Towards Multi-Method Support for Product Search and Recommending

Timm Kleemann
University of Duisburg-Essen
Duisburg, Germany
timm.kleemann@uni-due.de

Benedikt Loepp
University of Duisburg-Essen
Duisburg, Germany
benedikt.loepp@uni-due.de

Jürgen Ziegler
University of Duisburg-Essen
Duisburg, Germany
juergen.ziegler@uni-due.de

ABSTRACT

Today, online shops offer a variety of components to support users in finding suitable items, ranging from filters and recommendations to conversational advisors and natural language chatbots. All these methods differ in terms of cognitive load and interaction effort, and, in particular, in their suitability for the specific user. However, it is often difficult for users to determine which method to use to reach their goal. Moreover, as the settings are not propagated between the methods, there is a lack of support for switching components. In this paper, we study the reasons for using the different components in more detail and present an initial proposal for a multi-method approach that provides a more seamless experience, allowing users to freely and flexibly choose from all available methods at any time.

CCS CONCEPTS

• **Human-centered computing → User interface design**; *HCI theory, concepts and models*; **Empirical studies in HCI**; • **Information systems → Recommender systems**; **Search interfaces**; *Online shopping*.

KEYWORDS

Decision Aids, Faceted filtering, Conversational recommender systems, Product Advisors, Chatbots.

ACM Reference Format:
Timm Kleemann, Benedikt Loepp, and Jürgen Ziegler. 2022. Towards Multi-Method Support for Product Search and Recommending. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '22 Adjunct), July 4–7, 2022, Barcelona, Spain.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3511047.3536408

1 INTRODUCTION AND BACKGROUND

E-commerce websites offer an overwhelming selection of products, making it difficult for users to identify relevant content. To counteract this information overload and to facilitate finding suitable products in large item spaces, *search and filtering mechanisms* are almost always available [5, 33]. However, on the side of the users, they require a well-defined search goal and profound knowledge about the product domain. At the same time, modern techniques such as faceted filtering provide a high degree of control, allowing users to narrow down even very large result sets in an effective manner [16]. Yet, users need to invest a considerable amount of interaction effort and must understand the meaning of individual filter attributes and their relevance with respect to the current objective [25].

On the other end of the spectrum, *recommender systems* guide users more actively towards relevant content. Widely used in industry [4, 28], they reduce both cognitive load and interaction effort [15, 24]. As the items are automatically recommended based on historical user preference data [9], the users' influence on the process is, however, strongly limited [15, 19]. In addition, there is often a lack of transparency, making if difficult to comprehend why certain products are suggested [22], in particular, if domain knowledge is low or the presented information not detailed enough [18]. While most online shops still rely on these one-shot recommendations, more interactive approaches gain increasing attention, offering opportunities to adjust or critique the recommendations according to the current situation [10, 19].

Beyond that, conversational systems enable users to respond on a more human-like level [8, 29]. These include *dialog-based advisors*, which originate in early rule-based approaches where users had to answer a given sequence of questions based on a set of predefined answers [32]. In today's web, the questions are usually formulated on an application-oriented level rather than on the level of (technical) product features. This is beneficial for users with less domain knowledge, but raises problems for experts who (to a certain extent) know what they are looking for. In this case, *chatbots* may be more helpful: Building on recent advances in deep learning, they mimic a natural language conversation, thus providing users more freedom in how the dialog evolves [23]. Nevertheless, the underlying strategy to move the conversation forward is typically system-driven, based on a "system asks, user answers" pattern [36]. This feigned freedom of natural language interaction may leave users disappointed when they realize that in fact they cannot actively steer the conversation, e.g., by asking for clarifications [8]. Moreover, users are less efficient due to the effort of writing answers, while the systems may have difficulties interpreting the utterances in these answers [7].

Overall, this shows that the methods available in contemporary online shops support different user needs and cognitive stages in the decision-making process. In addition, it becomes visible that there is a lack of systems that assist users in a holistic fashion: Whenever several methods are offered on an e-commerce website, users typically do not know which one to use best. Moreover, information requested through one of the methods is rarely used by the other components to adapt the results. This separation of techniques is also mirrored in research, except for a few, very specific approaches that try to combine, e.g., selected interactive mechanisms [20] or conversational agents with faceted filtering [12].

