User Needs for Explanations of Recommendations: In-depth Analyses of the Role of Item Domain and Personal Characteristics

Thi Ngoc Trang Tran*
ttrang@ist.tugraz.at
Institute of Software Technology, Graz University of Technology
Graz, Styria, Austria

Alexander Felfernig
alexander.felfernig@ist.tugraz.at
Institute of Software Technology, Graz University of Technology
Graz, Styria, Austria

Viet-Man Le
vietman.le@ist.tugraz.at
Institute of Software Technology, Graz University of Technology
Graz, Styria, Austria

Thi Minh Ngoc Chau
ctmngoc@hueuni.edu.vn
School of Hospitality and Tourism, Hue University
Hue, Thua Thien Hue, Vietnam

Thu Giang Mai
mtgiang@hce.edu.vn
Hue University of Economics, Hue University
Hue, Thua Thien Hue, Vietnam

ABSTRACT

Explanations can be provided with different goals, such as clarifying how the system works, how well the recommended item meets the user's preferences, and how an explanation helps the user select an item faster. Although extensive research has been conducted in this research line, not much attention is paid to investigating user needs for explanations. To the best of our knowledge, no studies provide related insights, especially from the perspectives of item domain and personal characteristics. Up to now, it is not completely clear if user needs for explanations change across different item domains and vary according to user characteristics. To analyze these aspects, we developed three web-based prototype recommender systems for low-, average-, and high-involvement item domains and conducted a user study with 553 participants from different countries. Related results show that, in high-involvement item domains, users tend to have a look at explanations when they are not satisfied with the recommended items. An opposite tendency was found in low- and average-involvement item domains. Statistically, there is insufficient evidence to suggest correlations between users' needs for explanations and item domains or between users' needs and personal characteristics. However, the descriptive statistics show that users' need for explanations varies across different item domains. In this study, we also found the best explanation approaches to be used in a specific recommendation domain.

CCS CONCEPTS

• Information systems → Information systems applications; **Decision support systems**; **Recommender systems**; **• Human-centered computing → User models**; **User studies**.

*The main and corresponding author

UMAP '23, June 26–29, 2023, Limassol, Cyprus

ACM ISBN 978-1-4503-9932-6/23/06.
https://doi.org/10.1145/3565472.3592950

KEYWORDS

Recommender Systems, Explanations, Personal Characteristics, Item Domain, Item-based Explanations, Feature-based Explanations, Knowledge-based Explanations

ACM Reference Format:
Thi Ngoc Trang Tran, Alexander Felfernig, Viet-Man Le, Thi Minh Ngoc Chau, and Thu Giang Mai. 2023. User Needs for Explanations of Recommendations: In-depth Analyses of the Role of Item Domain and Personal Characteristics. In *UMAP '23: Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization (UMAP '23), June 26–29, 2023, Limassol, Cyprus.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3565472.3592950

1 INTRODUCTION

Recommender systems have become an indispensable tool in daily life activities and used in a variety of areas, such as entertainment activities [10, 16, 26], e-commerce [19, 28], tourism [7, 12], and healthcare [52, 54]. These systems assist users in making decisions more efficiently and choosing items that best match their preferences and needs. While recommender systems can support good choices, they are sometimes considered *"black boxes"* when no explanations of the recommended items are provided. To solve this issue, explanations of recommendations are generated to explain the underlying recommendation mechanism. One example explanation in the movie domain could be the following: *"The movie Avatar has been recommended to you since you watched similar movies before"*. Explaining recommendations contributes to the success of a recommender system in various ways, especially by helping users make better and more informed decisions [27, 33, 39, 42, 51, 62, 63]. Although extensive research has been performed, many aspects have not been sufficiently studied. In the following, we discuss open issues considered in this paper.

Users' needs for explanations vs. item domains. Earlier studies often discuss the role of explanations in recommender systems [21, 49]. It is, however, still unclear *if users always appreciate explanations.* We argue that there could exist situations where a user does not want to invest additional efforts in checking an explanation of the recommended item. Besides, we assume that users' needs for explanations differ depending on the item domain. Item domains can be categorized according to the *involvement* that reflects the importance and interest of users in the related decision. This factor can trigger changes in the amount of information needed to make a decision [8]. *A low-involvement decision* (e.g., selecting a song to listen to) is typically inexpensive (compared to income), purchased regularly, and poses a low risk to users if they take a suboptimal decision. *A high-involvement decision* (e.g., purchasing a car or an apartment) shows the opposite side of the mentioned aspects where users may spend time comparing further aspects such as item features, prices, and warranties [8]. With the mentioned assumptions, we wanted to clarify two aspects: *"When do users need explanations?"* and *"Are there correlations between users' needs for explanations and item domains?"*.

Users' needs for explanations vs. personal characteristics. Personal characteristics such as *"age", "gender", "cultural background", "personality",* and *"expertise"* have been proven to affect user control and user interaction with recommender systems [24]. In recommender systems, Tang et al. [45] show that different cultural backgrounds also trigger differences in users' preferences for recommended items. Regarding explaining recommendations, we assume that personal characteristics could also affect users' needs for explanations. For instance, there are differences concerning curiosity between male and female users [4], which can lead to different needs for explanations. A relevant open question in this context is: *"Are there correlations between users' needs for explanations and personal characteristics?"*.

Favorite explanations vs. item domains. There exist various types of explanations, most of them are generated based on the selected recommendation algorithm[1], such as *user-based or item-based explanations* [17, 41, 62], *feature-based* or *content-based explanations* [11, 44, 65, 66], and *knowledge-based explanations* [13]. However, to some extent, it is still not completely clear what are the best explanations from the user point of view. We assume users' preferences for explanations vary depending on the item domain. For instance, users like item-based explanations of the recommended items in the restaurant domain, but prefer knowledge-based explanations in the accommodation domain. This brings us to another research question: *"What is the favorite explanation from the user point of view in a specific item domain?"*.

To investigate the mentioned aspects and answer the research questions, we have developed three prototype recommender systems for three item domains - *restaurant, tourism*, and *accommodation* corresponding to *low-, average-*, and *high-involvement* item domains respectively. We chose these domains since they refer to familiar items on daily-life decisions. The restaurant recommender system helps a user to select a restaurant for having lunch/dinner, which is related to a low-involvement decision. The tourism recommender system assists a tourist in deciding on tours for the upcoming holidays, which requires more decision effort than this in the restaurant domain [53]. Finally, the accommodation recommender system supports a decision-making scenario where a tenant finds a shared room/an apartment for the next couple of years. Making such a decision is more effortful than deciding on a tour and even more on a restaurant [53]. We are aware that the tourism item domain is not always low-stake for all users, especially for those who are anxious or scared of traveling. In this study, we assume a traveling scenario where a user wants to visit a specific destination for a short time (two or three days) and has reserved only a small budget for the trip, which links to a lower-involvement decision compared to the mentioned accommodation decision.

On the basis of our prototype recommender systems, we conducted a user study with a large number of participants. Elaborated analyses from the perspectives of item domain and personal characteristics were then performed to examine our assumptions. This study is in line with the work done by Tran et al. [56], but brings further contributions. Different from the existing studies that look for a binary answer (*"yes"/"no"*) to the question *"Do users need an explanation?"*, our work moves one step further to point out *scenarios* where explanations are appreciated by users as well as *relationships* with item domains and user characteristics. Moreover, we find the best explanation types (from the user point of view) to be included in a recommender system for a specific item domain.

The remainder of the paper is organized as follows. In *Section 2*, we present related work on explanations in recommender systems. In *Section 3*, we introduce the explanation types used in our study, display styles to visualize, and metrics to evaluate them. Next, we define research questions and present the crucial steps of our user study in *Section 4.* The data analysis results and related discussions are presented in *Section 5.* Finally, we conclude the paper, present limitations, and discuss open topics for future work (*Section 6*).

2 RELATED WORK

2.1 Users' Needs for Explanations vs. Item Domains

Plenty of studies have been done to point out the crucial role of explanations in recommender systems and present various approaches to explanation generation/categorization. Tintarev and Masthoff [47, 49, 50] show that explanations facilitate users' decision-making processes, increase conversion rates, and lead to more satisfaction and trust in the recommender system. McSherry [29] briefly reviews explanations and discusses an approach to generate explanations based on case-based reasoning recommendation techniques. Nunes and Jannach [33] provide a taxonomy considering various facets such as explanation objective, responsiveness, content, and presentation. Zhang et al. [62] propose a chronological research timeline of explainable recommendations and a taxonomy to classify explanation approaches. Although extensive research on explaining recommendations has been done, only two studies currently exist analyzing users' needs for explanations and scenarios where explanations are appreciated. Hoeve et al. [46] examine if users want explanations for news recommendations. They find out that users want to see explanations but do not strongly prefer explanation visualization. Tran et al. [56] analyze users' needs for explaining recommendations and show that users look at explanations when they are less satisfied with the recommended items. However, this work is only limited to the movie domain, requiring further investigation into other item domains.

