Università degli Studi di Torino

DIPARTIMENTO DI INFORMATICA
Dottorato di Ricerca in Informatica
Ciclo: 36

Tesi di Dottorato

A Holistic Approach to Services Recommender Systems: Service Integration and Multimodal Justification

Tesi presentata da:
Zhongli Filippo Hu
Matricola 812739

Tutor:
Prof.ssa Liliana Ardissono

Coordinatrice del dottorato:
Prof.ssa Viviana Patti

Anni Accademici 2020 – 2023
Settore Scientifico-Disciplinare di Afferenza: INF/01 INFORMATICA

Colophon

I declare that the undersigned as well as the author of document is responsible for its content, and for parts taken from other works, these are expressly declared citing the sources.

This work is made with LaTeX.

The trade names and registered trademarks mentioned in the thesis belong to their respective owners.

Contact:

Zhongli Filippo Hu – zhonglifilippo.hu@unito.it

Acknowledgments

I would like to acknowledge the support of ChatGPT, an AI language model by OpenAI, and Grammarly, for their assistance in grammar and language refinement in this work. While the entirety of the content and research is my own, these two instruments enhanced the linguistic preciseness of this document.

Contents

Abstract

This thesis addresses the evolving landscape of recommender systems in the context of artificial intelligence and data-driven decision-making. Focusing on the integration of service models and multimodal information, this research aims to enhance user decision-making and experience with recommender systems. Traditional recommender systems, primarily centered on item-centric metrics, often overlook the context of user experiences and service interactions. In response, this thesis proposes a novel approach that incorporates service-based integration and the utilization of multimodal data, particularly images, to provide a holistic and user-centric recommendation process.

The methodology adopted includes the development of a new recommender system model, the justification of the results it generates, and its empirical validation through a series of user studies. These models integrate service-oriented aspects and multimodal data, moving beyond conventional textual and quantitative data to include image analysis for a richer understanding of items. The research also explores the impact of these advancements on user awareness, satisfaction, and trust.

Key findings indicate that integrating service models and multimodal information significantly enhances the quality and relevance of recommendations, contributing to improved user satisfaction and trust in the system.

This thesis contributes to the field of computer science by advancing recommender systems through service-based integration and multimodal information, offering a new paradigm for user-centric and explainable AI in decision-making systems.

Chapter 1

Introduction

In the evolving landscape of artificial intelligence and data-driven decision-making, recommender systems significantly influence user decisions in a multitude of sectors ranging from e-commerce to home-booking. Traditionally centered on item-centric metrics like ratings and features, these systems have efficiently navigated users through various options. Yet, this efficiency comes at the cost of a limited lens that often misses the broader context of user experiences and service interactions.

Recently, the demand for predictable and accountable AI has intensified, especially in light of regulations like the European General Data Protection Regulation, which underscores the need for transparency in intelligent systems. This regulatory landscape has transformed recommender systems, urging a shift from mere algorithmic performance to a more holistic approach encompassing the user's entire journey with an item. This includes not only the tangible attributes of products but also the often-overlooked experiential elements.

Furthermore, current explanation and justification models in recommender systems predominantly focus on textual and quantitative ratings, neglecting the rich information images can provide about an item's various experiential aspects. Recognizing this gap, our research extends these models by integrating multimodal information, including object recognition in images, to offer a more nuanced, service-oriented presentation of items. This integration allows users to filter and compare recommendations based on detailed evaluation dimensions, enhancing their decision-making process and overall experience with the system.

1.1 Main goals of the research

This thesis aims to advance the recommender systems field by integrating service models and multimodal information, thereby enhancing user decision-making and overall experience. The research aims to:

1. Develop a novel family of recommender systems that not only consider item properties but also the holistic consumer experience in item fruition stages.

2. Incorporate the stages of item fruition into the justification models of recommender systems, explicitly showcasing to users the diverse feedback and sentiment associated with each stage, thereby offering a more comprehensive understanding of each recommendation.

3. Extend traditional explanation and justification models by incorporating multimodal data, including images, to present a more comprehensive view of items.

4. Investigate how these advancements in recommender systems and their presentation can improve user awareness, satisfaction, and trust in the recommendations provided.

The concept of *"item fruition stages"* captures the comprehensive lifecycle of user interaction with a service, encompassing various dimensions of the experience. For example, within the restaurant domain, a patron might enjoy the ambiance and the quality of service provided by the waitstaff, yet find the culinary offerings lacking. By acknowledging and integrating these nuanced stages into our model, we aim not only to improve the precision of our recommendations but also to ensure their relevance and personal resonance for each user, thereby significantly enhancing the overall user experience.

1.2 Research Questions

Building upon the main goals, this thesis seeks to explore the following overarching research questions:

1. **RQ1:** How does the integration of a service-based representation of items, which explicitly models the stages of item consumption, impact the quality of recommendations, compared to systems that rely solely on local item properties and overall ratings?

2. **RQ2:** How does service-based justification of recommendations influence user awareness and confidence in evaluating these recommendations, and what is its effect on user satisfaction regarding the presentation of item-related information in recommender systems?

3. **RQ3:** How does a multimodal service-based presentation of images and textual data, possibly enhanced by keyword filtering, impact the effectiveness of the recommendation comparison process against traditional, non-stage-specific presentations?

4. **RQ4:** How does image-based filtering, considering various levels of detail, contribute to the effectiveness of information filtering in item lists, and in what ways does it impact user awareness and facilitate decision-making processes regarding item selections?

This thesis builds upon preliminary work (Hu, 2022), where a novel approach to service-aware recommendation systems was introduced. To make them service-aware, we have incorporated service modeling techniques. Specifically, our approach involves the application of Service Journey Maps (Richardson, 2010) and Service Blueprints (Bitner et al., 2008) to model the service underlying the interaction with items, from their exploration to post-sales. These models serve as tools for understanding and capturing the service processes. Service Journey Maps allow us to visualize the end-to-end experiences of users as they interact with various service stages. Similarly, Service Blueprints provide a detailed diagrammatic representation of the service process, highlighting the relationships between different service components, including front-stage (customer-facing) and back-stage (operational) activities. More details in Chapter 3.

1.3 Methodology overview

To achieve the goals outlined, our approach involves the development of a novel recommender system and three distinct justification models. Each justification model is designed to tackle unique aspects of user interaction and decision-making processes within recommender systems, offering a multifaceted understanding of how different types of explanations can impact user satisfaction and trust. Specifically, the justification models vary in their approach to integrating service-based information, multimodal data, and user feedback mechanisms, thereby allowing us to comprehensively evaluate the effectiveness of each strategy in enhancing the recommendation experience.

To validate the effectiveness of our proposed recommender system and the justification models, we have designed four user studies, each corresponding to the research presented in Chapters 6 through 8. These user studies are crafted to test the hypotheses in varied contexts and user scenarios. The first study focuses on the initial acceptance of service-based recommendations, and the second one focuses on the refinement of justification models, integrating fine-grained evaluation dimensions. The third study explores the integration of multimodal information, and the fourth study focuses on image-based information filtering.

Each of these user studies plays a fundamental role in empirically validating the effectiveness of our proposed models and methodologies. By methodically examining the impacts of service-based representations, justification models, and multimodal explanations, these studies collectively contribute to our understanding of enhancing user satisfaction in recommender systems.

The visualization models we designed to support justification take inspiration from guidelines of faceted user interfaces (Hearst, 2006) and adhere to Shneiderman (1992)'s guiding principle of "Overview first, zoom and filter, and details on demand."

A key aspect of our research involved logging and analyzing user interactions with the justification models. By examining these interactions, we gained valuable insights into user behavior, decision-making processes, and overall engagement with the system.

Consistent with research indicating that the effectiveness of explanations in recommender systems varies with users' cognitive styles, personality, and domain knowledge (Millecamp et al., 2019, 2020; Kouki et al., 2019), in the user studies we carried out, we exploited validated questionnaires to evaluate user's need for cognition (Coelho et al., 2020) and curiosity levels (Kashdan et al., 2009). Moreover, we elicited users's experience with online booking platforms. This information allowed us to correlate the results with personality traits.

Through online user studies, we validated the proposed recommendation and justification models to retrieve information about users' experience and satisfaction with the offered information exploration and item selection support. All the user studies we carried out have been approved by the Ethics Committee of the University of Torino (Protocol Number: 0421424).

1.4 Thesis structure

This thesis is structured to methodically explore the integration of service models and multimodal information in recommender systems. Below is an overview of the contributions of each chapter.

In **Chapter 2: Background and Related Work on Recommender Systems**, the thesis lays a comprehensive groundwork, beginning with an exploration of recommender systems. This chapter navigates through various facets of these systems, starting from an introduction to recommender systems, delving into the critical data sources such as items, users, and interactions, and examining the spectrum of recommendation techniques, including collaborative filtering, content-based, and review-based systems. It also explores the emerging role of image analysis in recommender systems, leading up to a discussion on explanation and justification methods.

Moving forward, **Chapter 3: Background on Service Models** shifts the focus to two well-known and largely applied service models, i.e., Service

Journey Maps and Service Blueprints. The chapter describes these models, providing a crucial framework for the rest of the research.

Chapter 4: Dataset outlines the dataset used in the research, to validate the proposed service-based and multimodal recommender systems.

Chapter 5: Preliminary Experiment presents the preliminary study to evaluate the effectiveness of our proposed visual model in aiding decision-making processes within the home recommendation systems. By integrating both quantitative and qualitative data in a layered mockup presentation, this experiment serves as an essential step in assessing how our service-based justification model impacts user choices.

In **Chapter 6: Integration of Service Model in Recommender Systems**, the thesis introduces the innovative concept of service-aware recommender systems. This chapter is dedicated to exploring how integrating service models into recommender systems can not only enrich the quality of recommendations but also significantly enhance the user experience.

In **Chapter 7: Enhancing the Justification of Results in Service-aware Recommender Systems**, we delve into the refinement of justification models within recommender systems. This chapter is dedicated to integrating fine-grained evaluation dimensions, aiming to enrich the user's comprehension and interaction with the recommendations.

Moving forward, **Chapter 8: Multimodal Interfaces** explores the integration of multimodal information, particularly images, into service-based justification models. This chapter highlights the significant role of visual elements in enhancing user experience and decision-making processes. The second part explores using images as a primary element in information filtering and user interface design within recommender systems. Focusing on the approach of utilizing recognized scenes to categorize images and user feedback, we examine how image-based information filtering can improve the justification of experience goods.

Chapter 9: Lessons Learned summarizes the insights and significant understandings derived from the user studies, while **Chapter 10: Conclusion** provides a comprehensive conclusion, weaving together the thesis's contributions, limitations, and future research directions.

Chapter 2

Background and related work on recommender systems

2.1 Premises

The landscape of recommender systems has evolved significantly, with advances in technology and methodology continuously reshaping how users interact with digital platforms.

Historically, early recommender systems were largely opaque, prioritizing algorithmic performance over user comprehension. This “black-box” approach often left users, developers, and regulators in the dark about the rationale behind the system’s recommendations. Recognizing the limitations of this approach, Herlocker et al. (2000) pioneered the importance of explicability in recommender systems. They pointed out that explaining the results of these systems is crucial for user acceptance and trust.

Since then, there has been a concerted effort in the research community to enhance the transparency and understandability of personalized recommendations. Key contributions in this area include those by Nunes and Jannach

(2017); Tintarev and Masthoff (2012, 2022); Jannach et al. (2019) among others. These studies helped to better understand how explanations can be integrated into recommender systems.

This chapter sets the stage for our research by exploring the development and current state of recommender systems, particularly focusing on aspects most pertinent to our thesis objectives. Understanding this background is crucial because it provides a comprehensive overview of the methodologies that have historically driven the field, including the strengths and limitations of existing models. This context is indispensable as we aim to develop a novel family of recommender systems by incorporating the service model.

Lately, we have explored the roles of image analysis and the different formats of recommendation explanations within recommender systems. This investigation set the groundwork for the justification models we propose.

2.2 Introduction on recommender systems

Recommender systems (RSs) are sophisticated software tools and techniques which provide suggestions for items aligned with users interests. These systems are helpful to various decision-making processes, in different domains. For instance, in e-commerce, they help customers find products tailored to their preferences; in entertainment, such as streaming services, they suggest movies or music based on past choices; and in information services, they aid users in discovering relevant articles or news content (Ricci et al., 2022).

These systems play a crucial role in assisting individuals who may find it overwhelming to navigate through a vast array of items offered by websites, such as e-commerce platforms. For instance, Amazon.com employs an RS to personalize the shopping experience for each customer, showcasing the power of recommendation in the online retail space (Linden et al., 2003; Smith and Linden, 2017).

Recommender systems have evolved from simple personalized suggestions, often provided as ranked lists, to more sophisticated methods that predict suitable products or services based on user preferences and constraints.

With the exponential growth of information and options available online, RSs have become essential in aiding users to avoid cognitive overload, which can lead to decision paralysis and diminished satisfaction (Schwartz, 2004). By guiding users toward new or relevant items, RSs help mitigate the information overload problem, using data about users, available items, and past interactions.

2.3 Data sources of the algorithms

Recommender systems (RSs) are complex information processing systems that actively gather a variety of data to construct recommendations. The data utilized in RSs usually pertains to three primary entities: items, users, and the interactions between them.

2.3.1 Items

In recommender systems, we refer to the entities being recommended as "items". They can range from tangible products to digital content. The data associated with these items is crucial and varies based on its accessibility and the nature of the items themselves. For instance, image recommendation require specialized image analysis algorithms for feature extraction from their raw content. This is essential for systems that recommend multimedia content, as it allows for a deeper understanding and classification of visual data (Lops et al., 2011).

Text-based items, including news articles and product reviews, present a different set of challenges. They necessitate the use of natural language processing (NLP) techniques to effectively parse and extract meaningful insights from textual content.

Furthermore, the advent of semantic-aware systems has markedly enhanced the capabilities of RSs in handling textual content. Unlike early models that struggled with a comprehensive understanding of text and semantic relationships, modern semantic-aware systems can interpret and give meaning to natural language. This deeper comprehension of textual content significantly improves the effectiveness of RSs. Two primary approaches are utilized in semantic representation: endogenous methods, which focus on keyword distribution within documents, and exogenous methods that use external knowledge sources like taxonomies and ontologies to address challenges of synonymy and polysemy in language (Musto et al., 2022).

These advancements in processing item data not only enhance the accuracy of recommendations but also help in addressing the cold-start problem, where new items and users lack sufficient interaction data. By extracting features from reviews and other descriptive content, RSs can better understand and recommend new or less-known items (Perano et al., 2021).

2.3.2 Users

Users are central to the functioning of recommender systems, and each user has unique characteristics. This information constitutes what is known as the user model, a structured representation that encapsulates the user's preferences and needs (Fischer, 2000).

User models in RSs are diverse, reflecting the variety of user modeling approaches employed across different systems. RSs can be seen as tools that generate recommendations by constructing and leveraging these user models (Berkovsky et al., 2008). The choice of data included in the model is influenced by a lot of factors: the specific recommendation technique used, the availability of user data, and the practicality of finding and processing this information.

For instance, in collaborative filtering systems, usually, user models are the matrix $\mathcal{U} \times \mathcal{I}$ with $u \in \mathcal{U}$ and $i \in \mathcal{I}$, where each cell r_{ui} has the rating that user u gave to the item i. This approach focuses on gauging user preferences based on their interactions with specific items (Herlocker et al., 2000). In contrast, context-aware RSs enrich the user model with contextual information, enhancing the relevance of recommendations in specific situations.

Content-based RSs, on the other hand, model users based on the content features of items they have interacted with. This could include a range of structured and unstructured data, like textual user-generated reviews, to provide a richer, more detailed user profile (Perano et al., 2021).

In summary, the complexity of user models in RSs are dictated by the nature of the RS, the type of content it deals with, and the personalization level of the recommendations.

2.3.3 Interactions

Interactions between users and items are a critical component of recommender systems. These systems record the interactions in log-like data structures, capturing a wealth of information vital for generating accurate recommendations. This data encompasses various elements of user-item engagement, ranging from the items purchased to the sequence of actions leading up to a purchase, such as browsing and adding items to a basket. It also includes explicit user feedback, like ratings, and implicit feedback inferred from user behaviors.

Explicit feedback, generally more reliable than implicit feedback, conveys a user's level of preference for an item. This can manifest as numerical ratings (like the 1–5 stars on Amazon.com), ordinal ratings (choices like "strongly agree" to "strongly disagree"), or binary ratings (up or down thumbs).

However, a significant challenge with this type of feedback is the inherent sparsity of the data. In a typical recommender system, where there is often a large number of items, it is improbable that every user has interacted with and consequently provided feedback on all items. If we consider the previous user-item matrix $\mathcal{U} \times \mathcal{I}$, the majority of the cells in this matrix remain empty, illustrating the sparsity problem. This scenario is common in real-world systems, where the sparsity of the user-item space poses challenges in effectively generating accurate recommendations.

Implicit feedback, on the other hand, is derived from the user's interactions with items, like clicking, purchasing, or adding to a basket. While this type of feedback is more abundant, it typically provides a weaker signal of preference, as it does not explicitly indicate the user's opinion about the item. Moreover, the lack of interaction with an item should not be interpreted as negative feedback, for the same reason described before for the sparsity problem: in a real-world scenario a user can not interact with all the items.

In context-aware RSs, the system takes into account not only the preferences of the users (as persistent information), but dynamically take into account contextual factors like time, location, and mood, modeling the changing needs of users.

2.4 Recommendation techniques

The primary function of a recommender system is to suggest items of potential interest to users. To reach this goal, an RS needs to estimate the utility of various items. This process typically involves predicting or comparing the perceived benefits of different items to the user.

Various RS types exist, each differing in their application domains and the algorithm used for utility prediction.

2.4.1 Collaborative filtering

Collaborative Filtering is a popular technique in recommender systems, relying on the collection and analysis of user feedback, such as ratings, to make recommendations (Ning et al., 2015; Koren and Bell, 2015). Collaborative Filtering can be further divided into two main types: User-Based and Item-Based.

- **User-Based Collaborative Filtering:** This approach employs the concept of similarity among users to make recommendations and recommends items based on the preferences of similar users. It operates on the assumption that if users A and B rated items similarly in the past, they will have similar tastes in the future (Herlocker et al., 2000).

- **Item-Based Collaborative Filtering:** Contrary to user-based, this method focuses on the similarity between items. If a user likes an item, the system recommends items similar to it, based on user ratings (Sarwar et al., 2001). In other words, this method leverages item similarity based on user interactions.

Both approaches are supported by the concept of “people-to-people cor-

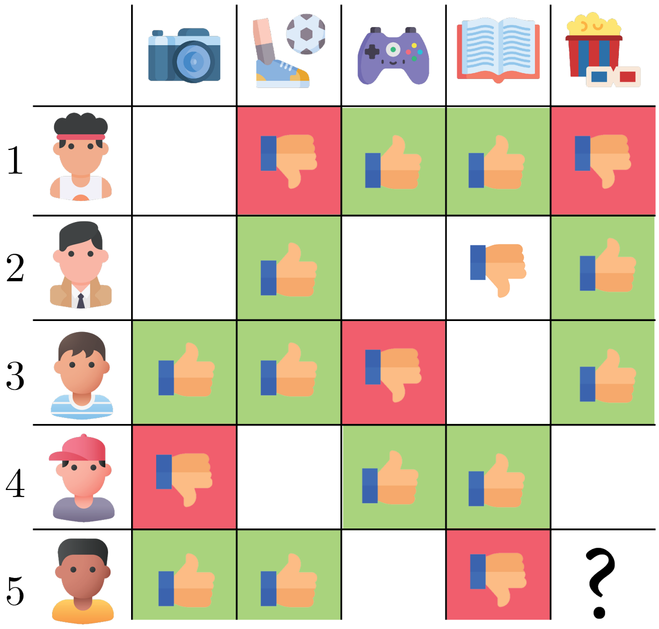


Figure 2.1: Example of prediction of user's rating with a Collaborative Filtering. The system forecasts a user's rating for an item they haven't yet rated, basing these predictions on the ratings provided by other users who have shown similar preferences to the active user. For example, in this particular scenario, the system has predicted that the active user is likely to not favor the video, drawing on the rating patterns of users with comparable tastes.

relation" in Collaborative Filtering, where the essence of recommendation lies in understanding and leveraging the patterns in user-item interactions. Collaborative filtering's popularity in recommender systems is largely due to its ability to provide personalized suggestions. This is achieved by modeling user preferences either directly (in user-based methods) or indirectly (in item-based methods) through the analysis of interaction patterns.

One of the most crucial decisions impacting both the accuracy of rating predictions and the computational efficiency in recommender systems is the selection between user-based and item-based methods. In recommender systems where the number of users surpasses the quantity of items, opting for item-based methods is often more advantageous, for higher accuracy in recommendations, better computational efficiency, and reduced need for frequent updates. Conversely, user-based approaches are known for generating more unique and novel recommendations, potentially enhancing the overall user experience by offering unexpected but satisfying choices(Ekstrand et al., 2014).

However, as told in 2.3.3, Collaborative Filtering is not without its challenges, such as the sparsity of user-item interaction data and the cold-start problem, where new users or new items have insufficient interaction data. To alleviate these issues, dimensionality reduction and graph-based methods were used.

Dimensionality reduction techniques, for instance, seek to condense the user-item matrix into a more manageable form, capturing the most significant features of users and items. A popular technique is matrix factorization, such as Singular Value Decomposition (Koren, 2008, 2009).

Graph-based methods will be explained in Section 2.4.5.

Recent developments in Collaborative Filtering have embraced deep learning, leveraging its capability to handle various data types and structures.

2.4.2 Content-based recommender systems

Content-Based Recommender Systems recommend items by analyzing the content of the items and the user's preferences (Ricci et al., 2022). In particular, the focus is on suggesting items to users based on the similarity of these items to those the user has shown a preference for in the past.

This system calculates the similarity by analyzing the features associated with the items. For instance, if a user has favorably rated films within the comedy genre, the system adapts to recommend more movies from this category (Lops et al., 2011; Pazzani and Billsus, 2007).

The foundation of traditional content-based recommendation lies in aligning the characteristics of the user's profile with the attributes of various items. Typically, these item attributes are identified through keywords extracted from their descriptions. However, to address the limitations inherent in simple keyword-based analysis, semantic indexing techniques have been developed. These techniques use conceptual representations for both items and user profiles, moving beyond mere keywords.

Semantic indexing is broadly categorized into two approaches: exogenous and endogenous. Exogenous techniques incorporate external knowledge sources, such as ontologies, encyclopedic databases like Wikipedia, and information from the Linked Data cloud. Endogenous techniques, conversely, employ a more nuanced semantic representation, positing that the meanings of words are shaped by their contextual usage across extensive textual data sets (Musto et al., 2022).

Content-Based Recommender Systems has some advantages compared to collaborative filtering methods. One of their primary strengths is user independence. Unlike collaborative filtering which requires interactions from various users to find "nearest neighbors", content-based RSs use only the ratings from the user to construct his/her profile.

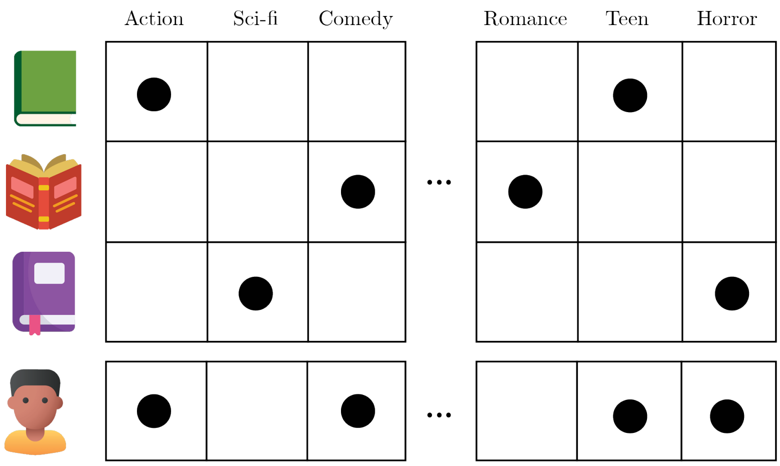
Another advantage of Content-based RSs is their inherent transparency. They can offer users clear explanations for recommendations by linking them to specific content features that led to the item's inclusion in the recommendation list. In contrast, collaborative systems, especially with matrix-factorization or deep learning methods, often lack this clarity.

One of the noteworthy strengths of these models lies in their ability to accurately capture the unique and specific interests of an individual user. Unlike other types of systems, it possesses the capability to recommend niche items: these items might be of interest to a relatively small user base, often ignored by broader recommendation algorithms. If the system identifies that a user has a preference for a certain type or genre of items, particularly those that are not widely popular, it will then focus on suggesting similar items within that same genre. This targeted approach ensures that users receive recommendations that align closely with their specific tastes, even if those tastes are niche or less common, and facilitates the discovery of less mainstream or popular content.

Content-based RSs are also useful in the cold-start problem for new items. They can handle the new items that no user has yet rated. In Collaborative filtering, all the recommended items are based on existing user interactions.

However, Content-Based RSs are not without their limitations. The quality of recommendations in this type of RS is linked to the number and type of features associated with the items. If descriptive features of the items are lacking, these systems struggle to make appropriate suggestions.

The lack of diversity in suggestions is another notable limitation. The



	Action	Sci-fi	Comedy		Romance	Teen	Horror
	●					●	
			●	…	●		
		●					●
	●		●	…		●	●

Figure 2.2: Example of a Content-Based Recommender System: Here, the user demonstrates a liking for specific features in content. The system, recognizing these preferences, suggests a first item that closely matches the user's favored features, illustrating the system's ability to provide personalized recommendations based on individual tastes.

system tends to recommend items similar to those already rated by the user. This is not a big problem if the user has rated a lot of different items, but if the user has rated only one type of item the suggestions will not help to discover new types of items.

Although they are useful for handling new items, Content-based RSs also suffer from the cold start problem for the new users. They require a sufficient number of ratings to understand user preferences accurately. In cases where few ratings are available, such as with new users, Content-Based RSs struggle to provide reliable recommendations.

2.4.3 Review-based recommender systems

Review-based recommender systems have emerged as a significant development in the field, harnessing the power of consumer feedback from online reviews (Chen et al., 2015). These systems delve into the rich semantic content of user-generated reviews, which offer a more detailed and nuanced understanding of user preferences than mere ratings or implicit interactions. Users often provide explanations for their ratings and express opinions, making reviews a source of information for the next customers.

In e-commerce platforms, where the practice of writing reviews is highly encouraged, these systems capitalize on textual data to offer personalized recommendations. Recent trends have seen a surge in applying deep learning techniques to efficiently process and interpret these reviews(Zheng et al., 2017).

Despite these advancements, review-based systems often struggle with contextualizing data in relation to the various stages of item fruition. As pointed out in Margaris et al. (2020), these systems face challenges in managing and interpreting the vast and heterogeneous aspects contained in reviews. The complexity arises in aligning user feedback with different stages of their experience with an item, from initial use to long-term satisfaction.

Recognizing that users may focus on different aspects of an item, Guan et al. (2019) designed an aspect-aware method. This approach extracts various aspects from reviews and employs an attention network to dynamically determine the relevance of each aspect. Such a method is particularly useful in scenarios like restaurant recommendations, where tastes, locations, or ambiances might be of varying importance to different users.

2.4.4 Graph-based recommender systems

Graph-Based Recommender Systems, such as those by Amal et al. (2019), Wang et al. (2018), and Musto et al. (2019), utilize graph structures to model the relationships between users and items. By representing users and items as nodes and their interactions as edges, graph-based RSs can effectively capture complex relations and dependencies. This approach is particularly powerful in uncovering hidden patterns and connections within the data, leading to more insightful recommendations.

Knowledge-aware recommendation systems leverage structured knowledge bases (e.g., knowledge graphs) to enhance the recommendation process. These systems can provide more accurate, explainable, and content-rich recommendations by understanding the relationships between items beyond traditional user-item interactions. Recent developments in KARS have shown significant improvements in recommendation performance, especially in addressing challenges such as cold start and data sparsity (Wang et al., 2023).

Recent methods based on text embeddings leverage natural language processing techniques to understand and utilize the textual content associated with items, such as descriptions or reviews, for improving recommendation quality. Graph embedding techniques, on the other hand, encode the structural information of graphs (e.g., user-item interaction networks or knowledge graphs) into low-dimensional vectors. These embeddings capture complex relationships and can be used to enhance recommendation algorithms (Zhang et al., 2020).

2.4.5 Other methods

In addition to the conventional content-based and collaborative filtering approaches, recommender systems employ a variety of other methods.

Demographic Recommender Systems generate recommendations based on the demographic profile of the user, such as age, gender, location, or language. The underlying assumption is that users within certain demographic groups will have similar preferences, thus recommendations are tailored to these groups. For instance, a website might present different content to users based on their country or customize suggestions according to the user's age group.

Community-Based Recommender Systems work on the principle of social influence and homophily, which is the tendency for individuals to associate and bond with similar others. The concept, central to homophily theory, posits that people are more likely to trust and be influenced by those within their social circle (McPherson et al., 2001). In the context of recommender systems, this translates to prioritizing recommendations based on the preferences and behaviors of a user's friends or social connections. The assumption is that suggestions from friends, who likely share similar tastes and interests due to homophily, will be more relevant and persuasive than those from unknown individuals (Golbeck and Hendler, 2006). This approach has gained traction with the proliferation of social networks, offering a rich dataset of user relationships and interactions to enhance recommendation accuracy.

Hybrid systems combine the strengths of different recommendation techniques to mitigate their individual weaknesses and improve the accuracy (Di Sciascio et al., 2019; Cardoso et al., 2019). For example, they might blend collaborative filtering and content-based methods to counter the new-item problem, where items without ratings are difficult to recommend. The integration of multiple techniques in hybrid systems enables them to be more versatile and effective across various scenarios (Burke, 2002). The advent of deep learning has further enhanced the capabilities of hybrid RSs, allowing for more sophisticated and adaptive recommendation strategies.

Multi-modal recommendation systems incorporate various types of data (e.g., text, images, audio) to make recommendations. These systems can provide a richer understanding of items and user preferences, leading to more personalized and accurate recommendations. Recent attempts in this area have explored how to effectively fuse information from different modalities to improve the recommendation process (Zhou, 2023).

Multi-Criteria Recommender Systems expand the traditional single-criterion approach to incorporate multiple dimensions of user preferences. These systems acknowledge that user satisfaction often hinges on a complex set of factors rather than a singular metric. By integrating multiple criteria into the recommendation process, these systems align closely with the multifaceted nature of user preferences (Adomavicius et al., 2011).

2.5 Image analysis in recommender systems

In the field of recommender systems, the integration of image analysis has opened new avenues for multimodal information filtering. This approach aligns with the concept of faceted exploration, where various aspects or "facets" of items are extracted and utilized to refine recommendations (Hearst et al., 2002; Mauro et al., 2020b). The application of image analysis in recommender systems is multifaceted and varied, ranging from feature extraction in fashion to ingredient identification in culinary contexts.

One notable application is in the domain of fashion, where systems like MMFashion analyze images to extract detailed features of clothing items

(Liu et al., 2021). This analysis allows for the creation of more nuanced and tailored recommendations, catering to the specific style preferences of users Geninatti Cossatin et al. (2023, 2024). Similarly, in the culinary world, recommender systems employ image analysis to identify ingredients in food recipes, offering suggestions based on visual cues (Kawano et al., 2013).

Beyond the extraction of item features, image analysis is also employed in building user profiles. By analyzing the images that users interact with or prefer, recommender systems can develop a deeper understanding of user preferences and behaviors (Kitamura and Itoh, 2018). This approach is particularly effective in domains where visual characteristics play a crucial role, such as in fashion or art.

Furthermore, the technology is instrumental in identifying sets of similar items. For instance, in fashion recommender systems, image analysis is used to find clothing items that are visually similar to those selected by the user or to suggest pairings (Chakraborty et al., 2021; Deldjoo et al., 2022). This capability significantly enhances the personalization aspect of recommendations, providing users with options that closely match their visual preferences.

Recommender systems are also leveraging image analysis to suggest images themselves. By analyzing the features of images, these systems can predict which images a user might appreciate, enhancing the user experience in platforms where visual content is predominant (Kobyshev et al., 2021).

In more specialized applications, image processing combined with recommender system techniques has been used to personalize rankings in critical domains such as healthcare. For example, Brandao et al. Brandão et al. (2021) utilized this approach for the personalized ranking of cancer drugs, showcasing the potential of image analysis in high-stakes decision-making contexts.

Moreover, hybrid recommender systems have employed feature extraction from images to enhance recommendation algorithms. An instance of this is seen in the work of Chu and Tsai (Chu and Tsai, 2017), where image features are utilized in conjunction with Matrix Factorization techniques to suggest favorite restaurants, demonstrating the versatility and effectiveness of combining image analysis with traditional recommendation algorithms.

2.6 Explanation and justification of recommender systems results

2.6.1 Premises

An "explanation" can be understood as a collection of statements that serve to elucidate a concept or to offer reasons behind a particular action or belief. Explanations hold a fundamental place in our daily lives, as they are instrumental in shaping and preserving our understanding of the world and events we encounter (Hu et al., 2021).

The evolution of recommender systems has witnessed a significant shift from being algorithm-centric "black boxes" to becoming more transparent and more attentive to the users. This evolution reflects a significant shift in the broader field of intelligent systems, emphasizing the importance of not just the decisions made by automated systems but also the ability to elucidate the reasoning behind these decisions. Initially, the focus in recommender systems was predominantly on refining algorithmic performance, which often came at the cost of user understanding and trust.

This paradigm began to change with the work of Herlocker et al. (2000), who highlighted the crucial impact of clear explanations on enhancing user acceptance and trust in recommender systems. Recognizing the importance of explanations marked a turning point in the field, leading to a renewed focus in which both the effectiveness of recommendations and their comprehensibility to users are prioritized.

Following this insight, a wave of research efforts emerged aimed at deepening the understanding of how recommendations are formulated and presented. Notable contributions in this area include the works of Nunes and Jannach (2017), Tintarev and Masthoff (2012, 2022), and Jannach et al. (2019), among others.

Their work has been instrumental in refining the strategies for improving user comprehension of personalized recommendations, thereby contributing to a more user-friendly and trustworthy interaction with recommender systems.

An explanation in recommender systems typically aims to clarify the processes or logic behind the generation of recommendations. It can be divided into two categories based on its purpose:

- **"How"**: These explanations focus on the operational aspects of the recommendation process. They provide insights into the mechanics of how a specific item was selected for recommendation, revealing the inner workings of the recommendation algorithm. This type of explanation is closely tied to enhancing the transparency of the system, but it can be challenging for users who are not familiar with the technicalities of the algorithm.

- **"Why"**: In contrast, "why" explanations delve into the rationale behind suggesting a specific item, without necessarily detailing the workings of the recommendation algorithm. These explanations rely on other available information to rationalize the recommendations made. This kind of explanation that generates post-hoc the rationale for suggesting a specific item without knowing the recommendation algorithm is called justification.

Starting from the purpose of the explanation we can classify three different approaches: The first one unravels the inner mechanics of the recommender system itself, shedding light on the algorithms and data processing techniques at play (Conati et al., 2021; Kouki et al., 2020); the second one includes methods that fuse the processes of generating recommendations and offering explanations for them, ensuring that users not only receive suggestions but also understand the rationale behind these suggestions (Dong and Smyth, 2017; Lu et al., 2018; Rana et al., 2022); the latter one are strategies that provide post-hoc justifications for the recommendations. These methods stand apart in their independence from the specific algorithms used for generating recommendations, offering insights after the recommendation process (Musto et al., 2021; Ni et al., 2019).

Among these, the third category is noteworthy for its versatility and algorithm-agnostic nature. This approach to justification is not limited by the constraints of the underlying recommendation algorithm, making it adaptable to a variety of systems.

2.6.2 Techniques

Various methods have been developed to facilitate the comprehension of recommendation systems.