UMAP '22 Adjunct, July 4–7, 2022, Barcelona, Spain

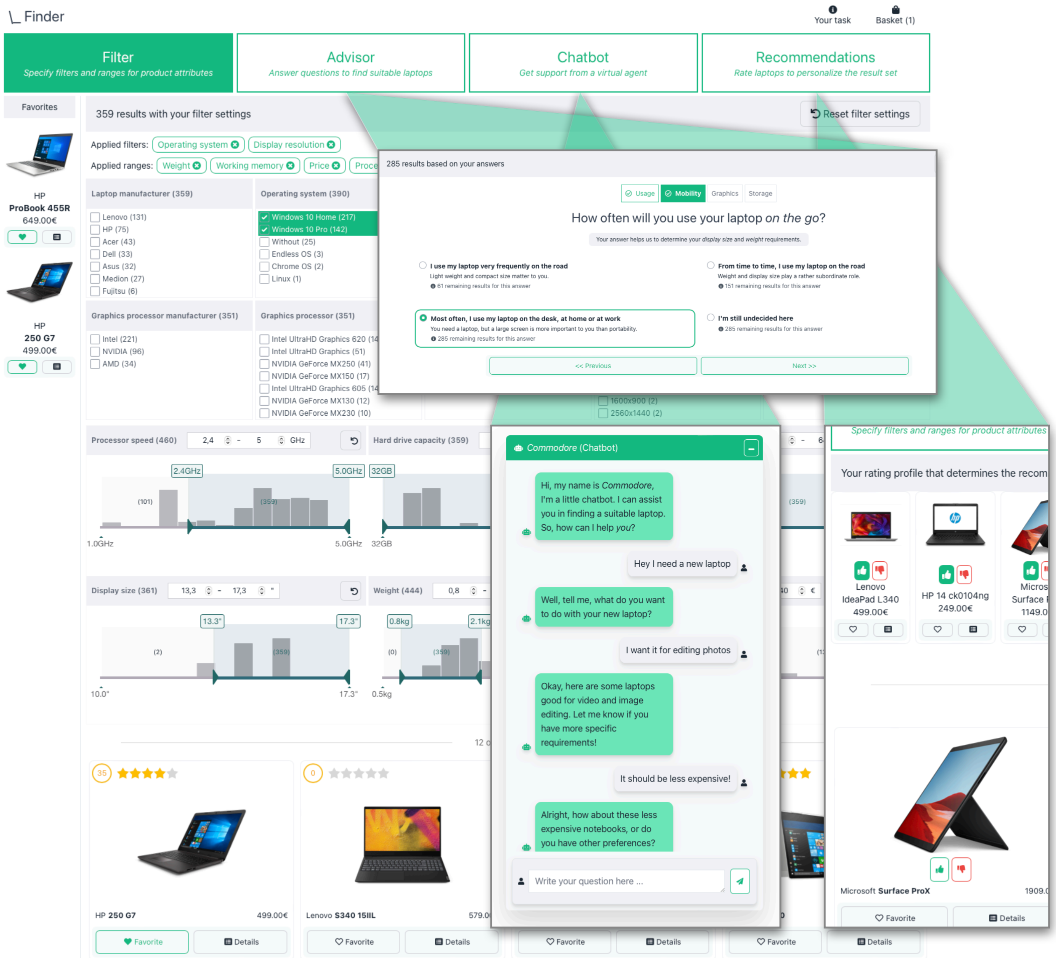

Figure 1: Screenshot of our fictitious online shop, showing the four components to support users in finding suitable items.

Related to that, it has been shown that aspects such as domain knowledge moderate which interaction possibilities are preferred [1, 13]. However, only few attempts have been made to explore user preferences in scenarios where a wider range of methods is available [11, 26, 30]. This is particularly unfortunate since it is well known that users, in practice, use several methods before settling on a product [26, 30], each with a different effect on the progress towards their goal [3, 6]. Consequently, we aim at exploring scenarios that involve using methods from the entire spectrum described above, and at exploiting their specific advantages. In this paper, we report an empirical user experiment with 100 participants who used a system that allowed to choose freely from all components. The results provide first insights as to which components users like to use when looking for a product, and for what reasons they switch to others. Based on these findings, we make an initial proposal for a multi-method approach that integrates the components in a way that each action performed in one component consistently affects the others, thereby avoiding that users lose progress or any input entered before.

2 EMPIRICAL USER EXPERIMENT

We assumed that users would have different reasons for switching between methods as well as for not using them at all. We further expected that user experience would be affected if multiple components are available. As this could appear more complicated at first sight, we expected effects on comprehensibility and intention to use the components again. We assumed that aspects such as prior experience with the methods would have a moderating effect.