On the other hand, *domain-sensitive recommendation* has become an emerging research topic in recommender systems, which investigates correlations between users' preferences and item domains [9]. Most of these studies focus on cross-domain recommendations

[1]There could be the case that the explanation design does not strongly tie to any specific recommendation algorithm [32, 36].

that address cold-start problems [1, 5], mitigate the sparsity problem [25, 35], and identify the relationships between items in two different domains [9]. However, to the best of our knowledge, no studies analyze relations between item domains and users' needs for explaining recommendations.

2.2 Users' Needs for Explanations vs. Personal Characteristics

Along with the development of *humanized recommender systems* [55], researchers have paid more attention to analyzing the impacts of human factors on explaining recommendations [2, 20, 30]. Existing studies show that each user looks for explanations with different aims, expectations, backgrounds, and needs [20]. They also expect to receive different explanations depending on their characteristics [2, 15, 30, 59, 62]. Alslaity and Tran [2] and Mille et al. [30] prove that personal characteristics affect user perception and user interaction with the recommender system when recommendations are explained. Other related studies on music recommender systems [22, 23, 31] also confirm that personal characteristics such as *"musical sophistication"* and *"visual memory capacity"* influence user acceptance concerning recommended items and user trust in recommender systems. Following this research line, we conduct our study with a different focus in this work. Instead of digging more deeply into the correlations between personal characteristics and explanation generation, we explore relationships between these factors and users' needs for explanations. *We examine if users with different personal characteristics have different explanation needs.*

2.3 Favorite Explanation Types vs. Item Domains

Explanations can be categorized into two types, model-intrinsic and post-hoc explanations [27, 62]. Another approach is to classify explanations based on the adopted recommendation algorithm. Some examples thereof are *user-based or item-based explanations (IBExp)* [17, 41, 62], *feature-based explanations (FBExp)* [11, 44, 65, 66], and *knowledge-based explanations (KBExp)* [13, 14]. While various types of explanations exist, it is still unclear what are the best explanations from the user point of view. Besides, users' preferences for explanations are assumed to change across different item domains, which has not been examined up to now. In this work, we analyze users' preferences for different explanation types in various item domains to bridge the mentioned gaps.

3 EXPLANATION TYPES, DISPLAY STYLES, AND EVALUATION METRICS

In our study, three post-hoc explanations[2] (*IBExp, FBExp,* and *KBExp*) were chosen to analyze users' needs for recommendation explanations. These explanation types were selected based on their popularity and effectiveness in various recommender systems. *IBExp* is created based on the *item-based collaborative filtering approach* [17, 41, 66]. *FBExp* is created based on the *feature-based recommendation* approach that reveals how relevant a recommendation is to the user in terms of item features [64]. *KBExp* is generated based on the *knowledge-based recommendation approach* [13, 14] that provides suggestions, especially for complex items (e.g., tours and apartments in our study). All selected explanations are adapted versions of those presented in [56], with changes (in terms of content and display styles) made to suit the selected item domains (see further details in the follow-up subsections).

Explanation goals can be used as *metrics* to evaluate the generated explanations, which are listed as follows: *transparency* (explaining how the system works), *scrutability* (telling the system it is wrong), *trust* (inspiring the trust and loyalty of users), *satisfaction* (increasing the system utility and users' joy concerning recommended items), *effectiveness* (assisting users on making good decisions), *efficiency* (accelerating users' decision-making processes), and *persuasiveness* (convincing users to consume the recommended items) [48]. However, the current literature has proven that each explanation type links to only some specific goals, but *"not all"* [62]. In the following subsections, we will present explanation goals (used as *evaluation metrics*) related to each explanation type.

3.1 Item-Based Explanations (*IBExp*)

Explanation creation. Item-based explanations can be created based on *item-based collaborative filtering* [41] (see also *Section 4.2.3*), recommending an item that is similar to the ones the target user has rated before and has the highest predicted rating. An example item-based explanation in the tourism domain is the following: ***"The tour Seaside Park** has been recommended to you since you wanted **to visit Japan and take Nature, Parks, and Zoos tours**. Besides, the recommended **tour** is similar to the **tours** you have selected earlier and has the highest predicted rating"*. A similar formulation is applied to this explanation type in other item domains, where the bold texts are adapted accordingly. The same rule is applied to other explanation types presented in the next subsections.

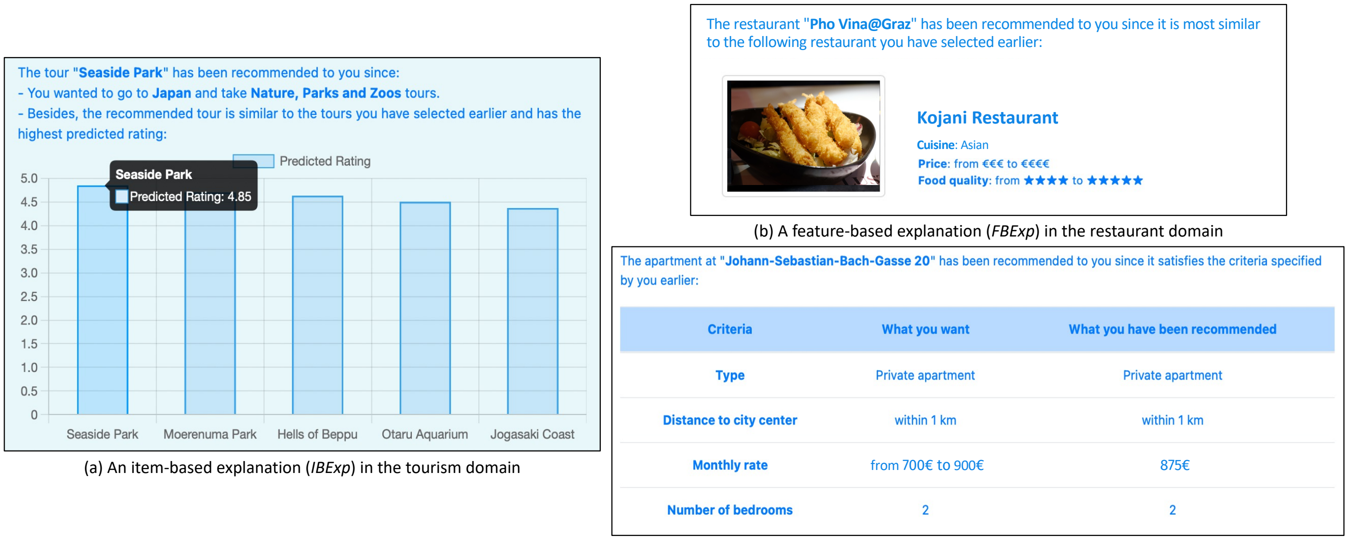
Explanation representation. Herlocker et al. [17] show that an item-based explanation can be effectively represented using a *bar chart* explaining how similar items have been rated. We also use a bar chart to represent the predicted rating of the candidate items. An example item-based explanation in the tourism domain is visualized as in *Figure 1(a)*.

Explanation goals. According to the study of Wang et al. [61], *IBExp* does not connect to all of the mentioned explanation goals, but only the *efficiency* goal. By showing how much the recommended item is better than other items of the same item type, an item-based explanation helps the target user make the decision faster.

3.2 Feature-Based Explanations (*FBExp*)

Explanation creation. Feature-based explanations can be created by *feature-based recommendation* (see also *Section 4.2.3*) that provides suggestions based on the information of item features. The item features are domain-dependent, i.e., features characterizing items vary depending on the item domain [62]. For instance, *"cuisine", "price",* and *"food quality"* are the features describing a restaurant. An example of this explanation type in the restaurant domain is the following: ***"The restaurant Pho Vina@Graz** has been recommended to you since it is most similar to the following **restaurant** you have selected earlier: **Kojani Restaurant, cuisine: Asian, price: €€€- €€€€; food quality: 4 - 5 (stars)**"*.

[2] A *post-hoc explanation* is based on an explanation model and generated after a decision has been made [37].



(a) An item-based explanation (*IBExp*) in the tourism domain

(b) A feature-based explanation (*FBExp*) in the restaurant domain

(c) A knowledge-based explanation (*KBExp*) in the accommodation domain

Figure 1: The visualization of all explanation types used in our study.