Single-algorithm systems primarily focus on making the recommendation process explicit through an exposition of the inference logic. These systems offer a step-by-step breakdown of how and why each item was recommended (Herlocker et al., 2000; Nunes and Jannach, 2017). These systems allow users to follow the algorithm's reasoning, providing a clearer understanding of why certain items are recommended. By revealing the inner workings of the algorithm, these systems enhance transparency, making the recommendation process less of a "black box" and more of an open book that users can understand and trust.

Aspect-based systems, on the other hand, identify and underscore specific features of items that match or conflict with the user's preferences (Muhammad et al., 2016).

Some systems categorize items based on their advantages and disadvantages. This comparative approach, as seen in works by Pu and Chen (2007), Chen et al. (2014) and Chen and Wang (2017), helps users in making more informed decisions by providing a balanced view of each item and facilitating user comparison.

Systems supporting information exploration may visualize how items are relevant to the search query keywords (Chang et al., 2019; Di Sciascio et al., 2016).

Meanwhile, graph-based systems utilize user-item connections to rationalize recommendations (Amal et al., 2019; Wang et al., 2018; Musto et al., 2019).

A few studies propose merging the processes of recommending and explaining, based on the premise that both are driven by similar logic (Dong and Smyth, 2017; Lu et al., 2018), or leveraging the strength of potential explanations to guide the recommendation process (Rana et al., 2022).

Hybrid systems provide multifaceted explanations to show how various integrated systems collectively influence item ranking. For instance, MyMovieFinder (Loepp et al., 2015) displays individual recommenders backing a suggestion, while RelevanceTuner (Tsai and Brusilovsky, 2019a) employs stackable bars for visual integration. TalkExplorer (Verbert et al., 2016) and IntersectionExplorer (Cardoso et al., 2019) use bidimensional graphs and grid layouts, respectively, to illustrate different relevance dimensions. Venn diagrams are utilized by Kouki et al. (2017) and in combination with color bars by Parra and Brusilovsky (2015) to differentiate the contributions of integrated recommenders in evaluating items.

2.6.3 Formats of explanation presentation

This section categorizes explanation formats into two primary types:

- **Textual Explanations** – Explanations delivered in natural language, such as text or audio.
- **Visual Explanations** – Explanations presented using visual elements like graphs or images.

Textual explanations Commonly, people exchange explanations verbally. In systems, this approach utilizes natural language processing to generate explanatory sentences. In recommender systems, a straightforward method is listing options and item features (McAuley and Leskovec, 2013). More complex methods include canned texts or sentence templates that incorporate specific topics (Lei et al., 2016; Krening et al., 2017).

For *black-box* models, like those based on ensemble classifiers (Wyner et al., 2015) or deep neural networks (Gilpin et al., 2018), the focus shifts to interpreting complex models post-hoc. This could involve generating *rationales* (Lei et al., 2016) or employing reinforcement learning (Krening et al., 2017). A recent approach by Musto et al. (2021) starts from item reviews to generate justifications.

White-box models, such as linear models (Ribeiro et al., 2016), decision trees (Lakkaraju et al., 2016), or rule-based systems, derive explanations directly from outputs. These might include inference traces, vocal explanations (Terano et al., 1989), query results (Basu and Ahad, 1992), or OWL knowledge generation (Sherchan et al., 2008). Iovine et al. (2020) introduced a Conversational Recommender System, providing explanations through a chatbot, potentially leading to voice assistant applications.

Visual explanations Visual explanations, particularly effective for comparing system results, have been used in recommender systems to provide the user with a summary of results. O'Donovan et al. (2008) designed PeerChooser for movie recommendations with an intuitive graphical interface. In music services, Gou et al. (2011) used graphs for social friend recommendations. Interactive Venn diagrams (Parra et al., 2014) and scatter plots (Tsai and Brusilovsky, 2019b) have been employed for research talks or articles.

In the restaurant domain, Kouki et al. (2017) utilized Venn diagrams and concentric circles to explain collaborative filtering suggestions. Millecamp et al. (2020) applied bar diagrams in the musical domain, showing songs in the context of user preferences. For real estate, Mauro et al. (2020a) introduced INTEREST, a model using bar diagrams for consumer feedback representation. Additionally, Kouki et al. (2019) employed dendrograms to represent item information in a tree structure.

2.7 Service-centric perspective in recommender systems

Contemporary recommender systems predominantly focus on user-centric data or item-centric data. These approaches overlook the service dimension of products, an aspect that is crucial in shaping customer satisfaction and decision-making. For instance, the quality of post-sales customer service, which varies significantly among retailers, introduces a critical consideration for users that extends beyond the mere functionalities or features of the product. Current service-agnostic review analyses fail to adequately capture this information.

Platforms like Airbnb (2022) and TripAdvisor (2017) do provide summaries of customer feedback, highlighting overall ratings and key aspects such as cleanliness and location (in the case of Airbnb), or cuisine and atmosphere (for TripAdvisor). These summaries, along with individual reviews, offer valuable insights; however, they often lack a direct linkage to specific aspect values.

While Chen et al. (2014) do extract and present item aspects from reviews in both quantitative and qualitative forms, they don't adequately model the service interaction stages, resulting in a flattened, item-centric aspect overview requiring user interpretation for experience evaluation.

Moreover, while there are efforts to extract and present item aspects from reviews in both quantitative and qualitative forms (Chen et al., 2014), these approaches do not sufficiently model the various stages of service interaction. This results in a flattened, item-centric aspect overview that relies heavily on user interpretation for evaluating the entire experience. These considerations point to the need for a more holistic approach to recommender systems.

Chapter 3

Background on service models

3.1 Introduction

The work of this thesis is centered on the integration of service models within recommender systems to deliver personalized and effective recommendations. This chapter centers on the essential concepts and influential research pertaining to service models, with a particular emphasis on two key frameworks: service journey maps and service blueprints.

This chapter will explore these two models. Service journey maps offer a visual representation of a customer's interaction with a service, highlighting the key touchpoints and the user's emotional journey. On the other hand, service blueprints provide a more detailed view, mapping out the operational processes and interactions behind the scenes that support the customer journey. This dual perspective is essential for comprehensively understanding how services function and how they can be optimized in the context of recommender systems.

Service journey maps and service blueprints were specifically chosen due to their capabilities to offer a holistic view of the user experience of services.

As described in Chapter 6 and Chapter 7, we choose Service journey maps for our proposed recommender system and Service Blueprint for the proposed explanation models. The rationale behind this choice is to reflect the distinct needs associated with each aspect of the recommender system.

For the recommender algorithm, the detail provided by service journey maps is deemed sufficient. These maps capture the overarching narrative of a user's interaction with a service, emphasizing the sequence of touchpoints and the user's emotional and experiential journey. By focusing on the broader user experience stages, service journey maps facilitate the identification of critical dimensions where personalized recommendations can have the most impact, thus enhancing user engagement and satisfaction without necessitating granular operational insights.

Conversely, for the explanation models, a deeper dive into the service's operational details becomes crucial. Here, service blueprints offer deeper granularity and allow the development of explanation models that can address specific user queries or concerns with greater precision. Users seeking to understand the rationale behind certain recommendations may be required to go deeper. Service blueprints enable the creation of explanation models that can provide detailed justifications that enhance transparency and build trust in the recommender system.

3.2 Service journey maps

Service Journey Maps (SJMs) (Richardson, 2010), also known as Customer Journey Maps, are a fundamental tool in the development and enhancement of both physical and digital services, focusing primarily on the customer's experience. These maps are visual narratives that chart the customer's course

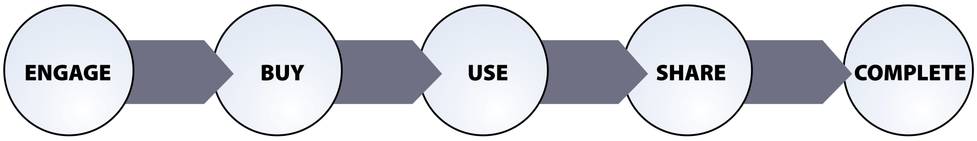


Figure 3.1: Example of a Service Journey Map, from Richardson (2010).

through various interactions with a service, from initial contact to the final stage of the service cycle. The primary objective of SJMs is to gain a deeper understanding of the customer's needs and experiences, thereby facilitating improvements in service design. Figure 3.1 from Richardson (2010) illustrates an example of an SJM, which is adapted from a model representing a typical customer journey timeline, which includes stages such as initial engagement with the customer (perhaps through advertising or in a store), purchasing the product or service, using it, sharing experiences with others (either in person or online), and eventually concluding the journey, which could involve upgrading, replacing, or opting for a competitor's offering (thereby initiating a new journey with another company).

A notable aspect of SJMs is their chronological representation of a service experience. This temporal perspective is crucial for capturing the sequence of customer interactions and emotions at each stage of the service journey. It helps businesses understand the customer's perspective, identifying points of satisfaction and frustration. For example, in the hotel industry, this could range from the initial booking process on a website to the post-stay feedback loop.

In the context of service-based integration and recommender systems, SJMs offer a unique opportunity. They allow the mapping of customer experiences in a way that traditional recommender systems may overlook.

By integrating SJMs into the analysis of recommender systems, it is possible to understand not just what products or services are being recommended, but also how they fit into the overall customer journey. This approach leads to more nuanced recommendations that are aligned with the customer's expectations and experiences at different stages of their journey.

Furthermore, SJMs can be instrumental in understanding and categorizing customer feedback. By analyzing reviews and feedback within the framework of the service journey, it becomes possible to pinpoint specific aspects of the service that need improvement. For instance, in a home-booking service, an SJM might include stages such as visiting the website, check-in, stay, and check-out.

3.3 Service blueprints

Service Blueprints (Bitner et al., 2008) are comprehensive visual tools designed to facilitate the design and development of products and services with a keen focus on the customer's perspective. These blueprints are instrumental in mapping out the entire journey a customer takes, from the initial point of contact (such as entering a website or physical store) to the final stage of customer care. Widely adopted in various domains, especially in e-commerce, Service Blueprints offer a structured approach to service and product modeling.

Figure 3.2 illustrates a Service Blueprint. Key components include:

1. **Physical Evidence**: This element encompasses all tangible aspects that a customer interacts with during the service experience. In the context of a home-booking service, physical evidence might include the design and usability of the booking platform's website, the physical aspects of the home (like key lock-boxes, keypad entries, or smart locks),

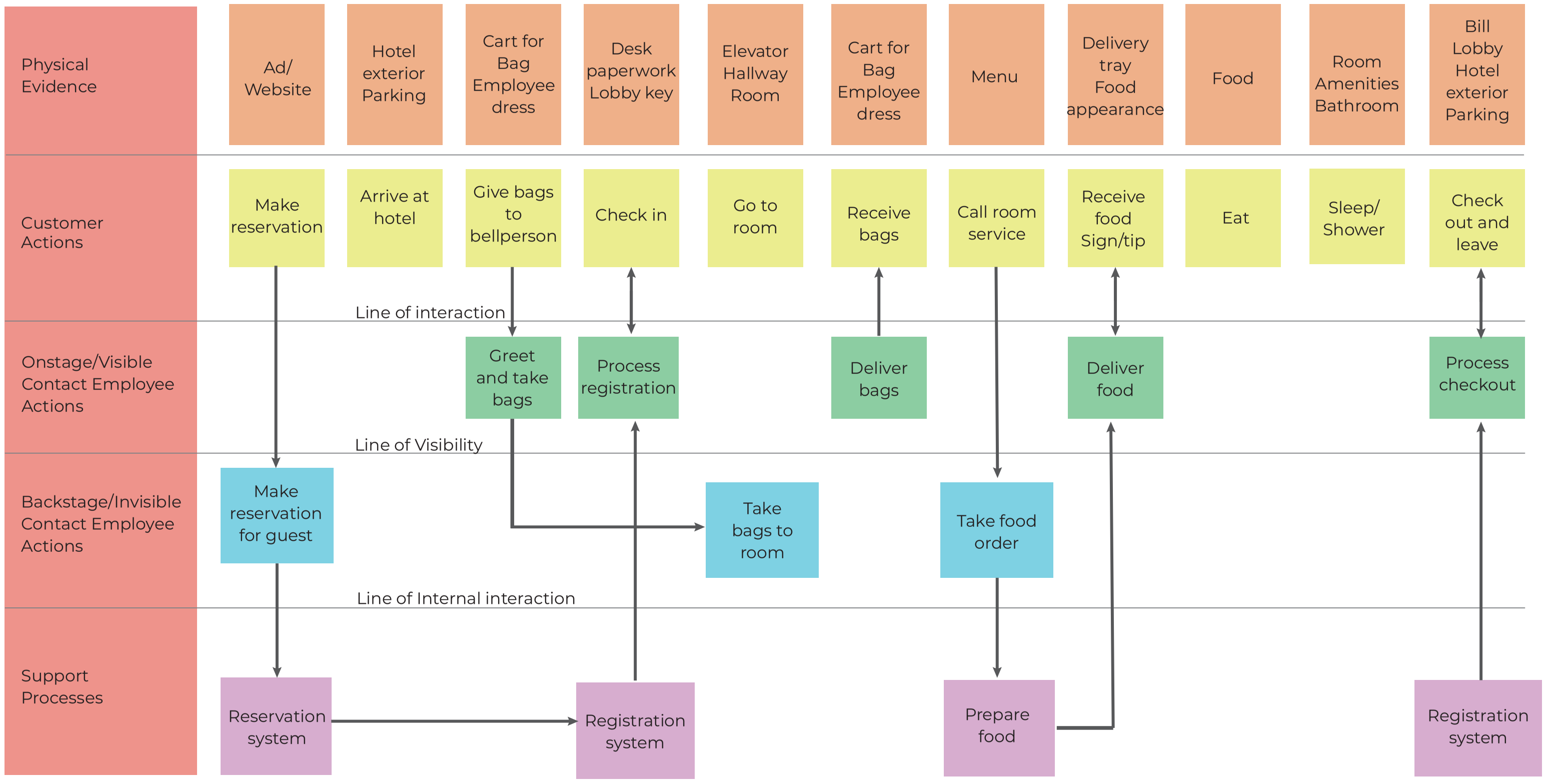


Figure 3.2: Example of a Service Blueprint that represent a hotel service Bitner et al. (2008).

and the quality of amenities and surrounding environment of the rental property.

2. **Customer Actions**: These are the steps or actions undertaken by customers as they navigate through the service. For a home-booking scenario, this could involve actions such as making a reservation, arriving at the property, and various activities related to personal care and managing the rented home during their stay.

3. **Onstage/Visible Contact Employee Actions**: This layer focuses on the actions performed by service providers that are visible to the customer. In a home-booking context, this might include the interactions during check-in, such as processing registration and welcoming the guest to the home.

4. **Backstage/Invisible Contact Employee Actions**: These are actions by employees that the customer doesn't directly see but are essential for delivering the service. Backstage actions support onstage activities and ensure that customer needs are met efficiently. In our hotel example, this might include staff preparing a room, kitchen staff preparing meals, or administrative staff managing bookings.

5. **Support Processes**: These are additional internal actions and processes that are necessary to support service delivery but are typically removed from direct customer interaction. Support processes often involve interactions with other departments or external agents. For instance, a hotel's relationship with suppliers, maintenance of IT systems, and internal communications between departments are all part of support processes.

6. **Line of Interaction**: This line separates the customer actions from the employee actions. It visually depicts the direct interactions between customers and service employees.

7. **Line of Visibility**: This line separates all service actions that are visible to the customer from those that are not (backstage actions). It helps in understanding which parts of the service process are exposed to the customer and which are hidden.

8. **Line of Internal Interaction**: This line separates the contact employees' actions from the support processes. It delineates the boundary between the front-office and back-office operations.

Service Blueprints are used to conceptualize and improve the design of services. However, we can use them to enhance the analysis and presentation of customer feedback, as well. In this context, they can guide the organization of the reviews by aligning the sentences with specific service stages. Additionally, they can help structure the presentation of various aspects of a service.

Chapter 4

Dataset

In this chapter, we focus on the dataset used in our research.

According to Nelson (1974)'s classification, we have three types of goods: search goods (products that can be evaluated before purchase, such as clothing), experience goods (the quality is learned through the fruition of the goods, such as a movie), and the credence goods (difficult to evaluate also after the purchase, such as legal advice).

The goal of our work is to develop a novel family of recommender systems that consider the service model and the development of justification models suitable for this domain. Moreover, we aim to extend traditional explanation and justification models by incorporating multimodal data using images.

For these reasons, we need to focus on experience goods rather than search goods, focusing on service-based systems. For our research, we need a dataset with several reviews and images. A good candidate we found was the home booking.

To validate the service-aware recommender system we developed and the justification of their results, we used an Airbnb dataset of homes and reviews.

This dataset includes detailed textual data extracted from Airbnb (2022).

Table 4.1: Descriptive statistics of the filtered dataset.

	Min	Max	Mean	Standard Deviation
Words per review	1	1002	47.00	46.41
Reviews per listing	1	648	20.80	35.96
Amenities per listing	0	66	20.98	7.85

To address the visual component, which is crucial for Chapter 8, the dataset has been further enriched with a set of additional images. This integration allowed us to explore and implement multimodal justification of recommender systems' results that not only consider text but also utilize visual information to enrich and enhance user awareness.

4.1 Airbnb London dataset

The primary dataset used across various studies in this thesis is a public dataset of Airbnb reviews concerning homes located in London. This dataset was downloaded from `http://insideairbnb.com/get-the-data.html` in January 2021.

It encompasses a broad spectrum of information, detailing not only the homes listed (referred to as "listings") but also the profiles of their administrators ("hosts") and the various features of these accommodations ("amenities"). A key component of this dataset is the list of reviews written by guests, which started on December 21st, 2009.

In light of the noticeable decline in home rentals in January 2020, likely linked to the onset of the COVID-19 pandemic, we decided to exclude reviews posted after the first day of that month. This temporal delimitation was necessary to maintain the consistency and relevance of the data with normal renting patterns.

Through language detection, the dataset was filtered by selecting only those reviews that were composed in English, thus ensuring clarity and uniformity in the data analyzed. Another filtering criterion was the exclusion of listings that had not been reviewed since 2018, allowing the focus to remain on homes that were relevant and recently engaged in the rental market.

The refined dataset Δ, as a result of these filtering processes, comprises 764,958 individual guests, 906,967 unique reviews, and a total of 43,604 listings. Table 4.1 presents a comprehensive array of descriptive statistics, providing a quantitative overview of this dataset.

A notable characteristic of the dataset is the diversity in the length and depth of guest reviews. This variance ranges from brief, succinct expressions of satisfaction (e.g., “Amazing location!”) to more detailed and descriptive accounts. These longer reviews often include a holistic evaluation of the guests’ experiences, touching upon various elements such as the quality and range of amenities offered, the cleanliness and comfort of the listings, as well as the demeanor and responsiveness of the hosts. Additionally, many reviews extend their scope to comment on the neighborhood and its suitability for different types of travelers, whether families with young children or solo adventurers. An illustrative example of such a comprehensive review is:

> “A warm and private place ideal for exploring London. Location was perfect and felt very safe. We stayed with our young children and they had space to stretch out with their toys, the lift was convenient and check-in was a breeze! Very clean and comfortable, we would stay here again!”

These reviews provide a multi-dimensional perspective on guest experiences. This aspect allows us to perform analysis to extract aspects in a quantitative and qualitative way, as will be detailed in the following chapters.

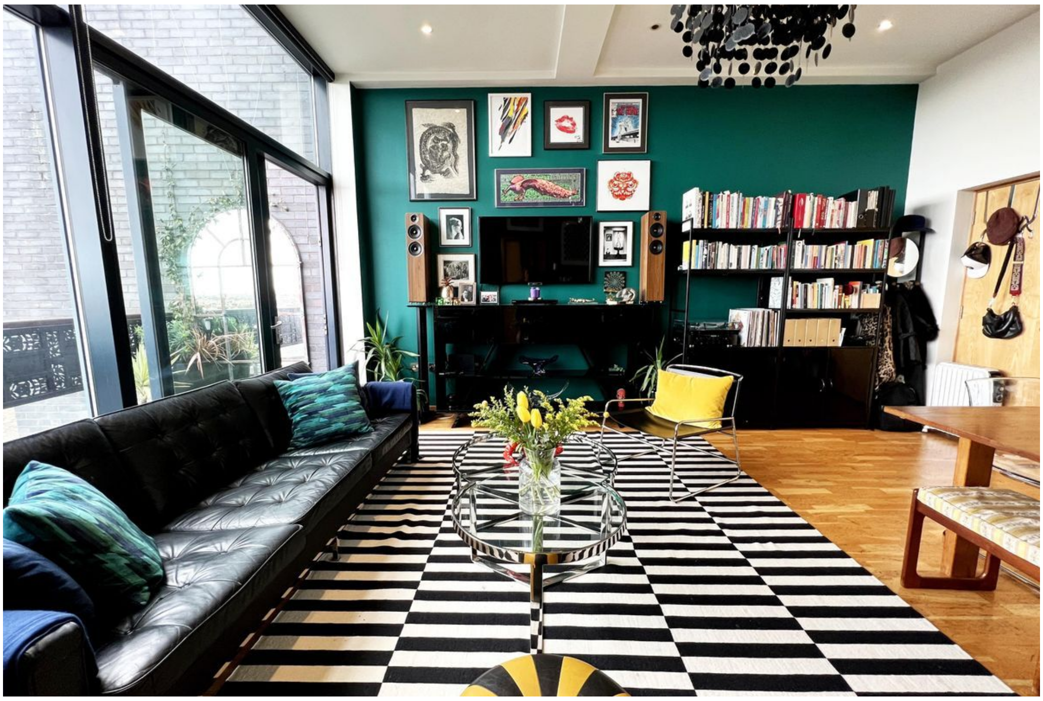


Figure 4.1: Example of a photograph from Airbnb (2022) we used for our user tests.

Regarding the reviews, we have the following information: listing's ID, date, reviewer's name, comments

For each home $h \in \Delta$, the dataset has more than 70 different fields, which we report in Appendix A.

The more important that we discuss in this thesis include its title, a link to the primary image, the list of amenities (like TV, balcony, etc.), the host's Airbnb webpage, the price, the number of rooms.

4.2 Images of the homes

An important component of our dataset involved the visual representation of Airbnb listings. Understanding that images play a significant role in shaping user perception and decision-making, in Chapter 8 we used photographs of the homes listed on Airbnb.

In September 2022, we undertook a targeted data collection process to enrich our dataset with image data. This process involved selective web scraping of Airbnb listings.

- **Selection Criteria:** To ensure a sufficient amount of information about homes for the recommendation task, we prioritized homes $h \in \Delta$ with a higher number of reviews, sorting them in descending order. This approach was adopted under the assumption that homes with more reviews are likely to provide a richer dataset for analysis, including diverse user feedback and detailed experiences.

- **Filtering:** We filtered out the homes having less than 15 photographs. The rationale for requiring a minimum number of photographs was to ensure a comprehensive visual representation of each home, allowing for a detailed analysis of various aspects like interior design, amenities, and overall ambiance.

In the next chapters, a significant focus will be placed on the preprocessing steps undertaken for each user study. These chapters will detail the preparation and processing of the data, highlighting how we tailored our approach to meet the specific needs and objectives of each user study.

Chapter 5

Preliminary experiment

This chapter delves into a preliminary study conducted to evaluate the efficacy of our proposed visual model in enhancing decision-making within the context of home recommendation systems. Our model (Figure 5.1), detailed in (Mauro et al., 2021), integrates quantitative and qualitative data, presenting this information incrementally to facilitate a nuanced analysis of home selections. This preliminary experiment serves as a foundational step in assessing the impact of our service-based justification model on user decision-making.

5.1 Study design and objectives

In Section 2.6.3, we described various visual explanations. Among the possible presentations for representing the ratings of each evaluation dimension, we decided to use the bar graph. We preferred this representation among the others, such as TagCloud, because we have a fixed set of evaluation dimension and we aim to let the user be aware of the sentiment of each dimension, through the value of the bar graph.

The study was designed to explore bar graphs depicting various evaluation dimensions, and the inclusion of qualitative data, such as aspects derived from item reviews, influence user preferences and decision-making processes.

The study primarily aimed to answer the following research questions:

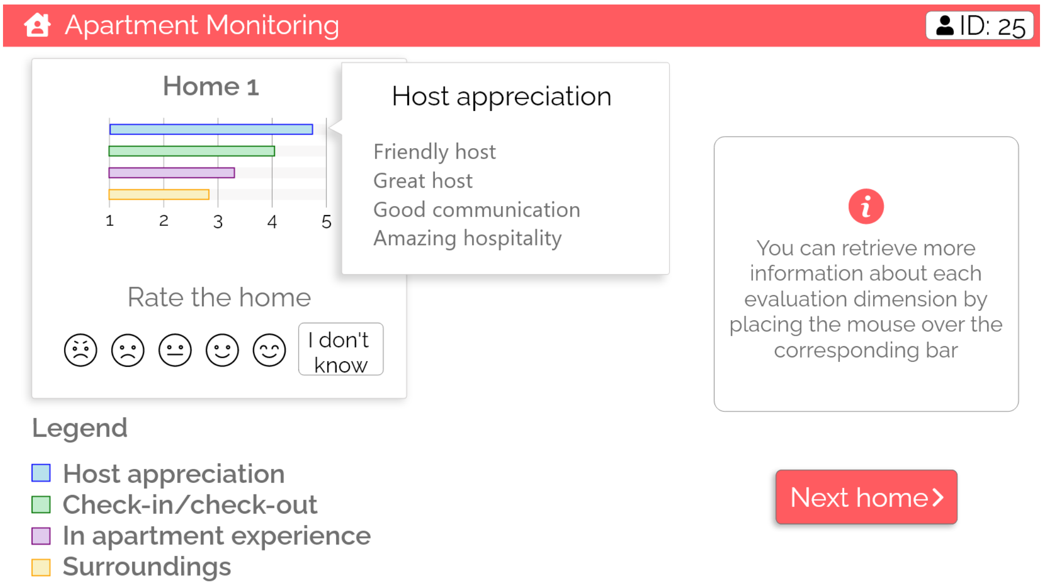
RQ5.1: Does the bar graph representation of a home h enable users to quickly assess the perceptions of previous guests based on their reviews about h?

RQ5.2: Does the overall visual model, which includes bar graphs and qualitative data consisting of aspects, assist users in focusing on the most promising homes within a recommendation list?

5.2 Methodology

The experiment involved a prototype interface showcasing our visual model. Participants engaged in tasks requiring the evaluation of Airbnb homes, presented through mock-up interfaces described in Figure 5.1, and outlined in (Mauro et al., 2021). These tasks were designed to measure the effectiveness of quantitative data alone versus the combined impact of quantitative and on-demand qualitative data on user decision-making. Notice that, at this stage of the research, the aspects of homes presented on demand, and the values for the bars in the visual model were handcrafted, as the strategy for extracting these values from data had not yet been developed.

The study involved only 11 participants, ranging in age from 19 to 57, with varying backgrounds and levels of technological proficiency. The reduced number of participants is due to the pandemic situation during the user study. The primary objective of this preliminary investigation was to swiftly gauge the efficacy of our chosen approach within a limited timeframe. It was our



You can retrieve more information about each evaluation dimension by placing the mouse over the corresponding bar

Figure 5.1: User interface of the preliminary model, showcasing the bar graph representation and on-demand qualitative information, from Mauro et al. (2021).

intention to expand the participant base in subsequent studies, aiming for a broader demographic representation and more comprehensive insights.

Participants were divided into two tasks:

- **Task 1 (`T1`):** Viewing only the bar graphs for coarse-grained evaluation dimensions.
- **Task 2 (`T2`):** Accessing both bar graphs and on-demand qualitative data related to the dimensions.

5.3 Findings and insights

Post-task feedback revealed that 54.54% of participants in `T1` felt the information provided was insufficient for rating the homes. Conversely, in `T2`, only

27.27% desired more information, with the rest finding the data adequate for evaluation. This highlights that while bar graphs alone fall short in supporting decision-making due to their lack of qualitative depth, including on-demand qualitative data significantly enhances the decision-making process.

Participants' remarks offered further insights into their interaction with the mock-up:

- *Practical Application of Bar Graphs:* Participants noted the utility of bar graphs in quickly filtering out homes that performed poorly in dimensions critical to them. For example, one participant mentioned discarding a home with a low `Host Appreciation` score to avoid potentially difficult interactions.
- *Qualitative Data Impact:* Several participants expressed that qualitative details about evaluation dimensions (i.e., the on-demand aspects) allowed them to "adjust" their perception of the bar graph values. For instance, if negative aspects of a home were irrelevant to a user, they might implicitly rate the home higher than its bar graph score suggested.

5.4 Conclusion

The initial findings from this user study were promising. The model demonstrated success in assisting users to efficiently filter the information space, providing them with an effective, holistic overview of consumer feedback. Furthermore, the incorporation of on-demand qualitative data about previous consumers' experiences significantly enhanced user awareness and understanding of the items.

This preliminary study laid a foundational understanding that guided the subsequent development of our visual model.

Chapter 6

Integration of service model in recommender systems

6.1 Introduction

This chapter focuses on the first research question we posed: "RQ1 How does the integration of a service-based representation of items, which explicitly models the stages of item consumption, impact the quality of recommendations, compared to systems that rely solely on local item properties and overall ratings?"

As mentioned in Section 2.4, traditional content-based, feature-based, and collaborative filtering recommender systems primarily use item properties and item ratings for generating suggestions. These methods, focusing on catalog features and user ratings, often overlook detailed consumer feedback brought by reviews. This gap results in a limited contextual understanding of the various stages of item consumption, thus constraining the effectiveness of recommendations.

To address this limitation, our research introduces an innovative approach that integrates service modeling techniques with recommender systems. This strategy aims to encompass the consumer experience throughout different stages of item consumption, enhancing the quality of recommendations and enriching user awareness in decision-making.

To evaluate the impact of this approach and reply to the first research question, we asked ourselves two questions:

- Can integrate a service-based representation of items, which explicitly models the stages of item consumption, improve the quality of recommendations in Top-N recommender systems compared to relying solely on local item properties and overall ratings?
- Does incorporating service-based information about item consumption stages, in addition to traditional item properties, enhance user awareness and confidence in their selection decisions, as opposed to presenting only item properties?

In our quest to answer these questions, this chapter of the thesis introduces a new category of recommender systems, which we have termed "service-aware recommender systems."

Despite the advancements in recommender systems, including Multi-Criteria Recommender Systems (MCRS) that enrich the recommendation process by considering multiple dimensions of user preferences (Adomavicius et al., 2011), our approach introduces a distinct dimension to recommendation accuracy and user experience. While MCRS effectively address the complexity of user preferences by integrating multiple criteria, they still primarily focus on the aggregation of these criteria to form recommendations. In contrast, our service-aware recommender systems extend beyond the aggregation of diverse user preferences by explicitly modeling the service consumption stages of items. We aim to provide a more contextual and informed recommendation, to improve user awareness and confidence in their decisions.

Following this distinction, the research and findings presented further in this chapter leverage the Service Journey Maps model, described in Section 3.2.

As highlighted in Mauro et al. (2020a), user experience in service-based systems can be modeled in stages. For example, in online product sales, the experience contains stages from searching for products on a retailer's website to post-purchase assistance. These stages help in identifying key evaluation dimensions for item selection.

By abstracting from the granular details found in item reviews, these dimensions provide a holistic summary of past consumer experiences.

This abstraction process allows for aggregating diverse consumer feedback into a more structured and analyzable form, facilitating the generation of a comprehensive and multifaceted summary of consumer experiences.

The research and findings presented in this chapter are elaborated in Mauro et al. (2022b). For this work, we performed a preliminary study (Mauro et al., 2021) which is described in Chapter 5.

The following sections describe the Service Journey Maps we built for this work and the analysis we performed on the reviews. Next, we describe the user test that we performed to answer the research questions, and at last, we discuss the results.

6.2 From service model to evaluation dimensions

The idea driving this research lies in the concept that by categorizing information extracted from item reviews into service-based evaluation dimensions, the recommender system can suggest and present the suggestion better.

Firstly, it improves suggestions based on the stages most important to the user. By focusing on specific stages of the consumer journey, the system can tailor recommendations more precisely to the user's current needs and preferences. This stage-based approach allows for a more targeted and relevant set of recommendations, enhancing user satisfaction and decision accuracy.

Secondly, the system enhances the presentation of user opinions. Instead of aggregating feedback into a generic overall rating as seen in traditional star-rating interfaces, our system delineates user opinions according to different stages of their experience. This stage-wise breakdown of feedback presents a clearer and more organized view of previous users' thoughts and experiences, making it easier for potential users to understand the various facets of the item or service. Such a structured feedback presentation aids users in comprehending the strengths and weaknesses of an item at each stage of the service journey, leading to a more informed decision-making process.

This approach stems from the understanding that traditional recommender systems, while effective in leveraging basic item properties and user ratings, often fail to capture the depth and breadth of the consumer experience.

The service-aware recommender system we propose is based on (Mauro et al., 2020a), in which the authors developed a Service Journey Map to represent the guest experience in home booking, primarily from the perspectives of the customer and the apartment owner.

This section delves into how this model is translated into evaluation dimensions.

The initial SJM outlined four stages: *Visit website, Check-in, Stay in apartment,* and *Check-out.* The authors aimed to associate each stage with a unique evaluation dimension and developed thesauri based on existing

Evaluation Dimension	Keywords
Host appreciation	host, owner, renter, interaction, people, relation, hospitality, manner, language, communication
Check-in/Check-out	entrance, entry, term, suggestion, welcome, key, reception, check-in, check-out, luggage, ...
In-apartment experience	room, lighting, fridge, home, appliances, washer, refrigerator, dishwasher, freezer, tv, security, ...
Surroundings	noise, music, sound, voice, disturbance, bell, city, beach, transport, airport, café, club ...

Table 6.1: Part of thesauri evaluation dimensions/keywords for the home booking domain. See Appendix B for the full thesauri.

literature on home and hotel booking service reviews. The thesauri include the most frequently occurring keywords retrieved from consumer feedback and associated with specific service stages. Table 6.1 shows an extract of these thesauri.

The cited analysis revealed several key insights:

- Experiences during *Check-in* and *Check-out* were frequently interconnected in guest reviews, sharing several common keywords.
- The *Stay in apartment* stage covered a wide array of keywords. Reviews often distinguish between interior aspects (like furniture and comfort) and external factors (such as location and nearby amenities).
- The role of host interaction was identified as a crucial evaluation dimension, intersecting all service stages.

Consequently, the authors revised the original evaluation dimensions to include: *Host appreciation, Search on website, Check-in/Check-out, In-apartment experience,* and *Surroundings.*

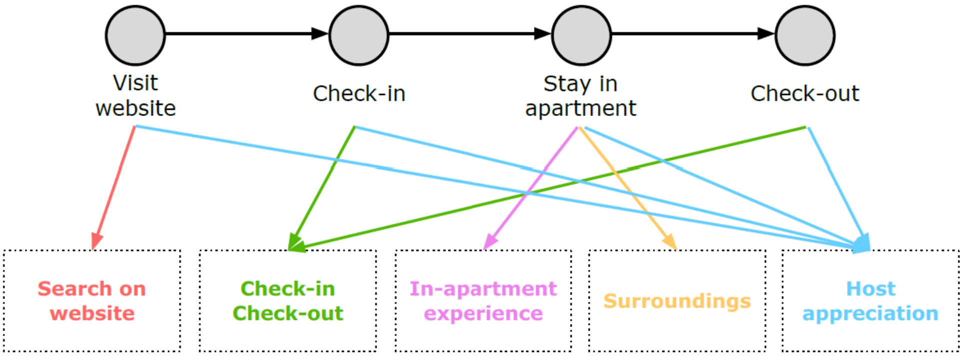


Figure 6.1: The top section of the figure illustrates the Service Journey Map, which details the stages of the home booking experience. Each stage in the map is linked to its corresponding evaluation dimension(s). Image taken from Mauro et al. (2020a).

In our work, we used the full lists of keywords for each dimension's thesaurus since these lemmatized keywords are crucial for indexing review sentences with the corresponding evaluation dimensions.