2.1 Method

To investigate these assumptions, we conducted an online user study with a fictitious online shop to capture interaction behavior and to collect both quantitative and qualitative feedback on the perception of the system and its components.

2.1.1 Prototype. We implemented the fictitious online shop for the domain of laptops, i.e. typical search products [27] where a variety of (often not obvious) features is usually available, even for single models from a specific brand. We crawled a dataset from *NBB.com*, a German notebook retailer, consisting of 1 269 products and information on 17 features, prices, and average ratings on a

5-star scale. As shown in Figure 1, the system offered four different components: a faceted filtering interface, a dialog-based product advisor, a natural language chatbot implemented using *Google Dialogflow*, and a content-based recommender based on likes and dislikes. Each component was presented on an individual page, accessible via a (randomized) navigation menu that was permanently visible at the top. Therefore, participants had to actively switch between components, even if they were only interested in seeing a component rather than interacting with it. While this may have affected their behavior in comparison to a real-world system, we deemed it necessary to identify in the log data when they engaged with a component. Other potential methods for detecting switching the focus such as eye tracking were considered less informative and also had hygiene restrictions due to COVID-19. The method allowed us to ask participants about their reasons for switching at the exact moment. The result set was always displayed in the lower part of the window. In addition, we implemented a product detail page, and, for the purpose of the study, a shopping basket.

2.1.2 Task. Participants were asked to put at least two laptops into the shopping basket (without any time constraints), based on one of the following scenarios to which they were assigned to in counterbalanced order: a *goal-driven* scenario, mentioning requirements for a laptop for work; an *explorative* scenario, just asking to look for laptop for a friend. In both cases, explicit product specifications were not provided so that it was not possible to simply translate them into filter criteria or chatbot queries. For the sake of space, we do not detail on the scenarios in the first analysis.

2.1.3 Questionnaires and Interaction Data. Before and after completing the task, a questionnaire was presented via an online tool. All items had 5-point Likert response scales.

Before interaction: First, we asked for demographic details and measured domain knowledge with self-constructed items. In addition, we presented screenshots of real-world instances of the components featured in the prototype, to ask about prior experience with and attitude towards the respective method (e.g. Amazon recommendations).

During interaction: Whenever participants switched between components, a popup was shown directly in the prototype, asking them about the reason for doing so. To keep this interruption as short as possible, to make it easier to respond, and not to affect the interaction too much, we provided a list of responses (R1–R6), of which at least one had to be selected. If none of the options was applicable, participants were also allowed to enter a reason themselves (R7).[1]

After interaction: Finally, we assessed the general usability by means of the *System Usability Scale* (SUS) [2], and measured satisfaction with the laptops chosen and the perceived difficulty of making this choice using the constructs and items described in [14]. Next, separately for each component, we used established constructs to measure the corresponding usage effort [14], ease of use [21], usefulness [21], perceived control [17], and understandability [17], as well as a self-generated item to ask participants whether they would use it again. Participants could also comment on the component or indicate that they did not use it. In this case, none of the above constructs was shown for this component, but the reason for not using it had to be explained. We logged all actions performed in the prototype.

2.1.4 Participants. We recruited 110 participants on *Prolific*. To ensure quality of the results, prescreening was done based on: first language, success rate on Prolific (≥ 98 %), number of completed studies (≥ 30), and used device (no smartphone or tablet). 101 participants finished the study. Given the average duration of 21.22 minutes (*SD*=6.27), we compensated them with £3.75. To ensure that only participants were included in the analysis who took the interaction with the prototype seriously, we considered only those with at least 5 actions in the system. This left us with $n = 100$, of which 47 participants were female and 2 identified themselves as non-binary. Age ranged from 21 to 75 (*M*=35.36, *SD*=12.60). The majority had a university degree (58.0 %) and were employed or self-employed (69 %). Most participants were from the UK (41.0 %) or the US (32.0 %). Domain knowledge was in an upper medium range (*M*=3.73, *SD*=0.92).

2.2 Results and Discussion

General usability was assessed as *OK*, with a SUS score of 64.88. The difficulty of choosing suitable items was on a moderate level (*M*=2.86, *SD*=1.23), but satisfaction with the final items was high (*M* = 4.25, *SD* = 0.73). The rather average SUS score could be a result of the artificial separation of the tools we did for the specific purpose of this study and the resulting complexity of the interface. Nevertheless, the results for choice satisfaction, one of the most important dimensions of user experience [14], shed a positive light on the support participants received in their decision making.