Explanation representation. Previous studies used *tags* [60] and *radar charts* [18] to visualize feature-based explanations. However, these display styles show some limitations, especially when they are used to represent items with a large number of features. To provide users with user-friendly and intuitive explanations, we use *text* and *images* to represent the explanations. Zhang et al. [64] show that textual feature-based explanation increases user interaction with the recommender system. Besides, we believe that a textual explanation with an image helps the target user easily find out to what extent the recommended item is similar to the ones he/she liked earlier. The visualization of this explanation type in the restaurant domain is depicted in *Figure 1(b)*.

Explanation goals. A feature-based explanation clarifies how the recommender system works, which links to the *transparency* goal [11]. Besides, Sinha et al. [43] claim that this goal has a close relationship with *trust*, which shows that explaining the underlying recommendation mechanism helps to increase the trust of users in the recommender system.

3.3 Knowledge-Based Explanation (*KBExp*)

Explanation creation. A knowledge-based explanation is generated by *knowledge-based recommendation* [13, 14] (see our recommendation approach in *Section 4.2.3*), showing the item features specified by the user and his/her preferences for these features. The explanation tells the user how well the recommended item meets his/her interests or requirements. An example of this explanation type in the accommodation domain is the following: *"**The apartment at Johann-Sebastian-Bach-Gasse 20** has been recommended to you since it satisfies the criteria specified by you earlier: **private apartment, within 1 km to the city center, the monthly rate from €700 to €900, and two bedrooms**".*

Explanation representation. Existing studies do not discuss a specific way to represent knowledge-based explanations. In this work, we propose representing a knowledge-based explanation using a three-column table. The first column shows the item features that the user is interested in. The second column shows the user's preferences for the features of interest. The last column shows the recommended item information. The visualization of a knowledge-based explanation in the accommodation domain is depicted in *Figure 1(c)*, including the text mentioned above, the participant's criteria, and the recommended item's features shown in a table.

Explanation goals. A knowledge-based explanation shows how well the recommended item meets the user's requirements, which connects to the *persuasiveness* goal [38, 58]. This explanation, therefore, also increases the user's *satisfaction* with the recommended item [11].

4 RESEARCH QUESTIONS AND USER STUDY

In this section, we define the relevant research questions and present the main steps of our user study.

4.1 Research Questions

According to the open aspects that have been discussed in *Section 1*, we define the following research questions:

RQ1: When do users need explanations of the recommended items?

RQ2: Are there correlations between users' needs for explanations and the item domain?

RQ3: Are there correlations between users' needs for explanations and personal characteristics?

RQ4: What is the favorite explanation from the user point of view in a specific item domain?

Answering the research questions helps to find out *(1)* scenarios where users appreciate explanations, *(2)* which user groups would need explanations, and *(3)* the best explanations to be generated in a recommender system for a specific item domain. Knowledge about these aspects contributes to advancing the state of the art in explaining recommendations.

4.2 User Study

In the following subsections, we present essential steps for designing and conducting our user study as well as details regarding *participants' demographic information* and *privacy concerns*.

4.2.1 ***Dataset Preparation***. Two datasets are needed for each prototype recommender systems: an *item dataset* and a *user-rating dataset*. The item dataset is used for generating feature-based and knowledge-based recommendations, while the user-rating dataset is for item-based recommendations. Existing datasets, such as *Yelp*[3], *TripAdvisor*[4], and *Inside Airbnb*[5] were considered but deemed unsuitable due to their *unavailability, inappropriateness,* and *copyright-related complexity*. The *Yelp* and *Inside Airbnb* datasets had overall item ratings instead of ratings for item features, and the *Inside Airbnb* dataset was for short stays instead of long stays. The *TripAdvisor* dataset was suitable but not free, potentially causing copyright issues. *Synthetic datasets* were therefore created for our study, which are described in the following.

Item datasets: The details of the item datasets are shown in *Table 1*. Each item dataset consists of the general information of items (see *column 2*), the features used in the recommendation process (see *column 3*), the domain values of the features (see *column 4*), and the cardinality/magnitude of a feature domain (see *column 5*). To ensure that the recommender system can always find at least one recommended item based on the user's requirements (related to the features), the number of entries for an item dataset is the multiplication of the cardinalities of the feature domains. For instance, the number of entries of the restaurant dataset is $4 \times 5 \times 5 = 100$. Besides, we created *ten* additional items that are shown to the participants to learn their preferences of items.

The domain values of a feature are re-used from the existing systems. For instance, the values of the features in the restaurant and tour datasets are defined based on similar features used in *TripAdvisor*[6] and the related studies [3, 57]. Besides, for the restaurant domain, a lunch/dinner selection scenario was assumed, while the tourism domain was assumed a nearby-country tour planning scenario to four selected countries in Asia: *China, India, Japan,* and *Thailand*. We selected these four countries since they are the best destinations in Asia according to *Touropia*[7]. In the accommodation domain, we assumed a scenario where a user is moving to a city for studying/working and wants to stay in a shared room/private apartment in the next couple of years. The features' domain values are based on the related information on the existing housing webpages/datasets[8].

User-item datasets: Each dataset is represented by a *full matrix* where users' preferences for *"all"* items are specified. A user's preference for an item is the average of his/her preferences for all features of the item. A user-item dataset is generated as follows. For each feature f_i of item I, we randomly selected X% of the users who "like" item I concerning feature f_i. The remaining $Y\%(Y = 100 - X)$ indicates the number of users who "dislike" item I with regard to feature f_i. As the best practice, X is assigned to *randbetween(30, 70)*, returning random values that guarantee equivalent numbers of users who like and dislike an item across different features. Given the five-star rating scale [1..5], we assigned $4+\theta$ to a "like" value and $2+\theta$ to a "dislike" value, where $\theta \in [-1, 1]$ (in our study, we selected $theta = +/-0.3$). An example of these assignments is presented in the following (see also *Table 2*):

- For feature f_1, we ran $X\% = randbetween(30, 70) = 60\%$, randomly selected *three* out of *five* users and assigned "like values" $(4 + \theta)$ to them. The remaining Y = 40% of users are assigned with "dislike values" $2 + \theta$.

- Performing similar steps for feature f_2, we had $X\% = randbetween(30, 70) = 40\%$ (i.e., *two out of five users* are assigned to "like values" $(4 + \theta)$) and the remaining Y = 60% of users are assigned to "dislike values" $(2 + \theta)$.

4.2.2 ***Task Distribution***. We developed three web-based prototype recommender systems that simulate real systems in three item domains, each with three recommendation algorithms and corresponding explanations - *IBExp, FBExp,* and *KBExp* - yielding nine experimental settings. A between-subjects user study was conducted at Graz University of Technology (Austria), School of Hospitality and Tourism (Vietnam), and Hue University of Economics (Vietnam), with links sent via email to three participant groups (around 100 participants each), comprising students and staff members. Participants were not required to have any specific domain knowledge. Each participant received *exactly one* experimental setting. The study ensured equivalent numbers of participants across explanation types. Student participants earned bonus points for the course, while staff members were offered a chance to win an Amazon voucher.

4.2.3 ***Recommendation Generation***. To generate recommendations, we asked the participants for personal information such as *age, gender, nationality, profession,* and *item type*. *Item type* identifies the item type that a participant likes, which was collected differently depending on the item domain. The participants had to select (from a predefined list) *a cuisine* in the restaurant domain, *a tourism destination* and *a tour type* in the tourism domain, or *an apartment/a shared room* in the accommodation domain. To avoid potential decision biases, items in the predefined lists were randomly shown to the participants. Besides, to learn the participants' preferences, we asked them to perform an additional task according to the provided explanation type. The participants who received the explanations *IBExp* or *FBExp* had to select at least *five items* they liked from a given list of items related to the selected item type. The participants who received the explanation *KBExp* had to specify their requirements for the desired item (e.g., *food quality = 4* and *price = 3* using a *5-level scale* (1: *the worst/most expensive*; 5: *the best/cheapest*)).