In this study, we exclude the `Visit website` stage from our analysis as our focus is not on assessing the Airbnb platform. Thus, the dimensions considered in the present work are the following:

1. `Host appreciation` reflects the guests' viewpoint of the host and their interactions throughout the service period.

2. `Check-in/Check-out` encapsulates the guest's experiences during their stay's initial and final phases, focusing on elements like promptness.

3. `In-apartment experience` pertains to the guests' impression of the apartment's interior, encompassing factors like its cleanliness and overall comfort.

4. `Surroundings` characterizes the guests' impression of the home's neigh-

Table 6.2: A subset of the aspects extracted from the reviews of a sample Airbnb home.

Aspect	Adjective	Evaluation	Frequency	Evaluation Dimension
host	wonderful	0.8929	4	`Host appreciation`
host	friendly	0.7172	2	`Host appreciation`
communication	lovely	0.7715	3	`Check-in/Check-out`
check-in	easy	0.7184	2	`Check-in/Check-out`
ambiance	nice	0.7554	16	`In-apartment experience`
bed	comfortable	0.7277	5	`In-apartment experience`
bus	good	0.7851	10	`Surroundings`
neighborhood	nice	0.7554	9	`Surroundings`

borhood, including the availability of services, transportation, and the level of tranquility.

6.3 Analysis of the reviews of the homes

To translate the textual information provided by item reviews into numerical values representing their evaluation with respect to the overall service, we structure the feedback from reviews based on the previously listed evaluation dimensions. This involves a detailed analysis of reviews for each home $h \in \Delta$, broken down into a three-stage process outlined in the following sections.

6.3.1 Aspect extraction and sentiment analysis from a home's reviews

In this phase, we identify aspects and their associated adjectives within the reviews of a home h, applying dependency parsing analysis of the sentences. Given a review r, we extract the list of $< aspect, adjective >$ found in r.

Subsequently, for each home $h \in \Delta$ we count the number of occurrences ($frequency$) of each $< aspect, adjective >$, to determine the regularity of the expressed opinion. Additionally, the sentiment associated with each aspect is calculated as the average output from the TextBlob (Loria, 2020) and Vader (Hutto and Eric, 2014) sentiment analysis tools. This value is normalized to get an *evaluation* in [0, 1].

The output of this step is a list of

$$< aspect, adjective, evaluation, frequency >$$

for each aspect-adjective pair that appears in the reviews of h. In Table 6.2, we posted a sample elaboration of a home of our dataset.

A critical consideration is the handling of lexical variations and synonyms to ensure that similar consumer sentiments are accurately aggregated, regardless of the specific wording used in the reviews. To address this, our NLP pipeline incorporates lemmatization techniques, which processes the adjectives and the aspects to their base or dictionary form, ensuring that different tenses or variations of a word are recognized as the same entity.

As part of our ongoing efforts to refine and enhance the accuracy of our service-aware recommender systems, future work will explore the integration of more sophisticated NLP techniques. Specifically, the incorporation of a synonym resolution system leveraging lexical databases such as WordNet represents a promising direction. This approach aims to further homogenize sentiment analysis by ensuring that varying expressions of sentiment related to the same aspect are treated equivalently.

6.3.2 Classification of the extracted aspects in evaluation dimensions

In line with the methodology proposed by Mauro et al. (2020a), the next step is grouping various aspects extracted from the reviews of h into the evaluation dimensions. An example of the results of this classification is illustrated in the fifth column of Table 6.2, where aspects are classified with their corresponding experience evaluation dimensions.

To accomplish this task, we employ four specialized dictionaries. Each dictionary encompasses a wide array of terms individuals commonly use when describing their experiences. For example, keywords such as “kitchen,” “bed,” and “bathroom” are present in the `In-apartment Experience` dictionary.

In this user test, the creation and use of specialized dictionaries was an effort that required manual engineering. We recognize the need for scalability and adaptability in this approach. Possible approaches involve the application of unsupervised machine learning techniques, such as clustering algorithms, to identify common themes within the reviews.

6.3.3 Computation of the values of the experience evaluation dimensions of h

In this step, we calculate the numeric rating of each evaluation dimension d, for each home h to understand the satisfaction, in the following way.

Within this dimension, we focus on the set AA_{dh}, which includes all $< aspect, adjective >$ pairings extracted from the reviews of home h and relevant to dimension d, calculated in the previous step. We use a weighted average to calculate the value of dimension d in home h (notated as $value_{dh}$). This average is based on the evaluations of each pair p in AA_{dh}. The weighting factor for each pair is determined by its occurrence frequency in the reviews of h. This approach allows us to adjust the impact of each pair based on the prevalence of similar opinions among reviewers:

$$value_{dh} = \frac{\sum\limits_{p \in AA_{dh}} frequency_p * evaluation_p}{\sum\limits_{p \in AA_{dh}} frequency_p} \tag{6.1}$$

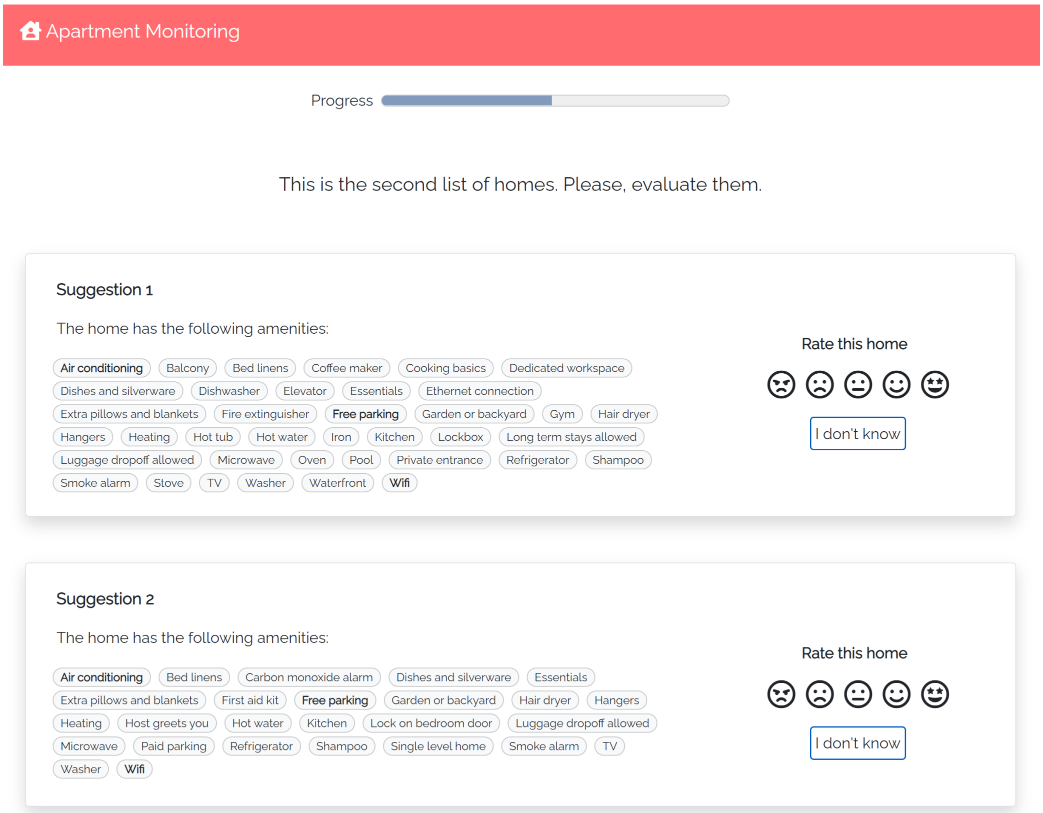
where $frequency_p$ represents how often the pair p appears in the reviews of h, and $evaluation_p$ is the evaluation assigned to p, derived from the sentiment of the aspect in p. For example, in assessing the `Host appreciation` dimension, we calculate the weighted average using the evaluation and frequency of pairs like <host, wonderful> and <host, friendly>, as referenced in Table 6.2.

Furthermore, our preliminary user study (see Chapter 5) suggests that reviewers generally view the absence of information about a home in a negative light. Consequently, if no reviews mention aspects related to a dimension d, or if a home lacks any reviews at all, we assign a default value of 0.1 to that dimension.

6.4 Recommendation models

In this section, we outline the service-aware recommendation models developed for this study, as well as the baseline models against which they are compared. Each model is detailed regarding its algorithmic foundation and the user interface designed for its evaluation through user interaction. Initially, we describe the baseline models, which are subsequently integrated into some of our service-aware recommendation systems to create a hybrid approach. Our discussion utilizes the following notations:

- Let $\mathcal{I}$ denote the collection of items (in this context, homes), and $\mathcal{U}$



This is the second list of homes. Please, evaluate them.

Suggestion 1

The home has the following amenities:

Air conditioning Balcony Bed linens Coffee maker Cooking basics Dedicated workspace
Dishes and silverware Dishwasher Elevator Essentials Ethernet connection
Extra pillows and blankets Fire extinguisher Free parking Garden or backyard Gym Hair dryer
Hangers Heating Hot tub Hot water Iron Kitchen Lockbox Long term stays allowed
Luggage dropoff allowed Microwave Oven Pool Private entrance Refrigerator Shampoo
Smoke alarm Stove TV Washer Waterfront Wifi

Suggestion 2

The home has the following amenities:

Air conditioning Bed linens Carbon monoxide alarm Dishes and silverware Essentials
Extra pillows and blankets First aid kit Free parking Garden or backyard Hair dryer Hangers
Heating Host greets you Hot water Kitchen Lock on bedroom door Luggage dropoff allowed
Microwave Paid parking Refrigerator Shampoo Single level home Smoke alarm TV
Washer Wifi

Figure 6.2: User interface for the presentation of the suggestions in FEATURES and CBF recommender systems.

represent the group of users.

- For each $i \in \mathcal{I}$ and $u \in \mathcal{U}$, we define vectors $\mathbf{i}$ and $\mathbf{u}$. These vectors encapsulate the respective profiles of the item and the user.
- $\hat{r}_{ui}$ is the predicted rating that user u gives to the item i.
- D is used to represent the set of experience evaluation dimensions that are taken into account in this study.

6.4.1 FEATURES

Model

FEATURES is a feature-based recommender system and is the baseline for our study.

In our study, we utilize the amenities of each home h in our dataset as the defining features of our items. These amenities are represented by the set $a_1, \ldots, a_z$, encompassing a range of characteristics that define each home.

The item profile, denoted as $\mathbf{i}$, is expressed as $< a_1, \ldots, a_z >$. It reflects the availability of each feature in the home. Here, for each feature a_j where j belongs to $\{1, \ldots, z\}$, we assign $f_j = 1$ if the home offers the respective amenity, and 0 if it does not.

The user profile, symbolized as $\mathbf{u}$, is captured as $< p_1, \ldots, p_z >$. This profile stores a user's preference levels for different home features. For any preference $j \in \{1, \ldots, z\}$, the value p_j is set as 1 to indicate "It's very important", 0 for "I don't like it", and 0.5 for "I don't care" (the default).

Our algorithm primarily focuses on the features that the user prefers or dislikes u. It calculates the estimated rating $\hat{r}_{ui}$ by normalizing the cosine similarity between the projections of vectors $\mathbf{u}$ and vector $\mathbf{i}$ (notated as $\vec{\mathbf{u}}$ and

$\vec{\mathbf{i}}$) on the feature components that have values of either 0 or 1. The formula for this calculation is as follows:

$$\hat{r}_{ui} = 1 + 4 * \frac{\vec{\mathbf{i}} \cdot \vec{\mathbf{u}}}{\|\vec{\mathbf{i}}\|_F \ * \|\vec{\mathbf{u}}\|_F} \tag{6.2}$$

in which $\cdot$ symbolizes the scalar vector product, $\| \cdot \|_F$ represents the Frobenius Norm, and * is used for the decimal product.

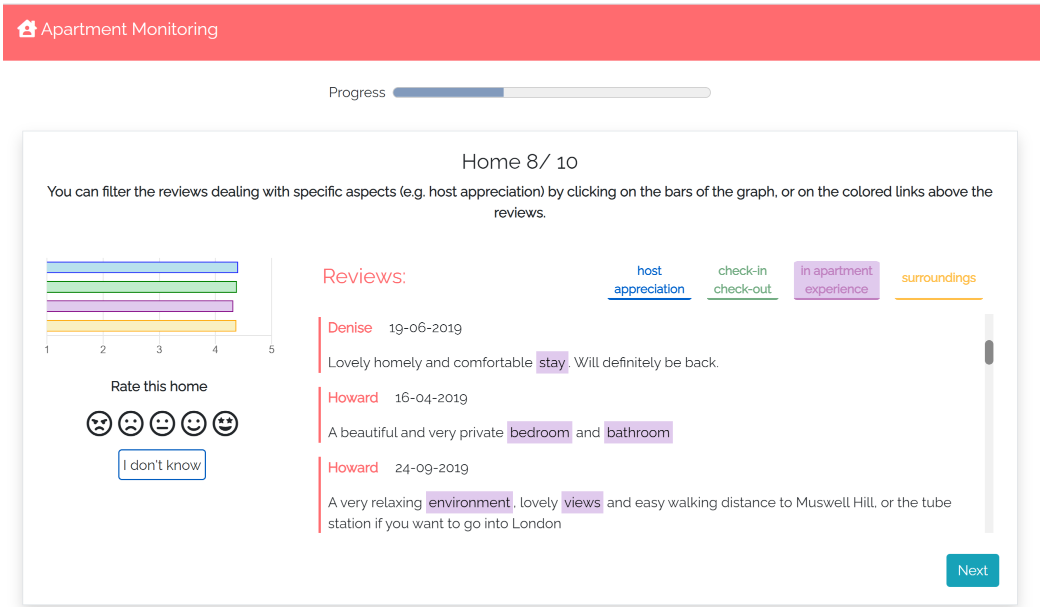
In cases where the vector $\vec{\mathbf{u}}$ is empty, the algorithm defaults to a standard popularity-based recommendation method (POP). This method prioritizes items based on the volume of reviews they have accumulated.

Finally, we convert the calculated rating in a $[1, \ldots, 5]$ scale.

User interface

Before recommending, this recommender system needs to build the user profile. For this purpose, the system displays some homes and the available amenities, allowing user u to indicate their preference levels for these features. On the interface's right sidebar, u can choose missing amenities in the current home that are available in others. Selecting any of these amenities automatically categorizes them as "It's very importan" in u's preferences. However,u could identify an amenity as favorable and unfavorable across different homes. This conflicting input is resolved by assigning a neutral "I don't care" preference in the user profile $\mathbf{u}$.

The page's lower section includes a rating elicitation tool, which, while not directly linked to FEATURES, is integrated due to its relevance in CBF (as detailed in Section 6.4.2). This element features a range of emoticons corresponding to a rating scale of [1, 5], alongside an "I don't know" option for users unsure about evaluating a home, as opting out. To avoid item assessment bias, we exclude details such as the home's name, cost, guest capacity, and



Home 8/ 10

You can filter the reviews dealing with specific aspects (e.g. host appreciation) by clicking on the bars of the graph, or on the colored links above the reviews.

Reviews:

Denise 19-06-2019

Lovely homely and comfortable stay. Will definitely be back.

Howard 16-04-2019

A beautiful and very private bedroom and bathroom

Howard 24-09-2019

A very relaxing environment, lovely views and easy walking distance to Muswell Hill, or the tube station if you want to go into London

Figure 6.3: User interface to collect the user's preferences in the STAGES recommender system.

images, in line with Tintarev and Masthoff (2015)'s recommendations. Refer to Figure 6.5.

Regarding the presentation of recommendations, illustrated in Figure 6.2, the user interface supports both the display of recommended items and their subsequent evaluation. Features that the user has marked as either positive or negative are highlighted in bold.

6.4.2 CBF

Model

To test our recommendation algorithm, we implemented a content-based recommendation model, as outlined in Lops et al. (2011), and we will use it as baseline.

The item profile **i**, is denoted by $< a_1, \ldots, a_z >$, encapsulating the feature values for the item. For each $j \in \{1, \ldots, z\}$, we assign $a_j = 1$ if the item includes the specific amenity, and 0 if it does not.

The user profile, represented as **u**, is expressed as $< p_1, \ldots, p_z >$. This profile records the user's liking for different item features. In alignment with the approach in Saia et al. (2016), each element of **u** is set to 1 if the user has given a high rating (between 4 and 5) to at least one item with the respective feature, and 0 in all other cases.

If a user u, has given a high rating to any item, the CBF method is employed to assess i by calculating the cosine similarity between **u** and **i**, with the results normalized to fit within a [1, 5] range. In scenarios where no high ratings are given by u, the system defaults to the POP method.

User Interface

The user interface for **collecting user preferences** is shown in Figure 6.5. The **delivery of recommendations** employs an interface akin to that shown in Figure 6.2. However, it does not highlight any features in bold, as it does not rely on explicit user preferences for its operations.

6.4.3 STAGES (service-aware)

Model

This model integrates insights about customer experiences during various stages of service fruition to create tailored recommendations. It assesses items based on a selection of experience evaluation dimensions $D = \{d_1, \ldots, d_n\}$, and their perceived significance to the user. In this study, D includes dimensions like {`Host appreciation`, `Check-in/Check-out`, `In-apartment experience`, `Surroundings`$\}$.

For each user $u \in \mathcal{U}$ and item $i \in \mathcal{I}$, the item profile $\mathbf{i}$ is represented as $< value_1, \ldots, value_n >$, capturing the metrics of the evaluation dimensions for item i, derived from its reviews using Equation 6.1. Here, for each j in $\{1, \ldots, n\}$, $value_j$ corresponds to the score of dimension d_j, as shown in Table 6.2.

The user profile $\mathbf{u}$ is represented as $< importance_1, \ldots, importance_m >$, reflecting the user's perceived importance of each dimension $d_1, \ldots, d_m$ in their decision-making process.

For each j in $\{1, \ldots, m\}$, $importance_j$ is deduced by scaling the Pearson correlation between u's overall ratings and the dimension d_j's values in reviewed items to a [0, 1] range. Ratings marked as "I don't know" are excluded for their lack of informativeness.

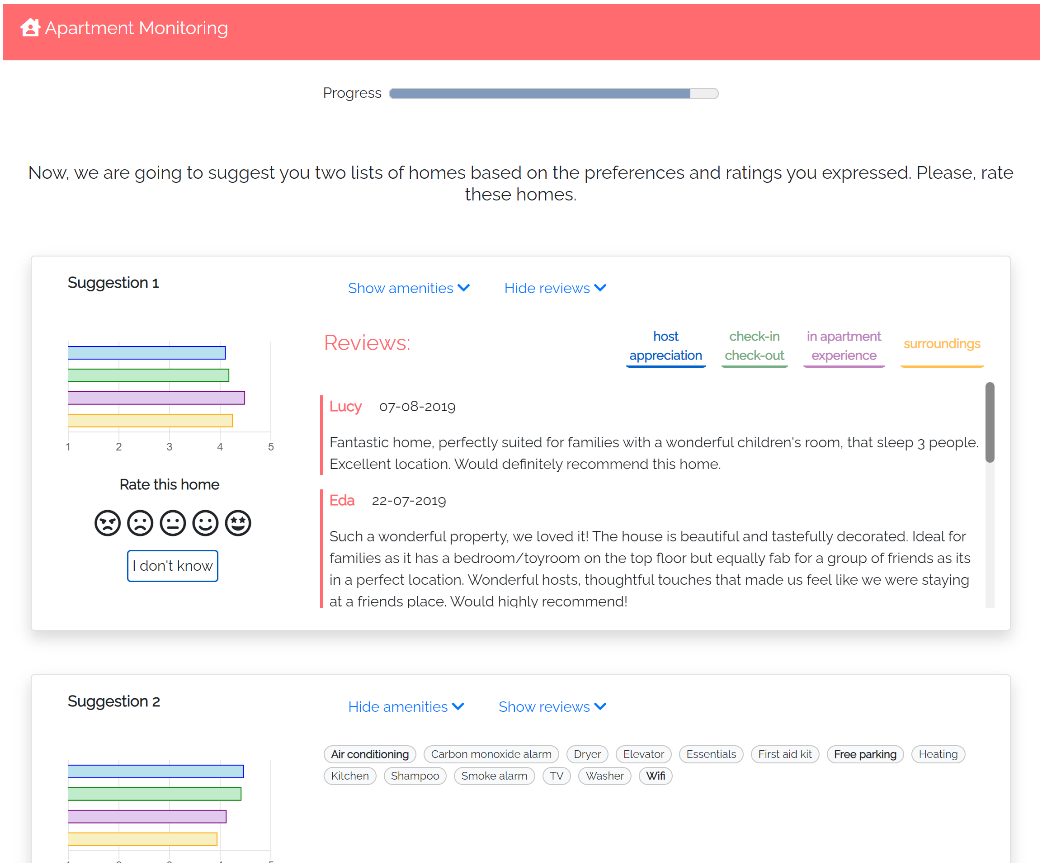
Essentially, if u consistently rates items higher when they score well in d_j and lower when they score poorly, it suggests a high relevance of d_j to u. Conversely, inconsistent ratings relative to d_j values indicate a lower interest in that dimension.

The rating of item i by user u is calculated as follows:

$$\hat{r}_{ui} = 1 + 4 * \prod_{j=1}^{m} (imp_{ju} * value_{ji} + 1 - imp_{ju}) \tag{6.3}$$

where imp_{ju} is the importance of dimension d_j in $\mathbf{u}$ and $value_{ji}$ is its score in $\mathbf{i}$. The formula $(imp * value + 1 - imp)$ modifies the terms of the product: it diminishes the effect of low scores for dimensions that u is indifferent to while preserving the influence of significant dimensions through the "$1 - imp$" component.

In cases where u has not provided item evaluations during the user prefer-



Now, we are going to suggest you two lists of homes based on the preferences and ratings you expressed. Please, rate these homes.

Lucy 07-08-2019

Fantastic home, perfectly suited for families with a wonderful children's room, that sleep 3 people. Excellent location. Would definitely recommend this home.

Eda 22-07-2019

Such a wonderful property, we loved it! The house is beautiful and tastefully decorated. Ideal for families as it has a bedroom/toyroom on the top floor but equally fab for a group of friends as its in a perfect location. Wonderful hosts, thoughtful touches that made us feel like we were staying at a friends place. Would highly recommend!

Figure 6.4: User interface for displaying recommendations in FEATURES-STAGES and CBF-STAGES systems.

ence acquisition phase, STAGES defaults to a POP-based rating estimation method.

User interface

Figure 6.3 displays the **collection of user preferences** of this model. It facilitates the assessment of the importance of various evaluation dimensions for each home h, displaying:

- A graphical representation, specifically a bar chart, that encapsulates the consumer experience with h as derived from its reviews. Each evaluation dimension is represented by a distinct colored bar. While these values originally range between [0, 1], they are adjusted to the [1, 5] scale for consistency with the home rating system.

- An element for gathering user ratings concerning the home with the range of emoticons with the scale [1, 5].
- The list of reviews of home h. The interface includes a feature allowing users to filter information by selecting one or multiple evaluation dimensions. This can be done by clicking either the corresponding bars or the dimension list above the reviews. Following selection, the system highlights comments that reference the chosen dimension(s), using color-coding for easy identification. For example, the figure illustrates reviews focusing on `In-apartment experience`. The categorization of aspects into dimensions is facilitated through the use of dictionaries, as discussed in Section 6.3.

Regarding the **presentation of the suggestion**, the interface is closely aligned with that in Figure 6.3, with an emphasis on showcasing personalized suggestions from the system. It includes a bar graph offering a summary of consumer experiences with suggested items. Additionally, users can explore specific reviews for more detail. In this context, details about the home's amenities are not displayed.

6.4.4 FEATURES-STAGES (service-aware)

Model

This model merges details about item characteristics with insights into customer experiences from a service-oriented viewpoint, presenting a comprehensive item analysis to the user. It achieves this by averaging the rating predictions from both FEATURES and STAGES, thus blending feature-focused and service-focused recommendation approaches.

User interface

For the **collection of user preferences**, our experimentation utilized the user profiles generated through the interfaces of FEATURES and STAGES. These interfaces supply the necessary preference data for FEATURES-STAGES.

Regarding the **display of recommendations**, the user interfaces from FEATURES and STAGES are unified using a tabbed layout. This setup allows users to navigate and inspect both types of data seamlessly. The two homes in Figure 6.4 exemplify this integration.

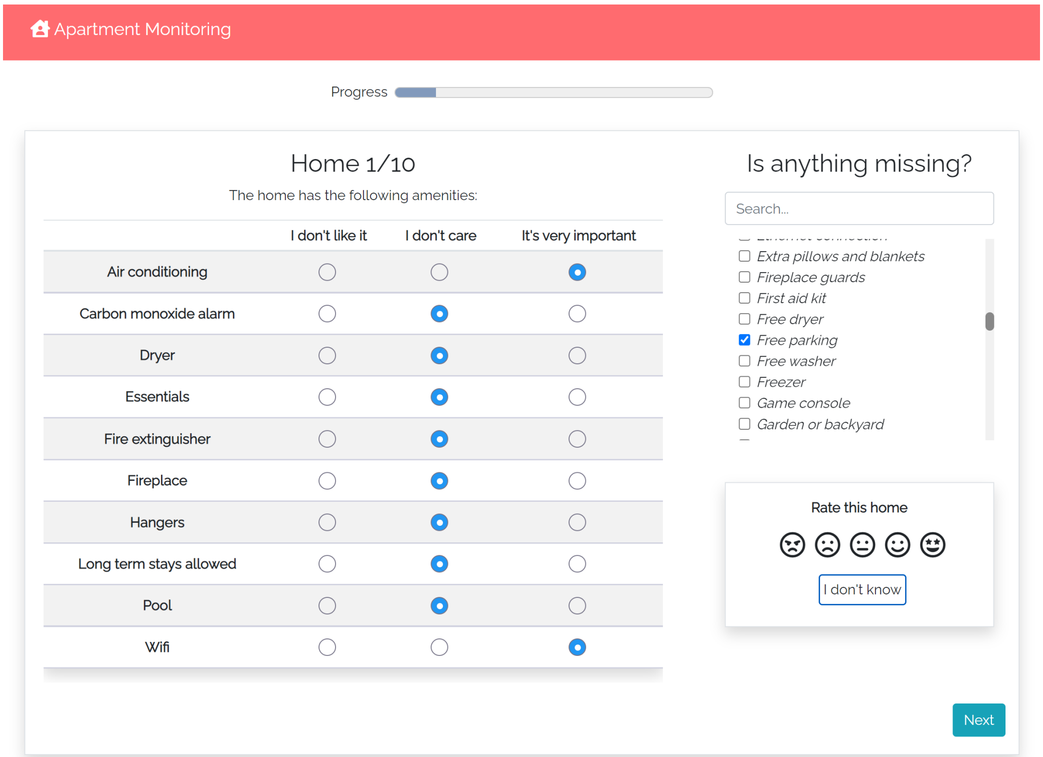
6.4.5 CBF-STAGES (service-aware)

Model

This algorithm integrates content-based filtering with service-aware recommendation techniques. It determines item ratings by calculating the arithmetic mean of the ratings predicted by CBF and STAGES.

User interface

The CBF-STAGES interface, for both acquiring user preferences and showcasing system recommendations, is identical to that of FEATURES-STAGES.



Home 1/10

The home has the following amenities:

	I don't like it	I don't care	It's very important
Air conditioning	○	○	●
Carbon monoxide alarm	○	●	○
Dryer	○	●	○
Essentials	○	●	○
Fire extinguisher	○	●	○
Fireplace	○	●	○
Hangers	○	●	○
Long term stays allowed	○	●	○
Pool	○	●	○
Wifi	○	○	●

Is anything missing?

Search...

- □ *Extra pillows and blankets*
- □ *Fireplace guards*
- □ *First aid kit*
- □ *Free dryer*
- ☑ *Free parking*
- □ *Free washer*
- □ *Freezer*
- □ *Game console*
- □ *Garden or backyard*

Rate this home

😠 ☹ 😐 🙂 🤩

Figure 6.5: User interface to collect the user's preferences in the FEATURES and CBF recommender systems.

6.5 User study

Our objective is to evaluate the effectiveness of the recommendation algorithms and their ability to aid in decision-making, as outlined in the five models presented in Section 6.4.

The user study was conducted using an interactive application that allowed participants to navigate through the experiment's stages autonomously, as detailed in Section 6.4. To glimpse this system's user interface, refer to Figure 6.5 through 6.4.

6.5.1 Participant recruitment and context

We reached out to potential participants through social media and email lists, specifically targeting those with prior experience using online platforms to book homes or hotels.

Participants voluntarily joined the study without any form of compensation, providing informed consent for their involvement.

Our user study was designed in adherence to established ethical guidelines for controlled experiments[1] (Kirk, 2013). Participants, through the test application's user interface, were informed about their rights, including the freedom to withdraw at any time, and inquire about the experiment's purpose and outcomes. Before the experiment, participants were required to: (i) read and understand a consent form outlining the nature of the experiment and their rights, (ii) formally acknowledge their understanding and agreement through the test application, and (iii) confirm they were 18 years or older. Uniform instructions were provided to all participants before the experimental tasks commenced. To maintain confidentiality, we did not collect names but

[1] https://www.tech.cam.ac.uk/research-ethics/school-technology-research-ethics-guidance/controlled-experiments

Table 6.3: Post-task questionnaire. Statements are grouped by user experience construct. Participants answered in the {Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree} scale.

Construct	Factor	Statement
Perceived Quality of Recommendations (Q)	Q1	The items recommended to me matched my interests.
	Q2	This system gave me good suggestions.
	Q3	The items recommended to me are similar to each other.
Perceived User-Awareness Support (U)	U1	This system explains why the products are recommended to me.
	U2	I understood why the items were recommended to me.
	U3	This recommender system made me more confident about my decision.
Interface Adequacy (I)	I1	The labels of this recommender system interface are clear.
	I2	Finding an item to book with the help of this recommender system is easy.
	I3	The information provided for the recommended items is sufficient for me to make a booking decision.

used anonymous codes throughout the study and in subsequent data analysis.

6.5.2 Procedure

Our user study employed a within-subjects design approach. We treated each experimental condition as an independent variable, with every participant experiencing all conditions. To reduce biases and mitigate the effects of learning or fatigue, we varied the sequence of tasks within the test application.

Recognizing the diversity in users' backgrounds and technological proficiency, we set no time constraints for completing the study. The study comprised three main phases:

1. Initially, the test application verified if users were at least 18 years old and gathered their consent for participation. Continuation in the test was contingent upon affirmative responses to these preliminary questions.[2] Subsequently, participants were prompted to complete a questionnaire gathering basic demographic data, cultural background, experience with booking platforms, and their general propensity to trust with limited information. This questionnaire was a modified version of the ResQue questionnaire for recommender systems (Pu et al., 2011).

2. In the next phase, the application focused on obtaining participants' preferences to create their user profiles. Since some recommendation models utilize the same interface for profile creation, the first model chosen automatically gathered user preferences and shared them with subsequent models. Participants rated ten homes at two separate times during the study, as depicted in Figures 6.5 and 6.3. They also evaluated homes in a recommendation list for each algorithm tested (shown in

[2] The consent form text is accessible here: `https://bit.ly/3jjYlEa`.

Table 6.4: Post-test questionnaire. Participants answered the questions in the {Very little, Little, I don't care, Important, Very important} scale.

#	Question
1	How much is the host appreciation important in your choices?
2	How much are check-in/check-out important in your choices?
3	How much is the in-apartment experience important in your choices?
4	How much are the surroundings important in your choices?
5	How much is the visualization of the amenities offered by the home (e.g. WiFi, washing machine, etc.) important in your choices?
6	How much is the visualization of the bar graph characterizing the home (e.g. host appreciation, surroundings, etc.) important in your choices?

Figures 6.2 and 6.4), each containing five homes to be rated for rental suitability using a star-based rating system.

Following each list evaluation, participants completed a post-task questionnaire, expressing their level of agreement with statements listed in Table 6.3. This questionnaire, derived from the ResQue framework, included statements classified under three constructs: *Perceived Quality of Recommendations* (Q), *Perceived User-Awareness Support* (U), and *Interface Adequacy* (I).

3. The final phase examined the hypothesis, similar to Tintarev and Masthoff (2012), that detailed information and feedback are crucial in high-investment domains like home-booking. Here, participants responded to a post-test questionnaire (Table 6.4) assessing the importance of amenities visualization and consumer experience summarization in evaluating the system's recommendations.

6.6 Experimental results

6.6.1 Demographic data and background

We performed a power analysis to determine the minimal number of participants needed for statistically significant results. This analysis involves four key parameters:

- *Alpha ($\alpha = 0.05$)*: This is the p value determining the threshold probability for rejecting the null hypothesis in the absence of a significant effect, representing the Type I error rate.

- *Power = 0.80*: The probability of correctly accepting the alternative hypothesis when it is true, which addresses the Type II error rate.

- *Effect size = 0.40*: The anticipated effect size, indicating the expected magnitude of a result within the population; our aim was to detect medium-sized effects.

- *Sample size N*: The necessary number of participants to achieve the desired statistical power. Our sample size estimation indicated that N = 42 participants would be required to maintain a statistical power of 80%.

Based on this analysis, we established a minimum sample size of N = 42. Consequently, we recruited 48 participants (N = 48) for the user study, conducted from May 15 to June 15, 2021. The average duration of the experiment was approximately 36.79 minutes, with a standard deviation of 19.83 minutes.

The demographic breakdown of the participants was as follows:

- Gender: 20 females, 28 males, 0 non-binary, 0 undisclosed.

- Age Distribution: 1 participant aged 18-20, 30 aged 21-30, 11 aged 31-40, 1 aged 41-50, 4 aged 51-60, and 1 older than 60.
- Education Level: 4 with high school education, 34 with university education, and 10 with a Ph.D. Backgrounds included 17 in technical fields, 22 in scientific fields, 5 in humanities and languages, 3 in economics, and 1 in another field.
- Computer Proficiency: 37 advanced users, 9 average users, and 2 beginners.
- Usage of E-commerce or Online Booking Services: 15 participants a few times a month, 8 between 1-3 times a week, 11 daily, and 14 a few times a year.
- Trust Propensity: 4 participants were very likely to trust with little knowledge, 15 somewhat likely, 23 somewhat unlikely, and 6 very unlikely to trust.

6.6.2 Evaluation of recommendation quality

The effectiveness of the recommendation algorithms was assessed primarily based on their ranking performance, as positioning high-quality solutions at the top of a recommendation list is crucial for user decision support. Additionally, we evaluated the accuracy of the algorithms in terms of minimizing rating estimation errors. The metrics we used for this evaluation are:

- **NDCG (Normalized Discounted Cumulative Gain)**: This metric evaluates the quality of ranking. It aggregates the gain of items from the top to the bottom of the list, applying a logarithmic discount to lower-ranked items.

- **RMSE (Root-Mean-Square Error)** and **MAE (Mean Absolute Error)**: These metrics are employed to calculate the error between the algorithm's predicted ratings and the actual ratings provided by participants in the user study.

- **Utility**: This metric provides an accuracy score for the entire list (not just individual items) based on user ratings. The value of suggestions diminishes at lower ranks in the list. The utility for a list of five items is calculated as follows:

$$Ut_u = \sum_{j=1}^{5} \frac{\max(r_{ui_j} - n, 0)}{2^{\frac{j-1}{\alpha-1}}} \tag{6.4}$$

  where r_{ui_j} is the rating a user u assigns to the item at the j^{th} position; n denotes the neutral vote (set to 3 in our study); α is a half-life parameter indicating the list position at which there is a 50% likelihood of the item being inspected and rated by the user. In our experiments, as all five items in the list were rated by the users, we set α to 5.

The results of these evaluations are presented in Table 6.6. We conducted a one-way ANOVA analysis to compare the performance of the algorithms, focusing on NDCG and Utility metrics since RMSE and MAE are calculated across the entire set of ratings, not per individual user. Our analysis revealed significant differences in both NDCG [F(232,4) $=$ 4.31; $p <$ 0.003] and Utility [F(232,4) $=$ 7.58; $p <$ 0.001] metrics.