2.2.1 Prior Experience. Participants had different levels of experience with real-world instances of the four components. However, the results are well in line with the general availability of the methods in today's web: Participants use filtering interfaces very often (*M*=4.58, *SD*=0.82), but recommendation components (*M*=2.81, *SD*=1.29), chatbots (*M*=2.69, *SD*=1.29), and advisors (*M*=2.19, *SD*=1.29), much less frequently. Similarly, they are generally satisfied, with filtering interfaces again being rated most positively (*M*=4.04, *SD*=0.78). Still, recommendations (*M*=2.97, *SD*=1.09), advisors (*M*=2.91, *SD*=1.05), and chatbots (*M*=2.59, *SD*=1.23), are perceived quite positively.

2.2.2 General Usage and Perception. Only few participants used just one component while solving the task (6.0 %). Almost half of them (45.0 %) used all four components at least once, and 26.0 % still three. Detailed results on the usage are reported in Table 1. The upper part shows, among others, that the majority started with the filter component.

Given the above results on prior experience, we suspect that participants initially chose the component they were most familiar with, even if it was not necessarily the most appropriate one. Instead, the component that was used when participants finished the task, more likely was the most suitable one. Although also in this case the filter component was used most often, the differences to the other components were much smaller than at the beginning.

[1] As this was done only 12 times, and the comments were mainly related to minor usability issues or technical problems, we omit a detailed analysis.

Table 1: Statistics and questionnaire results (with *M, SD*). Best values are bold. Group sizes differ as not everyone used all components.

		Filter (n=98)	Rec. (n=84)	Advisor (n=83)	Chatbot (n=73)
Usage statistics	Used at least once	**84.0 %**	71.0 %	82.0 %	69.0 %
	Started with	**54.0 %**	17.0 %	21.0 %	8.0 %
	Ended end	**33.0 %**	25.0 %	26.0 %	16.0 %
Questionnaire constructs	Perceived effort	3.29 (1.15)	3.09 (1.20)	**3.89** (1.06)	3.03 (1.22)
	Ease of use	3.91 (0.98)	3.20 (1.18)	**4.07** (0.90)	2.97 (1.21)
	Usefulness	**3.83** (1.06)	2.81 (1.23)	3.80 (1.17)	2.58 (1.38)
	Perceived control	**4.06** (0.86)	2.77 (1.02)	3.54 (0.97)	2.54 (1.04)
	Understandability	**4.21** (0.81)	3.23 (1.13)	4.13 (0.86)	3.20 (1.76)
	Intent. to use again	**3.93** (1.22)	2.61 (1.23)	**3.93** (1.29)	2.60 (1.41)

Some participants did not make use of certain components at all. For example, they commented that: "I am knowledgeable enough that I do not need recommendations." Others gave justifications based on their common behavior: "Never use it [the advisor] in real life." Moreover, several participants indicated that they in general do not like the way of interacting with certain components. This was particularly the case due to negative experiences they had with chatbots, because these "are unreliable," require "a lot of time typing," and at the end, "get things wrong too often."

The questionnaire results are shown in the lower part of Table 1. For all dependent variables, filter and advisor were rated higher than the other components. With respect to perceived effort and ease of use, the advisor performed best. However, as expected, perceived effort was the only dimension in which the filter component received a much lower score (still higher than the score for recommendations and chatbot). The chatbot always received the lowest scores. Participants missed a predefined structure and "expected it to ask leading questions." Also in terms of control and understandability, advisor and chatbot received the lowest scores, and participants were rather undecided if they would use these automated methods again. Nevertheless, the scores were acceptable, and with the high standard deviations, it appears that there were participants who suffered from the lack of control, and participants who enjoyed the reduced effort. Concerning filter and advisor, however, all participants expressed a very positive opinion.

2.2.3 Switching Behavior. Almost all participants switched between components at least once (94.0 %), which already suggests that providing only a single method to support decision making is usually insufficient. One participant explicitly mentioned it "would be better if you could use all components in combination." Nearly half of them (49.0 %) switched components 1 to 3 times over the entire course of the interaction, 33.0 % did so 4 to 5 times. More than 6 switches were made by 12.0 %, the maximum was 10. Table 2 details the reasons participants had for doing so. Trying out the components was the response that was chosen most frequently (R4). This, however, might be different if the system is used again, since the artificial situation of the study and the first contact with the prototype likely motivated participants to try out all components. This also explains that there were only few participants left who switched components because they (still) had the expectation a different one would be more suitable (R5).