As mentioned earlier, we developed prototypes in which simple recommendation algorithms are running to simulate real recommender systems. The basic idea of the recommendation algorithms is presented as follows:

Item-based recommendation: The system identifies *five* items that are most similar to those selected by a participant U. The *Cosine similarity* presented in [41] is used to calculate the similarity between two items I_i and I_j. Based on the calculated similarity, the system predicts the participant's rating for similar items using

[3]https://www.kaggle.com/datasets/yelp-dataset/yelp-dataset
[4]https://www.tripadvisor.com/
[5]http://insideairbnb.com
[6]https://www.tripadvisor.com
[7]https://www.touropia.com/best-countries-to-visit-in-asia/
[8]https://housinganywhere.com; https://www.kaggle.com/datasets/

	General information	Feature	Domain values	|Domain|
Restaurant	name, opening hours, image	cuisine	"Austrian and European"; "Asian"; "Italian, Mediterranean, and Spanish"; "African, American, and Caribbean"	4
		price	€; €€; €€€; €€€€; €€€€€	5
		food quality	[1..5] (values in the 5-star rating scale)	5
Tour	name, description, image	destination	"China"; "India"; "Japan"; "Thailand"	4
		tour type	"nature, parks, and zoos"; "cultural tours and sightseeing"; "food, drinks, and nightlife"; "sports and outdoor activities"	4
		duration	"less than 3 hours"; "from 3 to 5 hours"; "more than 5 hours"	3
		price	"less than €50"; "from €50 to €100"; "more than €100"	3
Shared room	name, size, facilities, amenities, preferred tenants, available from, service fee, deposit	distance to city center	"within 1 km"; "within 5km"; "within 10km"	3
		monthly rate	"less than €350"; "from €350 to €450"; "more than €450"	3
		number of housemates	[1..5]	5
Private apartment	name, size, facilities, amenities, preferred tenants, available from, service fee, deposit	distance to city center	"within 1 km"; "within 5km"; "within 10km"	3
		monthly rate	"less than €700"; "from €700 to €900"; "more than €900"	3
		number of bedrooms	[1..3]	3

Table 1: The information of the item datasets used in our prototype recommender systems.

user	rating(f_1)	rating(f_2)	overall rating
u_1	4.3	2.3	3.3
u_2	2.3	4.3	3.3
u_3	1.7	1.7	1.7
u_4	3.7	3.7	3.7
u_5	4.3	2.3	3.3

Table 2: An example of assigning rating values to a list of users. We assume that there are five users (u_1 ... u_5) and each item is described by two features (f_1 and f_2). The overall rating is the average of the ratings of features f_1 and f_2.

Formula 1 and selects an item with the highest predicted rating as the recommended item.

$$rating(U, I_i) = \frac{\sum_i (U, I_j) \times Sim(I_i, I_j)}{\sum_j Sim(I_i, I_j)} \quad (1)$$

Feature-based recommendation: A recommendation in the restaurant domain can be generated as follows[9]. The user has selected five *Asian* restaurants with *food quality* ∈ [3..5] and *price* ∈ [3..5]. The system filters out a list of restaurants with the specified features, on which one restaurant is randomly chosen as the recommended item.

Knowledge-based recommendation: A recommendation in the accommodation domain can be generated as follows[10]. The participant wanted to stay in an apartment *within 1 km to city center*, has *two bedrooms* and *the monthly rent from €700 to €900.* Based on these requirements, the system first finds all apartments fulfilling the specified criteria and then randomly selects one apartment as the recommended item.

4.2.4 *Recommendation Representation.*

The recommendation to a participant is represented in *Figure 2.* By default, the recommendation shows only the basic information of the recommended item, but a participant could see more details by clicking "*See further information*". To be aware of the participants' preferences for the recommended item, we asked them to rate the recommended item using a *5-point scale* (*1 - do not like at all, 5 - really like*).

4.2.5 *Explanation Generation.*

In each item domain, the *three* mentioned explanations were generated, formulated, and visualized using the templates and display styles presented in the *Section 3.* The participants could check the explanation of the recommendation by clicking on *"Why this <**item**>?".* The explanation is then shown on the same page, right below the recommendation rating section. Having a look at the explanation was *not a mandatory task*, i.e., the participants could decide to watch the explanation or not.

4.2.6 *Explanation Evaluation.*

Each explanation type links to specific explanation goals which are not always the same (see *Section 3*). This shows that the mentioned explanation types are evaluated based on different metrics. In order to achieve standard evaluations for all explanation types, we proposed three additional metrics - *understandability, design,* and *overall satisfaction* (i.e., these metrics were used to evaluate each explanation type). The statements of the additional metrics are tailored as follows: *understandability* - "The explanation is understandable", *display style* - "I like how the explanation is represented", and *overallSatisfaction* - "The explanation helps to increase my satisfaction with the recommender system". The participants evaluated each metric by providing feedback on the corresponding statement using a *5-point scale* (1 - totally disagree, 5 - totally agree).

[9] *A feature-based recommendation* in the remaining item domains is generated in a similar fashion.
[10] *A knowledge-based recommendation* in the remaining item domains is generated in a similar fashion.

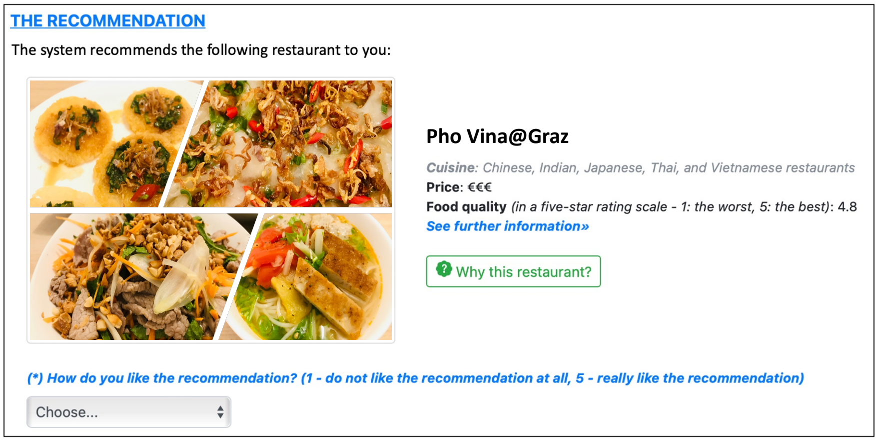


Figure 2: A recommendation in the restaurant domain. The participants can see further information by clicking on *"See further information"*. They can also see the explanation of the recommendation by clicking on the button *"Why this restaurant?"*.

If a participant did not have a look at the explanation, he/she was then asked to select one or multiple reasons from a list of pre-defined reasons: *(1)* I did not care about how/why the recommendation has been generated; *(2)* I was not satisfied with the recommendation; *(3)* I was not interested in the explanation; *(4)* I could guess how the recommendation has been generated; *(5)* I was satisfied with the recommendation; and *(6)* I did not have time to have a look at the explanation. To avoid potential biases, the mentioned reasons were *randomly* shown to the participants. The participants could also enter their own reasons in a textbox.

4.2.7 ***Participants and Demographic Information.*** After two months of running the user study, we were able to collect 581 participants. We excluded 28 participants who did not complete the user study or gave inconsistent answers. One example of an inconsistent answer is the following: A participant was not satisfied with the recommended item (i.e., provided a rating of 1 or 2), and he/she did not have a look at the explanation with the reason *"I was satisfied with the recommendation"*. We then used the dataset of 553 participants (male: 35.081%, female: 64.195%, and others: 0.723%) from 18 to 60 years old for further analysis. Among these, 71.067% of the participants are Vietnamese, and the remaining are European from different countries such as Austria, Bosnia, Croatia, Germany, Italy, Romania, Russia, and Slovenia. Most of our participants are students (94.033%) from 18 to 25 years old. The remaining are staff members who work in different areas, such as lecturers, researchers, officers, designers, software developers, and marketing experts. Due to the significant differences in the distribution of the participants according to *age* and *profession*, our analysis concerning personal characteristics was done only for the *gender* and *nationality* aspects.

4.2.8 ***Privacy concerns.*** Before participating in the user study, the participants were informed about the privacy policy and agreed with this policy before starting the user study. To preserve the participants' privacy, the collected information was used only for research purposes and was not shared with anyone else. Also, we guaranteed to delete this information as soon as we have completed this study.

5 DATA ANALYSIS METHODS, RESULTS, AND DISCUSSIONS

This section presents the details of data analyses, results, and related discussions concerning the research questions.

5.1 RQ1 - When do users need explanations of the recommended items?

Method: We assume that users' needs for explanations depend on their preferences for the recommended items. To examine our assumption, we analyzed the participants' needs for explanations in correlation with their ratings for the recommended item. In each item domain, we collected a dataset containing two variables: *seeExp* encoded as integer numbers $\in [0, 1]$ (see explanations or not) and *recItemRat* $\in [1..5]$ representing ratings for the recommended items. We ran *Bivariate Correlation* tests ($\alpha = 0.05$) to examine if there exist correlations between these two variables. Besides, we created two-way tables representing the number of participants who saw/did not see explanations corresponding to the rating values and then ran *Chi-square* tests ($\alpha = 0.05$). Also, we performed *descriptive statistics* to calculate the percentage of the participants who had a look at explanations in each item domain. Moreover, we split the rating values into layers and analyzed how the participants rated for the recommended item with a rating value rat_i. Finally, to learn the root causes, we analyzed the feedback of the participants who did not take a look at explanations.