Table 6.5: Post-task questionnaire results for each recommender system. For each mean value, the asterisks denote the statistical significance of the difference between the best-performing algorithm and the other ones. Significance levels: (***)$p < 0.001$, (**)$p < 0.05$.

Construct	Factor	Recommendation Algorithm				
		FEATURES Mean(SD)	CBF Mean(SD)	STAGES Mean(SD)	FEATURES-STAGES Mean(SD)	CBF-STAGES Mean(SD)
Perceived Quality of Recommendations (Q)	Q1	**4.31(0.72)**	3.83(0.97)	4.15(0.68)	4.27(0.71)	3.54(1.13)
	Q2	**4.29(0.62)**	3.85(0.99)	4.15(0.82)	4.25(0.84)	3.58(1.05)
	Q3	3.92(0.79)	3.88(0.87)	**3.94(0.76)**	3.85(0.74)	3.56(0.85)
	Mean	**4.17(0.73)**	3.85(0.94)	4.08(0.76)	4.12(0.78)	3.56(1.01)***
Perceived User-Awareness Support (U)	U1	3.29(1.20)	3.07(1.21)	3.05(1.17)	**3.62(1.09)**	3.49(0.92)
	U2	3.94(0.79)	3.94(0.92)	3.52(0.82)	**4.19(0.68)**	3.72(0.88)
	U3	3.60(1.05)	3.35(1.14)	3.50(0.90)	**3.81(0.89)**	3.54(1.07)
	Mean	3.61(1.04)	3.46(1.14)	3.36(0.99)**	**3.88(0.93)**	3.59(0.96)
Interface Adequacy (I)	I1	3.81(0.92)	3.74(1.03)	3.87(1.01)	3.96(0.81)	**3.98(0.82)**
	I2	3.60(1.09)	3.38(1.12)	3.69(0.78)	**3.92(0.87)**	3.69(0.99)
	I3	3.27(1.16)	3.02(1.34)	3.15(0.95)	**3.71(0.97)**	3.63(1.06)
	Mean	3.56(1.08)	3.38(1.20)**	3.57(0.96)	**3.86(0.88)**	3.76(0.97)

Following the initial analysis, we performed a *post-hoc* comparison using the Tukey HSD test to further investigate the differences between algorithms. Our findings are as follows:

- In terms of NDCG, STAGES emerged as the superior model, showing statistically significant improvements over CBF ($p < 0.05$) and CBF-STAGES ($p < 0.003$).
- With respect to the Utility metric, STAGES again outperformed the others, achieving significant results compared to CBF ($p < 0.01$) and CBF-STAGES ($p < 0.001$).
- When evaluating the minimization of error in rating estimation, FEATURES was identified as the most effective algorithm.

Additionally, the last column of Table 6.6 details the instances of opting-out (marked by "I don't know" ratings) during the home evaluations. This phenomenon was observed more frequently with CBF (10 instances), STAGES (9 instances), and FEATURES (6 instances). In contrast, CBF-STAGES and FEATURES-STAGES, which both present item reviews along with the ratings, did not record any "I don't know" responses.

6.6.3 Analysis of user feedback

User experience with the recommender systems

The user experience with each of the tested recommender algorithms was assessed using the results from the post-task questionnaire, as shown in Table 6.5, categorized by user experience constructs. A *post-hoc* comparison using the Tukey HSD test indicated:

Table 6.6: Recommendation performance of algorithms. The best results are in boldface. For each evaluation metric, (*) denotes different levels of the statistical significance of the difference between the best-performing algorithm, and the other ones. The last column shows the number of "I don't know" evaluations provided by participants when using the algorithms.

Algorithm	RMSE	MAE	NDCG	Utility	#Opting out
FEATURES	**0.6919**	**0.5170**	0.9792	5.1805	6
CBF	0.8857	0.7219	0.9669*	3.9482*	10
STAGES	0.8561	0.7393	**0.9847**	**5.3388**	9
FEATURES-STAGES	0.7225	0.5883	0.9736	4.7379	0
CBF-STAGES	0.9829	0.7900	0.9612*	3.4969*	0

- For *Perceived Quality of Recommendations* (Q), FEATURES scored highest (M=4.17, SD=0.73), showing significant superiority over CBF-STAGES. FEATURES-STAGES, with a mean score of 4.12 (SD=0.78), ranked second but performed best in other constructs.

- For *Interface Adequacy* (I), FEATURES-STAGES (M=3.86, SD=0.88) was superior, significantly outdoing CBF ($p < 0.05$).

- In terms of *Perceived User-Awareness Support* (U), FEATURES-STAGES led (M=3.88, SD=0.93), significantly outperforming STAGES ($p < 0.05$).

Furthermore, a one-way ANOVA was conducted to compare user experiences across the recommendation algorithms, revealing significant differences in all constructs:

- *Perceived Quality of Recommendations* (Q) [F(235,4) = 7.34; $p < 0.001$];

- *Interface Adequacy* (I) [F(235,4) = 2.53; $p < 0.05$];

- *Perceived User-Awareness Support* (U) [$F(235{,}4) = 2.69$; $p < 0.05$].

Structured equation model analysis

We utilized the Structured Equation Model analysis (Ullman and Bentler, 2012) to deepen the user experience with the five recommenders. This analysis helps reveal relationships between latent variables using observable variables. We associated two constructs from the post-task questionnaire 6.3 (*Perceived User-Awareness Support* and *Perceived Quality of Recommendations*) with Decision-making Support (DS) aspects; one construct (*Interface Adequacy*) with User Interfaces aspects, and tested five Algorithms (ALG) as dummy variables (CBF, FEATURES, STAGES, CBF-STAGES, and FEATURES-STAGES, shown in Figure 6.6).

The Confirmatory Factor Analysis confirmed the validity of these constructs:

1. For convergent validity, the Average Variance Extracted (AVE) of each construct exceeded 0.50.
2. For discriminant validity, the squared root of AVE was less than the correlation value for each construct.

The constructs met the criteria:

- *Perceived User-Awareness Support*: $AVE = 0.5463$, $\sqrt{AVE} = 0.7391$, largest correlation = 0.410.
- *Perceived Quality of Recommendations*: $AVE = 0.5913$, $\sqrt{AVE} = 0.7690$, largest correlation = 0.410.
- *Interface Adequacy*: $AVE = 0.5983$, $\sqrt{AVE} = 0.7735$, largest correlation = 0.337.

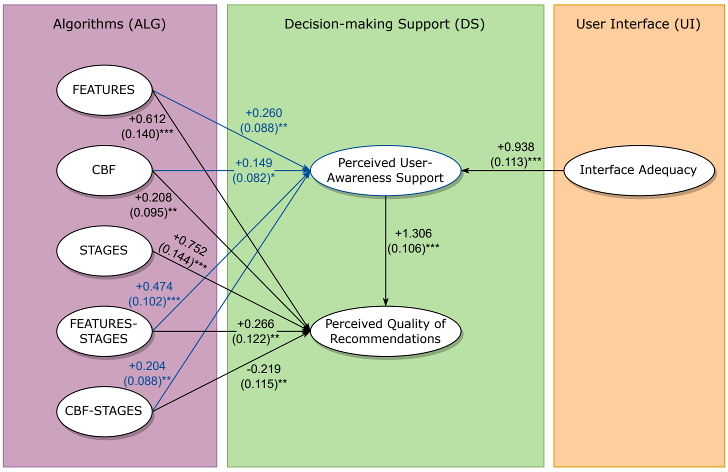


Figure 6.6: Structural Equation Model. Significance levels: (***)$p < 0.001$, (**)$p < 0.05$, (*)$p < 0.1$. The numbers on the arrows represent the β-coefficients (and standard error) of the effect.

Figure 6.6 illustrates the Structural Equation Model, highlighting dependencies, β-coefficients, and standard errors to represent the correlations between various constructs. The analysis reveals several noteworthy correlations:

- *Interface Adequacy* demonstrates a positive effect (+0.938; $p < 0.001$) on *Perceived User-Awareness Support*. This implies that the way items are presented in the system considerably influences user-awareness support.
- A positive correlation exists between *Perceived User-Awareness Support* and *Perceived Quality of Recommendations* (+1.306; $p < 0.001$), suggesting that comprehensive item information enhances users' perception of recommendation quality.

Notably, FEATURES-STAGES shows the strongest correlation with *Perceived User-Awareness Support* (+0.474; $p < 0.001$), likely due to its comprehensive presentation of amenities, consumer feedback, and reviews, which facilitates informed decision-making.

Interestingly, all algorithms, except for STAGES, positively influence *Perceived User-Awareness Support*. This finding suggests that relying solely on consumer feedback might not be sufficient for making rental decisions, as such feedback may not adequately assure that the home possesses the amenities needed by the user.

Regarding the *Perceived Quality of Recommendations*, all algorithms except CBF-STAGES positively correlate with this construct. The negative correlation of CBF-STAGES (-0.219; $p < 0.05$) is attributed to its lower performance in accuracy, ranking, and error estimation, as detailed in Section 6.6.2. Conversely, STAGES displays the highest correlation (+0.752; $p < 0.001$) with *Perceived Quality of Recommendations*, underscoring the value of consumer feedback in generating effective recommendations.

Post-test results

The outcomes of the post-test questionnaire are summarized in Table 6.7. Dimensions such as `In-apartment experience` and `Surroundings` were identified as most critical in decision-making. The importance of `Host` and `Check-in/Check-out` varied among participants. For information visualization, amenities were deemed more important than bar graphs, likely because users prioritize verifying that the selected homes offer the features they value.

6.7 Discussion

The findings from our user study shed light on the effectiveness of the models we evaluated, focusing on both item recommendation and result visualization

Table 6.7: Post-test questionnaire results.

Importance of dimensions in users' choices (number of users)					
	Very little	Little	I don't care	Important	Very important
Host	9	9	4	16	10
Check-in/Check-out	9	10	4	22	3
In-apartment experience	1	3	0	17	27
Surroundings	8	2	0	21	17
Importance of visualization of information in users' choices (number of users)					
Amenities	1	3	2	16	26
Bar graphs	3	6	4	24	11

aspects.

In terms of the recommendation algorithm, the measures indicate that STAGES, which bases its suggestions solely on the assessment of user experiences during item utilization stages, generate the most accurate item rankings. This result underscores the value of using experience evaluation dimensions in recommender systems. Meanwhile, FEATURES stands out in minimizing rating estimation errors. However, this aspect is secondary to our primary objective, which is to elevate the visibility of high-quality items in the recommendation lists.

Regarding the presentation, models that combine different types of information, specifically CBF-STAGES and FEATURES-STAGES, enhance user confidence in evaluating items. In contrast, models relying on a singular type of information, whether it be user experience data (STAGES) or features (CBF and FEATURES), encountered some instances of opting out by users. However, with the integrated visualizations in CBF-STAGES and FEATURES-STAGES, all users were able to rate the homes.

FEATURES-STAGES, effectively combines service-aware and feature-based data, emerges as the second-best algorithm in rating estimation, and also delivers commendable NDCG scores.

Regarding our research question, the combination of service-oriented item representation with feature data can enhance the accuracy of recommendations, according to the experiment results. Notably, the finest results were attained with STAGES, focusing solely on service-related item details. However, this approach occasionally led to some users feeling less confident about their decision-making. Therefore, a balanced approach combining consumer experience and item features in suggestion presentation, as implemented in FEATURES-STAGES, seems most effective. This system not only ranked second in performance but also experienced no opting-outs.

Moreover, users' perceptions after the interaction with the systems were analyzed. Concerning the *Perceived Quality of Recommendations* (Q), FEATURES was perceived as the top model in terms of generating the most relevant suggestions, probably from its alignment with user preferences, recommending homes that match the amenities identified as important during preference elicitation and accentuating these in the results display.

However, FEATURES-STAGES stands out in *Perceived User-Awareness Support* (U) and *Interface Adequacy* (I), which relate to users' understanding of the recommendation logic, awareness of suggestions, and confidence in decision-making. This superior perception can be attributed to its comprehensive display strategy, which includes amenities, bar graphs, and reviews, enabling a more effective analysis and comparison of potential homes than a simple presentation of amenities.

The outcomes of the Structural Equation Model reinforce these insights. The *Perceived User-Awareness Support* (U) positively affects the *Perceived Quality of Recommendations* (Q), indicating the necessity of ample item information for users to regard the recommendations as high-quality. Additionally, the *Perceived User-Awareness Support* (U) is positively impacted by

Interface Adequacy (I), implying that a more detailed presentation of item data enhances user confidence in evaluating available options.

Post-test questionnaire responses reveal a preference for data visualization regarding amenities over consumer experience summaries. However, combining this with the observation that FEATURES-STAGES is perceived as offering superior user-awareness support, we deduce that both amenities information and consumer feedback are crucial in decision-making processes.

These findings allow us to affirmatively respond to the second question: A recommender system that presents both item features and service-aware data improves user awareness of choices and confidence in decision-making. This enhancement stems from providing comprehensive information for evaluating both item features and the overall item experience.

Chapter 7

Enhancing the justification of results in service-aware recommender systems

7.1 Introduction

The complexity of consumer experience in modern service landscapes, as detailed in prior studies (Stickdorn et al., 2011), necessitates an advanced recommender system approach. For example, in sectors like home-booking, the experience extends beyond the tangible attributes of the product to include nuanced aspects like host interactions and shared space dynamics (Lee, 2022). Such complex consumer-service interactions demand a more sophisticated model for recommendation justification.

This chapter seeks to enhance user understanding and confidence in recommender systems by acknowledging the variability in service levels, as shown in research focusing on diverse service providers (Yi et al., 2020). We aim to provide users with a comprehensive view of recommended items, encompassing all aspects of their potential experiences.

The core of this chapter is the development of an advanced service-based justification approach for recommender systems. Building on the foundations laid in the previous chapter and in (Mauro et al., 2022b), this approach utilizes consumer feedback more effectively, extracting and employing both broad and nuanced dimensions of consumer experiences to justify recommendations.

In contrast to Service Journey Maps, which were previously discussed, we now integrate the Service Blueprints model (Bitner et al., 2008) into our methodology. This decision stems from the limitations of SJMs in their ability to classify the keywords into detailed evaluation dimensions of the service. SJMs, being linear and directly connecting stages to dictionaries, do not support the detailed classification needed for our analysis. Service Blueprints, on the other hand, offer a more detailed and structured analysis of service interactions, allowing for a more precise categorization of consumer reviews into distinct evaluation dimensions, thereby enriching the justification process in recommender systems.

The primary objective of this chapter is to assess the impact of this service-based justification on user perception and satisfaction. The research question we aim to address is RQ2: "How does service-based justification of recommendations influence user awareness and confidence in evaluating these recommendations, and what is its effect on user satisfaction regarding the presentation of item-related information in recommender systems?"

To address this question, we have developed models that use both coarse-grained and fine-grained evaluation dimensions from consumer feedback. These models present item aspects in distinct ways, but both aim to provide a thorough and incremental exploration of data based on various interests in the evaluation dimensions.

Following this introduction, we will detail the specifics of Service Blueprints, the methodology employed, and the developed justification models. Subsequent sections will present our findings, the structure of our user study, and its results. The chapter concludes with a discussion of the implications. The work in this chapter is described in detail in (Mauro et al., 2022a).

7.2 Defining evaluation dimensions

The first step in developing our service-based justification models involves defining the evaluation dimensions of consumer experiences.

At the forefront of our methodology is the task of identifying the key dimensions that evaluate a user's experience with service-based items. While the SJM described in Section 6.2 supports the definition of coarse-grained evaluation dimensions, we are also interested in defining finer-grained evaluation dimensions to be used in a detailed evaluation of items. For this reason, this process begins with the creation of a Service Blueprint, tailored to encapsulate the user experience in the home-booking context. Our blueprint, inspired by and extending upon the works of Bitner et al. (2008), Ren et al. (2016), and Cheng and Jin (2019), offers a comprehensive view of customer interactions and expectations in home-booking scenarios, particularly focusing on Airbnb.

In our quest to construct an effective justification model for recommendations, our attention is primarily centered on two critical layers of the Service Blueprint: the Customer Actions and Physical Evidence layers. These layers are fundamental as they chronologically map out the user's journey, from initial interaction with the Airbnb website to the final steps of check-out, encompassing all tangible and intangible aspects of the experience.

For instance, the Customer Actions layer describes the user's journey,

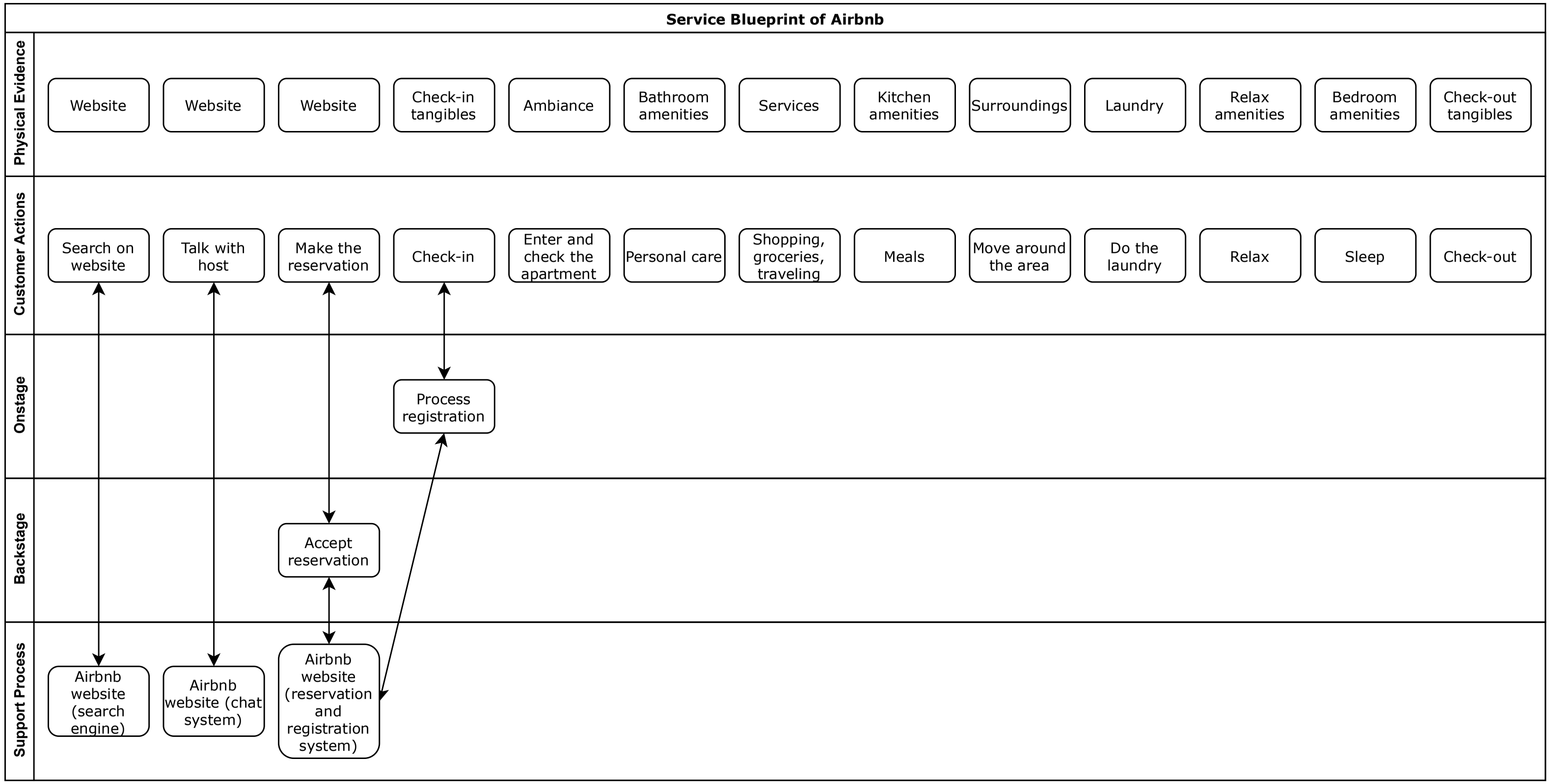


Figure 7.1: Service Blueprint we defined to describe the home-booking domain (Airbnb).

Table 7.1: Coarse-grained and fine-grained evaluation dimensions with references to physical evidence and keywords.

Coarse-grained dimensions	Fine-grained dimensions	Physical Evidence	Dictionary
Host appreciation	Host	-	advice, communication, host, tip, ...
Check-in/Check-out	Check-in	Check-in tangibles	arrival, access, check-in, wait, key, ...
	Check-out	Check-out tangibles	check-out, departure, goodbye, ...
In-apartment experience	Ambiance	Ambiance	air conditioning, atmosphere, smell, ...
	Bathroom	Bathroom amenities	towel, shower, soap, hair-dryer, ...
	Kitchen	Kitchen amenities	kitchen, fridge, microwave, oven, ...
	Laundry	Laundry	dryer, ironing board, washer, ...
	Relax	Relax amenities	balcony, wi-fi, tv, swimming pool, ...
	Bedroom	Bedroom amenities	bed, pillow, wardrobe, blanket, ...
Surroundings	Surroundings	Surroundings	attraction, gym, lake, street, sunset, ...
	Services	Services	transportation, atm, bus, grocery, ...

including activities like checking in, engaging in local activities, and the eventual check-out process. Each of these actions involves interaction with various tangible elements and, at times, the host, which all impact guest satisfaction.

To holistically represent these experiences, we introduce an additional layer in our blueprint that correlates Customer Actions with specific evaluation dimensions. This approach allows us to capture the nuances of the user experience, associating each step of the customer journey with relevant experience metrics.

Our model delineates two primary types of evaluation dimensions:

- **Fine-grained Evaluation Dimensions:** These dimensions concern the specifics of user interactions with tangible elements and human actors at each journey step. This granularity allows us to capture detailed aspects of the experience, such as the guest's perception of the host or the ambiance of the accommodation. For instance, the check-in process may be a critical dimension if a guest encounters issues with a late host.

- **Coarse-grained Evaluation Dimensions:** In contrast, these dimensions offer a broader summary of the user experience, aggregating multiple fine-grained dimensions. They provide a general view of the experience, like summarizing all aspects of in-apartment interactions under a single dimension and corresponding to the dimensions extracted from the Service Journey Map in Section 6.2.

As derived from our Service Blueprint, mapping these evaluation dimensions effectively covers all elements of the Physical Evidence layer, ensuring a comprehensive representation of the user's journey in the home-booking service context. Table 7.1 shows these two types of evaluation dimensions.

7.3 Extraction and organization of item aspects

In this section, we detail the process of extracting and categorizing aspects of homes from customer reviews. This method is an adaptation of the approach of the previous chapter. In this version, we focus on classifying according to the coarse-grained and fine-grained evaluation dimensions of the user experience.

1. **Aspect-Adjective Pair Analysis:** We do the same steps described in Section 6.3.1. From the reviews of a particular item $i \in \Delta$ (REV_i), we extract occurrences of aspect-adjective pairs. Different from the previous chapter, we divide the review into sentences, and for every identified pair, we generate a tuple $< aspect, asp\#r, adjective, asp\_adj\#r, evaluation >$. This tuple structure helps in quantifying and normalizing consumer opinions, where:

- $asp\#r$ indicates the count of review $r \in REV_i$ mentioning the aspect.
- $asp\_adj\#r$ is the number of review $rinREV_i$ that include the specific aspect-adjective combination.
- *evaluation*, is calculated as the previous chapter (see Section 6.3.1).

This approach allows for a structured representation of consumer feedback, moving beyond simple frequency counts.

2. **Classification of Aspects:** The next step involves classifying the extracted aspects into relevant fine-grained evaluation dimensions. This classification uses entity recognition to pinpoint references to specific entities like people or places.

   Additionally, we utilize the thesauri we used in the previous chapter, described in Section 6.2, which contain terms related to each dimension. To adapt these thesauri for this work, we sectioned the thesauri for the new fine-grained dimensions. An example of these thesauri is shown in Table 7.1. This classification links aspects to specific interaction stages, facilitating a granular organization of feedback and the summarization of consumer experiences.

3. **Aggregation and Evaluation of Dimensions:** We calculate the score for each coarse-grained evaluation dimension by computing the weighted mean of evaluations for aspect-adjective pairs classified under that dimension. The weight for each pair is determined by its frequency of mention ($asp\_adj\#r$), ensuring that the evaluation reflects the prevalence of user opinions. In cases where a dimension lacks data, we assign it a value of 0, indicating an absence of knowledge about that aspect.

Table 7.2: Sample aspects extracted from the reviews of a sample Airbnb home.

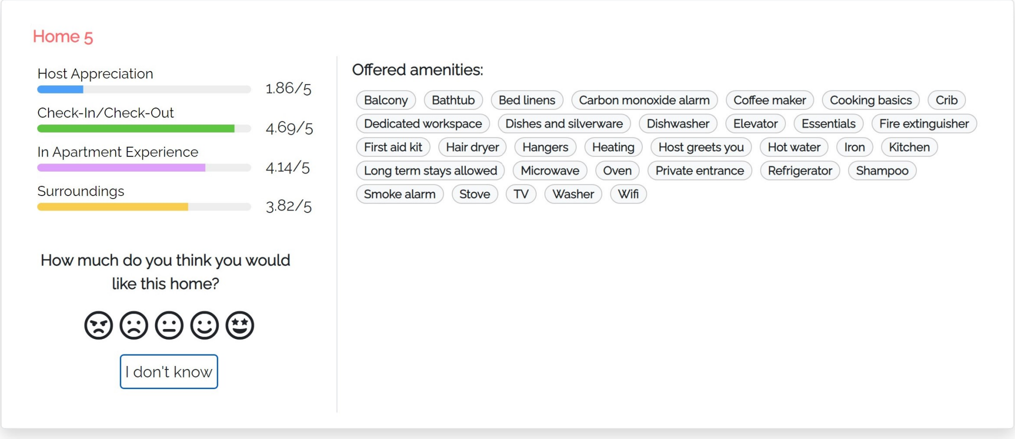
aspect	asp#r	adjective	asp_adj#r	evaluation	dimension
location	23	great	6	4.42	ambiance
location	23	excellent	2	4.57	ambiance
location	23	good	2	4.14	ambiance
location	23	convenient	1	3.00	ambiance
host	22	great	7	4.42	host-prop
host	22	friendly	4	3.87	host-prop
host	22	excellent	2	4.57	host-prop
host	22	lovely	2	4.09	host-prop
place	9	lovely	3	4.09	ambiance
place	9	great	2	4.42	ambiance
place	9	airy	1	3.00	ambiance
bed	4	comfortable	2	3.91	bedroom
bed	4	superb	1	4.62	bedroom
restaurant	4	cool	1	3.67	surroundings
restaurant	4	lovely	1	4.09	surroundings
restaurant	4	nice	1	4.02	surroundings

Table 7.2 illustrates an example of the results of this analysis, showing how data is aggregated according to the fine-grained evaluation dimensions. Additionally, during this analysis, we index review sentences based on aspect-adjective pairs to facilitate their retrieval for justifying recommendations.

7.4 Service-based justification models

We propose two justification models: M-THUMBS and M-ASPECTS. Below, we describe the key components of our service-based justification models as they appear in the user interface.

- **Central Display Area:** This section highlights the essential features of the item under review. In the context of home booking, this includes



Home 5

Host Appreciation 1.86/5

Check-In/Check-Out 4.69/5

In Apartment Experience 4.14/5

Surroundings 3.82/5

How much do you think you would like this home?

😠 ☹ 😐 🙂 😍

I don't know

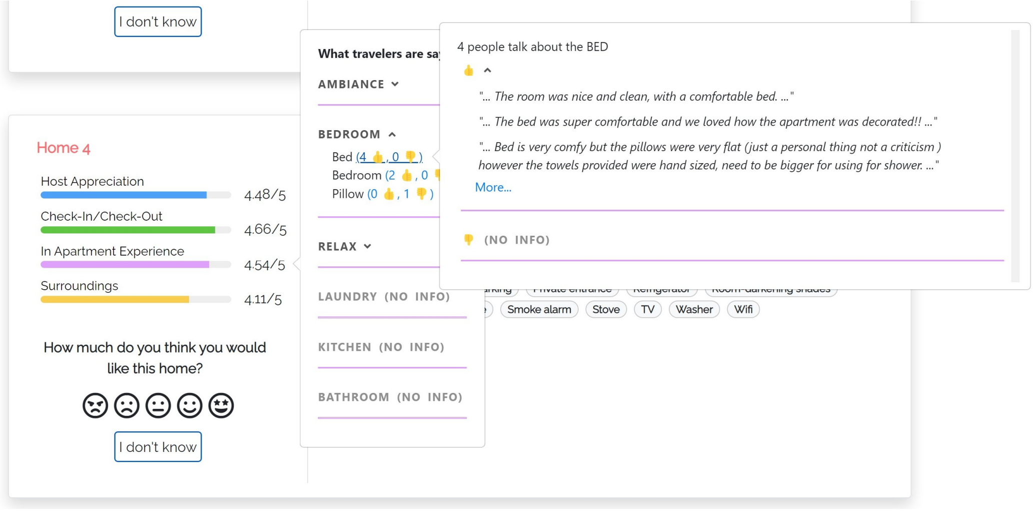
Offered amenities:

Balcony, Bathtub, Bed linens, Carbon monoxide alarm, Coffee maker, Cooking basics, Crib, Dedicated workspace, Dishes and silverware, Dishwasher, Elevator, Essentials, Fire extinguisher, First aid kit, Hair dryer, Hangers, Heating, Host greets you, Hot water, Iron, Kitchen, Long term stays allowed, Microwave, Oven, Private entrance, Refrigerator, Shampoo, Smoke alarm, Stove, TV, Washer, Wifi

Figure 7.2: Snapshot of the user interface used in our justification models.

amenities that are binary (present or absent) and are displayed accordingly. Similar to the previous experiment, following recommendations by Tintarev and Masthoff (2022), we exclude potentially bias-inducing information like price, room count, images, and property names, which can affect user judgment. For instance, some homes have long names that mention their location or the view: “Beautiful Flat - Near London Eye”, or “Penthouse, huge terrace near Picadilly Circus”.

- **Consumer Experience Summary:** On the left side of the interface, colored bar graphs represent the summarized experiences of previous consumers, similar to the interfaces of the previous chapter. Each bar graph corresponds to a different coarse-grained evaluation dimension such as `Host Appreciation`, `Check-in/Check-Out`, `In Apartment Experience`, and `Surroundings` and reflects the evaluations assigned in the aspect extraction process. A greyed-out bar graph indicates no available feedback, distinguishing it from a low evaluation score. Users can click on these graphs for more detailed insights into the consumer experiences.



"... The room was nice and clean, with a comfortable bed. ..."

"... The bed was super comfortable and we loved how the apartment was decorated!! ..."

"... Bed is very comfy but the pillows were very flat (just a personal thing not a criticism) however the towels provided were hand sized, need to be bigger for using for shower. ..."

Figure 7.3: Interface of the justification M-THUMBS model.

- **Interactive Rating Component:** Additionally, a rating component, displayed as a range of emoticons, allows users to rate the item on a scale of 1 to 5. This feature, while not directly contributing to our study's recommendation performance, serves dual purposes: attracting user attention to the presented data and gathering implicit feedback on user confidence in evaluating the items. An "I don't know" option is included to skip the evaluation.

Next, we detail the unique features of our two justification models, highlighting the distinct information provided when users interact with the bar graphs.

7.4.1 M-THUMBS model

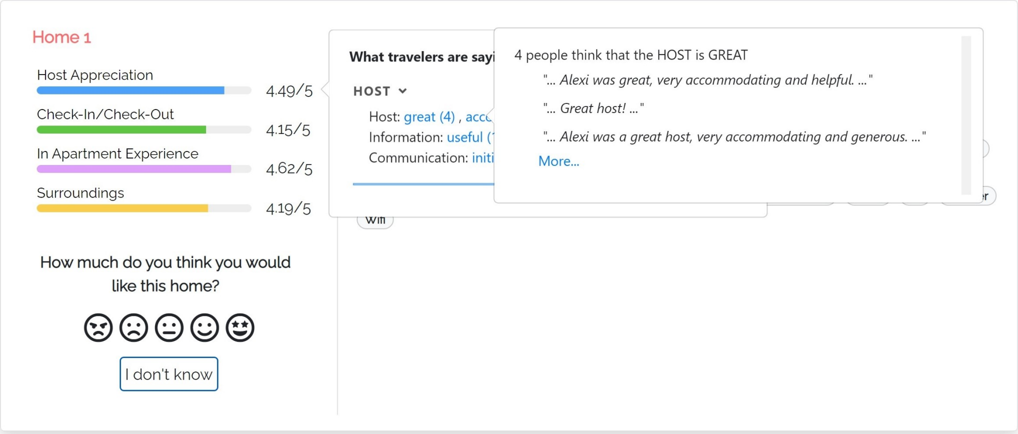
In the M-THUMBS model (illustrated in Fig. 7.3), user interaction with the coarse-grained dimension bar D triggers the display of the "What travelers are saying" component. This section showcases fine-grained dimensions $d \in D$, such as “AMBIANCE” under the “In Apartment Experience” category. These dimensions are ordered based on the frequency of aspect mentions in the item’s reviews. Dimensions with no aspect mentions are shown in light grey and marked with a “NO INFO” tag, indicating they cannot be expanded.

The model prioritizes the display of aspects within each fine-grained dimension, sorted by their relevance (number of mentions, denoted as *asp#r* in Table 7.2). When a dimension is expanded, up to three of the most relevant aspects are shown, along with an option to view the full list. Each aspect is accompanied by a thumbs-up or thumbs-down icon, indicating positive or negative feedback based on the reviews, derived from Table 7.2. Clicking on these icons allows users to view specific review quotes related to the aspect.

7.4.2 M-ASPECTS model

The M-ASPECTS model (see Fig. 7.4) shares the “What travelers are saying” component’s organization with the M-THUMBS model. However, it differs in how it presents information for each fine-grained dimension. Here, aspects are associated with a list of the most relevant adjectives as per their mentions in the item reviews.

The relevance of an adjective is determined by the *asp_adj#r* value of its aspect-adjective pair in Table 7.2. Each adjective, along with its relevance value, is clickable, enabling users to access quotes from reviews that mention this specific adjective. This feature provides a more detailed perspective on consumer feedback, focusing on the descriptive quality of the reviews.



Home 1

Host Appreciation 4.49/5

Check-In/Check-Out 4.15/5

In Apartment Experience 4.62/5

Surroundings 4.19/5

How much do you think you would like this home?

What travelers are sayi

Host: great (4) , accc

Information: useful (1

Communication: initi

4 people think that the HOST is GREAT

"... *Alexi was great, very accommodating and helpful. ...*"

"... *Great host! ...*"

"... *Alexi was a great host, very accommodating and generous. ...*"

Figure 7.4: Interface of the M-ASPECTS model.

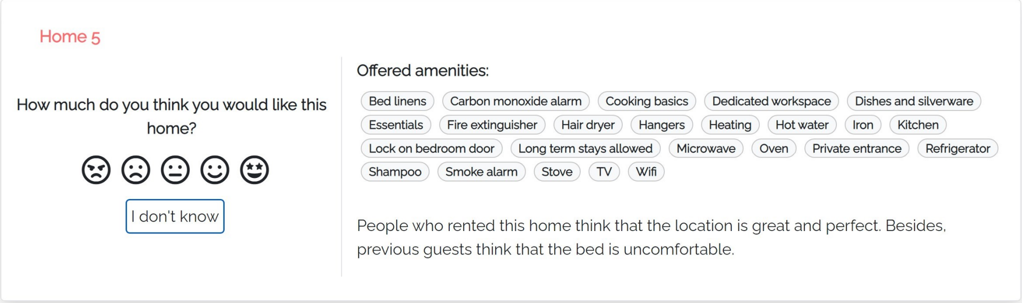
7.5 Baseline models

In contrast to our service-based justification models, we introduce a set of baselines to provide a comparative perspective. These models adopt more conventional approaches to item representation and user interaction, akin to what is typically seen in e-commerce and home-booking platforms. By evaluating these baselines against our proposed service-based models, we aim to highlight the advantages and potential shortcomings of each approach. Below, we detail two such baseline models, M-SUMMARY, M-OPINIONS, and M-REVIEWS.