The other reasons participants had for switching between components were in line with the characteristics of the techniques: Participants switched from a component because they did not find a suitable item (R1) more often when the interaction was guided only a little (chatbot) or too much limited (recommendations). If, on the other hand, the interaction was strongly guided (advisor), this reason was selected less frequently. Also the filter component was rarely left for this reason (R1), here, however, because participants knew how to use this component, and were thus able to apply at least simple filter criteria. In contrast to these results, participants were less likely to leave chatbot or recommendations in order to verify the results (R2). Mostly, they selected this reason when they switched from advisor or filter to more automated methods. Apparently, they expected that these components would confirm their selection or help them in a similar way as an employee in a brick-and-mortar store. On the other hand, when participants felt the need to further constrain the result set, they switched to the filter—or to the advisor, if they were currently using the filter (R3). As opposed to R1, the latter two reasons imply that the result set needs to be maintained when switching to other components, which clearly calls for an integrated approach to support decision making.

In contrast to some qualitative feedback reported above, several participants found that they "did not know enough about laptops to truly decide what was good and bad" in the recommendation component. Others "expected [the chatbot] to ask leading questions." Accordingly, participants mostly indicated that they hoped the manual components would be more suitable when they left the recommendation, and, in particular, the chatbot component (R5). While participants with high domain knowledge likely continued to the rich mechanisms of the filtering interface, and those with less expertise to the guided advisor dialog, a quantitative analysis in terms of domain knowledge is still required. Then, we will also take other user characteristics such as personality factors and decision-making style into account.

3 TOWARDS A MULTI-METHOD APPROACH

As discussed in Section 1, each method comes with different advantages, and is therefore more or less suitable for users in their respective situation. The study results reported in Section 2 confirm the benefits of having multiple components available. For a more seamless experience, however, the actions performed in one component should automatically update the state of the others. This would make it possible for users to switch between the methods without losing their progress, and to move continuously towards their goal. Next, we present an initial proposal for integrating the methods more closely with each other. This proposal is conceptual, describing the integration from a user perspective, leaving out specific implementation details.

First, it is necessary to ensure that each component can still be used independently, as not all users want or need to use multiple methods. But, as soon as users start to use filter, advisor, or chatbot, these components mutually influence each other, as shown in Figure 2 (a). In line with our study results, users are thus supported in further constraining the result set whenever they switch to a faceted filtering component, namely, by suggestions of filter values and ranges that were not considered so far (in addition to criteria set earlier based on previous responses to advisor or chatbot).

Table 2: Overview of reasons given for switching between components. Most frequent target components are highlighted bold.

Transitions			Frequency	R1 Nothing found	R2 Verify results	R3 Constrain results	R4 Try out	R5 More suitable	R6 Unintentional	R7 Other
Filter	→	**Rec.**	**52 (47.7 %)**	13.4 %	29.9 %	7.5 %	40.3 %	3.0 %	0.0 %	6.0 %
	→	Advisor	33 (30.3 %)	19.4 %	25.8 %	16.1 %	32.3 %	0.0 %	6.5 %	0.0 %
	→	Chatbot	24 (22.0 %)	10.7 %	28.6 %	7.1 %	50.0 %	3.6 %	0.0 %	0.0 %
Rec.	→	Filter	38 (37.6 %)	25.7 %	17.1 %	5.7 %	28.6 %	11.4 %	5.7 %	5.7 %
	→	**Advisor**	**43 (42.6 %)**	37.5 %	10.4 %	4.2 %	35.4 %	8.3 %	2.1 %	2.1 %
	→	Chatbot	20 (19.8 %)	18.5 %	22.2 %	14.8 %	40.7 %	0.0 %	0.0 %	3.7 %
Advisor	→	**Filter**	**29 (34.1 %)**	14.0 %	16.3 %	14.0 %	44.2 %	2.3 %	7.0 %	2.3 %
	→	**Rec.**	**29 (34.1 %)**	17.9 %	30.8 %	7.7 %	33.3 %	7.7 %	2.6 %	0.0 %
	→	Chatbot	27 (31.8 %)	13.5 %	24.3 %	8.1 %	45.9 %	2.7 %	2.7 %	2.7 %
Chatbot	→	Filter	20 (31.7 %)	37.9 %	3.4 %	10.3 %	20.7 %	20.7 %	3.4 %	3.4 %
	→	**Rec.**	**28 (44.4 %)**	30.8 %	28.2 %	2.6 %	20.5 %	7.7 %	7.7 %	2.6 %
	→	Advisor	15 (23.8 %)	17.4 %	13.0 %	8.7 %	30.4 %	21.7 %	8.7 %	0.0 %