Results and discussions: The Bivariate Correlation tests show no statistically significant correlations between the participants' ratings and their *"seeing explanation"* behavior (restaurant: *Pearson Correlation = 0.052, p = 0.64*; tourism: *Pearson Correlation = 0.037, p = 0.775*; accommodation: *Pearson Correlation = 0.063, p = 0.606*). The Chi-Square tests show a similar outcome, where no significant correlations could be found between the needs of users for explanations and the rating values ($p_{restaurant} = 0.463; p_{tourism} = 0.138; p_{accommodation} = 0.348$). However, further outcomes were found based on descriptive statistics. In the accommodation domain, the participants were more likely to take a look at the explanations

Rating value (rat_i)	Restaurant domain	Tourism domain	Accommodation domain
1	44.444%	41.667%	**67.222%**
2	52.778%	50.000%	**79.365%**
3	57.925%	46.337%	44.872%
4	**67.722%**	**67.944%**	60.138%
5	**64.135%**	**60.352%**	60.702%

Table 3: The percentage of the participants who saw the explanations and rated recommended items using the rating value rat_i. Each number in columns 2 - 4 represents the average percentage computed in the explanation types in a specific item domain.

if they were not satisfied with the recommended items (see the last column of *Table 3*). This can be explained by the fact that items in high-involvement domains are closely related to important decisions. Therefore, the dissatisfaction of the participants triggers a higher need for explanations. This behavior can also be explained by the *consumer dissatisfaction theory* found in psychological studies [6]. Having a look at explanations for recommendations in high-involvement item domains would help users to avoid sub-optimal decisions. On the other hand, the feedback of the participants who did not look at the explanations provides further arguments supporting the observed behavior. We found the top reason given by *33.333%* of the participants (in the accommodation domain) showing that they did not want to have a look at the explanations since *"they were satisfied with the recommendation"*. However, the analysis results in the restaurant and tourism domains show an opposite tendency, i.e., the participants wanted to see the explanations when they were satisfied with the recommended item. This tendency could be explained by the fact that, in less important decisions (e.g., low- and average-involvement) where a quick decision-making process is needed, it could be the case that the participants just want to check the explanation of the satisfying recommended options rather than spending additional time just for looking at the explanations of unsatisfying recommended options.

Take-home messages: In high-involvement item domains, users are more likely to check explanations for unsatisfying recommendations. Low- and average-involvement item domains show the opposite, where users tend to check the explanation of satisfying recommendations. Being aware of these tendencies can help recommender developers learn the satisfaction of users with the recommendation by analyzing their interactions with explanations.

5.2 RQ2 - Are there correlations between users' needs for explanations and the item domain?

Method: We collected a dataset consisting of two *categorical* variables: *seeExp* showing if a participant took a look at the explanation or not (*"yes"*/*"no"*) and *dom* indicating item domains. We ran *Chi-Square test of independence* ($\alpha = 0.05$) to examine the association between these two variables (i.e., whether the variables are independent or related). We also performed *descriptive statistics* to further analyze these correlations.

Results and discussions: The Chi-Square test does not show a statistically significant result. In particular, *Pearson Chi-Square Value* $\chi(2) = 0.847$, *p (2-sided) = 0.655*, and *Phi and Cramer's V = 0.034* draw a very weak correlation between the two mentioned variables. However, the descriptive statistics (in *Table 4* and also *Table 3* in *Section 5.1*) show differences in terms of the participants' needs for explanations across different item domains (see the related discussion in *Section 5.1*). In the restaurant and tourism domain, *FBExp* attracted more attention from the participants compared to other explanation types. This tendency can be explained by the needs of users for receiving recommendations generated by content-based recommendation approaches that have also been proven in related work [3, 40]. Besides, the participants were interested in checking *IBExp* in the accommodation domain, which surprised us pretty much since we expected, in high-involvement item domains, *KBExp* would have been checked more often by the participants.

	Restaurant domain	Tourism domain	Accommodation domain
IBExp	61.667%	57.143%	**61.538%**
FBExp	**68.852%**	**71.233%**	56.364%
KBExp	61.538%	56.364%	57.627%

Table 4: The percentage of the participants who had a look at the explanations in three item domains.

Take-home messages: For low- and average-involvement item domains, content-based recommendation approaches should still be promoted as discussed in the existing studies [3, 40]. The corresponding explanations, such as feature-based explanations, should also be provided to show the similarity between the recommended and rated items. In high-involvement item domains, besides the clarification of the knowledge-based recommendation approach (as discussed in [13, 14]), explanations that show the popularity of the recommended item could be of interest to users.

5.3 RQ3 - Are there correlations between users' needs for explanations and personal characteristics?

Method: As mentioned in *Section 4.2.7*, we analyzed only two personal characteristics, namely *gender* and *nationality.* For gender, we investigated the differences in the needs for explanations between *male* and *female* participants (we did not analyze the other genders due to a tiny number of participants (0.685%)). Regarding nationality, we analyzed the differences in the needs for seeing explanations between *Vietnamese* and *European* participants. We also conducted further analyses taking into account both gender and nationality characteristics at the same time. For example, we compared the needs for seeing explanations between *Vietnamese male/female* and

	Restaurant	Tourism	Accommodation
Vietnamese vs. European	p = 0.639, $\chi(1)$ = 0.221, Phi and Cramer's V = 0.034	p = 0.892, $\chi(1)$ = 0.018, Phi and Cramer's V = 0.010	p = 0.716, $\chi(1)$ = 0.133, Phi and Cramer's V = 0.027
Male vs. Female	p = 0.222, $\chi(1)$ = 1.489, Phi and Cramer's V = 0.078	p = 0.449, $\chi(1)$=0.573, Phi and Cramer's V = 0.056	p = 0.850, $\chi(1)$ = 0.036, Phi and Cramer's V = 0.015
European male vs. Vietnamese male	p = 0.533, $\chi(1)$ = 0.389, Phi and Cramer's V = 0.075	p = 0.353, $\chi(1)$ = 0.862, Phi and Cramer's V = 0.128	p = 0.960, $\chi(1)$=0.002, Phi and Cramer's V = 0.007
European female vs. Vietnamese female	p = 0.495, $\chi(1)$ = 0.466, Phi and Cramer's V = 0.063	p = 0.651, $\chi(1)$ = 0.205, Phi and Cramer's V = 0.040	p = 0.813, $\chi(1)$=0.056, Phi and Cramer's V = 0.022

Table 5: Chi-square tests for examining correlations between user needs for explanations and personal characteristics.

Kruskal-Wallis test (alpha = 0.05)

	Exp	Mean Rank	p
U	*IBExp*	**69.01**	0.086
	FBExp	52.80	
	KBExp	59.23	
D	*IBExp*	**65.88**	0.231
	FBExp	53.36	
	KBExp	61.54	
S	*IBExp*	**70.76**	**0.034**
	FBExp	51.25	
	KBExp	59.24	

Mann-Whitney U test (alpha = 0.0167)

	S
IBExp vs. *FBExp*	**0.012**
IBExp vs. *KBExp*	0.111
FBExp vs. *KBExp*	0.253

(a) Restaurant

Kruskal-Wallis test (alpha = 0.05)

	Exp	Mean Rank	p
U	*IBExp*	85.99	**0.036**
	FBExp	**106.14**	
	KBExp	86.51	
D	*IBExp*	88.40	0.172
	FBExp	**102.92**	
	KBExp	88.14	
S	*IBExp*	89.07	0.206
	FBExp	**102.40**	
	KBExp	88.06	

Mann-Whitney U test (alpha = 0.0167)

	U
IBExp vs. *FBExp*	0.020
IBExp vs. *KBExp*	0.986
FBExp vs. *KBExp*	0.043

(b) Tourism

Kruskal-Wallis test (alpha = 0.05)

	Exp	Mean Rank	p
U	*IBExp*	**90.57**	0.992
	FBExp	89.57	
	KBExp	89.86	
D	*IBExp*	**94.68**	0.629
	FBExp	86.75	
	KBExp	87.87	
S	*IBExp*	**94.52**	0.610
	FBExp	89.20	
	KBExp	85.76	

(c) Accommodation

Figure 3: The results of Kruskal-Wallis tests ($\alpha = 0.05$) and Mann-Whitney U tests ($\alpha = 0.0167$) in all item domains. The explanations were analyzed based on three dimensions: Understandability (U), Display Style (D), and Overall Satisfaction (S).

European male/female participants. To address all the mentioned aspects, we ran *Chi-Square tests of independence* ($\alpha = 0.05$) with the datasets consisting of two categorical variables: *seeExp* - see explanations or not (*"yes"/"no"*) and *perChar* - gender, nationality, or nationality-gender. Besides, we performed *Mann-Whitney U tests* to look into relationships between recommended item ratings and personal characteristics.

