski, M. Millecamp, K. Verbert, Visual, textual or hybrid: the effect of user expertise on different explanations, in: 26th International Conference on Intelligent User Interfaces, 2021, pp. 109–119.

[41] D. Norman, The design of everyday things: Revised and expanded edition, Basic books, 2013.

[42] Why you only need to test with 5 users, https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/, ???? Accessed: 2022-05-20.

[43] P. Pu, L. Chen, R. Hu, Evaluating recommender systems from the user's perspective: survey of the state of the art, User Modeling and User-Adapted Interaction 22 (2012) 317–355.

[44] V. Braun, V. Clarke, Using thematic analysis in psychology, Qualitative research in psychology 3 (2006) 77–101.