Alternatively, more suitable values and ranges are highlighted automatically. Conversely, based on criteria set by the user in the filter component, the advisor selects the next meaningful question and the chatbot starts (or continues) the conversation accordingly, in this way helping to look for items that appear more suitable than the results obtained with the more specific filter component. Moreover, when switching to advisor or chatbot, users are supported in verifying the current results. This is achieved by highlighting the answer in the advisor dialog that matches best the previously specified filter values, and by enabling the chatbot to answer clarifying questions about the result set.

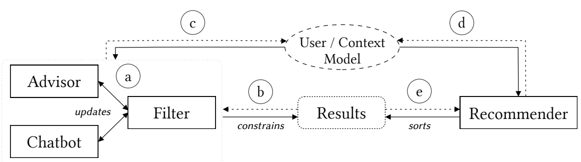

Figure 2: Visualization of the data flow (solid lines) between different components and their mutual effects (dashed lines).

For the integration of these three components as shown in Figure 2, we propose to exploit a knowledge graph, as already done in more specific cases, e.g., question-answering systems [34], search [31], and recommendation [35]. The knowledge graph should represent domain-related concepts and item features as well as weights on these aspects produced by the user's actions. Based on these weights, the result set can be constrained in a generic fashion (b). In turn, the advisor can display information about the number of items remaining for a particular answer, and the chatbot can refer to features of the items the user has clicked on or expressed a preference for. Beyond that, user and context model allow to automatically adapt these components to the individual user based on earlier responses, filter criteria set in previous sessions, or situational factors (c). The same applies to the recommendation component (d), which influences the result set obtained as described above by means of a re-ranking mechanism (e). Besides, it can work in the background to show other recommendations, e.g., on a product detail page, or may be extended by a critiquing mechanism to let users actively influence the recommendations.

The current prototype, however, uses a rule-based approach to realize the integration. Accordingly, as a next step, we plan to fully implement the above approach. Moreover, we will perform a more thorough statistical analysis to uncover possible relationships between personal characteristics such as domain knowledge, and the usage of specific components, in particular, depending on the nature of the task, i.e. goal-driven or explorative. From this, we will investigate how to assist specific users in determining the method to start with or to switch to, e.g., by providing cues about the component that is currently most suitable (and the reason why). Finally, one must note that the present results stem from a specific implementation of the components in a single domain, which calls for further studies to confirm our findings. Nevertheless, with respect to all of the aforementioned aspects, we consider our work a promising first step towards better support of decision making in online shopping.

ACKNOWLEDGMENTS

This work was partially supported by the Eurostars project ACODA (grant no. 01QE1946C).

REFERENCES

[1] A. Bodaghi and E. Homayounvala. 2018. Personalization of Interactive Recommender Systems for Expert Users. In *Proceedings of the 4th International Conference on Web Research (ICWR '18)*. 58–62. https://doi.org/10.1109/ICWR.2018.8387238

[2] John Brooke. 1996. SUS: A "Quick and Dirty" Usability Scale. *Usability Evaluation in Industry* 189 (1996).

[3] Sylvain Castagnos, Nicolas Jones, and Pearl Pu. 2009. Recommenders' Influence on Buyers' Decision Process. In *Proceedings of the 3rd ACM Conference on Recommender Systems (RecSys '09)*. ACM, New York, NY, USA, 361–364. https://doi.org/10.1145/1639714.1639788

[4] Carlos A. Gomez-Uribe and Neil Hunt. 2016. The Netflix Recommender System: Algorithms, Business Value, and Innovation. *ACM Transactions on Management Information Systems* 6, 4 (2016), 13:1–13:19. https://doi.org/10.1145/2843948

[5] Marti A. Hearst. 2009. *Search User Interfaces*. Cambridge University Press. https://doi.org/10.1017/CBO9781139644082