Results and discussions: The results of the mentioned Chi-Square tests are summarized *Table 5*. The *p* values, *Pearson Chi-Square* values ($\chi(1)$), and *Phi and Cramer's* values show that there was not enough evidence to suggest an association between the needs for explanations and gender/nationality. Although differences in gender and cultural background do not bring significant impacts on the needs of the participants for explaining recommendations, further analysis in the recommended item ratings tells us interesting information. The *Mann-Whitney U tests* ($\alpha = 0.05$) in the restaurant and accommodation item domains show significant differences concerning the rating behavior of participants from different cultural backgrounds ($p_{restaurant} = 0.001, p_{accommodation} = 0.000$). The mean rank (MR) and average rating (AVG) values in these two domains show that Vietnamese participants were more satisfied with the recommended items compared to European participants (Restaurant: $MR_{Vietnam}$ = **104.34**, MR_{EUROPE} = 77.24, $AVG_{Vietnam}$ = **4.016**, AVG_{EUROPE} = 3.419; Accommodation: $MR_{Vietnam}$ = **101.00**, MR_{EUROPE} = 62.40, $AVG_{Vietnam}$ = **3.961**, AVG_{EUROPE} = 3.039).

Take-home messages: Users' needs for explanations are independent of their gender and nationality. In low- and high-involvement item domains, Asian users are more satisfied with the recommended item than European users.

5.4 RQ4 - What is the favorite explanation from the user point of view in a specific item domain?

Method: We analyzed the evaluations of the participants for each explanation type according to *three* metrics: *understandability, display style,* and *overall satisfaction*. For each explanation type of a specific item domain, we collected *three* evaluation sets corresponding to the mentioned metrics. Please note that the mentioned datasets were collected from the participants who had a look at the explanations. The evaluations in each set share the same characteristics, such as *ordinal variables* $\in [1..5]$, *independent* of each other (the evaluations of one explanation did not rely upon those of other explanations), and *not normally distributed* (*Shapiro-Wilk* tests, $\alpha = 0.05$, $p - values < \alpha$). For this reason, we selected the *Kruskal-Wallis* test ($\alpha = 0.05$) to analyze our data. In each item domain, we ran *Kruskal-Wallis* tests for the three mentioned metrics. For each metric, the test examined if there were statistically significant differences in the participants' evaluations across different explanation types. We also inspected the mean ranks generated in the *Kruskal-Wallis* tests to identify the best explanation. *The higher the mean rank, the better the explanation in terms of the mentioned metrics.* Besides, we performed pair-wise tests (*Mann-Whitney U* tests) to find significant differences between the two explanation types. Since *three* Mann-Whitney U tests are needed for three explanation types, to avoid *Type I* errors[11], we applied a *Bonferroni adjustment* [34] to adapt the significance level: $\alpha' = 0.05/3 = 0.0167$.

Results and discussions: The Kruskal Wallis tests in the accommodation item domain do not show significant differences in the

[11] In hypothesis testing, a Type I error involves rejecting the null hypothesis (e.g., there are no differences among the groups) when it is true [34]

participants' evaluations with regard to the mentioned evaluation metrics (see *Figure 3*). However, we found significant results in the restaurant and tourism domains. In the restaurant domain, *IBExp* is the favorite explanation of users. The mean rank values of the Kruskal-Wallis tests show that *IBExp* achieves the highest mean rank values in all evaluation metrics. Statistically, the Kruskal Wallis test in the *"overallSatisfaction"* metric shows a significant result, which leads to the conclusion that *IBExp* best helps to increase the overall satisfaction of users with the recommender system. The participants' preferences for this explanation type could be explained by the fact that the participants tend to look for items liked by similar users or highly rated by the community. This tendency is clearly shown in low-involvement items domains when users select items such as songs, movies, and restaurants. For instance, a user might visit *Restaurant X* since it is on the list of top five restaurants in the city. In the tourism domain, the participants preferred *FB-Exp* over other explanation types. The Kruskal-Wallis test shows a significant result in the *"understandability"* metric, meaning that explanations that show the similarity (in terms of feature-wise) between the recommended item and the consumed items would achieve a high level of the participants' understandability.

Take-home messages: Users' preferences for explanations change across different item domains, especially in low- and average-involvement item domains. In low-involvement item domains, showing how well the community has rated the recommended item would help to improve user satisfaction with the recommended system. In average-involvement item domains, showing how the recommended item is similar to the previously-consumed items is an excellent way to formulate understandable explanations.

6 CONCLUSIONS, LIMITATIONS, AND FUTURE WORK

Analyzing the needs of users for explanations is an essential task that helps the developers of recommender system be aware of when explanations should be shown to users. We conducted a user study in different item domains and analyzed users' needs for explanations from the item domain and personal characteristics perspectives. We also found ways to design explanations with a high level of understandability and overall satisfaction with the recommender system.

Although the paper has provided elaborate analyses from different perspectives, it has some limitations. The *first* limitation lies in the synthetic datasets, which could affect users' experiences and their evaluations of the recommended items. For future work, we will collect real datasets from the existing systems with further consideration about copyright-related complexities. The *second* limitation was about the relativity of the involvement of the selected item domain. We are aware that, depending on the scenario, decisions in a specific item domain can be considered with different involvement levels. For instance, deciding on a tour does not always link to an average-involvement decision but could also be referred to as a high-involvement decision in scenarios when users are often scared of traveling. Similarly, although the accommodation domain is usually related to high-involvement decisions compared to the restaurant and tourism domains, other scenarios, such as purchasing a house/apartment, would reflect more clearly high-involvement decisions. We plan to conduct further user studies focusing on high-involvement decision scenarios and extend them to other item domains to increase the generalization of the results from the item domain perspective. The *last* limitation was the unequal distributions of the participants in terms of *nationality*, *age range*, and *profession* (more than 90% of the participants are students from 18 - 25 years old and more than 70% of participants are Vietnamese). For the following steps, we will extend our user study to various user groups to further analyze the impacts of personal characteristics on users' needs for explaining recommendations.

REFERENCES

[1] Fabian Abel, Eelco Herder, Geert-Jan Houben, Nicola Henze, and Daniel Krause. 2013. Cross-system user modeling and personalization on the Social Web. *User Modeling and User-Adapted Interaction* 23, 2-3 (2013), 169–209. https://doi.org/10.1007/s11257-012-9131-2

[2] Alaa Alslaity and Thomas Tran. 2020. The Effect of Personality Traits on Persuading Recommender System Users. In *Proceedings of the 7th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems co-located with 14th ACM Conference on Recommender Systems (RecSys 2020), Online Event, September 26, 2020 (CEUR Workshop Proceedings, Vol. 2682)*, Peter Brusilovsky, Marco de Gemmis, Alexander Felfernig, Pasquale Lops, John O'Donovan, Giovanni Semeraro, and Martijn C. Willemsen (Eds.). CEUR-WS.org, 48–56. http://ceur-ws.org/Vol-2682/paper5.pdf

[3] Asmae Bentaleb, Younès El Bouzekri El Idrissi, and Ayoub Ait Lahcen. 2021. A Review on Content Based Recommender Systems in Tourism. In *Digital Technologies and Applications*, Saad Motahhir and Badre Bossoufi (Eds.). Springer International Publishing, Cham, 527–541.

[4] Hasida Benzur and Moshe Zeidner. 1988. Sex differences in anxiety, curiosity, and anger: A cross-cultural study. *Sex roles* 19 (September 1988), 335–347. https://doi.org/10.1007/BF00289841

[5] Shlomo Berkovsky, Tsvi Kuflik, and Francesco Ricci. 2008. Mediation of user models for enhanced personalization in recommender systems. *User Model User-Adap Inter* 18, 245–286 (2008), 169–209. https://doi.org/10.1007/s11257-007-9042-9

[6] Portia Opoku Boadi, Li Guoxin, Andrew Adjah Sai, and Philip Antwi. 2017. A critique of the impact of customer dissatisfaction. In *International Conference on Transformations and Innovations in Management (ICTIM '17), Shanghai, China*, X. Zheng (Ed.). Atlantis Press, 559–567.

[7] Luchiana Brodeala. 2020. *Online Recommender System for Accessible Tourism Destinations*. Association for Computing Machinery, New York, NY, USA, 787–791. https://doi.org/10.1145/3383313.3411450

[8] Cristina Calvo-Porral and John Stanton. 2015. *Principles of Marketing* (3rd ed.). University of Minnesota Libraries Publishing, 309 19th Avenue South Minneapolis, MN 55455.