7.5.1 M-SUMMARY Model

The M-SUMMARY model (illustrated in Fig. 7.5) offers a streamlined approach to presenting item features and review summaries. This model concisely summarizes the item's features, focusing on the most prominent aspects and adjectives derived from the item's reviews.

Similar to (Musto et al., 2021), the M-SUMMARY model employs a Backus-



Home 5

How much do you think you would like this home?

Offered amenities:

People who rented this home think that the location is great and perfect. Besides, previous guests think that the bed is uncomfortable.

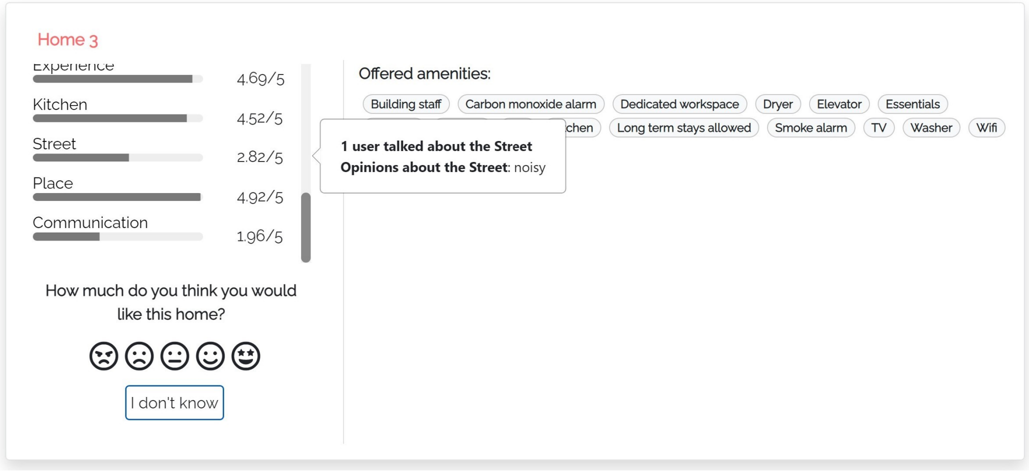
Figure 7.5: Interface of the M-SUMMARY model.

Naur Form (BNF) grammar to generate varied textual summaries dynamically. This system ensures a diverse range of sentence constructions. The selection of aspects and adjectives for inclusion in these summaries is based on their relevance, measured by the frequency of their mention in reviews ($asp\#r$ and $asp\_adj\#r$). Unlike methods that rely on Kullback–Leibler divergence (KL) for relevance determination, our approach prioritizes the direct frequency of mentions. This decision derives from the limitation of KL in accommodating the wide range of expressions commonly used by guests in their reviews, particularly those terms that are important in describing home experiences but might be absent from standard dictionaries.

7.5.2 M-OPINIONS Model

The M-OPINIONS model (depicted in Fig. 7.6) enhances the item feature display by incorporating an evaluative summary of the most pertinent aspects derived from item reviews ($asp\#r$), organized in order of their significance.

In this model, each aspect is represented by a gray bar graph, accompanied by a numerical rating within the range of 1 to 5. Clicking on these bars unveils detailed information, such as the frequency of mentions of the aspect and the descriptive adjectives used by guests. The numerical rating for each aspect is



Home 3

Experience 4.69/5

Kitchen 4.52/5

Street 2.82/5

Place 4.92/5

Communication 1.96/5

How much do you think you would like this home?

Offered amenities:

Building staff, Carbon monoxide alarm, Dedicated workspace, Dryer, Elevator, Essentials, chen, Long term stays allowed, Smoke alarm, TV, Washer, Wifi

1 user talked about the Street

Opinions about the Street: noisy

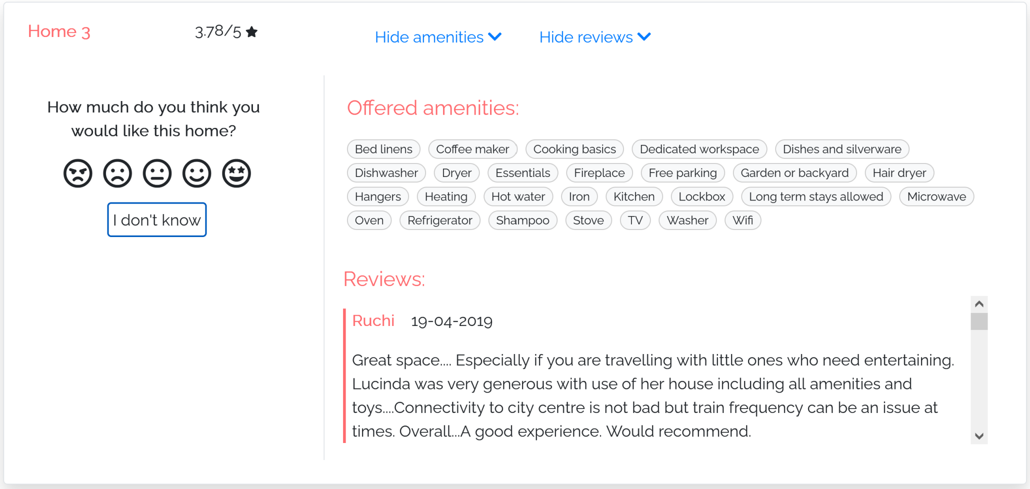
Figure 7.6: Interface of the M-OPINIONS model.

derived as a weighted average of the scores of related aspect-adjective pairs (from Table 7.2), where the frequency of mentions ($asp\_adj\#r$) serves as the weight.

This model's design parallels the concept of the opinion bar chart by Chen et al. (2014), but with a simpler bar chart format to align with the aesthetic of our service-based justification models.

7.5.3 M-REVIEWS Model

The M-REVIEWS model (illustrated in Fig. 7.7) presents a conventional approach, similar to what is typically seen on e-commerce and home-booking platforms. It displays standard information such as the item's features, the average rating given by consumers, and their reviews.



How much do you think you would like this home?

😠 ☹ 😐 🙂 😍

Offered amenities:

Bed linens · Coffee maker · Cooking basics · Dedicated workspace · Dishes and silverware · Dishwasher · Dryer · Essentials · Fireplace · Free parking · Garden or backyard · Hair dryer · Hangers · Heating · Hot water · Iron · Kitchen · Lockbox · Long term stays allowed · Microwave · Oven · Refrigerator · Shampoo · Stove · TV · Washer · Wifi

Reviews:

Ruchi 19-04-2019

Great space.... Especially if you are travelling with little ones who need entertaining. Lucinda was very generous with use of her house including all amenities and toys....Connectivity to city centre is not bad but train frequency can be an issue at times. Overall...A good experience. Would recommend.

Figure 7.7: Interface of the M-REVIEWS model.

7.6 User study

To assess the user experience with our justification models detailed in 7.4 with the baselines (Section 7.5), particularly focusing on their effectiveness and satisfaction (Tintarev and Masthoff, 2022), we conducted a comprehensive user study. This study also sought to understand how our service-based models compare in helping users make informed decisions and enhancing their user experience.

In line with the findings of Tsai and Brusilovsky (2021), we recognize that users often explore both high-ranked and lower-ranked items in recommendation lists. Therefore, our study needed to present a range of items, spanning from high to low quality, ensuring that our models could accurately represent each item's attributes, regardless of their ranking.

7.6.1 Participant recruitment and context

Similar to the previous experiment detailed in Section 6.4.5, we initiated the participant recruitment process by distributing an invitation across various public mailing lists and social networks. The invitation specifically mentioned a preference for individuals with prior experience in using online booking or e-commerce platforms. Participation in the study was entirely voluntary and offered no monetary incentives.

The study was conducted using an interactive web application. This application was designed to guide participants through the study's various stages intuitively. To protect participant privacy, no personal identifying information was collected. Instead, each participant was assigned a unique numerical identifier to anonymize the data collected during their session.

7.6.2 Procedure

Our user study adopted a within-subjects design similar to our previous experiment (see Section 6.5.2 for the details). Each participant experienced all treatment conditions, which were managed as independent variables. To mitigate the effects of order bias, fatigue, and practice, the sequence of tasks was counterbalanced by the test application. No time restrictions were imposed during the test, allowing participants ample freedom to interact with and explore the provided information. The study was structured as follows:

1. **Informed Consent and Age Verification:** see Section 6.5.2.

2. **Demographic Questionnaire and Personal Characteristics:** The demographic questionnaire where similar to the Section 6.5.2.

   Additionally, we collected data on Personal Characteristics (PC), focusing on Trust in Booking Systems and General Trust in Technology (Tsai and Brusilovsky, 2021), as shown in Table 7.3. Participants responded to these items on a 5-point Likert scale (Strongly Disagree to Strongly Agree), which we converted to a numerical scale of [1, 5]. The table also includes the mean values of participant responses.

3. **Interaction with Justification Models:** Participants were then presented with the justification models described in Sections 7.4 and 7.5, in a counterbalanced order. For each model, they were asked to explore and rank five homes. Post-interaction, participants completed a questionnaire (Table 7.4) derived from Pu et al. (2011); Di Sciascio et al. (2019); Lewis and Sauro (2009), assessing their experience with each model. This questionnaire, also using a 5-point Likert scale, focused on three constructs: Perceived User Awareness Support, Interface Adequacy, and Satisfaction. These constructs are integral to our Structural Equation Model analysis discussed in Section 7.7.4.

During the study, participants also responded to the Curiosity and Exploration Inventory-II (CEI-II) (Kashdan et al., 2009) and the Need for Cognition questionnaire (Coelho et al., 2020). CEI-II helped gauge participants' motivation for seeking knowledge and new experiences, including their openness to novel and unpredictable aspects of daily life. The Need for Cognition questionnaire assessed participants' propensity for engaging in and enjoying thoughtful activities.

7.7 Experimental results

Our user study, conducted between November 15 and December 15, 2021, involved 66 individuals. However, we filtered out 7 participants for failing attention checks, leading to a final count of 59.

The participants are considered sufficient for statistical significance with $\alpha = 0.05$, *power* = 0.80, and *effect size* = 0.35, as determined by power analysis. The average duration of the experiment was approximately 35.45 minutes.

7.7.1 Demographic and background information

The demographic and background details of the participants were as follows:

- **Gender and Age Distribution:** The participant pool comprised 24 females, 33 males, 2 who preferred not to disclose their gender, and no non-binary individuals. Their ages were distributed as follows: under 20 years (2 participants), 21-30 years (43), 31-40 years (7), 41-50 years (2), 51-60 years (4), and over 60 years (1 participant).

- **Educational and Professional Background:** Of the participants, 13 had completed high school, 40 had university degrees, and 6 held PhDs. Their professional backgrounds varied: 17 in technical fields, 31 in science, 6 in humanities and languages, 2 in economics, and 4 in other areas. Regarding computer proficiency, 46 considered themselves advanced users, 10 average, and 3 beginners.

- **Experience with Online Platforms:** In terms of familiarity with online booking or e-commerce platforms, 18 participants used these platforms several times a week, 26 a few times a month, 14 occasionally in a year, and 1 had never used such platforms.

- **Trust in Booking Systems (PC1):** As indicated in Table 7.3, there was moderate agreement among participants in trusting booking system recommendations (statement 1: Mean = 3.10, SD = 0.82). Most felt the need to read through home reviews (statement 3: Mean = 4.12, SD

Table 7.3: Questionnaire about personal characteristics and mean values of the answers.

Construct	Factor	Statement	M(SD)
Trust in *Booking Systems* (PC1)	1	I tend to trust the suggestions generated by booking systems.	3.10(0.82)
	2	I think that the ratings given by other users are enough to book homes.	3.19(0.86)
	3	I need to inspect the reviews given by other users to book homes.	4.12(0.74)
	4	I need to inspect the description of the home to book it.	4.31(0.70)
Trust in Technology (PC2)	1	I feel technology never works.	1.66(0.58)
	2	I'm less confident in doing things when I use supporting technology.	1.80(0.94)
	3	The usefulness of technology is highly overrated.	1.92(0.86)
	4	I tend to trust a person/thing, even though I have little knowledge of it.	2.83(0.89)

Table 7.4: Post-task questionnaire. Statements are grouped by user experience construct.

Construct	Factor	Statement
Perceived User Awareness Support (U)	U1	The information provided was sufficient for me to understand what previous users think about the homes.
	U3	The information about the homes was easy to interpret and understand.
	U4	I quickly found the information about the homes.
Interface Adequacy (I)	I1	It was easy to understand why some homes were good and others not.
	I2	I found the user interface very intuitive.
	I3	The user interface was sufficiently informative.
Satisfaction (S)	S1	I think that I would like to frequently use this system to evaluate homes.
	S3	I thought this system to evaluate homes was easy to use.
	S4	I felt very confident using this system to evaluate homes.

= 0.74) and descriptions (statement 4: Mean = 4.31, SD = 0.70) before booking, but were only somewhat reliant on user ratings (statement 2: Mean = 3.19, SD = 0.86).

- **General Trust in Technology (PC2):** The participants generally viewed technology favorably. Statements questioning technology trust received low mean scores. However, a sense of skepticism was noted towards unfamiliar people or technologies (statement 4: Mean = 2.83, SD = 0.89).

7.7.2 Evaluating the justification models' impact on users

The responses from the post-task questionnaire (Table 7.5) provided insightful data on user experiences with the different justification models. A Kruskal-Wallis test indicated significant variances in user experience aspects across the five models:

- *Perceived User Awareness Support* [$H = 13.40$, $df = 4$, $p < 0.008$];
- *Interface Adequacy* [$H = 10.21$, $df = 4$, $p < 0.035$];
- *Satisfaction* [$H = 8.07$, $df = 4$, $p < 0.084$].

Subsequent analysis, using the Mann-Whitney test for post-hoc comparisons, revealed the following insights:

- **Perceived User Awareness Support:** Model M-THUMBS, with a mean score of 3.66 and a standard deviation of 1.05, stood out as the most effective model, significantly outperforming M-SUMMARY, M-OPINIONS, and M-REVIEWS. Participants found M-THUMBS superior in terms of clarity and ease of understanding information about homes (U3) and efficiency in locating relevant data (U4). Contrastingly, M-REVIEWS, which does not summarize reviews, ranked lowest in these aspects. Despite this, M-REVIEWS was rated highest for understanding previous guests' opinions about the homes (U1), with M-THUMBS closely following.
- **Interface Adequacy:** Also in this model, M-THUMBS (Mean = 3.52, SD = 1.10) was perceived as the top model, particularly in helping users discern the pros and cons of homes. It surpasses both M-SUMMARY

and M-OPINIONS in statistical significance. However, M-SUMMARY was considered the most intuitive due to its simplified text-based summary of information.

- **Satisfaction:** M-THUMBS (Mean = 3.47, SD = 1.12) emerged as the leading model in terms of user satisfaction, significantly surpassing both M-SUMMARY and M-OPINIONS. Participants expressed a preference for using M-THUMBS regularly for home evaluations and reported higher confidence in using it compared to other models. Interestingly, M-SUMMARY was seen as the easiest to use, likely due to its minimalistic interface, with M-THUMBS following closely in ease of use.

7.7.3 Participant decision confidence

In our study, the 59 participants collectively rated 295 homes. An analysis of their decision-making confidence, indicated by "I don't know" responses, produced the following opting-out rates:

- M-THUMBS: 10 instances (3.39%);
- M-ASPECTS: 15 instances (5.08%);
- M-SUMMARY: 30 instances (10.17%);
- M-OPINIONS: 33 instances (11.19%);
- M-REVIEWS: 32 instances (10.85%).

These opting-out rates align with the perception of M-THUMBS as the most effective model in terms of *Perceived User Awareness Support.* Higher opting-out rates for M-SUMMARY and M-REVIEWS suggest a lack of confidence

Table 7.5: Post-task questionnaire results describing participants’ experience with the justification models. Results are grouped by user experience construct. For each construct, three rows show the values obtained for the questions of Table 7.4 (factors). The “Average” row reports the mean value of the factors. The highest values are in boldface. Stars denote the statistical significance of the difference between the best-performing model and the other ones. Significance levels: (***) $p < 0.01$, (**) $p < 0.05$, (*) $p < 0.1$.

Construct	Factor	Justification Model				
		M-THUMBS	M-ASPECTS	M-SUMMARY	M-OPINIONS	M-REVIEWS
		M(SD)	M(SD)	M(SD)	M(SD)	M(SD)
Perceived User Awareness Support	U1	3.61(0.95)	3.44(1.10)	2.59(1.12)	3.08(1.16)	**3.73(1.08)**
	U3	**3.58(1.04)**	3.42(1.04)	3.56(1.10)	3.07(1.19)	3.07(1.22)
	U4	**3.80(1.16)**	3.53(1.09)	3.63(1.07)	3.20(1.17)	2.80(1.28)
	Average	**3.66(1.05)**	3.46(1.07)	3.26(1.19)**	3.12(1.17)***	3.20(1.25)**
Interface Adequacy	I1	**3.46(1.06)**	3.34(1.14)	2.98(1.18)	3.12(1.13)	2.93(1.26)
	I2	3.46(1.25)	3.36(1.06)	**3.81(1.01)**	3.44(1.10)	3.53(1.02)
	I3	**3.64(0.98)**	3.47(1.01)	2.41(1.02)	3.14(1.14)	3.39(0.89)
	Average	**3.52(1.10)**	3.39(1.07)	3.07(1.21)***	3.23(1.13)*	3.28(1.09)
Satisfaction	S1	**3.29(1.22)**	3.10(1.18)	2.58(1.10)	2.86(1.17)	3.00(1.08)
	S3	3.61(0.98)	3.41(0.89)	**3.85(0.96)**	3.31(1.00)	3.51(1.02)
	S4	**3.51(1.14)**	3.37(0.91)	2.90(1.21)	3.05(1.09)	3.27(1.05)
	Average	**3.47(1.12)**	3.29(1.01)	3.11(1.22)**	3.07(1.10)**	3.26(1.07)

in making evaluations with these models. This might be attributed to M-SUMMARY ’s limited representation of guest opinions and M-REVIEWS ’s requirement for users to read through full reviews, whereas M-THUMBS balances summary and detailed feedback effectively.

7.7.4 Structural equation model analysis

Also in this case, we conducted a Structural Equation Model (SEM) analysis to explore the deeper relationships between various unobserved constructs (latent variables) and observable variables.

We linked two constructs—*Perceived User Awareness Support* and *Inter-*

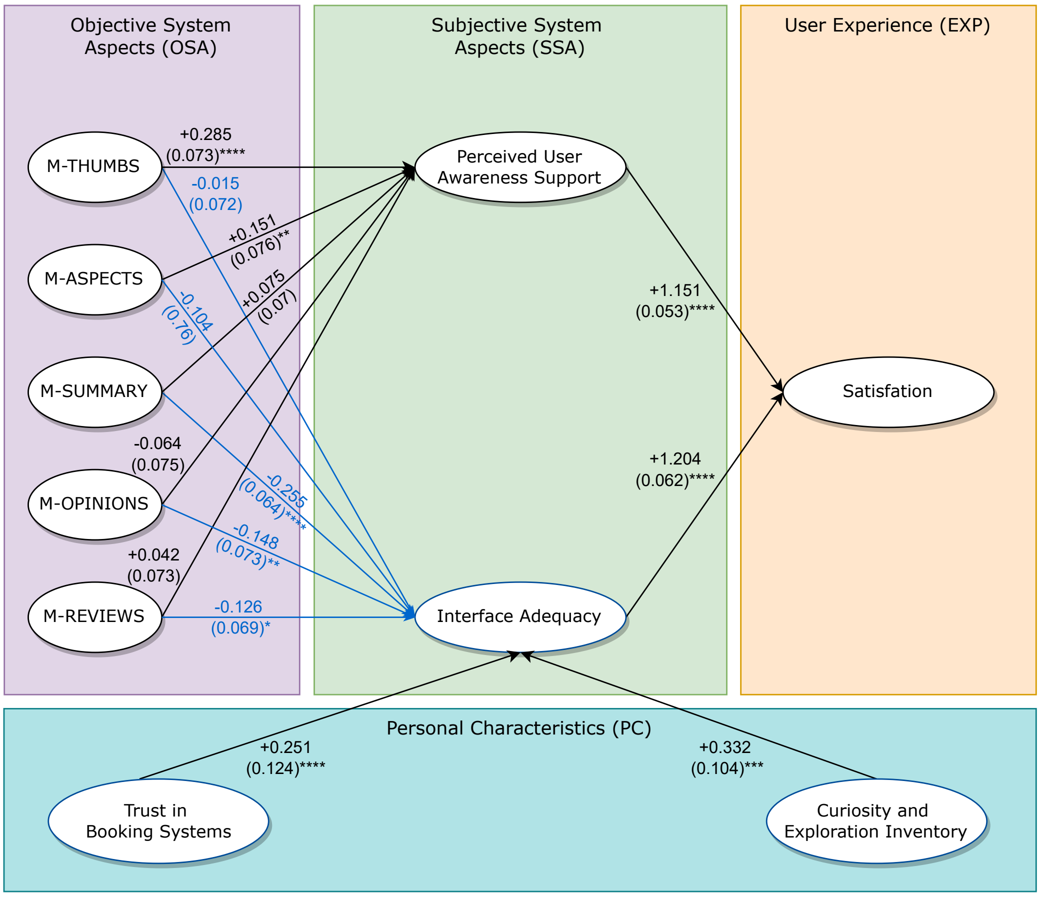


Figure 7.8: Structural Equation Model. The numbers on the arrows represent the β-coefficients and standard error of the effect. Significance levels: (****)$p < 0.001$, (***)$p < 0.01$, (**)$p < 0.05$, (*)$p < 0.1$.

face Adequacy—to Subjective Systems Aspects (SSA), and *Satisfaction* to User Experience Aspects (EXP). Five justification models were represented as dummy variables (M-THUMBS, M-ASPECTS, M-SUMMARY, M-REVIEWS, and M-OPINIONS) and considered Objective System Aspects. Additionally, we incorporated constructs such as *Trust in Booking Systems*, *Trust in Technology*, *Curiosity and Exploration Inventory*, and *Need for Cognition* for Personal Characteristics (PC).

The Confirmatory Factor Analysis validated these constructs. We assessed convergent validity using the Average Variance Extracted (AVE), which needed to exceed 0.50. Discriminant validity was evaluated by ensuring the largest correlation value was less than the square root of the AVE value of both factors. All constructs met these criteria:

- *Perceived User Awareness Support*: AVE = 0.64, $\sqrt{AVE} = 0.80$, largest correlation = 0.59;
- *Interface Adequacy*: AVE = 0.47, $\sqrt{AVE} = 0.69$, largest correlation = 0.68;
- *Satisfaction*: AVE = 0.62, $\sqrt{AVE} = 0.79$, largest correlation = 0.70.

Our SEM (illustrated in Fig. 7.8) shows dependencies, β-coefficients, and standard error values, elucidating the correlations between constructs. *Trust in Technology* and *Need for Cognition* were excluded due to high p values.

The analysis revealed that all models positively affected *Perceived User Awareness Support*, except for M-OPINIONS. M-THUMBS showed the strongest positive correlation, consistent with the post-task questionnaire results. There was also a notable positive correlation between *Perceived User Awareness Support* and *Satisfaction*, suggesting that M-THUMBS effectively enhances user satisfaction.

All models negatively impacted *Interface Adequacy*, indicating some complexity in understanding consumer feedback. However, M-THUMBS was the least negatively impacted, hinting at a better user interface. The positive correlation between *Interface Adequacy* and *Satisfaction* suggests models with better interface adequacy lead to higher user satisfaction.

Additionally, positive correlations were observed between the users' curiosity level (measured through the CEI-II and denoted as Curiosity and

Table 7.6: Post-task questionnaire results grouped by CEI-II value. The highest values for each group of participants are in boldface. Stars denote the statistical significance of the difference between the best-performing model and the other ones. Significance levels: (***)$p < 0.01$, (**)$p < 0.05$, (*)$p < 0.1$.

		CEI-II<3.5	CEI-II>=3.5
M-THUMBS	*Perceived User Awareness Support*	**3.67**	**3.66**
	Interface Adequacy	**3.64**	3.44
	Satisfaction	**3.51**	**3.44**
M-ASPECTS	*Perceived User Awareness Support*	3.29	3.58
	Interface Adequacy	3.29	**3.46**
	Satisfaction	3.21	3.35
M-SUMMARY	*Perceived User Awareness Support*	3.11**	3.36
	Interface Adequacy	3.08**	3.06**
	Satisfaction	2.94**	3.22
M-OPINIONS	*Perceived User Awareness Support*	2.92***	3.26*
	Interface Adequacy	3.06**	3.35
	Satisfaction	2.92**	3.18
M-REVIEWS	*Perceived User Awareness Support*	3.24	3.17
	Interface Adequacy	3.26	3.30
	Satisfaction	3.13	3.35**

Exploration Inventory in Figure 7.8) and *Interface Adequacy*, and between *Trust in Booking Systems* and *Interface Adequacy*, indicating that those with higher trust in booking systems and a propensity for exploration and curiosity perceive the user interfaces more positively and are more satisfied.

7.7.5 Analyzing user experience according to personality traits

We analyzed their responses considering their personality traits to gain deeper insights into how users perceive the justification models.

Influence of the curiosity trait

Participants were categorized based on their scores from the Curiosity and Exploration Inventory-II (CEI-II) questionnaire(Kashdan et al., 2009), which assesses the drive to acquire knowledge and new experiences (Stretching) and the readiness to accept the uncertain and unpredictable aspects of life (Embracing). Those scoring below 3.5 on the CEI-II scale (24 participants) were considered to have lower curiosity levels, whereas those scoring 3.5 or above (35 participants) were seen as more curious. This division was based on the distribution of responses in our sample.

The user experience findings were as follows:

- **Participants with Lower Curiosity Levels:** This group rated M-THUMBS as the most effective model across all user experience dimensions, as shown in the column labeled "CEI-II $<$ 3.5" in Table 7.6. The performance of M-THUMBS was significantly better than M-SUMMARY and M-OPINIONS. M-ASPECTS was identified as the second-best model for these constructs.

- **Participants with Higher Curiosity Levels:** For this group, the results were more varied but still favored our service-based models. M-THUMBS scored higher than the baselines in all areas, but M-ASPECTS was particularly strong in *Interface Adequacy.*

These outcomes suggest that M-THUMBS consistently ranks as the most preferred model, independent of the user's level of curiosity. However, there are intriguing nuances in the *Interface Adequacy* ratings that warrant further examination. The primary distinction between M-THUMBS and M-ASPECTS lies in their approach to summarizing consumer feedback: M-THUMBS uses

Table 7.7: Post-task questionnaire results grouped by Need for Cognition (NfC) value. We use the same notation as in Table 7.6.

		NfC<3.5	NfC>=3.5
M-THUMBS	*Perceived User Awareness Support*	**3.73**	**3.61**
	Interface Adequacy	**3.68**	3.39
	Satisfaction	**3.60**	3.36
M-ASPECTS	*Perceived User Awareness Support*	3.63	3.33
	Interface Adequacy	3.51	3.29
	Satisfaction	3.44	3.18
M-SUMMARY	*Perceived User Awareness Support*	3.19**	3.31
	Interface Adequacy	2.99***	3.13*
	Satisfaction	3.00**	3.19
M-OPINIONS	*Perceived User Awareness Support*	3.21**	3.05***
	Interface Adequacy	3.36	3.13
	Satisfaction	3.04**	3.10
M-REVIEWS	*Perceived User Awareness Support*	3.03**	3.33
	Interface Adequacy	3.08**	**3.44**
	Satisfaction	3.05**	**3.42**

bar graphs and thumbs-up/down icons, whereas M-ASPECTS integrates these with interactive filters for individual adjectives linked to aspects.

This variation in user preferences can be explained by the differing needs of our two participant groups. Those with lower curiosity levels tend to favor simpler, more schematic summaries of feedback, aligning with M-THUMBS's design. In contrast, highly curious users are more engaged with detailed, fine-grained information, as provided by M-ASPECTS's interface.

Influence of Need for Cognition

Participants were categorized based on their scores from the Need for Cognition (NfC) questionnaire(Coelho et al., 2020), which evaluates an individual's inclination towards engaging in and enjoying intellectual activities. The categorization split participants into two groups: those with an NfC score below 3.5 (26 individuals) and those with a score of 3.5 or higher (33 individuals).

An analysis of how these groups interacted with the justification models produced the following insights, as detailed in Table 7.7:

- **Participants with Lower NfC Scores:** In this category, M-THUMBS was identified as the most effective model, followed by M-ASPECTS. These models surpassed the baseline models in all user experience constructs, with several significant differences noted between M-THUMBS and the baseline models.
- **Participants with Higher NfC Scores:** Among these participants, M-THUMBS also led in terms of *Perceived User Awareness Support*, with M-ASPECTS and M-REVIEWS following closely. Interestingly, M-REVIEWS was favored for *Interface Adequacy* and *Satisfaction*, indicating a preference for this model among those with a higher propensity for cognitive engagement.

In summary, both groups found the service-based justification models, particularly M-THUMBS, to be more effective in enhancing user awareness compared to baseline models. For participants with lower NfC scores, M-THUMBS stood out as the preferred choice, likely due to its straightforward and structured presentation of consumer feedback, such as bar graphs and simple thumbs-up/down indicators. On the other hand, participants with higher NfC scores showed a preference for models like M-REVIEWS, which allow for more independent data analysis and interpretation, catering to their tendency for deeper cognitive engagement.

Discussion

M-THUMBS stands out for its ability to present information clearly and easily digestibly, making it especially suitable for participants with lower levels of curiosity or cognitive engagement (Need for Cognition). Its strength lies in providing a concise overview of consumer feedback while allowing users to inspect more detailed aspects of items as needed. This dual capability contributes to its wide appreciation among users, irrespective of their CEI-II and NfC scores, particularly in terms of enhancing awareness.

In contrast, participants with a higher Need for Cognition showed a preference for models like M-REVIEWS and M-OPINIONS. These models provide to users direct and unsummarized information, aligning with their desire for a less guided and more autonomous exploration of data. Models like M-REVIEWS, which present unfiltered review texts, or M-OPINIONS, which offer a comprehensive list of features, enable such users to engage more actively with the system. They have the flexibility to navigate through the information at a speed that is comfortable for them, extracting the details most relevant to their decision-making process.

This distinction suggests that while simplicity and ease of access to information are crucial, the degree of user control and freedom in data interaction plays a significant role in user satisfaction, especially for those inclined towards deeper cognitive processing.

7.7.6 Log data

The analysis of logged user actions, as detailed in Tables 7.8 and 7.9, provides insights into participant engagement with each justification model, both for the overall participant group and when segmented by CEI-II or NfC scores.

This analysis includes the average time spent exploring homes during the study and the mean number of interactions with various elements of the test application. Key interaction metrics are:

- **"#clicks on the bar graphs"** (M-THUMBS, M-ASPECTS): Average number of times participants clicked to reveal fine-grained dimensions linked to a selected coarse-grained category.
- **"#clicks on fine-grained dimensions"** (M-THUMBS, M-ASPECTS): Average clicks to display the aspects within a fine-grained dimension.
- **"#clicks to view more aspects"** (M-THUMBS, M-ASPECTS): Average clicks to see more aspects within a fine-grained dimension.
- **"#clicks on thumbs up/down"** (M-THUMBS): Average number of clicks to view positive/negative review excerpts for a specific aspect.
- **"#clicks on aspects"** (M-ASPECTS): Average clicks to view review excerpts mentioning a particular aspect.
- **"#visualized aspects"** (M-OPINIONS): Average number of aspects displayed, considering the user's ability to scroll through the list.
- **"#visualized reviews"** (M-REVIEWS): Average number of reviews viewed, acknowledging the scrollable nature of the review list.

Analysis of the entire participant group

M-REVIEWS prompted the most extended interaction durations, as participants needed to read and analyze numerous reviews to form opinions about the homes. About 30 reviews (or 6 per home) were viewed on average. This high

Table 7.8: Log analysis. The Total column reports mean values for all the participants of the user study. The last two columns refer to the CEI-II groups.

		Total	CEI-II<3.5	CEI-II>=3.5
M-THUMBS	Time spent to explore 5 homes	175.58	205.79	154.86
	#clicks on the bar graphs	38.17	36.83	39.09
	#clicks on fine-grained dimensions	15.41	18.83	13.06
	#clicks to view more aspects	14.36	17.13	12.46
	#clicks on thumbs up/down	24.20	27.08	22.23
M-ASPECTS	Time spent to explore 5 homes	169.03	193.92	151.97
	#clicks on the bar graphs	29.00	36.25	24.03
	#clicks on fine-grained dimensions	13.24	17.29	10.46
	#clicks to view more aspects	12.36	15.92	9.91
	#clicks on the aspects	24.56	26.88	22.97
M-SUMMARY	Time spent to explore 5 homes	76.14	89.08	67.26
M-OPINIONS	Time spent to explore 5 homes	152.54	185.88	129.69
	#clicks on aspects	59.61	64.08	56.54
	#visualized aspects	80.66	88.25	75.46
M-REVIEWS	Time spent to explore 5 homes	270.07	310.71	242.20
	#visualized reviews	30.24	32.29	28.83

number, considering that only 3 reviews are typically visible without scrolling, indicates that participants actively scrolled through reviews to gather more insights.

On the other hand, M-SUMMARY, which simplifies review insights into a summary, required significantly less time for interaction.

M-THUMBS and M-ASPECTS fell in the middle regarding engagement time, slightly longer than M-OPINIONS. Notably, M-THUMBS saw marginally more interaction time compared to M-ASPECTS. This difference is explained through click analysis:

- In M-THUMBS, users expanded bar graphs about 8 times per home on average (total mean: 38.17), suggesting frequent comparisons between

Table 7.9: Log analysis. We use the same indicators and notation as in Table 7.8. Participants are grouped by NfC.

		Total	NfC<3.5	NfC>=3.5
M-THUMBS	Time spent to explore 5 homes	175.58	124.96	215.45
	#clicks on the bar graphs	38.17	25.73	47.97
	#clicks on fine-grained dimensions	15.41	11.46	18.52
	#clicks to view more aspects	14.36	10.96	17.03
	#clicks on thumbs up/down	24.20	15.81	30.82
M-ASPECTS	Time spent to explore 5 homes	169.03	139.73	192.12
	#clicks on the bar graphs	29.00	22.23	34.33
	#clicks on fine-grained dimensions	13.24	12.23	14.03
	#clicks to view more aspects	12.36	11.54	13.00
	#clicks on aspects	24.56	19.15	28.82
M-SUMMARY	Time spent to explore 5 homes	76.14	69.92	81.03
M-OPINIONS	Time spent to explore 5 homes	152.54	104.58	190.33
	#clicks on the aspects	59.61	46.85	69.67
	#visualized aspects	80.66	72.46	87.12
M-REVIEWS	Time spent to explore 5 homes	270.07	225.04	305.55
	#visualized reviews	30.24	28.81	31.36

homes. They accessed widgets for specific fine-grained dimensions (average total: 15.41 times) and expanded the list of aspects about 3 times per home (average total: 14.36). The mean number of clicks on thumbs up/down (24.20, approximately 5 per home) indicates that users found these icons, showing the number of supportive reviews, as a useful summary of guest perceptions, reducing the need to explore many review quotes.

- For M-ASPECTS, coarse-grained dimension exploration occurred about 6 times per home (total mean: 29). Clicks on fine-grained dimensions and aspects were roughly similar to M-THUMBS's thumbs interactions. The frequency of aspect clicks (about 5 per home, total mean: 24.56) suggests that users selectively inspected quotes linked to specific adjectives, given that each aspect might have multiple adjectives.

- With M-OPINIONS, participants viewed numerous aspects (around 16 per home, total mean: 80.66), as this model directly displays a list of aspects without grouping by fine-grained dimension. This necessitated checking multiple aspects to form an opinion, with users seeking out the most relevant ones.