[6] Gerald Häubl and Valerie Trifts. 2000. Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids. *Marketing Science* 19, 1 (2000), 4–21. https://doi.org/10.1287/mksc.19.1.4.15178

[7] Andrea Iovine, Fedelucio Narducci, and Giovanni Semeraro. 2020. Conversational Recommender Systems and Natural Language: A Study Through the ConveRSE Framework. *Decision Support Systems* 131 (2020), 113250. https://doi.org/10.1016/j.dss.2020.113250

[8] Dietmar Jannach, Ahtsham Manzoor, Wanling Cai, and Li Chen. 2021. A Survey on Conversational Recommender Systems. *Comput. Surveys* 54, 5, Article 105 (2021), 36 pages. https://doi.org/10.1145/3453154

[9] Gawesh Jawaheer, Peter Weller, and Patty Kostkova. 2014. Modeling User Preferences in Recommender Systems: A Classification Framework for Explicit and Implicit User Feedback. *ACM Transactions on Interactive Intelligent Systems* 4, 2 (2014), 8:1–8:26. https://doi.org/10.1145/2512208

[10] Michael Jugovac and Dietmar Jannach. 2017. Interacting with Recommenders – Overview and Research Directions. *ACM Transactions on Interactive Intelligent Systems* 7, 3 (2017), 10:1–10:46. https://doi.org/10.1145/3001837

[11] Timm Kleemann, Magdalena Wagner, Benedikt Loepp, and Jürgen Ziegler. 2021. Modeling User Interaction at the Convergence of Filtering Mechanisms, Recommender Algorithms and Advisory Components. In *Proceedings of Mensch & Computer (MuC '21)*. ACM, New York, NY, USA, 499–511. https://doi.org/10.1145/3473856.3473859

[12] Timm Kleemann and Jürgen Ziegler. 2019. Integration of Dialog-Based Product Advisors into Filter Systems. In *Proceedings of Mensch & Computer (MuC '19)*. ACM, New York, NY, USA, 67–77. https://doi.org/10.1145/3340764.3340786

[13] Bart P. Knijnenburg, Niels J.M. Reijmer, and Martijn C. Willemsen. 2011. Each to His Own: How Different Users Call for Different Interaction Methods in Recommender Systems. In *Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11)*. ACM, New York, NY, USA, 141–148. https://doi.org/10.1145/2043932.2043960

[14] Bart P. Knijnenburg, Martijn C. Willemsen, and Alfred Kobsa. 2011. A Pragmatic Procedure to Support the User-Centric Evaluation of Recommender Systems. In *Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11)*. ACM, New York, NY, USA, 321–324. https://doi.org/10.1145/2043932.2043993

[15] Joseph A. Konstan and John Riedl. 2012. Recommender Systems: From Algorithms to User Experience. *User Modeling and User-Adapted Interaction* 22, 1-2 (2012), 101–123. https://doi.org/10.1007/s11257-011-9112-x

[16] Jonathan Koren, Yi Zhang, and Xue Liu. 2008. Personalized Interactive Faceted Search. In *Proceedings of the 17th International Conference on World Wide Web (WWW '08)*. ACM, New York, NY, USA, 477–486. https://doi.org/10.1145/1367497.1367562

[17] Yu Liang and Martijn C. Willemsen. 2021. Interactive Music Genre Exploration with Visualization and Mood Control. In *Proceedings of the 26th International Conference on Intelligent User Interfaces (IUI '21)*. ACM, New York, NY, USA, 175–185. https://doi.org/10.1145/3397481.3450700

[18] Benedikt Loepp, Tim Donkers, Timm Kleemann, and Jürgen Ziegler. 2018. Impact of Item Consumption on Assessment of Recommendations in User Studies. In *Proceedings of the 12th ACM Conference on Recommender Systems (RecSys '18)*. ACM, New York, NY, USA, 49–53. https://doi.org/10.1145/3240323.3240375

[19] Benedikt Loepp, Tim Donkers, Timm Kleemann, and Jürgen Ziegler. 2019. Interactive Recommending with Tag-Enhanced Matrix Factorization (TagMF). *International Journal of Human-Computer Studies* 121 (2019), 21–41. https://doi.org/10.1016/j.ijhcs.2018.05.002

[20] Benedikt Loepp and Jürgen Ziegler. 2019. Towards Interactive Recommending in Model-Based Collaborative Filtering Systems. In *Proceedings of the 13th ACM Conference on Recommender Systems (RecSys '19)*. ACM, New York, NY, USA, 546–547. https://doi.org/10.1145/3298689.3346949