[9] Iván Cantador, Ignacio Fernández-Tobías, Shlomo Berkovsky, and Paolo Cremonesi. 2015. *Cross-Domain Recommender Systems*. 919–959. https://doi.org/10.1007/978-1-4899-7637-6_27

[10] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA) *(RecSys '16)*. Association for Computing Machinery, New York, NY, USA, 191–198. https://doi.org/10.1145/2959100.2959190

[11] Henriette Cramer, Vanessa Evers, Satyan Ramlal, Maarten Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wielinga. 2008. The Effects of Transparency on Trust in and Acceptance of a Content-Based Art Recommender. *User Modeling and User-Adapted Interaction* 18, 5 (nov 2008), 455–496. https://doi.org/10.1007/s11257-008-9051-3

[12] Elif Erbil and Wolfgang Wörndl. 2021. Generating Multi-Day Round Trip Itineraries for Tourists. In *Proceedings of the Workshop on Web Tourism co-located with the 14th ACM International WSDM Conference (WSDM 2021), Jerusalem, Israel, March 12, 2021 (CEUR Workshop Proceedings, Vol. 2855)*, Catalin-Mihai Barbu, Ludovik Coba, Amra Delic, Dmitri Goldenberg, Tsvi Kuflik, Markus Zanker, and Julia Neidhardt (Eds.). CEUR-WS.org, 1–7. http://ceur-ws.org/Vol-2855/main_short_1.pdf

[13] Alexander Felfernig, Bartosz Gula, and Erich Teppan. 2007. Knowledge-based Recommender Technologies for Marketing and Sales. *International journal of pattern recognition and artificial intelligence* 21, 2 (2007), 333–354. https://doi.org/10.1142/S0218001407005417 I.

[14] Alexander Felfernig and Alfred Kiener. 2005. Knowledge-Based Interactive Selling of Financial Services with FSAdvisor. In *Proceedings of the 17th Conference on Innovative Applications of Artificial Intelligence - Volume 3* (Pittsburgh, Pennsylvania) *(IAAI'05)*. AAAI Press, 1475–1482.

[15] Sofia Gkika, Marianna Skiada, George Lekakos, and Panos Kourouthanassis. 2016. Investigating the Role of Personality Traits and Influence Strategies on the Persuasive Effect of Personalized Recommendations. In *Proceedings of the 4th Workshop on Emotions and Personality in Personalized Systems co-located with ACM Conference on Recommender Systems (RecSys 2016)* (Boston, MA, USA). CEUR Workshop Proceedings, Aachen, Germany, 9–17.

[16] Carlos A. Gomez-Uribe and Neil Hunt. 2016. The Netflix Recommender System: Algorithms, Business Value, and Innovation. *ACM Trans. Manage. Inf. Syst.* 6, 4, Article 13 (dec 2016), 19 pages. https://doi.org/10.1145/2843948

[17] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. 2000. Explaining Collaborative Filtering Recommendations. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work* (Philadelphia, Pennsylvania, USA) *(CSCW '00)*. Association for Computing Machinery, New York, NY, USA, 241–250. https://doi.org/10.1145/358916.358995

[18] Yunfeng Hou, Ning Yang, Yi Wu, and Philip S. Yu. 2019. Explainable Recommendation with Fusion of Aspect Information. *World Wide Web* 22, 1 (jan 2019), 221–240. https://doi.org/10.1007/s11280-018-0558-1

[19] Hyunwoo Hwangbo, Yang Sok Kim, and Kyung Jin Cha. 2018. Recommendation system development for fashion retail e-commerce. *Electronic Commerce Research and Applications* 28 (2018), 94–101. https://doi.org/10.1016/j.elerap.2018.01.012

[20] Yuichi Ishikawa, Akihiro Kobayashi, and Atsunori Minamikawa. 2019. Predicting Advertising Appeal from Receiver's Psychological Traits and Ad Design Features. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization* (Larnaca, Cyprus) *(UMAP'19 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 45–49. https://doi.org/10.1145/3314183.3324979

[21] Dietmar Jannach, Michael Jugovac, and Ingrid Nunes. 2019. Explanations and User Control in Recommender Systems. In *Proceedings of the 23rd International Workshop on Personalization and Recommendation on the Web and Beyond* (Hof, Germany) *(ABIS '19)*. Association for Computing Machinery, New York, NY, USA, 31. https://doi.org/10.1145/3345002.3349293

[22] Yucheng Jin, Nava Tintarev, and Katrien Verbert. 2018. Effects of Individual Traits on Diversity-Aware Music Recommender User Interfaces. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization* (Singapore, Singapore) *(UMAP '18)*. Association for Computing Machinery, New York, NY, USA, 291–299. https://doi.org/10.1145/3209219.3209225

[23] Yucheng Jin, Nava Tintarev, and Katrien Verbert. 2018. Effects of Personal Characteristics on Music Recommender Systems with Different Levels of Controllability. In *Proceedings of the 12th ACM Conference on Recommender Systems* (Vancouver, British Columbia, Canada) *(RecSys '18)*. Association for Computing Machinery, New York, NY, USA, 13–21. https://doi.org/10.1145/3240323.3240358

[24] Bart P. Knijnenburg, Martijn C. Willemsen, Hakan Soncu Gantner, and Chris Newel. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* 22 (2012), 441–504. https://doi.org/10.1007/s11257-011-9118-4

[25] Bin Li, Qiang Yang, and Xiangyang Xue. 2009. Can Movies and Books Collaborate? Cross-Domain Collaborative Filtering for Sparsity Reduction. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence* (Pasadena, California, USA) *(IJCAI'09)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2052–2057.

[26] Yu Liang and Martijn C. Willemsen. 2019. *Personalized Recommendations for Music Genre Exploration*. Association for Computing Machinery, New York, NY, USA, 276–284. https://doi.org/10.1145/3320435.3320455

[27] Zachary C. Lipton. 2018. The Mythos of Model Interpretability: In Machine Learning, the Concept of Interpretability is Both Important and Slippery. *Queue* 16, 3 (jun 2018), 31–57. https://doi.org/10.1145/3236386.3241340

[28] Chieh Lo, Hongliang Yu, Xin Yin, Krutika Shetty, Changchen He, Kathy Hu, Justin M Platz, Adam Ilardi, and Sriganesh Madhvanath. 2021. *Page-Level Optimization of e-Commerce Item Recommendations*. Association for Computing Machinery, New York, NY, USA, 495–504. https://doi.org/10.1145/3460231.3474242

[29] David McSherry. 2005. Explanation in Recommender Systems. *Artificial Intelligence Review* 24 (2005), 179–197. https://doi.org/10.1007/s10462-005-4612-x

[30] Martijn Millecamp, Nyi Nyi Htun, Cristina Conati, and Katrien Verbert. 2019. To Explain or Not to Explain: The Effects of Personal Characteristics When Explaining Music Recommendations. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 397–407. https://doi.org/10.1145/3301275.3302313

[31] Martijn Millecamp, Nyi Nyi Htun, Yucheng Jin, and Katrien Verbert. 2018. Controlling Spotify Recommendations: Effects of Personal Characteristics on Music Recommender User Interfaces. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization* (Singapore, Singapore) *(UMAP '18)*. Association for Computing Machinery, New York, NY, USA, 101–109. https://doi.org/10.1145/3209219.3209223

[32] Cataldo Musto, Fedelucio Narducci, Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. 2018. Linked Open Data-based Explanations for Transparent Recommender Systems. *International Journal of Human-Computer Studies* 121 (03 2018).

[33] Ingrid Nunes and Dietmar Jannach. 2017. A systematic review and taxonomy of explanations in decision support and recommender systems. *CoRR* abs/2006.08672 (2017). https://arxiv.org/abs/2006.08672

[34] Julie Pallant. 2007. *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows Version 15* (3rd ed.). Open University Press, Milton Keynes, UK, USA.

[35] Weike Pan, Evan W. Xiang, Nathan N. Liu, and Qiang Yang. 2010. Transfer Learning in Collaborative Filtering for Sparsity Reduction. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence* (Atlanta, Georgia) *(AAAI'10)*. AAAI Press, 230–235.

[36] A. Papadimitriou, P. Symeonidis, and Y. Manolopoulos. 2012. A Generalized Taxonomy of Explanations Styles for Traditional and Social Recommender Systems. *Data Mining and Knowledge Discovery* 24, 3 (2012), 555–583.

[37] Georgina Peake and Jun Wang. 2018. Explanation Mining: Post Hoc Interpretability of Latent Factor Models for Recommendation Systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (London, United Kingdom) *(KDD '18)*. Association for Computing Machinery, New York, NY, USA, 2060–2069. https://doi.org/10.1145/3219819.3220072

[38] Zhaochun Ren, Shangsong Liang, Piji Li, Shuaiqiang Wang, and Maarten de Rijke. 2017. Social Collaborative Viewpoint Regression with Explainable Recommendations. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (Cambridge, United Kingdom) *(WSDM '17)*. Association for Computing Machinery, New York, NY, USA, 485–494.