Impact of curiosity

The log data analysis for groups with varying levels of curiosity (low and high CEI-II values) is presented in Table 7.8. This analysis reveals that both time spent evaluating homes and interaction patterns with the user interfaces are consistent with the overall group trends previously discussed. Notably, participants with lower curiosity scores tended to spend more time evaluating homes and engaged in more interactions, such as clicking to explore aspects or reviews. This suggests that less curious users might require more time to locate the necessary information, leading to increased navigation time.

Influence of Need for Cognition

Table 7.9 provides log data analysis based on participants' Need for Cognition scores. Similarly, M-REVIEWS resulted in the longest interaction times, while M-SUMMARY had the shortest. Interestingly, differences emerged between M-THUMBS and M-ASPECTS. Participants with higher NfC scores showed a similar interaction pattern as the overall group. However, those with lower NfC scores spent less time but engaged in more clicks with M-THUMBS compared to M-ASPECTS. This might appear contradictory at first, but it could be that M-THUMBS's summary approach, using thumbs up/down for aspect evaluation, supports a more efficient home evaluation process than M-ASPECTS, which necessitates examining individual aspects.

Analysis

Across all participants and subgroups based on curiosity and cognitive style, the log analysis highlights that M-REVIEWS typically results in the most extended evaluation time for homes. This extended time is largely due to participants having to read through comprehensive review lists. In contrast, M-SUMMARY leads to the quickest evaluations by offering a summarized version of consumer feedback.

M-THUMBS and M-ASPECTS, with their incremental data access via bar graphs and interactive exploration tools (such as thumbs and clickable adjectives), engage users slightly longer than M-SUMMARY and M-OPINIONS. However, the click analysis indicates that users assessed the homes using a relatively small amount of data. Furthermore, M-THUMBS and M-ASPECTS were highly rated for their information awareness support across all participant groups. Therefore, the additional time spent interacting with their features can be interpreted as indicative of user interest and engagement in the evaluation process.

7.8 Discussion

Our study's primary finding is the higher *Perceived User Awareness Support* offered by our service-based justification models, particularly M-THUMBS, compared to the contemporary models we examined. This aspect is crucial for decision-making and remains algorithm-agnostic, suggesting that incorporating service-based justification in recommender systems is a promising direction.

However, the user experience in terms of interface adequacy and satisfaction with these models varies depending on individual curiosity traits and cognitive styles:

- Analysis of participant feedback and the Structural Equation Model, coupled with the lower rates of opting out in home evaluations, affirm the superior performance of M-THUMBS and M-ASPECTS over the baseline models. Log data also revealed that these service-based models facilitated informed decisions with a relatively modest amount of information.

- Participants with higher curiosity levels rated the *Interface Adequacy* of M-ASPECTS (which allows detailed inspection of individual adjectives of aspects) more favorably than M-THUMBS (which utilizes a thumbs up/down summary approach). Moreover, participants with higher Need for Cognition showed a preference for M-REVIEWS, which presents item reviews directly, in terms of both *Interface Adequacy* and *Satisfaction*.

These results indicate that users with lower curiosity or Need for Cognition appreciate models that efficiently organize and summarize data, providing swift access to relevant information. On the other hand, more curious individuals and those with a higher Need for Cognition value prefer more autonomy in analyzing consumer feedback.

Our findings suggest that to adequately support diverse user needs in making informed item selections, personalizing the user interface based on individual user characteristics can improve the service-based justification.

Chapter 8

Multimodal user interfaces

8.1 Introduction

The previous chapter introduced a service-based justification model for recommender system results based on Service Blueprint. This method understands that items are complex and affect customers through different stages and interactions with many services and people. Our approach considers multiple evaluation dimensions of experience. However, similar to other justification models, our proposal did not exploit item images.

In this chapter, we describe the extension of the previous work to manage *multimodal information* about items. The first idea we explore in a preliminary study is to use the service model to filter relevant multimodal information, including images, to enrich the support to decision-making.

The proposed model filters qualitative and quantitative data, including images, based on detailed dimensions reflecting the user experience at various interaction stages with the items. We categorize multimodal data according to service stages and relevant keywords by utilizing techniques like image recognition (Redmon et al., 2016) and a Service Blueprint (Bitner et al.,

2008) for representing the service fruition associated with using an item. This approach facilitates the exploration and comparison of recommendations, a critical aspect of the decision-making process that consumers engage in before finalizing their choices (Chen et al., 2014).

Our goal is to assess the effectiveness of this refined information-filtering approach in aiding user decision-making. This preliminary study aims to reply to RQ3: "How does a multimodal service-based presentation of images and textual data, potentially enhanced by keyword filtering, impact the effectiveness of the recommendation comparison process against traditional, non-stage-specific presentations?"

Similarly to the previous chapter, we conducted a user test with an experimental application that assists users in navigating and selecting from a list of available homes.

Our visualization model utilizes keywords and detailed experiential dimensions as filters, allowing users to selectively engage with the multimodal information most relevant to their interests and needs.

The work of this preliminary study is detailed in Hu et al. (2023b).

Following this preliminary study, the second part of this chapter extends the investigation of multimodal justification of results in services, by focusing on the strategic use of images as a primary mechanism for information filtering. Previous studies have explored service modeling, justification models, and multimodal data integration. Here, we shift our attention to the information that images carry in. Traditionally relying on item ratings, features, and textual reviews for suggestions and presentations, these systems have typically treated images as secondary. Our research aims to utilize them as visual supplements and central elements in information filtering and user interface design in recommender systems.

We propose and examine two user interfaces for image-based information presentation and filtering, contrasting these with traditional baselines like those used in services.

This second part of the chapter aims to reply to RQ4: "How does image-based filtering, considering various levels of detail, contribute to the effectiveness of information filtering in item lists, and in what ways does it impact user awareness and facilitate decision-making processes regarding item selections?"

This second part of the chapter represents a departure from the previous multimodal, service-oriented approach detailed in Hu et al. (2023b), which emphasized high-level experience dimensions. The focus now shifts to the scalability and practicality of image-based filtering, exploring how different image characteristics, such as the types of objects they show or the scenes they represent, influence user awareness and decision-making. This user study is presented in Hu et al. (2023a).

Section 8.2 delves into the preliminary study, laying the groundwork for understanding the role of multimodal information in enhancing user decision-making within recommender systems.

Beginning with Section 8.3, the chapter transitions to a detailed discussion of the data and image processing methods employed in the main study. This includes the pre-processing of images and textual data.

Section 8.4 introduces the justification models developed for the main user test, including scene-based information filtering, object-based information filtering, and a comparison with baseline models.

The subsequent Section 8.5 outlines the methodology of the main user study, including participant recruitment and context, and the procedure followed during the study.

Section 8.6 presents the experimental results of the main user test, detailing the demographic profile and initial responses of participants.

Finally, Section 8.7 concludes the chapter with a discussion of the findings from both the preliminary and main user studies, reflecting on the effectiveness of multimodal information integration in recommender systems.

8.2 Preliminary Study

This section describes the preliminary study to investigate the efficacy of incorporating multimodal information using the service model.

8.2.1 Dataset

In these two studies, we continued to utilize the dataset Δ comprising Airbnb home reviews situated in London, as detailed in the previous chapters. However, our focus was on a significantly smaller subset of this dataset because we were interested in examining, given a specific recommendation (which could be generated using any algorithm), which form of justification proves to be most effective. As we planned to do this in controlled user studies, few sample homes could be sufficient to carry out the test with users, and analyzing the entire London dataset was beyond the scope and necessity of our studies.

In line with this focused approach, our experimental design necessitated multiple images per home, and we targeted 15 homes, 5 for each model.

We performed the web scraping described in Section 4.2, sorting the homes $h \in \Delta$ by review count in descending order and then manually selecting the top 15 homes based on their continued listing status on Airbnb and the presence of at least 15 photographs. These images were then subjected to object recognition to identify and label various entities, such as beds, TVs, etc...

For object recognition of this preliminary study, we employed a YOLOv5 model (Jocher, 2022), adapting it through transfer learning using the Scene Understanding Database (SUN2012) (Xiao et al., 2010). SUN2012, comprising 16,873 images, annotates each with object types and their spatial coordinates. It covers an extensive array of object "classes," many of which were pertinent to our study. This adapted YOLOv5 model was then applied to the 15 images selected for our user study, generating vector representations that catalog identified classes. We ignored the coordinates of the recognized objects.

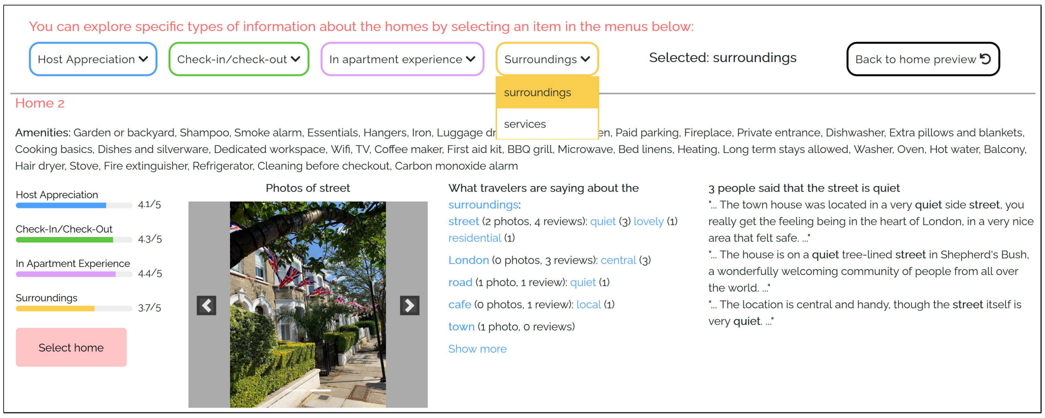
However, this model exhibited limited precision, as some SUN2012 classes were underrepresented (e.g., "bathtub", "parking"), leading to inaccurate detection in our Δ dataset images. Therefore, leveraging the small dataset we worked on, we manually reviewed and corrected the object labels for our user study. As shown in the next chapter, object detection can be improved by using more recent techniques, which we applied in our more recent work.

8.2.2 Service-based classification of multimodal information about items

To effectively justify recommendations in a service-oriented context, we have to:

- define the interaction stages involving tangible elements and actors in the item consumption process;
- identify the evaluative aspects of the user experience during these stages.

Furthermore, it's useful for the system to categorize information according to these evaluative aspects, allowing users to filter data based on their specific interests. For these reasons, we used in the work of this Chapter the same



You can explore specific types of information about the homes by selecting an item in the menus below:

Selected: surroundings

Home 2

Amenities: Garden or backyard, Shampoo, Smoke alarm, Essentials, Hangers, Iron, Luggage dr en, Paid parking, Fireplace, Private entrance, Dishwasher, Extra pillows and blankets, Cooking basics, Dishes and silverware, Dedicated workspace, Wifi, TV, Coffee maker, First aid kit, BBQ grill, Microwave, Bed linens, Heating, Long term stays allowed, Washer, Oven, Hot water, Balcony, Hair dryer, Stove, Fire extinguisher, Refrigerator, Cleaning before checkout, Carbon monoxide alarm

Host Appreciation	4.1/5
Check-In/Check-Out	4.3/5
In Apartment Experience	4.4/5
Surroundings	3.7/5

Photos of street

What travelers are saying about the surroundings:
street (2 photos, 4 reviews): quiet (3) lovely (1) residential (1)
London (0 photos, 3 reviews): central (3)
road (1 photo, 1 review): quiet (1)
cafe (0 photos, 1 review): local (1)
town (1 photo, 0 reviews)
Show more

3 people said that the street is quiet
"... The town house was located in a very **quiet** side **street**, you really get the feeling being in the heart of London, in a very nice area that felt safe. ..."
"... The house is on a **quiet** tree-lined **street** in Shepherd's Bush, a wonderfully welcoming community of people from all over the world. ..."
"... The location is central and handy, though the **street** itself is very **quiet**. ..."

Figure 8.1: A segment of the FILTER-WITH-IMG model's user interface.

approach in knowledge representation of the previous Chapter, as detailed in Section 7.2, and in these works (Mauro et al., 2022b,a).

In particular, we use the hierarchical classification of course-grained dimensions, such as "host appreciation," and fine-grained dimensions, such as "kitchen." For further data classification, we use the same range of dictionaries to specify terms relevant to each category. Table 7.1 includes examples of these keywords.

We can classify textual and visual item aspects by linking these dictionaries to the detailed evaluative dimensions. For image classification, we leverage the vector representations described in Section 8.2.1, aligning the identified classes from the SUN2012 dataset with the keywords in our dictionaries.

8.2.3 Justification models

FILTER-WITH-IMG

Across the various justification models we developed, our test application showcases a selection of five homes. Figure 8.1 illustrates a segment of the user interface for the **FILTER-WITH-IMG model**. We describe the components of our proposed model.

- **Central Display Area** Each home *h* featured in the application is accompanied by a list of its amenities (like air conditioning, WiFi, etc.) that are present in the home.
- **Summary Bar Graph:** A summary bar graph visualizes the aggregated experiences of past guests with home *h*. This graph offers a quick, intuitive understanding of the overall guest satisfaction and is a carry-over from the previous experiment. It summarizes key dimensions of `Host Appreciation`, `Check-in/Check-Out`, `In Apartment Experience`, and `Surroundings`.
- **Images of Home *h*:** For each home, the interface displays a list of images that give a visual tour of the property. This can include interior and exterior shots, providing a comprehensive view of what the home looks like.
- **Reviews of Home *h*:** Alongside the images, the application presents user reviews specific to each home. These reviews, offering firsthand accounts and experiences of past guests, are a critical component of the decision-making process. The reviews are displayed in a manner that allows easy browsing, and potentially, they are categorized or filtered to highlight comments on specific aspects like the home's cleanliness or the host's hospitality.
- **"Select Home" Option:** A prominent feature for each home is the "Select Home" button. This function allows users to mark a home as their preferred choice, facilitating the decision-making process. This option is particularly useful in scenarios where users compare multiple

properties and need a simple way to indicate their interest or final decision.

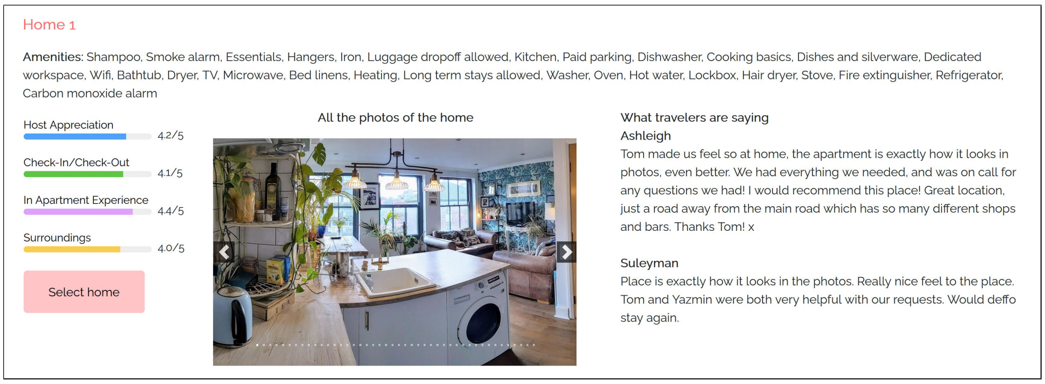
- **Excluding bias-inducing information:** Similar to Section 7.4, we exclude bias-inducing information. In this experiment, differently from the previous one, we added the images.

The interface's upper section contains filters for information, each representing a coarse-grained evaluation dimension (like "Surroundings" in Figure 8.1). Users can inspect into specific fine-grained dimensions d (for instance, "surroundings") by interacting with these filters. Selecting a filter refines the displayed content to focus on the chosen fine-grained evaluation dimension. There is the possibility of a lack of corresponding images and reviews. In this case, a standard "no information" image is displayed. For each home h, images classified under d are showcased in an interactive carousel, along with reviews about h and dimension d (denoted as R_{hd}). The interface also includes two scrollable areas:

- The left section offers keyword-based filters, narrowing the focus to specific aspects (nouns) and adjectives extracted from R_{hd} reviews.

- The right section displays these reviews. If a keyword is selected, the application highlights sentences from R_{hd} that reference the chosen term. For instance, the figure focuses on the "street" keyword within the "surroundings" category of h.

These filters also influence the image carousel: selecting "street," for example, limits the carousel only to show "Photos of the street."



Amenities: Shampoo, Smoke alarm, Essentials, Hangers, Iron, Luggage dropoff allowed, Kitchen, Paid parking, Dishwasher, Cooking basics, Dishes and silverware, Dedicated workspace, Wifi, Bathtub, Dryer, TV, Microwave, Bed linens, Heating, Long term stays allowed, Washer, Oven, Hot water, Lockbox, Hair dryer, Stove, Fire extinguisher, Refrigerator, Carbon monoxide alarm

What travelers are saying

Ashleigh

Tom made us feel so at home, the apartment is exactly how it looks in photos, even better. We had everything we needed, and was on call for any questions we had! I would recommend this place! Great location, just a road away from the main road which has so many different shops and bars. Thanks Tom! x

Suleyman

Place is exactly how it looks in the photos. Really nice feel to the place. Tom and Yazmin were both very helpful with our requests. Would deffo stay again.

Figure 8.2: A segment of the FULL-DATA model's user interface.

Baselines

Figure 8.2 illustrates a segment of the **FULL-DATA model**'s user interface. Similar to the FILTER-WITH-IMG model, this interface also presents the home's amenities in the central part of the interface and a graphical bar graph summary of customer experiences with the property. Unlike FILTER-WITH-IMG, however, it displays all photos in an interactive carousel and aggregates all reviews in a scrollable section. In other words, it lacks the capability to filter information by detailed evaluation dimensions or keywords.

The **FILTER-WITHOUT-IMG model**, not depicted here for brevity, takes its cue from FILTER-WITH-IMG but excludes the image carousel. This model provides users the functionality to sift through home reviews based on detailed experience evaluation dimensions and offers additional data refinement using keyword filters.

8.2.4 User study

Participant recruitment and context

A user study was designed to assess the effectiveness of the three proposed justification models. To facilitate this, we developed a test application as depicted in Figures 8.1 and 8.2. This application autonomously guided participants through the evaluation process, recording their interactions, such as clicks and scrolls, for behavioral analysis. Importantly, in alignment with privacy considerations, the application only collected non-identifiable data, using numerical identifiers for session tracking.

Participants were recruited via social networks and public mailing lists, targeting adults. The invitation included a link to the test application, and participation was voluntary and uncompensated.

Procedure

The study employed a within-subjects design. The conditions of the experiment (FILTER-WITH-IMG, FULL-DATA, FILTER-WITHOUT-IMG) were treated as independent variables, with each participant experiencing all conditions. Task order was varied among users to mitigate potential biases from fatigue or practice effects. No time constraints were imposed on task completion. The experiment involved the following steps:

1. Participants began by reviewing and consenting to an informed agreement form (available at `https://bit.ly/3X3Myg4`). They also confirmed being 18 years or older.

2. The application then gathered basic demographic information, cultural background, and familiarity with booking and e-commerce platforms.

3. Subsequently, participants interacted with the three justification models in a counterbalanced sequence. They explored five homes (identical across all participants to facilitate comparative analysis) and chose their preferred option for booking. Following each model interaction, a post-task questionnaire was administered, measuring user agreement with statements from Table 8.1, sourced from (Pu et al., 2011; Di Sciascio et al., 2019; Lewis and Sauro, 2009) and based on the ResQue recommender system questionnaire. These statements aimed to measure user interface experience and perceptions, with responses on a scale ranging from Strongly Disagree to Strongly Agree, corresponding to numerical values [1, 5].

Attention checks were included in the questionnaires to ensure diligent participation in the study.

8.2.5 Experimental results

Our user study was conducted between November 1st and December 20th, 2022. The average duration of participation was approximately 19.89 minutes, with a standard deviation of 10.36 minutes.

From the initial pool of 59 participants, 9 were excluded due to failing attention checks, leading to a final sample size of N = 50.

The participants are considered sufficient for statistical significance with $\alpha = 0.05$, *power* = 0.80, and *effect size* = 0.40, as determined by power analysis, since $N > 42$.

The participant demographics comprised 21 females, 29 males, with ages ranging from below 20 (13 participants) to between 41 and 50 (1 participant). Educational backgrounds varied, including high school (15), university (30), and Ph.D. (5) degrees. Participants came from diverse fields: 19 from technical,

Table 8.1: Post-task questionnaire results. We report the mean value of users' replies with Standard Deviation. The best values for each statement are in boldface (minimum for Q4, maximum for the other statements). Stars denote the statistical significance of the difference between the best-performing model and the other ones. Significance levels: (**)$p < 0.01$, (*)$p = 0.08$.

	FILTER WITH-IMG	FULL-DATA	FILTER WITHOUT-IMG
Q1: It was easy to understand why some homes were good and others not.	**3.52(0.81)**	3.34(1.02)	2.78(1.13)**
Q2: The system helped me to compare the homes.	**3.62(0.85)**	3.18(1.08)*	3.00(1.11)**
Q3: The system was sufficiently informative.	**3.76(0.72)**	3.72(0.90)	2.84(1.22)**
Q4: The system was cluttered or confusing.	2.50(0.95)	**2.32(1.17)**	2.90(1.15)**
Q5: The information about the homes was sufficient for me to select a home.	**3.88(0.77)**	3.86(0.86)	2.80(1.21)**
Q6: The information about the homes was easy to interpret and understand.	3.70(0.89)	**3.78(0.86)**	3.04(1.03)**
Q7: I found the information about homes quickly.	**3.62(0.78)**	3.52(0.91)	3.00(1.01)**
Q8: I think that I would like to frequently use this system to compare homes.	**3.30(0.95)**	3.16(1.13)	2.38(1.03)**
Q9: I felt very confident using this system to compare homes.	**3.36(0.85)**	**3.36(0.90)**	2.72(0.97)**

20 from scientific, 5 from humanities and languages, 2 from economics, and 4 from other areas. Regarding computer proficiency, 23 participants identified as advanced users, 24 as average, and 3 as beginners. Concerning familiarity with online booking or e-commerce platforms, 13 reported using them a few

times a week, 26 a few times a month, and 11 a few times a year.

Evaluation of the visualization models

The results from the post-task questionnaire are summarized in Table 8.1. A *post-hoc* Mann-Whitney test indicated marginal statistical significance in the difference between FILTER-WITH-IMG and FULL-DATA. However, a significant difference was observed in the perception of models that included images (FILTER-WITH-IMG and FULL-DATA) compared to FILTER-WITHOUT-IMG, which lacked them ($p < 0.01$).

Participants perceived FILTER-WITH-IMG as the most effective in facilitating understanding of why certain homes are preferable (Q1) and in aiding the comparison of homes (Q2, significantly different from FILTER-WITH-IMG, $p < 0.01$, and FILTER-WITHOUT-IMG, $p = 0.08$). FILTER-WITH-IMG was also seen as the most informative (Q3) and sufficient in terms of data for making a selection (Q5). These results suggest that FILTER-WITH-IMG, with its blend of information filtering and visual aids, excels in assisting with item selection and comparison. Conversely, FILTER-WITHOUT-IMG, lacking visual elements, was less favorably evaluated in these areas.

In terms of user interface perception, participants found FULL-DATA to be less cluttered and confusing (Q4) and easier to interpret (Q6). This preference for FULL-DATA may stem from its similarity to familiar platforms like Airbnb and Booking, which display comprehensive reviews without filtering tools. Despite the potential complexity of filtering reviews, participants recognized its value, stating that FILTER-WITH-IMG helped them locate information more swiftly (Q7).

The overall satisfaction with the models, especially regarding item comparison, was gauged through Q8 and Q9. Models featuring both images and

Table 8.2: Log analysis. Time is measured in seconds, # denotes the mean number of events per user.

	FILTER WITH-IMG	FULL-DATA	FILTER WITHOUT-IMG
Mean time spent to explore 5 homes	170.06	178.26	169.2
# scrolling on homes	28.84	20.92	29.18
# scrolling on reviews	45.56	33.06	54.74
# visualized reviews	32.30	17.00	32.49
# clicks on fine-grained dimensions	8.90	-	5.65
# clicks on keywords	1.62	-	2.1
# clicks on photos	15.56	48.50	-

text (FILTER-WITH-IMG and FULL-DATA) were comparably rated. Participants expressed a preference for frequently using FILTER-WITH-IMG for home comparisons (Q8) and felt equally confident in using both models for this task (Q9). These findings suggest that while FILTER-WITH-IMG may offer superior comparison support, the similarity in information provided by both models instills comparable confidence in decision-making.

Log analysis

Our test application recorded all the user interactions, such as clicks and scrolls, during the user study. The recorded scrolls can be categorized into two types: one for browsing the list of homes and the other for exploring the reviews. Table 8.2 presents key data we extracted, showing average user engagement with each justification model:

- “Average Time for Exploring 5 Homes” indicates the time users typically spent reviewing and choosing their preferred home.

- "Mean Scrolls on Homes" quantifies how often a home remained visible on-screen for over 2 seconds, highlighting the extent of scrolling to view the list of homes. Brief visibility under 2 seconds, possibly accidental while navigating, wasn't considered.
- "Mean Scrolls on Reviews" measures the frequency of review visibility exceeding 2 seconds, the minimum to consider a brief review read, excluding quick scrolls.
- "Mean Number of Reviews Viewed" tracks distinct reviews seen on-screen for more than 2 seconds.
- "Clicks on Fine-Grained Dimensions" denotes the average number of times participants used this filter for homes, an option available in FILTER-WITH-IMG and FILTER-WITHOUT-IMG only.
- "Clicks on Keywords" reflects the frequency of filtering home information based on aspects like "kitchen" or adjectives like "beautiful," also specific to FILTER-WITH-IMG and FILTER-WITHOUT-IMG.
- "Clicks on Photos" represents the average instances of participants interacting with home image carousels, a feature in FILTER-WITH-IMG and FULL-DATA.

The duration of interaction across all three interfaces was relatively similar, though FULL-DATA, lacking the filtering functions of the other models, had marginally higher engagement.

Notably, "Mean Scrolls on Homes" revealed more active navigation within FILTER-WITH-IMG and FILTER-WITHOUT-IMG compared to FULL-DATA, indicating a 38% higher engagement, suggesting more intensive comparison activity. Similarly, "Mean Scrolls on Reviews" and "Mean Number of Reviews

Viewed" were higher with FILTER-WITH-IMG and FILTER-WITHOUT-IMG, with the latter having the most scroll activity, likely compensating for the absence of visual cues.

Comparing models with information filters, participants clicked on fine-grained evaluation dimensions twice per home in FILTER-WITH-IMG and less frequently in FILTER-WITHOUT-IMG. This difference is attributed to FILTER-WITH-IMG's visual elements, which guide relevant filtering. Keyword filter usage was minimal and similar in both models, offering limited insights.

Finally, comparing FILTER-WITH-IMG and FULL-DATA, which both feature image carousels, users interacted more with FULL-DATA's photos, viewing approximately three times more images. This is explained by FILTER-WITH-IMG's ability to focus on relevant scenes, reducing the need for extensive photo browsing, unlike in FULL-DATA where users sifted through images more indiscriminately.

8.2.6 Discussion

The user study highlighted that FILTER-WITH-IMG and FULL-DATA, both of which display images of homes, were better received than FILTER-WITHOUT-IMG. This highlights the utility of integrating both textual and visual information for effective item comparison.

A key finding concerns the balance between the robustness of information filtering in multimodal data exploration and the complexity introduced by the service-based filters. FILTER-WITH-IMG, with its integration of images and service-aware information presentation, enables efficient data retrieval and simplifies review analysis. Despite being perceived as somewhat more cluttered than the baseline models, objective data from log analysis confirms that FILTER-WITH-IMG's fine-grained evaluation dimension filters significantly enhance comparison activities compared to FULL-DATA. This finding affirmatively answers research question.

However, the study indicates a lack of interest among participants in using keyword filters for detailed item aspects, suggesting that once information is filtered by fine-grained evaluation dimensions, users find the data sufficiently relevant, negating the need for further reduction. The perception of FILTER-WITH-IMG as moderately cluttered leads to the recommendation of simplifying its interface by minimizing keyword and aspect filters.

Following this section of the preliminary study, we will describe the main study of this chapter.

8.3 Data and image processing of the main study

This study utilizes the same dataset Δ as detailed in Section 8.2.1. Unlike our prior research, this work incorporates a significantly more precise technique for image analysis.

This section outlines the analysis we performed on the images to extract the scenes they represent (e.g., the bedroom of a home) and, similar to the previous chapter, the objects they show. Moreover, this section presents our method for processing textual data related to home amenities and consumer reviews by abstracting the service models used so far in our work.

8.3.1 Pre-processing of the images

Scene recognition

This step is dedicated to identifying images' specific context or scene. In the home-booking domain, this corresponds to identifying the rooms or environment of homes and surroundings shown in their photos.

For this purpose, an image-recognition model trained on home scenes is needed. We utilized the Places365-Standard dataset[1], which includes 1.8 million training images and 36,000 validation images, categorized into 365 distinct scene types (denoted as $SCENES_P$). Scene recognition was performed using the `ResNet50` model[2] Zhou et al. (2017), known for its accuracy in top-1 and top-5 scene recognition.

Given that Places365-Standard includes a wide range of scenes, some of which are not pertinent to the home-booking context (e.g., volcanoes or embassies), we refined $SCENES_P$ to suit our domain better. Similarly, we consolidated similar scene categories (e.g., merging "bedroom," "hotel room," and "child's room") into singular representations. Our result scene set, $SCENES$, thus comprises categories like kitchen, living room, bedroom, etc.

The classification process for each home image i, was as follows:

1. Apply `ResNet50` to obtain a list $S = [s_1, \ldots, s_5]$ (with $s_j \in SCENES_P$), sorted by likelihood in reverse order. This list represents the algorithm's recognized scenes for i.

2. Convert S into a new list $S' = [s_1, \ldots, s_k]$ (with $s_j \in SCENES$), using the predefined mappings. The most probable scene, $s_1 \in S'$, is then used to tag i if $S' \neq \emptyset$.

[1] `https://paperswithcode.com/dataset/places365`

[2] `https://github.com/CSAILVision/places365`

In cases where $S' = \emptyset$ (due to no scenes from S mapping to $SCENES$), the image was considered unclassifiable and excluded. For example, a luxurious building was mistaken as an embassy ($\notin SCENES$). We decided to exclude these problematic cases to support the automated management of images. The final set of categorized images is denoted as I.

Object recognition in images

Then, we focused on identifying objects within the images $i \in I$. For this task, we employed the `UODDM` (Unified Object Detection with Deep Models) by Shen and Stamos (2023)[3], a model demonstrating state-of-the-art performance in understanding indoor scenes. This model is a deep neural network pre-trained on COCO dataset (Lin et al., 2015)[4]. This comprehensive dataset contains over 320,000 images and 2.5 million labeled instances across 80 diverse object categories. It has been fine-tuned using datasets like SUN RGB-D (Song et al., 2015), which offer annotations for various indoor objects.

For image $i \in I$, we have a list of recognized classes of objects as a result of the analysis, which we to annotate the images with relevant objects.

Consequently, each image i in our study was represented by the following vector:

$$\vec{v_i} = [scene, [class_1, \ldots, class_n]]$$

where *scene* is the most confidently recognized scene and $class_k$ are the types of objects identified in i. For instance, a bedroom image ι might have a vector representation as follows:

$\vec{v_\iota}$ = [bedroom, [bed, pillow, window, chair, desk, cabinet, dresser, picture]]

[3] `https://github.com/liketheflower/UODDM`

[4] `https://cocodataset.org`

Table 8.3: Mappings Between Scenes and Lemmas, Used to Classify Amenities and Reviews by Scene (*SCENES-LEMMAS*).

Scene	**Lemmas**
kitchen	cooker, dishwasher, fridge, kettle, microwave, plate oven, ...
bedroom	bed, blanket, bunkbed, pillow, sheet, slipper, wardrobe, ...
bathroom	bathtub, towel, bidet, hair-dryer, shower, toiletry, ...
...	...

8.3.2 Pre-processing of textual data

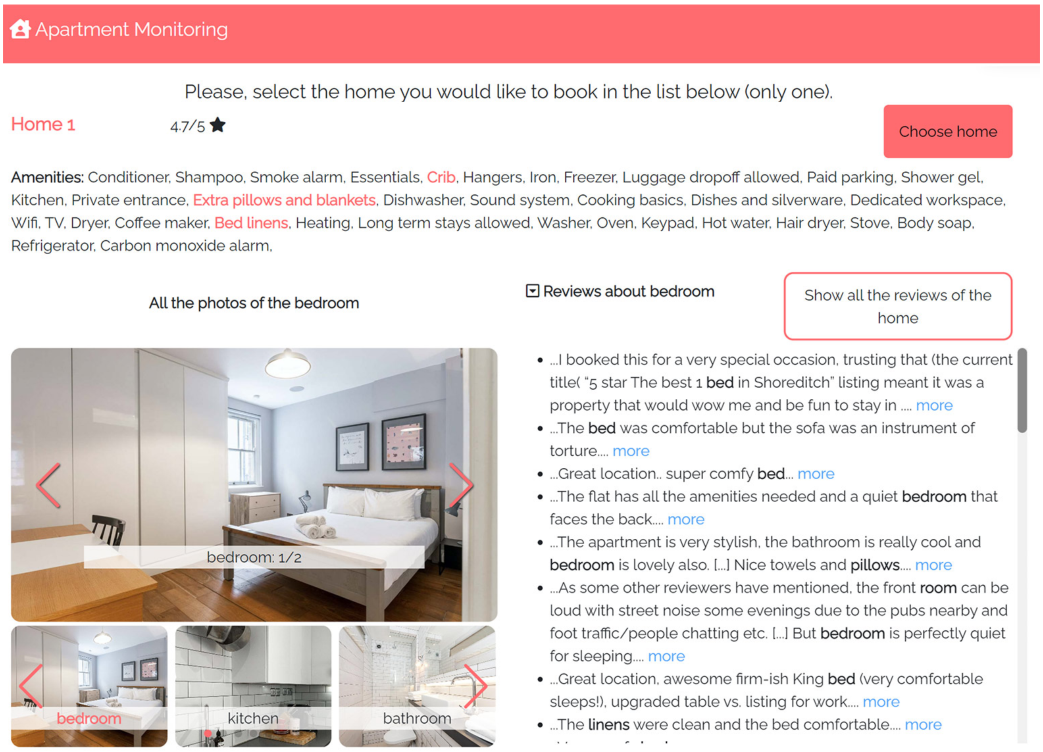
Regarding the pre-processing of the reviews, we followed the previous works described in Section 6.3.1.

Then, we performed the following analysis:

- **Classification by Scene:** The lemmas of amenities and review sentences were then classified according to predefined scene categories $SCENE$ (as detailed in Section 8.3.1). This was achieved by using `spaCy` for synonym matching, leading to the creation of the *SCENES-LEMMAS* mappings (Table 8.3).

- **Mapping to Object Classes:** We further mapped these lemmas to specific object classes defined in $CLASSES$, enhancing the granularity of our categorization. This process resulted in *CLASSES-LEMMAS* mappings.

- **Review Indexing:** The final step involved indexing both the complete reviews and their sentences. This indexing was based on lemmas and scenes, setting the foundation for efficient retrieval during user interaction with the system. Note that in this work we do not use the thesauri of the work described in Chapter 6, since we do not index by the service evaluation stages.



Please, select the home you would like to book in the list below (only one).