[21] Pearl Pu, Li Chen, and Rong Hu. 2011. A User-Centric Evaluation Framework for Recommender Systems. In *Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11)*. ACM, New York, NY, USA, 157–164. https://doi.org/10.1145/2043932.2043962

[22] Pearl Pu, Li Chen, and Rong Hu. 2012. Evaluating Recommender Systems from the User's Perspective: Survey of the State of the Art. *User Modeling and User-Adapted Interaction* 22, 4-5 (2012), 317–355. https://doi.org/10.1007/s11257-011-9115-7

[23] Kiran Ramesh, Surya Ravishankaran, Abhishek Joshi, and K. Chandrasekaran. 2017. A Survey of Design Techniques for Conversational Agents. In *Proceedings of the 2nd International Conference on Information, Communication and Computing Technology (ICICCT '17)*. Springer Singepore, Singapore, 336–350. https://doi.org/10.1007/978-981-10-6544-6_31

[24] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. *Recommender Systems Handbook*. Springer US, Boston, MA, USA, Chapter Recommender Systems: Introduction and Challenges, 1–34.

[25] Giovanni Maria Sacco and Yannis Tzitzikas. 2009. *Dynamic Taxonomies and Faceted Search: Theory, Practice, and Experience*. Vol. 25. Springer Science & Business Media. https://doi.org/doi.org/10.1007/978-3-642-02359-0

[26] James Schaffer, James Humann, John O'Donovan, and Tobias Höllerer. 2020. *Contemporary Research: Models, Methodologies, and Measures in Distributed Team Cognition*. CRC Press, Boca Raton, FL, USA, Chapter Quantitative Modeling of Dynamic Human-Agent Cognition, 137–186.

[27] Sylvain Senecal and Jacques Nantel. 2004. The Influence of Online Product Recommendations on Consumers' Online Choices. *Journal of Retailing* 80, 2 (2004), 159–169. https://doi.org/10.1016/j.jretai.2004.04.001

[28] Brent Smith and Greg Linden. 2017. Two Decades of Recommender Systems at Amazon.com. *IEEE Internet Computing* 21, 3 (2017), 12–18. https://doi.org/10.1109/MIC.2017.72

[29] Yueming Sun and Yi Zhang. 2018. Conversational Recommender System. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*. ACM, New York, NY, USA, 235–244. https://doi.org/10.1145/3209978.3210002

[30] Preeti Virdi, Arti D. Kalro, and Dinesh Sharma. 2020. Online Decision Aids: The Role of Decision-Making Styles and Decision-Making Stages. *International Journal of Retail & Distribution Management* 48, 6 (2020), 555–574. https://doi.org/10.1108/IJRDM-02-2019-0068

[31] Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. 2018. DKN: Deep Knowledge-Aware Network for News Recommendation. In *Proceedings of the 27th World Wide Web Conference (WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1835–1844. https://doi.org/10.1145/3178876.3186175

[32] Joseph Weizenbaum. 1966. ELIZA–A Computer Program for the Study of Natural Language Communication between Man and Machine. *Commun. ACM* 9, 1 (1966), 36–45. https://doi.org/10.1145/365153.365168

[33] Ka-Ping Yee, Kirsten Swearingen, Kevin Li, and Marti Hearst. 2003. Faceted Metadata for Image Search and Browsing. In *Proceedings of the 20th SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, ACM, New York, NY, USA, 401–408. https://doi.org/10.1145/642611.642681

[34] Wen-tau Yih, Ming-Wei Chang, Xiaodong He, and Jianfeng Gao. 2015. Semantic Parsing via Staged Query Graph Generation: Question Answering with Knowledge Base. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (IJCNLP '15)*. Association for Computational Linguistics, Beijing, China, 1321–1331. https://doi.org/10.3115/v1/P15-1128

[35] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. 2016. Collaborative Knowledge Base Embedding for Recommender Systems. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. ACM, New York, NY, USA, 353–362. https://doi.org/10.1145/2939672.2939673

[36] Yongfeng Zhang, Xu Chen, Qingyao Ai, Liu Yang, and W. Bruce Croft. 2018. Towards Conversational Search and Recommendation: System Ask, User Respond. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM '18)*. ACM, New York, NY, USA, 177–186. https://doi.org/10.1145/3269206.3271776