[39] Mireia Ribera Turró and Agata Lapedriza. 2019. Can we do better explanations? A proposal of User-Centered Explainable AI. In *IUI Workshops* (Los Angeles, USA) *(IUI '19)*. CEUR Workshop Proceedings, 1–7.

[40] Mohammad Salehan, Sonya Zhang, and Navid Aghakhani. 2017. A recommender System for Restaurant Reviews Based on Consumer Segment. In *Proceedings of the 23rd Americas Conference on Information Systems AMCIS' 2017*, Vol. 11. 3150–3154.

[41] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. 2001. Item-Based Collaborative Filtering Recommendation Algorithms. In *Proceedings of the 10th International Conference on World Wide Web* (Hong Kong, Hong Kong) *(WWW '01)*. Association for Computing Machinery, New York, NY, USA, 285–295. https://doi.org/10.1145/371920.372071

[42] Rajesh Sharma and Sanjog Ray. 2016. Explanations in Recommender Systems: An Overview. *Int. J. Bus. Inf. Syst.* 23, 2 (jan 2016), 248–262. https://doi.org/10.1504/IJBIS.2016.078909

[43] Rashmi Sinha and Kirsten Swearingen. 2002. The Role of Transparency in Recommender Systems. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA) *(CHI EA '02)*. Association for Computing Machinery, New York, NY, USA, 830–831.

[44] Panagiotis Symeonidis, Alexandros Nanopoulos, and Yannis Manolopoulos. 2008. Justified recommendations based on content and rating data. *WebKDD Workshop on Web Mining and Web Usage Analysis*.

[45] T. Yang Tang, Pinata Winoto, and Robert Ziqin Ye. 2011. Analysis of a multi-domain recommender system. In *Proceedings of the 3rd International Conference on Data Mining and Intelligent Information Technology Applications (ICMiA' 2011)*. 280–285.

[46] Maartje ter Hoeve, Mathieu Heruer, Daan Odijk, Anne Schuth, M. Spitters, and M. Rijke. 2017. Do News Consumers Want Explanations for Personalized News Rankings. In *The Workshop on Responsible Recommendation* (Como, Italy) *(FATREC '07)*.

[47] Nava Tintarev. 2007. Explanations of Recommendations. In *Proceedings of the 2007 ACM Conference on Recommender Systems* (Minneapolis, MN, USA) *(RecSys '07)*. Association for Computing Machinery, New York, NY, USA, 203–206. https://doi.org/10.1145/1297231.1297275

[48] Nava Tintarev. 2007. Explanations of Recommendations. In *Proceedings of the 2007 ACM Conference on Recommender Systems* (Minneapolis, MN, USA) *(RecSys '07)*. Association for Computing Machinery, New York, NY, USA, 203–206. https://doi.org/10.1145/1297231.1297275

[49] Nava Tintarev and Judith Masthoff. 2007. A Survey of Explanations in Recommender Systems, In Proceedings of the 2007 IEEE 23rd International Conference on Data Engineering Workshop. *Data Engineering Workshop, 2007 IEEE 23rd International Conference on*, 801–810.

[50] Nava Tintarev and Judith Masthoff. 2012. Evaluating the effectiveness of explanations for recommender systems: Methodological issues and empirical studies on the impact of personalization. *User Modelling and User-Adapted Interaction* 22, 4-5 (Oct. 2012), 399–439.

[51] Nava Tintarev and Judith Masthoff. 2015. *Explaining Recommendations: Design and Evaluation*. Springer US, Boston, MA, 353–382.

[52] Thi Ngoc Trang Tran, Alexander Felfernig, Christoph Trattner, and Andreas Holzinger. 2020. Recommender systems in the healthcare domain: state-of-the-art and research issues. *Journal of Intelligent Information Systems* (17 Dec. 2020), 1–31. https://doi.org/10.1007/s10844-020-00633-6

[53] Thi Ngoc Trang Tran, Müslüm Atas, Alexander Felfernig, Ralph Samer, and Martin Stettinger. 2018. Investigating Serial Position Effects in Sequential

Group Decision Making. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization* (Singapore, Singapore) *(UMAP '18)*. Association for Computing Machinery, New York, NY, USA, 239–243. https://doi.org/10.1145/3209219.3209255

[54] Thi Ngoc Trang Tran, Müslüm Atas, Alexander Felfernig, and Martin Stettinger. 2018. An Overview of Recommender Systems in the Healthy Food Domain. *J. Intell. Inf. Syst.* 50, 3 (jun 2018), 501–526. https://doi.org/10.1007/s10844-017-0469-0

[55] Thi Ngoc Trang Tran, Alexander Felfernig, and Nava Tintarev. 2021. Humanized Recommender Systems: State-of-the-Art and Research Issues. *ACM Trans. Interact. Intell. Syst.* 11, 2, Article 9 (July 2021), 41 pages. https://doi.org/10.1145/3446906

[56] Thi Ngoc Trang Tran, Viet Man Le, Muesluem Atas, Alexander Felfernig, Martin Stettinger, and Andrei Popescu. 2021. Do Users Appreciate Explanations of Recommendations? An Analysis in the Movie Domain. In *Fifteenth ACM Conference on Recommender Systems*. Association for Computing Machinery, New York, NY, USA, 645–650.

[57] Arun Tripathi and A.K. Sharma. 2020. Recommending Restaurants: A Collaborative Filtering Approach. In *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*. 1165–1169. https://doi.org/10.1109/ICRITO48877.2020.9197946

[58] Chun-Hua Tsai and Peter Brusilovsky. 2018. Explaining Social Recommendations to Casual Users: Design Principles and Opportunities. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion* (Tokyo, Japan) *(IUI '18 Companion)*. Association for Computing Machinery, New York, NY, USA, Article 59, 2 pages.

[59] Kosetsu Tsukuda and Masataka Goto. 2020. Explainable Recommendation for Repeat Consumption. In *Fourteenth ACM Conference on Recommender Systems* (Virtual Event, Brazil) *(RecSys '20)*. Association for Computing Machinery, New York, NY, USA, 462–467. https://doi.org/10.1145/3383313.3412230

[60] Jesse Vig, Shilad Sen, and John Riedl. 2009. Tagsplanations: Explaining Recommendations Using Tags. In *Proceedings of the 14th International Conference on Intelligent User Interfaces* (Sanibel Island, Florida, USA) *(IUI '09)*. Association for Computing Machinery, New York, NY, USA, 47–56. https://doi.org/10.1145/1502650.1502661

[61] Beidou Wang, Martin Ester, Jiajun Bu, and Deng Cai. 2014. Who Also Likes It? Generating the Most Persuasive Social Explanations in Recommender Systems. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada*, Carla E. Brodley and Peter Stone (Eds.). AAAI Press, 173–179.

[62] Yongfeng Zhang and Xu Chen. 2020. Explainable Recommendation: A Survey and New Perspectives. *Foundations and Trends® in Information Retrieval* 14, 1 (2020), 1–101.

[63] Yongfeng Zhang, Guokun Lai, Min Zhang, Yi Zhang, Yiqun Liu, and Shaoping Ma. 2014. Explicit Factor Models for Explainable Recommendation Based on Phrase-Level Sentiment Analysis. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval* (Gold Coast, Queensland, Australia) *(SIGIR '14)*. Association for Computing Machinery, New York, NY, USA, 83–92. https://doi.org/10.1145/2600428.2609579

[64] Zhirun Zhang, Li Chen, Tonglin Jiang, Yutong Li, and Lei Li. 2022. Effects of Feature-Based Explanation and Its Output Modality on User Satisfaction With Service Recommender Systems. *Frontiers in big Data* 5 (2022), 897381–897381.

[65] Wayne Xin Zhao, Yanwei Guo, Yulan He, Han Jiang, Yuexin Wu, and Xiaoming Li. 2014. We know what you want to buy: a demographic-based system for product recommendation on microblogs. In *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014*, Sofus A. Macskassy, Claudia Perlich, Jure Leskovec, Wei Wang, and Rayid Ghani (Eds.). ACM, 1935–1944.

[66] Lei Zheng, Vahid Noroozi, and Philip S. Yu. 2017. Joint Deep Modeling of Users and Items Using Reviews for Recommendation. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (Cambridge, United Kingdom) *(WSDM '17)*. Association for Computing Machinery, New York, NY, USA, 425–434. https://doi.org/10.1145/3018661.3018665