Amenities: Conditioner, Shampoo, Smoke alarm, Essentials, Crib, Hangers, Iron, Freezer, Luggage dropoff allowed, Paid parking, Shower gel, Kitchen, Private entrance, Extra pillows and blankets, Dishwasher, Sound system, Cooking basics, Dishes and silverware, Dedicated workspace, Wifi, TV, Dryer, Coffee maker, Bed linens, Heating, Long term stays allowed, Washer, Oven, Keypad, Hot water, Hair dryer, Stove, Body soap, Refrigerator, Carbon monoxide alarm,

- ...I booked this for a very special occasion, trusting that (the current title("5 star The best 1 **bed** in Shoreditch" listing meant it was a property that would wow me and be fun to stay in .... more
- ...The **bed** was comfortable but the sofa was an instrument of torture.... more
- ...Great location.. super comfy **bed**... more
- ...The flat has all the amenities needed and a quiet **bedroom** that faces the back.... more
- ...The apartment is very stylish, the bathroom is really cool and **bedroom** is lovely also. [...] Nice towels and **pillows**.... more
- ...As some other reviewers have mentioned, the front **room** can be loud with street noise some evenings due to the pubs nearby and foot traffic/people chatting etc. [...] But **bedroom** is perfectly quiet for sleeping.... more
- ...Great location, awesome firm-ish King **bed** (very comfortable sleeps!), upgraded table vs. listing for work.... more
- ...The **linens** were clean and the bed comfortable.... more

Figure 8.3: FILTER-BY-SCENE user interface.

These pre-processing steps are crucial for the subsequent image-based filtering process, enabling the system to correlate textual information with visual data accurately.

8.4 Justification models

This section outlines the justification models for our user study. We developed a web-based test application to handle the models we want to evaluate.

8.4.1 Scene-based information filtering

The scene-based model, hereafter referred to as FILTER-BY-SCENE, displays scenes in a given home h. An illustrative example is provided in Figure 8.3.

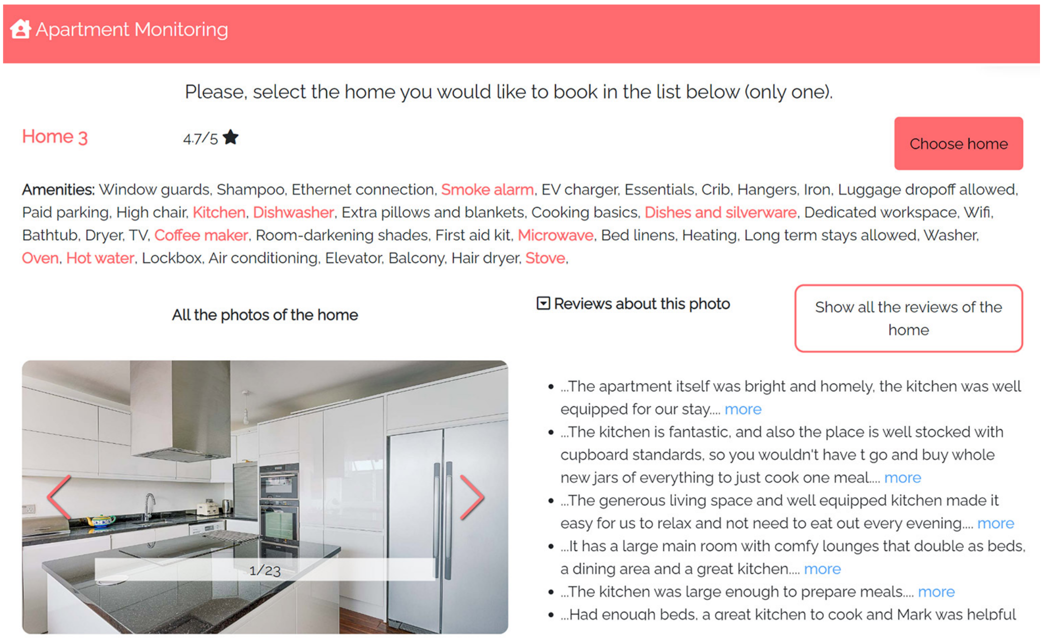
The model's primary features are:

- Two image carousels: the first one representing different scenes $sc \in SCENES$ dynamically generated based on the classification of h's images, user can select the scene that (s)he wants to inspect; the latter one representing the list of images $i \in sc$, with sc the selected scene.
- Adjacent to these carousels, the reviews of h are displayed. These reviews are selected through a scene-indexing process detailed in Section 8.3.1.

  For any given review r, our interface displays only those sentences that correlate with the currently selected scene sc (displayed in the carousel), with the words w that involve sc, according to the dictionaries (see *SCENES-LEMMAS* mapping in Section 8.3.2), displayed in bold. If we have more sentences in the same review, we will see the concatenation of these sentences with "[. . .]" between each sentence.

  Users have the option to access the full review r by selecting the "more" link. The click of this link shows the entire review with sentences associated with sc displayed in bold.

  Furthermore, users are provided with the functionality to see all reviews related to the property h, or they can choose to return to the reviews specifically about sc, through the "Show all reviews of the home" and "Reviews about *sc*" buttons, respectively.
- The interface showcases the average user rating of the home, providing a quick insight into its overall appeal. Additionally, it shows the list of amenities available at the property. To further aid in user navigation and relevance, amenities that are directly related to the current scene selected by the user are distinctively highlighted, enhancing the user's



Please, select the home you would like to book in the list below (only one).

Home 3 4.7/5 ★

Amenities: Window guards, Shampoo, Ethernet connection, Smoke alarm, EV charger, Essentials, Crib, Hangers, Iron, Luggage dropoff allowed, Paid parking, High chair, Kitchen, Dishwasher, Extra pillows and blankets, Cooking basics, Dishes and silverware, Dedicated workspace, Wifi, Bathtub, Dryer, TV, Coffee maker, Room-darkening shades, First aid kit, Microwave, Bed linens, Heating, Long term stays allowed, Washer, Oven, Hot water, Lockbox, Air conditioning, Elevator, Balcony, Hair dryer, Stove.

- ...The apartment itself was bright and homely, the kitchen was well equipped for our stay.... more
- ...The kitchen is fantastic, and also the place is well stocked with cupboard standards, so you wouldn't have t go and buy whole new jars of everything to just cook one meal.... more
- ...The generous living space and well equipped kitchen made it easy for us to relax and not need to eat out every evening.... more
- ...It has a large main room with comfy lounges that double as beds, a dining area and a great kitchen.... more
- ...The kitchen was large enough to prepare meals.... more
- ...Had enough beds, a great kitchen to cook and Mark was helpful

Figure 8.4: FILTER-BY-OBJ user interface.

ability to quickly discern features of particular interest in the context of the selected scene.

8.4.2 Object-based information filtering

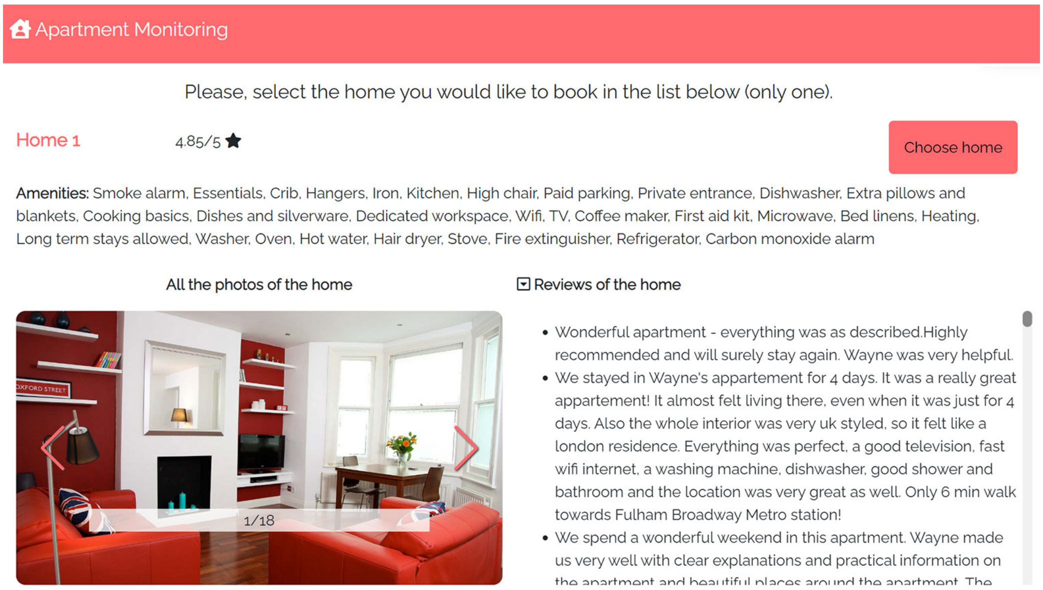
The FILTER-BY-OBJ model, illustrated in Figure 8.4, shares several elements with FILTER-BY-SCENE but introduces unique aspects:

1. A singular carousel displays all images of home h, unlike the scene-specific carousels in FILTER-BY-SCENE.

2. Reviews are filtered based on both the scene and objects identified within an image i. This is achieved through the vector representation $\vec{v_i} = [sc, [c_1, \ldots, c_m]]$, which captures the scene and identified objects in i. The review selection mechanism involves the following:

- Extracting object classes from $\vec{v_i}$ and mapping them to relevant lemmas using a *CLASSES-LEMMAS* mapping (see Section 8.3.2).
- Displaying and highlighting sentences from reviews that include lemmas related to the identified objects.

8.4.3 Baseline model



Apartment Monitoring

Please, select the home you would like to book in the list below (only one).

Home 1 4.85/5 ★

Choose home

Amenities: Smoke alarm, Essentials, Crib, Hangers, Iron, Kitchen, High chair, Paid parking, Private entrance, Dishwasher, Extra pillows and blankets, Cooking basics, Dishes and silverware, Dedicated workspace, Wifi, TV, Coffee maker, First aid kit, Microwave, Bed linens, Heating, Long term stays allowed, Washer, Oven, Hot water, Hair dryer, Stove, Fire extinguisher, Refrigerator, Carbon monoxide alarm

All the photos of the home

Reviews of the home

- Wonderful apartment - everything was as described.Highly recommended and will surely stay again. Wayne was very helpful.
- We stayed in Wayne's appartement for 4 days. It was a really great appartement! It almost felt living there, even when it was just for 4 days. Also the whole interior was very uk styled, so it felt like a london residence. Everything was perfect, a good television, fast wifi internet, a washing machine, dishwasher, good shower and bathroom and the location was very great as well. Only 6 min walk towards Fulham Broadway Metro station!
- We spend a wonderful weekend in this apartment. Wayne made us very well with clear explanations and practical information on the apartment and beautiful places around the apartment. The

Figure 8.5: Illustration of the baseline user interface.

Our baseline user interface, depicted in Figure 8.5, draws inspiration from a classical home booking service's design. It showcases a straightforward presentation of the home's photographs and reviews without implementing specific information filtering mechanisms. This approach offers users an unfiltered perspective of previous guests' experiences, giving them a comprehensive view of the home.

This baseline interface serves as a fundamental comparison point against the justification models developed in our research.

Table 8.4: User Trust and Interest in Booking System Features: A Questionnaire Overview (Scale: 1-5)

Statement	**Mean Value**
S1: Importance of photos in home comparison	4.718
S2: Importance of reviews in home comparison	4.465
S3: Trust in booking system suggestions	3.126
S4: Sufficiency of user ratings for booking decisions	3.014
S5: Necessity of inspecting user reviews before booking	4.309
S6: Necessity of inspecting home descriptions before booking	4.451
S7: Necessity of inspecting home photos before booking	4.662

8.5 User study

8.5.1 Participant recruitment and context

A user study was conducted to evaluate the effectiveness and user experience of the interfaces described in Section 8.4. Our test application autonomously administered the study, anonymously capturing user interaction data. Attention checks were incorporated within the questionnaires to ensure the validity of the responses.

Participants were recruited through social networks, mailing lists, and university channels. Participation was voluntary and uncompensated.

8.5.2 Procedure

The user study was structured as a within-subjects experiment. It involved three distinct interface models (FILTER-BY-SCENE, FILTER-BY-OBJ, and BASELINE) serving as independent variables. Each participant interacted with all three models, with the sequence of interaction counterbalanced to mitigate biases from fatigue and practice effects. The study comprised several phases:

1. Introduction to the experiment and collection of informed consent

Table 8.5: Post-task Questionnaire Results on the Whole Group of Participants. The Best Values for Each Statement Are in Boldface (Minimum Value for Q4, Q6, and Q10, Maximum for the Other Statements). Mean Values Are in [1, 5]. We Report the p-values of the Statistically Significant Results According to a Kruskal-Wallis Test.

Statement	**p-value**	**FILTER BY-SCENE**	**FILTER BY-OBJ**	**BASELINE**
Q1: It was easy to understand why some homes were good and others not.	0.011	**3.507**	3.085	3.056
Q2: The system helped me to compare the homes.	0.002	**3.577**	3.352	3.028
Q3: The system was sufficiently informative.	0.084	**3.831**	3.761	3.535
Q4: The system was cluttered or confusing.	0.019	**2.521**	2.901	3.014
Q5: The information about the homes was sufficient for me to select a home.	0.006	**3.986**	3.761	3.563
Q6: The system provided too much information about the homes.		**2.901**	3.099	3.211
Q7: The information about the homes was easy to interpret and understand.	0.090	**3.746**	3.451	3.394
Q8: I found the information about homes quickly.	0.002	**3.831**	3.493	3.239
Q9: I think that I would like to frequently use this system to compare homes.	0.047	**3.338**	3.070	2.930
Q10: I found this system to compare homes unnecessarily complex.		**2.592**	2.859	2.901
Q11: I thought this system to compare homes was easy to use.		**3.521**	3.465	3.394
Q12: I felt very confident using this system to compare homes.	0.035	**3.549**	3.282	3.169

(accessible at `https://bit.ly/42URUwE`). Participants confirmed their age (18 or above) to proceed.

2. Collection of demographic data, cultural background, and familiarity with booking/e-commerce platforms.

3. Administration of the Need for Cognition questionnaire Coelho et al. (2020), assessing participants' cognitive engagement preferences.

4. Completion of the Trust in booking system and interest in reviews and images questionnaire (see Table 8.4).

5. Interaction with the three interfaces in a counterbalanced order, involving selection tasks and subsequent questionnaire completion (Table 8.5). The questionnaire, based on (Pu et al., 2011; Di Sciascio et al., 2019; Lewis and Sauro, 2009), utilized a 5-point Likert scale to evaluate the interface experience.

6. A brief post-test questionnaire on the utility of scene and image-based filtering.

7. Completion of the TIPI questionnaire (Gosling et al., 2003) to assess personality traits.

Table 8.6: Post-task Questionnaire Results Grouped by Need for Cognition. We Use the Same Notation as in Table 8.5

Statement	**Low NfC group (NfC < 3.5; 35 participants)**				**High NfC group (NfC ≥ 3.5; 36 participants)**			
	p-value	**FILTER BY-SCENE**	**FILTER BY-OBJ**	**BASELINE**	**p-value**	**FILTER BY-SCENE**	**FILTER BY-OBJ**	**BASELINE**
Q1: It was easy to understand why some homes were good and others not.		**3.543**	3.286	3.229	0.031	**3.472**	2.889	2.889
Q2: The system helped me to compare the homes.	0.043	**3.771**	3.571	3.286	0.014	**3.389**	3.139	2.778
Q3: The system was sufficiently informative.		**3.914**	3.829	3.629		**3.750**	3.694	3.444
Q4: The system was cluttered or confusing.	0.033	**2.457**	2.771	3.114		**2.583**	3.028	2.917
Q5: The information about the homes was sufficient for me to select a home.	0.026	**4.086**	3.800	3.657		**3.889**	3.722	3.472
Q6: The system provided too much information about the homes.	0.089	**2.800**	3.229	3.314		3.000	**2.972**	3.111
Q7: The information about the homes was easy to interpret and understand.		**3.829**	3.543	3.486		**3.667**	3.361	3.306
Q8: I found the information about homes quickly.	0.041	**3.857**	3.743	3.286	0.030	**3.806**	3.250	3.194
Q9: I think that I would like to frequently use this system to compare homes.		**3.486**	3.257	3.200	0.091	**3.194**	2.889	2.667
Q10: I found this system to compare homes unnecessarily complex.		**2.543**	2.686	2.914		**2.639**	3.028	2.889
Q11: I thought this system to compare homes was easy to use.		**3.629**	3.686	3.486		**3.417**	3.250	3.306
Q12: I felt very confident using this system to compare homes.		**3.914**	3.571	3.543		**3.194**	3.000	2.806

8.6 Experimental results

The user study was conducted from May 2nd to May 20, 2023. Out of 79 initial participants, 71 completed the study successfully, with 8 being excluded due to failing attention checks. The average duration for completion was approximately 22 minutes. The sample size is deemed sufficient for statistical significance with $\alpha = 0.05$, *power* = 0.80, and *effect size* = 0.35, as determined by power analysis.

8.6.1 Demographic profile and initial responses

An overview of the 71 participants' demographics and responses is as follows:

- *Gender Distribution*: 31 female, 39 male, 1 preferred not to disclose.
- *Age Groups*: 60 between 18-31 years, 5 between 31-50 years, 6 over 50.
- *Education Levels*: 20 with high school, 47 with university degrees, 4 with doctorates.
- *Professional Backgrounds*: 33 in science, 14 in technical fields, 9 in humanities, 4 in languages, 3 in economics, 8 in other areas.
- *Computer Literacy*: 5 beginners, 39 intermediate, 27 advanced.
- *Usage Frequency of Booking and E-commerce Platforms*: 2 never, 7 occasionally, 27 a few times a year, 35 a few times a month.

Regarding their preferences (Table 8.4), participants rated the importance of photos (S1, Mean = 4.718) and reviews (S2, Mean = 4.465) in home selection highly. They showed moderate trust in booking system recommendations (S3, Mean = 3.126) and somewhat agreed that user ratings are generally sufficient for booking decisions (S4, Mean = 3.014). Reviews (S5, Mean = 4.309), descriptions (S6, Mean = 4.451), and photos (S7, Mean = 4.662) were considered essential for making a booking.

8.6.2 Post-task questionnaire

Table 8.5 presents the outcomes of the post-task questionnaire, assessing user experiences across different visualization models. The model FILTER-BY-SCENE emerged as superior in terms of user experience across all evaluated statements, with a significant number of these findings being statistically notable; FILTER-BY-OBJ was identified as the next best model.

The most pronounced statistically significant findings underscore that FILTER-BY-SCENE effectively aided participants in comparing different homes (Q2, $p = 0.02$) and enabled them to access relevant information swiftly (Q8, $p = 0.002$). Additionally, FILTER-BY-SCENE was instrumental in aiding users to discern the merits or drawbacks of homes (Q1, $p = 0.011$). Users reported feeling more confident in making comparisons between homes using this system (Q12, $p = 0.035$).

Other significant observations include FILTER-BY-SCENE's ability to be informative (Q3) and provide sufficient data for making selections (Q5), without leading to information overload or confusion (Q4). The data about homes was deemed easy to interpret and understand (Q7). Users also preferred using FILTER-BY-SCENE frequently in their home comparison tasks (Q9), citing increased confidence over other systems.

FILTER-BY-SCENE notably enhanced decision-making capabilities compared to the BASELINE model, which mirrors the format of traditional home-booking platforms. Both BASELINE and FILTER-BY-OBJ, by presenting users with a large amount of information, makes the comparison of homes more difficult.

A small subset of participants offered open-ended feedback in the questionnaire. Key observations include:

- FILTER-BY-SCENE received 12 comments, with users appreciating the categorization of images and reviews according to specific home areas. The scene labels helped locate desired information, and the overall filtering approach was well-received.
- FILTER-BY-OBJ garnered 11 comments, with 8 noting that its highly specific filtering sometimes hampered obtaining a quick overall impression. Overlapping of reviews across photos was also seen as a downside, increasing the volume of text to read.
- BASELINE also received 11 comments, primarily critiquing its presentation of an excessive, unfiltered bulk of reviews, making it challenging to gain an overview and compare homes effectively.

Further analysis was conducted on how participants' personalities, including Need for Cognition (NfC) and traits from the TIPI questionnaire, influenced their experience with the three user interfaces.

8.6.3 Post-task questionnaire - Split by Need for Cognition

The division of participant responses in the post-task questionnaire, by their Need for Cognition (NfC), is illustrated in Table 8.6. We categorized participants into two groups: one consisting of 35 individuals with NfC scores below 3.5, and the other comprising 36 individuals with NfC scores of 3.5 or higher. This classification aimed to explore whether a propensity to engage in and enjoy intellectual activities influences how participants perceive our models' visual information filtering features.

Observations from this NfC-based analysis are aligned with those from the overall participant data, with a reduced number of statistically significant differences. Both the low-NfC and high-NfC groups showed a preference for FILTER-BY-SCENE over the alternative models. Interestingly, participants with lower NfC scores rated FILTER-BY-SCENE more favorably compared to the aggregate participant group. In contrast, those with higher NfC scores assigned marginally lower ratings to FILTER-BY-SCENE.

8.6.4 Post-task questionnaire - Split by personality traits

This analysis examines how different personality traits among users influenced their experience with our models. We split the user sample based on their personality scores, using 4 as the cutoff point (scores ≤ 4 and scores > 4).

- Focusing first on the trait of *openness*, which encompasses attributes like creativity, curiosity, and a propensity for new experiences, we observed distinct preferences. Participants scoring high in openness showed a strong preference for FILTER-BY-SCENE across all questionnaire aspects, evidenced by 5 statistically significant outcomes. In comparisons between FILTER-BY-OBJ and BASELINE, these users leaned towards FILTER-BY-OBJ for aspects related to item comparison and discovery (Q1, Q2, Q5, Q8, $p < 0.03$), though they found FILTER-BY-OBJ more overwhelming than BASELINE (Q4, $p = 0.026$). Participants with lower openness scores showed mixed preferences, with some inclination towards FILTER-BY-OBJ. Nevertheless, FILTER-BY-SCENE was the top performer for them in Q2 and Q7, the only categories with significant differences ($p < 0.1$).

- Next, we assessed the *conscientiousness* trait, indicative of a person's tendency towards diligence, meticulousness, and organization. This trait is often linked to a preference for detailed information in decision-making. Similar to observations in *agreeableness*, participants from both conscientiousness subgroups favored FILTER-BY-SCENE across all questionnaire statements. Particularly, those with higher conscientiousness found FILTER-BY-SCENE's home information comprehensive enough for making decisions (Q5, 4.019, $p = 0.042$). On the statistically significant fronts, FILTER-BY-OBJ was seen as superior to BASELINE.

- For traits like *extroversion*, *agreeableness*, and *neuroticism*, we found a unanimous preference for FILTER-BY-SCENE in both subgroups. In these cases, FILTER-BY-OBJ was ranked second, especially in instances with statistically significant results.

8.6.5 Post-test questionnaire results

In the post-test questionnaire, participants were asked to provide their evaluations regarding the effectiveness of the different filtering methods used in the study. Specifically, they assessed the utility of filtering reviews based on scenes compared to filtering based on objects. The analysis of their responses revealed a notable preference for the scene-based filtering approach.

Participants rated the scene-based review filtering as significantly more useful, with a mean score of $M = 4.253$. This higher score suggests that participants found the scene-oriented approach is more intuitive and relevant in understanding and evaluating the homes.

Table 8.7: Log Analysis. "Time" Is the Mean Number of Minutes Users Spent on its User Interface; "#Reviews" ("#Images") Is the Mean Number of Visualized Reviews (Images).

Event	**FILTER-BY-SCENE**	**FILTER-BY-OBJ**	**BASELINE**
Time (minutes)	3.1	2.612	1.771
#Reviews (visualized)	94.254	79.746	21.746
#Images clicked	32.944	32.803	26.62

On the other hand, the object-based filtering method received a somewhat lower, yet still appreciable, mean score of $M = 3.648$. This score reflects a moderate level of usefulness perceived by the participants. It suggests that while this method was beneficial in providing detailed insights into specific features of the homes, it might not have been as immediately impactful or contextually rich as the scene-based approach.

8.6.6 Log analysis

In our user study, we tracked the users' actions with logs and analyzed them. Table 8.7 presents the results of the analysis, which provides insight into how participants interacted with the different visualization models. One of the key findings is the amount of time users spent engaging with each model. Notably, participants devoted a greater amount of time interacting with FILTER-BY-SCENE compared to FILTER-BY-OBJ. Furthermore, the least amount of time was spent on BASELINE, indicating a potentially lower level of engagement or exploration with this model.

This trend in time allocation aligns closely with the volume of information that participants explored during the study. For instance, while using FILTER-BY-SCENE, they reviewed approximately 94 different reviews, whereas, with FILTER-BY-OBJ, around 80 reviews were visualized. In stark contrast, only about 22 reviews were explored when participants used BASELINE. This significant difference in the number of reviews they read suggests a deeper level of information processing and consideration when participants had access to the scene-based filter.

Interestingly, the analysis of photo visualizations tells a different story. Across all three models (FILTER-BY-SCENE, FILTER-BY-OBJ, and BASELINE), participants viewed a similar number of photos, averaging about 10 images per home in both FILTER-BY-SCENE and FILTER-BY-OBJ, and 9 in BASELINE.

The methodology for estimating the visualization of reviews and images involved tracking the duration each item was displayed on the screen. We only considered visualizations that lasted more than 2 seconds as significant.

These log analysis results highlight an interesting aspect of user behavior. It appears that regardless of the model, participants consistently prioritized visual information (images of homes) in their comparison and selection process. Yet, when it comes to textual content, there was a marked increase in engagement with the scene-based filter and a smaller increase with the object-based filter compared to the baseline. This suggests that while images are a constant factor in decision-making, the availability and type of textual information filtering significantly influence the depth of review exploration.

8.7 Discussion

The investigations carried out through the preliminary and main user studies in this chapter have provided substantial insights into the effectiveness of integrating multimodal information within service-based recommender systems. The findings across both studies reveal a clear preference among users for interfaces that combine textual and visual information, underscoring the essential role of images in enhancing item comparison and decision-making processes.

From the preliminary study, it emerged that interfaces like FILTER-WITH-IMG, which adeptly integrate images with service-aware information presentation, not only facilitate efficient data retrieval but also significantly improve the user experience by simplifying review analysis. Despite introducing a perceived complexity, this approach was found to offer a more nuanced and effective comparison tool than the more straightforward FULL-DATA, thereby affirmatively addressing our research questions regarding the utility of multimodal presentation in recommender systems.

The main study further builds on these insights by highlighting a distinct user preference for the FILTER-BY-SCENE interface over FILTER-BY-OBJ and BASELINE models. This preference was particularly pronounced among participants with lower Need for Cognition (NfC) scores, who favored the effective information filtering offered by FILTER-BY-SCENE. The coarser-grained, scene-based approach was appreciated for its ability to aggregate multimodal information in a way that enhanced user awareness and confidence, suggesting that the granularity of information filtering plays a crucial role in user satisfaction.

Moreover, both studies together point to a significant finding: while users value the integration of multimodal information for item comparison, there is a nuanced balance to be struck between the robustness of information presentation and the interface's usability. The preliminary study's feedback on the cluttered perception of FILTER-WITH-IMG and the main study's identification of the modest performance ratings of the interfaces underscore a common theme—the need for simplification and user-centric design in multimodal recommender systems.

The results across the studies underscore the utility of multimodal information, particularly images, in enriching the recommendation process. However, they also call attention to the importance of presentation and filtering mechanisms. The effectiveness of scene-based filtering in FILTER-BY-SCENE, as opposed to the detailed, object-based approach or the baseline, suggests a promising direction for future research and system development. It highlights the potential of contextually aggregated multimodal information in enhancing the user experience, especially in domains like home booking where visual and textual content is abundant.

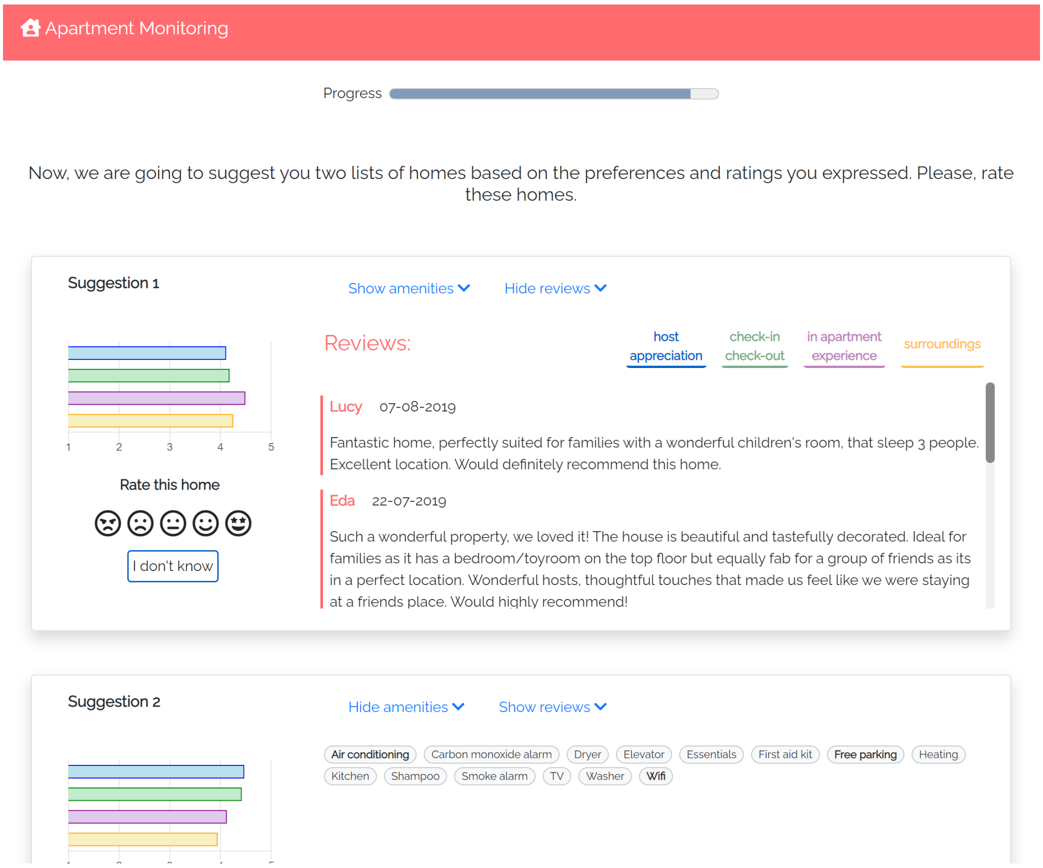
Chapter 9

Lessons learned

This chapter summarizes the significant insights and understandings derived from the user studies described in the previous chapters. Each study contributed a unique perspective, enriching our comprehension of service-aware recommender systems and multimodal user interfaces and addressing the goals described in Section 1.1.

9.1 Integration of service model in recommender systems

From Chapter 6, we learned that recommendations based purely on service experiences (STAGES) demonstrated high accuracy in item ranking, validating the importance of service-oriented data in recommender systems. However, a balanced approach, as seen in FEATURES-STAGES, which combines consumer experience with item features, emerged as the most effective. This model not only ranked high in performance but also improved user confidence and decision-making, positively answering our first research question (can the integration of service-oriented item representation with feature data enhance



Now, we are going to suggest you two lists of homes based on the preferences and ratings you expressed. Please, rate these homes.

Lucy 07-08-2019

Fantastic home, perfectly suited for families with a wonderful children's room, that sleep 3 people. Excellent location. Would definitely recommend this home.

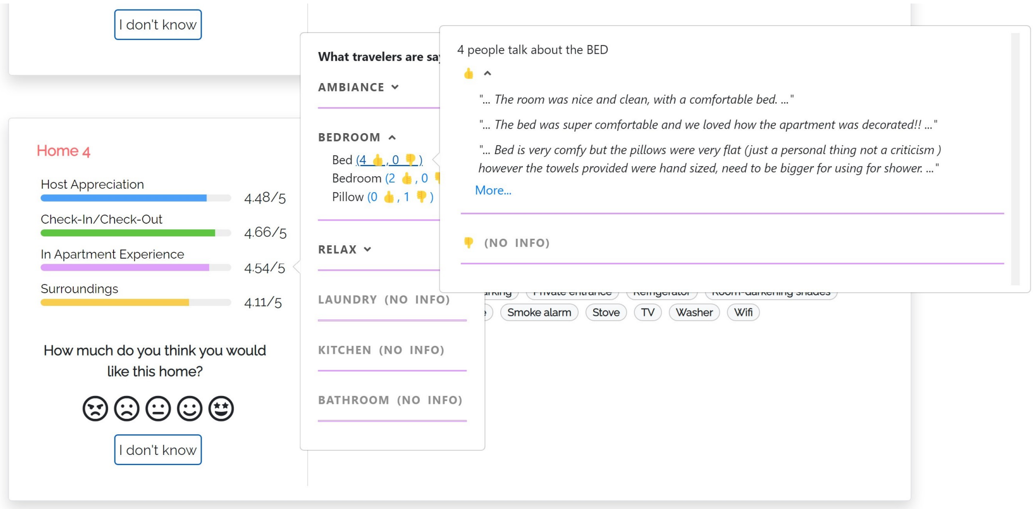
Eda 22-07-2019

Such a wonderful property, we loved it! The house is beautiful and tastefully decorated. Ideal for families as it has a bedroom/toyroom on the top floor but equally fab for a group of friends as its in a perfect location. Wonderful hosts, thoughtful touches that made us feel like we were staying at a friends place. Would highly recommend!

Figure 9.1: User interface of FEATURES-STAGES and CBF-STAGES systems.

the accuracy of recommendations?).

Moreover, in terms of presentation, the models combining different types of information, specifically CBF-STAGES and FEATURES-STAGES, substantially enhanced user confidence in evaluating items. This finding is crucial because it signals the importance of multimodal representation in recommendations – users feel more confident and supported when they are presented with a holistic view of the items, containing both quantitative data and qualitative experience feedback.



"... *The room was nice and clean, with a comfortable bed.* ..."

"... *The bed was super comfortable and we loved how the apartment was decorated!!* ..."

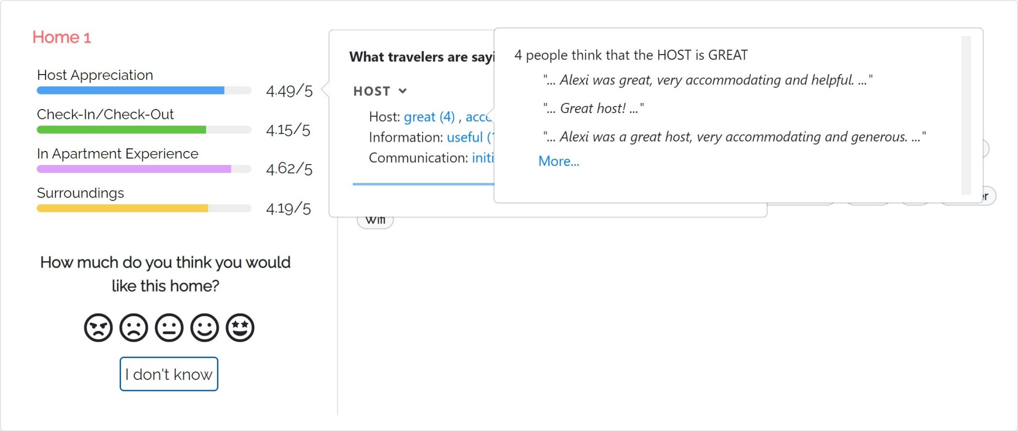
"... *Bed is very comfy but the pillows were very flat (just a personal thing not a criticism) however the towels provided were hand sized, need to be bigger for using for shower.* ..."

Figure 9.2: Interface of the justification M-THUMBS model.

9.2 Enhancing the justification models of results in service-aware recommender systems

Chapter 7's findings highlighted the effectiveness of our service-based justification models, particularly M-THUMBS (see Figure 9.2), in enhancing *Perceived User Awareness Support*. This study underscored the need for algorithm-agnostic, service-based approaches in recommender systems.

A significant aspect of this chapter was understanding how user experience in terms of interface adequacy and satisfaction varies with individual user traits, such as curiosity and cognitive styles. For instance, participants with higher curiosity levels expressed a clear preference for the detailed inspection capabilities of M-ASPECTS (see Figure 9.3). This model, which allows a granular examination of individual aspects of recommendations, was favored for providing a deeper dive into the data.


Home 1

Host Appreciation 4.49/5

Check-In/Check-Out 4.15/5

In Apartment Experience 4.62/5

Surroundings 4.19/5

How much do you think you would like this home?

I don't know

What travelers are sayi

HOST

Host: great (4) , acc

Information: useful (

Communication: initi

Wifi

4 people think that the HOST is GREAT

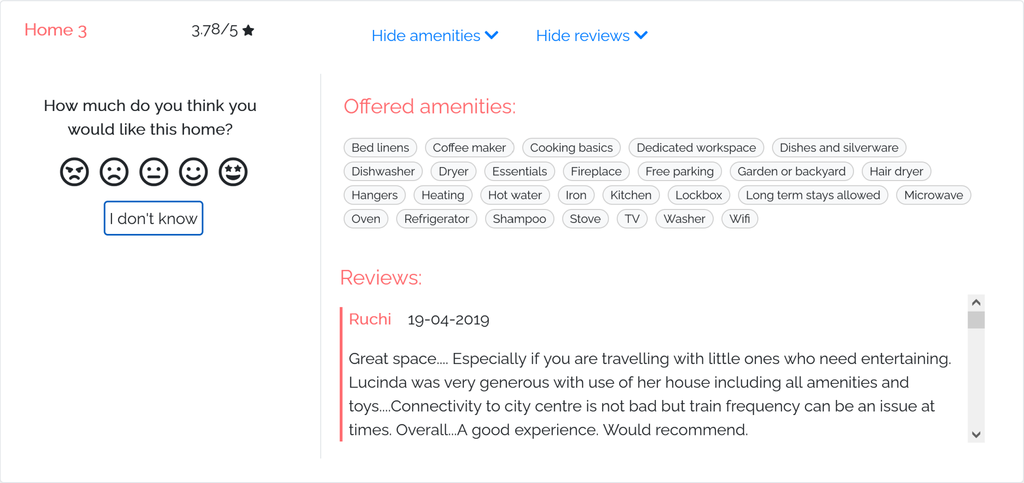
"... *Alexi was great, very accommodating and helpful. ...*"

"... *Great host! ...*"

"... *Alexi was a great host, very accommodating and generous. ...*"

More...

Figure 9.3: Interface of the M-ASPECTS model.


Home 3 3.78/5 ★

Hide amenities Hide reviews

How much do you think you would like this home?

I don't know

Offered amenities:

Bed linens · Coffee maker · Cooking basics · Dedicated workspace · Dishes and silverware · Dishwasher · Dryer · Essentials · Fireplace · Free parking · Garden or backyard · Hair dryer · Hangers · Heating · Hot water · Iron · Kitchen · Lockbox · Long term stays allowed · Microwave · Oven · Refrigerator · Shampoo · Stove · TV · Washer · Wifi

Reviews:

Ruchi 19-04-2019

Great space.... Especially if you are travelling with little ones who need entertaining. Lucinda was very generous with use of her house including all amenities and toys....Connectivity to city centre is not bad but train frequency can be an issue at times. Overall...A good experience. Would recommend.

Figure 9.4: Interface of the M-REVIEWS model.

Conversely, users with lower Need for Cognition (NfC) favored M-REVIEWS (see Figure 9.4), which presents item reviews directly.

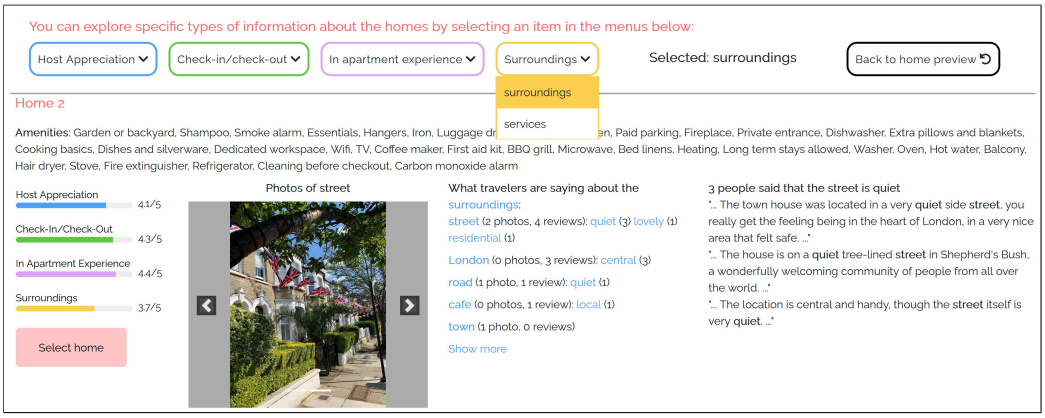
These findings indicate that recommender systems could be significantly enhanced by tailoring the user interface to individual user characteristics.

The Structural Equation Model analysis provided further validation, affirming that a more detailed presentation of item data, which enhances *Perceived User-Awareness Support* (U), positively impacts users' perceptions of the *Perceived Quality of Recommendations* (Q). This relationship is key to understanding user satisfaction with recommender systems: users are more likely to appreciate and trust recommendations when they are presented with ample and well-organized information.

Furthermore, the lower rates of opting out in home evaluations with M-THUMBS and M-ASPECTS suggest these models' effectiveness in facilitating informed decision-making. This is further supported by log data, which revealed that users were able to make decisions with a relatively modest amount of information when it was well-organized and relevant.

9.3 Multimodal user interfaces

The preliminary user study, as detailed in Chapter 8.2, offered critical insights into the effectiveness of incorporating both textual and visual information within recommender systems, specifically in the context of home comparisons on platforms like Airbnb. The integration of images, exemplified by models such as FILTER-WITH-IMG (see Figure 9.5) and FULL-DATA (see Figure 9.6), significantly enhanced users' ability to compare and evaluate potential accommodations, underscoring the vital role of visual elements alongside textual data.



You can explore specific types of information about the homes by selecting an item in the menus below:

Selected: surroundings

Home 2

Amenities: Garden or backyard, Shampoo, Smoke alarm, Essentials, Hangers, Iron, Luggage dr en, Paid parking, Fireplace, Private entrance, Dishwasher, Extra pillows and blankets, Cooking basics, Dishes and silverware, Dedicated workspace, Wifi, TV, Coffee maker, First aid kit, BBQ grill, Microwave, Bed linens, Heating, Long term stays allowed, Washer, Oven, Hot water, Balcony, Hair dryer, Stove, Fire extinguisher, Refrigerator, Cleaning before checkout, Carbon monoxide alarm

Host Appreciation 4.1/5

Check-In/Check-Out 4.3/5

In Apartment Experience 4.4/5

Surroundings 3.7/5

Photos of street

What travelers are saying about the surroundings:

street (2 photos, 4 reviews): quiet (3) lovely (1) residential (1)

London (0 photos, 3 reviews): central (3)

road (1 photo, 1 review): quiet (1)

cafe (0 photos, 1 review): local (1)

town (1 photo, 0 reviews)

Show more

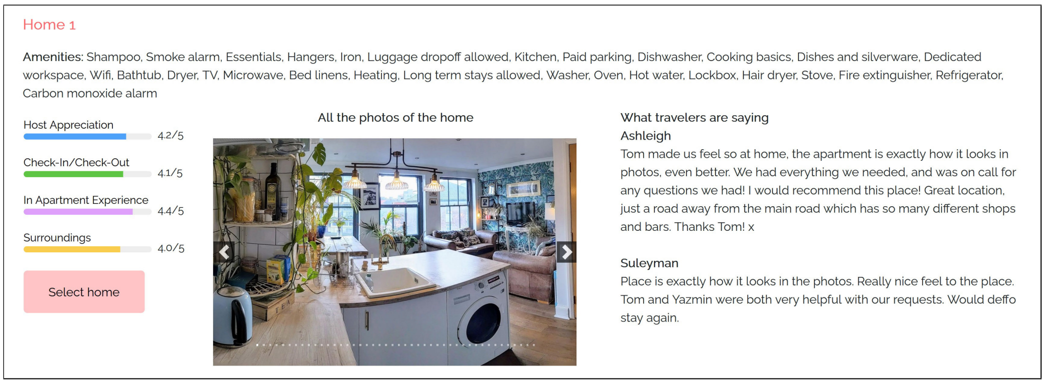
3 people said that the street is quiet

"... The town house was located in a very **quiet** side **street**, you really get the feeling being in the heart of London, in a very nice area that felt safe. ..."

"... The house is on a **quiet** tree-lined **street** in Shepherd's Bush, a wonderfully welcoming community of people from all over the world. ..."

"... The location is central and handy, though the **street** itself is very **quiet**. ..."

Figure 9.5: FILTER-WITH-IMG model's user interface.



Home 1

Amenities: Shampoo, Smoke alarm, Essentials, Hangers, Iron, Luggage dropoff allowed, Kitchen, Paid parking, Dishwasher, Cooking basics, Dishes and silverware, Dedicated workspace, Wifi, Bathtub, Dryer, TV, Microwave, Bed linens, Heating, Long term stays allowed, Washer, Oven, Hot water, Lockbox, Hair dryer, Stove, Fire extinguisher, Refrigerator, Carbon monoxide alarm

Host Appreciation 4.2/5

Check-In/Check-Out 4.1/5

In Apartment Experience 4.4/5

Surroundings 4.0/5

All the photos of the home

What travelers are saying

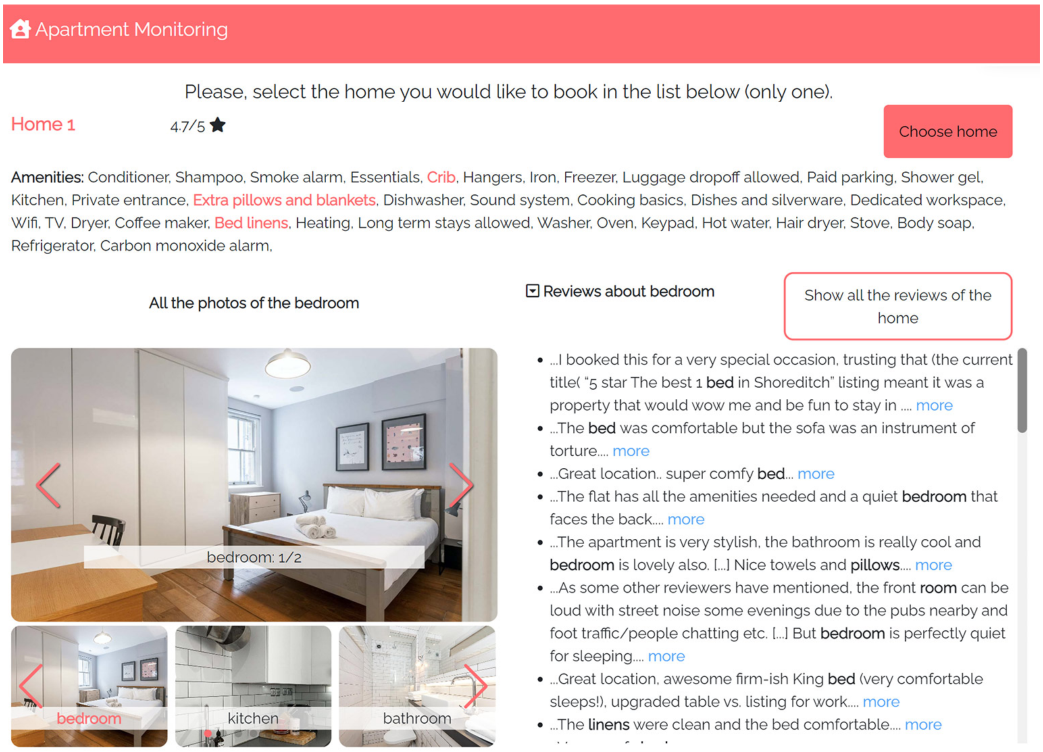
Ashleigh

Tom made us feel so at home, the apartment is exactly how it looks in photos, even better. We had everything we needed, and was on call for any questions we had! I would recommend this place! Great location, just a road away from the main road which has so many different shops and bars. Thanks Tom! x

Suleyman

Place is exactly how it looks in the photos. Really nice feel to the place. Tom and Yazmin were both very helpful with our requests. Would deffo stay again.

Figure 9.6: FULL-DATA model's user interface.

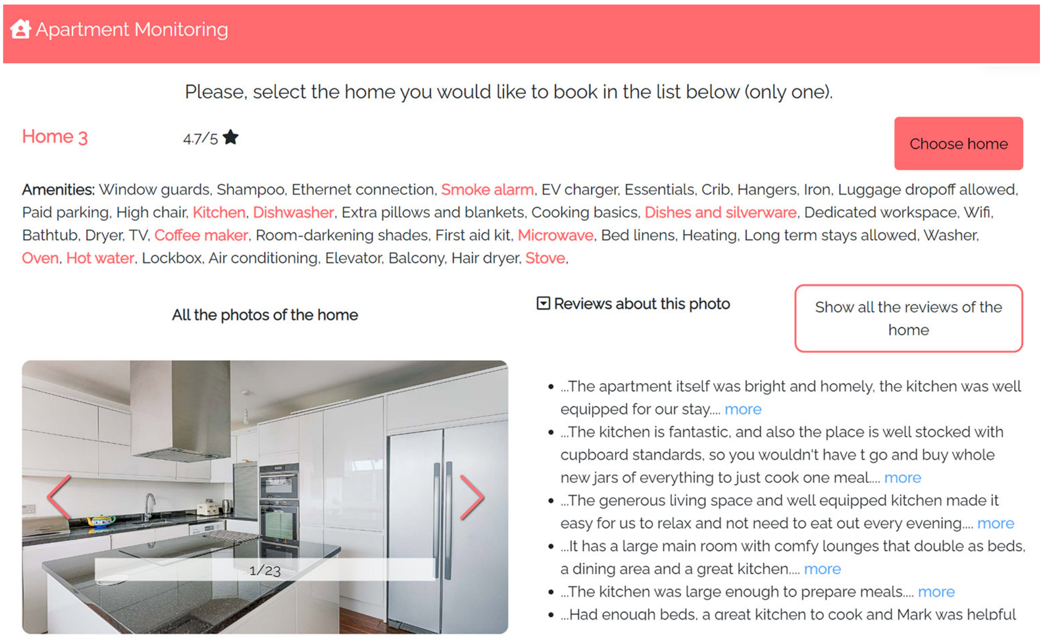


Please, select the home you would like to book in the list below (only one).

Amenities: Conditioner, Shampoo, Smoke alarm, Essentials, Crib, Hangers, Iron, Freezer, Luggage dropoff allowed, Paid parking, Shower gel, Kitchen, Private entrance, Extra pillows and blankets, Dishwasher, Sound system, Cooking basics, Dishes and silverware, Dedicated workspace, Wifi, TV, Dryer, Coffee maker, Bed linens, Heating, Long term stays allowed, Washer, Oven, Keypad, Hot water, Hair dryer, Stove, Body soap, Refrigerator, Carbon monoxide alarm,

- ...I booked this for a very special occasion, trusting that (the current title(“5 star The best 1 **bed** in Shoreditch” listing meant it was a property that would wow me and be fun to stay in .... more
- ...The **bed** was comfortable but the sofa was an instrument of torture.... more
- ...Great location.. super comfy **bed**... more
- ...The flat has all the amenities needed and a quiet **bedroom** that faces the back.... more
- ...The apartment is very stylish, the bathroom is really cool and **bedroom** is lovely also. [...] Nice towels and **pillows**.... more
- ...As some other reviewers have mentioned, the front **room** can be loud with street noise some evenings due to the pubs nearby and foot traffic/people chatting etc. [...] But **bedroom** is perfectly quiet for sleeping.... more
- ...Great location, awesome firm-ish King **bed** (very comfortable sleeps!), upgraded table vs. listing for work.... more
- ...The **linens** were clean and the bed comfortable.... more

Figure 9.7: FILTER-BY-SCENE user interface.



Please, select the home you would like to book in the list below (only one).

Home 3 4.7/5 ★

Amenities: Window guards, Shampoo, Ethernet connection, Smoke alarm, EV charger, Essentials, Crib, Hangers, Iron, Luggage dropoff allowed, Paid parking, High chair, Kitchen, Dishwasher, Extra pillows and blankets, Cooking basics, Dishes and silverware, Dedicated workspace, Wifi, Bathtub, Dryer, TV, Coffee maker, Room-darkening shades, First aid kit, Microwave, Bed linens, Heating, Long term stays allowed, Washer, Oven, Hot water, Lockbox, Air conditioning, Elevator, Balcony, Hair dryer, Stove.

All the photos of the home

Reviews about this photo

- ...The apartment itself was bright and homely, the kitchen was well equipped for our stay.... more
- ...The kitchen is fantastic, and also the place is well stocked with cupboard standards, so you wouldn't have t go and buy whole new jars of everything to just cook one meal.... more
- ...The generous living space and well equipped kitchen made it easy for us to relax and not need to eat out every evening.... more
- ...It has a large main room with comfy lounges that double as beds, a dining area and a great kitchen.... more
- ...The kitchen was large enough to prepare meals.... more
- ...Had enough beds, a great kitchen to cook and Mark was helpful

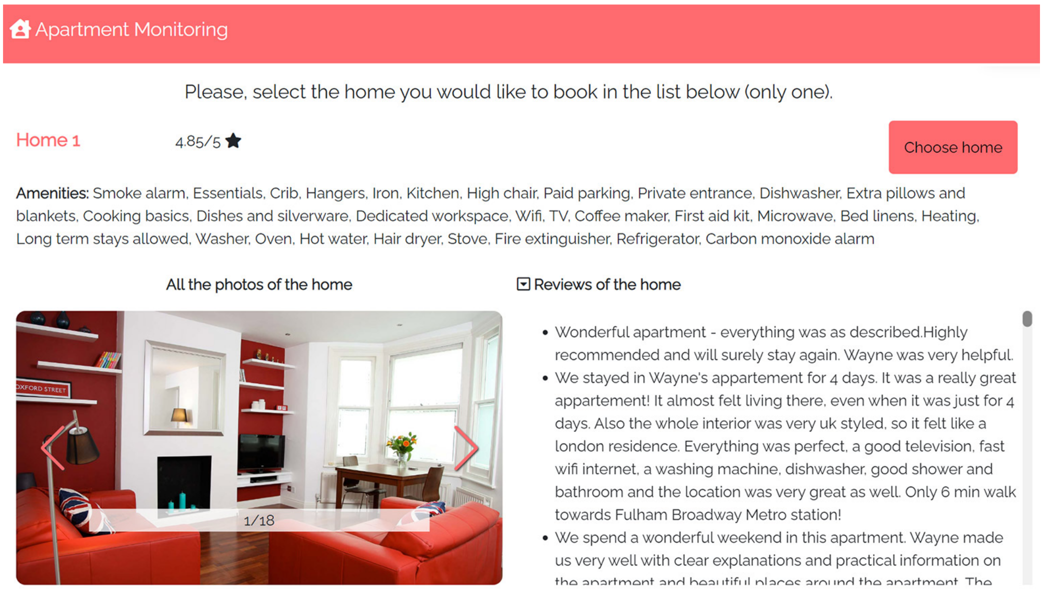
Figure 9.8: FILTER-BY-OBJ user interface.

A lesson from this study was the recognition of the value added by service-based filters in multimodal data exploration, replying positively to the third research question (How does a multimodal service-based presentation of images and textual data impact the effectiveness of the recommendation comparison process against traditional, non-stage-specific presentations), despite the perceived complexity these added to the user interface.

The perceived moderate clutter of FILTER-WITH-IMG leads to a recommendation for future development: simplifying the interface. This simplification could enhance the user experience by reducing complexity while maintaining the benefits of multimodal information integration.

The recommendation emerging to simplify the interface of FILTER-WITH-IMG underlines a crucial aspect of user experience: the balance between information richness and interface complexity.

Recognizing the need for simplification, the following main user study



Apartment Monitoring

Please, select the home you would like to book in the list below (only one).

Home 1 4.85/5 ★

Amenities: Smoke alarm, Essentials, Crib, Hangers, Iron, Kitchen, High chair, Paid parking, Private entrance, Dishwasher, Extra pillows and blankets, Cooking basics, Dishes and silverware, Dedicated workspace, Wifi, TV, Coffee maker, First aid kit, Microwave, Bed linens, Heating, Long term stays allowed, Washer, Oven, Hot water, Hair dryer, Stove, Fire extinguisher, Refrigerator, Carbon monoxide alarm

All the photos of the home

Reviews of the home

- Wonderful apartment - everything was as described.Highly recommended and will surely stay again. Wayne was very helpful.
- We stayed in Wayne's appartement for 4 days. It was a really great appartement! It almost felt living there, even when it was just for 4 days. Also the whole interior was very uk styled, so it felt like a london residence. Everything was perfect, a good television, fast wifi internet, a washing machine, dishwasher, good shower and bathroom and the location was very great as well. Only 6 min walk towards Fulham Broadway Metro station!
- We spend a wonderful weekend in this apartment. Wayne made us very well with clear explanations and practical information on the apartment and beautiful places around the apartment. The

Figure 9.9: BASELINE user interface.

tested simpler interfaces that omitted the detailed service model, instead integrating scene recognition. This approach aimed to retain the benefits of multimodal information integration while reducing interface clutter and enhancing user navigability and overall experience.

A significant outcome of this study was the clear preference among users, especially those with lower Need for Cognition (NfC) scores, for the FILTER-BY-SCENE interface (Figure 9.7) over the FILTER-BY-OBJ and BASELINE models (Figure 9.8 and Figure 9.9). This preference underscores the effectiveness and user-friendliness of scene-based filtering approaches. FILTER-BY-SCENE, which aggregates multimodal information (both visual and textual) by context (such as room types or services), was particularly appreciated for its ability to facilitate the decision-making process by providing contextually relevant information in a more digestible format.

In addressing the fourth research question, the study revealed that the utility of item images in filtering information is significant, underscoring their important role in enriching the user experience and effectiveness of recommender systems. Moreover, the granularity of filtering emerged as a key factor: the data indicated that scene-based filtering was more effective in enhancing user awareness and confidence in their selection choices compared to object-based filtering and the baseline.

However, one notable limitation identified through this study was the need for simplification. This feedback is crucial as it highlights a recurring theme from the previous chapters – the importance of balancing information richness with interface simplicity and usability. Users' preference for the FILTER-BY-SCENE model, with its context-based information aggregation, signals a promising direction for future interface designs in recommender systems, particularly those that involve complex and multifaceted choices like home booking.

In conclusion, Chapter 8's findings reinforce the importance of carefully considering how information is presented in recommender systems. The insights gained from these studies advocate for a user-centric approach in the design of recommender systems, where both the richness of information and the simplicity of its presentation are harmoniously balanced to enhance the overall user experience.

Chapter 10

Conclusion

The work of this thesis aims to enhance recommender systems by integrating service-based models and multimodal information in the inferences carried out by the systems and in the presentation of information to the user. Key contributions of this research include the development of a novel recommender system model that transcends traditional item-centric approaches, focusing on the holistic user experience in item fruition. This was achieved through the integration of service models and multimodal data, particularly image analysis, resulting in a more comprehensive presentation of items, thereby enhancing user decision-making and overall experience with the system.

The advancements proposed in this thesis contribute to the field of recommender systems. By incorporating service models and multimodal information, this research contributes to addressing the growing demand for transparent, accountable, and user-centric AI, as required by contemporary regulatory frameworks like the GDPR. This shift from algorithmic efficiency to user journey-centric approaches represents a significant paradigm shift in the design and functionality of recommender systems.

The methodology adopted, involving the development of a new recommender system model, new justification models, and their validation through user studies, has provided empirical evidence supporting the effectiveness of service-aware and multimodal recommender systems. The findings indicate that such systems not only improve the accuracy of recommendations but also significantly enhance user trust and satisfaction, affirming the importance of considering the user's cognitive styles and personal traits in recommender system design.

While this research has made steps, it also acknowledges certain limitations. These limitations stem primarily from the challenges associated with integrating multimodal data, which introduces a level of complexity in the user interfaces. There is a fine balance to be found between offering rich, multimodal information and ensuring the user interface remains intuitive and user-friendly. The complexity of handling diverse data types—such as text, images, and user ratings—requires to not overwhelm the user or compromise the system's usability.

Another noteworthy limitation of our proposed multimodal approach is the dependency on the object recognition models and scene detection capabilities. The efficacy of these systems is critically contingent upon their ability to accurately identify and understand the context of various objects within a scene. However, a notable challenge occurs with objects that are absent from the training dataset. Such instances lead to a gap in recognition, as the models lack the necessary data to identify and process these objects.

Furthermore, the construction of vocabularies involved manual engineering. This approach was chosen to ensure the high quality of the output. However, recognizing the potential for scalability and efficiency, future research will explore the integration of recent advancements in Large Language Models (LLMs) to automate aspects of this process. For example, LLMs could be employed to assist in the generation of domain-specific vocabularies or in the interpretation and classification of complex user feedback, thereby reducing manual effort while maintaining or even enhancing the quality of the system's outputs.

The validity of our findings also merits discussion. The user studies conducted as part of this research rely on hypothetical scenarios where users provide ratings based on their perceived experience with the items recommended by our system. For the chosen domain, it was not possible to let the user really experience the recommended items. This methodology, while common in the field, raises questions about the reliability of such hypothetical ratings. Future studies could incorporate methods to simulate the experience of item consumption more realistically. For instance, in the domain of real estate, virtual tours or immersive experiences could provide users with a closer approximation of staying in a home.

Lastly, the choice of a specific domain for this study was made to provide depth and focus to our research. However, the principles underlying service-based and multimodal recommender systems, as discussed in this thesis, have broad applicability across various domains. While the experiments were not replicated in other domains due to feasibility constraints, the underlying theories and methodologies suggest a high potential for generalization. For instance, other services such as restaurants and experiences can be used for testing the external validity of our findings.

Looking forward, future research directions could include efforts to simplify the user interfaces of the justification models. The goal would be to retain the depth and breadth of multimodal information without sacrificing ease of use. Simplifying these interfaces requires data presentation that does not dilute the quality or comprehensiveness of the information provided. Additionally, the scalability of these models across different domains warrants further exploration.

Another promising direction involves incorporating geographic GPS data and other auxiliary datasets to enhance the surrounding evaluation dimension of recommendations. This integration would allow recommender systems to consider not just the properties of the items or services themselves but also their context and environment. By providing a holistic view of items and services, such systems could significantly improve the overall user experience.

Building upon the successes of image-based filtering in this research, future systems could explore the use of object recognition to create interactive image reviews. These images, designed to be hoverable or clickable, could reveal user reviews or information related to the recognized objects within the images. This approach promises to offer users a more intuitive and informative way to explore and understand the recommended items or services, enhancing their decision-making process.

Furthermore, an opportunity exists to create recommender system models and justification approaches that personalize results based on the user's personality. Personalizing recommendations and justifications according to individual user profiles might lead to increased user satisfaction and trust in the system. This personalization would align recommendations more closely with individual preferences and cognitive styles, offering a novel dimension to user experience in recommender systems.

In conclusion, the integration of service models and multimodal information, as explored in this thesis, opens new horizons for the design of intelligent systems that not only recommend but also reason about users' needs and preferences, in a broader context of evaluation, ultimately leading to more informed and satisfactory user decisions.

Appendix A

Listings description

In this Appendix, we provide the list of the fields of the listings of the dataset downloaded (see Chapter 4)

Field	**Description**
id	Airbnb's unique identifier for the listing
listing url	
scrape id	Inside Airbnb "Scrape" this was part of
last scraped	UTC. The date and time this listing was "scraped".
name	Name of the listing
description	Detailed description of the listing
neighborhood overview	Host's description of the neighbourhood
picture url	URL to the Airbnb hosted regular sized image for the listing

host id	Airbnb's unique identifier for the host/user
host url	The Airbnb page for the host
host name	Name of the host. Usually just the first name(s).
host since	The date the host/user was created. For hosts that are Airbnb guests this could be the date they registered as a guest.
host location	The host's self reported location
host about	Description about the host
host response time	
host response rate	
host acceptance rate	That rate at which a host accepts booking requests.
host is superhost	
host thumbnail url	
host picture url	
host neighbourhood	
host listings count	The number of listings the host has (per Airbnb calculations)
host total listings count	The number of listings the host has (per Airbnb calculations)
host verifications	
host has profile pic	
host identity verified	

neighbourhood	
neighbourhood cleansed	The neighbourhood as geocoded using the latitude and longitude against neighborhoods as defined by open or public digital shapefiles.
neighbourhood group cleansed	The neighbourhood group as geocoded using the latitude and longitude against neighborhoods as defined by open or public digital shapefiles.
latitude	Uses the World Geodetic System (WGS84) projection for latitude and longitude.
longitude	Uses the World Geodetic System (WGS84) projection for latitude and longitude.
property type	Self selected property type. Hotels and Bed and Breakfasts are described as such by their hosts in this field
room type	All homes are grouped into the following three room types: Entire place, Private room, Shared room, Entire place
accommodates	The maximum capacity of the listing

bathrooms	The number of bathrooms in the listing
bathrooms text	The number of bathrooms in the listing. On the Airbnb web-site, the bathrooms field has evolved from a number to a textual description. For older scrapes, bathrooms is used.
bedrooms	The number of bedrooms
beds	The number of bed(s)
amenities	List of the amenities
price	daily price in local currency
minimum nights	minimum number of night stay for the listing (calendar rules may be different)
maximum nights	maximum number of night stay for the listing (calendar rules may be different)
minimum minimum nights	the smallest minimum night value from the calender (looking 365 nights in the future)
maximum minimum nights	the largest minimum night value from the calender (looking 365 nights in the future)
minimum maximum nights	the smallest maximum night value from the calender (looking 365 nights in the future)

maximum maximum nights	the largest maximum night value from the calender (looking 365 nights in the future)
minimum nights avg ntm	the average minimum night value from the calender (looking 365 nights in the future)
maximum nights avg ntm	the average maximum night value from the calender (looking 365 nights in the future)
calendar updated	
has availability	[t=true; f=false]
availability 30/60/90/365	avaliability x. The availability of the listing x days in the future as determined by the calendar. Note a listing may not be available because it has been booked by a guest or blocked by the host.
calendar last scraped	
number of reviews	The number of reviews the listing has
number of reviews ltm	The number of reviews the listing has (in the last 12 months)
number of reviews l30d	The number of reviews the listing has (in the last 30 days)
first review	The date of the first/oldest review
last review	The date of the last/newest review

review scores rating	
review scores accuracy	
review scores cleanliness	
review scores checkin	
review scores communication	
review scores location	
review scores value	
license	The licence/permit/registration number
instant bookable	[t=true; f=false]. Whether the guest can automatically book the listing without the host requiring to accept their booking request. An indicator of a commercial listing.
calculated host listings count	The number of listings the host has in the current scrape, in the city/region geography.
calculated host listings count entire homes	The number of Entire home/apt listings the host has in the current scrape, in the city/region geography
calculated host listings count private rooms	The number of Private room listings the host has in the current scrape, in the city/region geography
calculated host listings count shared rooms	The number of Shared room listings the host has in the current scrape, in the city/region geography

reviews per month	The number of reviews the listing has over the lifetime of the listing

Appendix B

Thesauri evaluation dimension/Keywords

Evaluation Dimension	Keywords
Host appreciation	host, owner, renter, interaction, people, relation, hospitality, manner, language, communication
Search on website	search, reservation, booking, arrangement, agreement, deal, line, sign, message, channel, mail, voice, information, info, stuff, example, program, website
Check-in/Check-out	entrance, arrival, entry, suggestion, term, conversation, understanding, welcome, regard, key, english, reception, check-in, check-out, query, wait, money, checkin, checkout, hour, check, help, direction, instruction, advice, luggage, access, bag, wheelchair, mobility, baggage, departure, time, delay, document, identification, code

In-apartment experience	visit, family, experience, dog, cat, animal, parking, room, space, night, morning, view, living, bed, bedroom, water, door, bathroom, bath, garden, floor, stair, shower, clean, step, call, kitchen, interior, exterior, decoration, amenity, amenity, wi-fi, wifi, shower, maintenance, cleaning, fixture, repair, support, sheet, cover, blanket, cookware, cooker, kettle, pot, air, conditioning, conditioner, lighting, fridge, home, appliances, washer, refrigerator, dishwasher, freezer, tv, pc, computer, laptop, meal, dish, tea, breakfast, dinner, snack, launch, smoking, smoke, air, breeze, gas, temperature, heat, smell, light, sun, sight, atmosphere, ambiance, sunlight, sunshine, ray, furniture, relax, safety, security, law, guard, lock, box, pool, balcony, cleanliness, material, phone, stay, cook, experience, party, meal, terrace, accommodation, porch, supply, fragrance, courtyard, beverage, snack, treat, speaker, towel, platter, air, stove, furnishing, bedspread, table, equipment, bunkbed, pleasure, size, area, coffee, insect, mosquito, ceiling, dryer, breakfast, library, bird, television, privacy, toiletry, guest, lack, terrasse, hallway, facility, house, accessibility, location, apartment, apt, place, home, block, suite, hostel, rooms, flat, construction, penthouse, base, view, architecture, garden, yard, backyard, grove, field, playground, design, decor, layout, order, color, style, paint, space, internet, mattress, window, curtain, heater, lamp, soap, shampoo

Surroundings	noise, music, sound, voice, disturbance, bell, quietness, city, beach, transport, airport, café, restaurant, walking, nearby, food, shops, bus, station, ferry, street, surrounding, attraction, crowd, town, cab, neighborhood, park, culture, walk, bakery, outskirt, transportation, downtown, center, ride, zone, trip, square, road, taxi, sunset, shop, store, museum, weather, eatery, traffic, distance, sport, gym, swimming pool, silence, mountain, lake, river, crops, sea, seaside, beach, shopping, neighbour, neighbor, neighbourhood, street, park, playground, pub, disco, club

Table B.1: Thesauri evaluation dimensions/keywords for the home booking domain from Mauro et al. (2020a)

Bibliography

Adomavicius, G., Manouselis, N., and Kwon, Y. (2011). *Multi-Criteria Recommender Systems*, pages 769–803. Springer US, Boston, MA.

Airbnb (2022). Airbnb. `https://airbnb.com`.

Amal, S., Tsai, C.-H., Brusilovsky, P., Kuflik, T., and Minkov, E. (2019). Relational social recommendation: application to the academic domain. *Expert Systems with Applications*, 124:182 – 195.

Basu, A. and Ahad, R. (1992). Using a relational database to support explanation in a knowledge-based system. *IEEE Transactions on Knowledge and Data Engineering*, 4(6):572–581.

Berkovsky, S., Kuflik, T., and Ricci, F. (2008). Mediation of user models for enhanced personalization in recommender systems. *User Model. User-Adapt. Interact.*, 18:245–286.

Bitner, M. J., Ostrom, A. L., and Morgan, F. N. (2008). Service blueprinting: A practical technique for service innovation. *California Management Review*, 50(3):66–94.

Brandão, L., Belfo, F., and Silva, A. (2021). Wavelet-based cancer drug recommender system. *Procedia Computer Science*, 181:487–494. CENTER-IS/ProjMAN/HCist 2020.

Burke, R. (2002). Hybrid recommender systems: survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4):331–370.

Cardoso, B., Sedrakyan, G., Gutiérrez, F., Parra, D., Brusilovsky, P., and Verbert, K. (2019). IntersectionExplorer, a multi-perspective approach for exploring recommendations. *International Journal of Human-Computer Studies*, 121:73 – 92.

Chakraborty, S., Hoque, M. S., Rahman Jeem, N., Biswas, M. C., Bardhan, D., and Lobaton, E. (2021). Fashion recommendation systems, models and methods: A review. *Informatics*, 8(3).

Chang, J. C., Hahn, N., Perer, A., and Kittur, A. (2019). SearchLens: composing and capturing complex user interests for exploratory search. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, IUI '19, pages 498–509, New York, NY, USA. ACM.

Chen, L., Chen, G., and Wang, F. (2015). Recommender systems based on user reviews: the state of the art. *User Modeling and User-Adapted Interaction*, 25(2):99–154.

Chen, L. and Wang, F. (2017). Explaining recommendations based on feature sentiments in product reviews. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, IUI '17, page 17–28, New York, NY, USA. Association for Computing Machinery.

Chen, L., Wang, F., Qi, L., and Liang, F. (2014). Experiment on sentiment embedded comparison interface. *Knowledge-Based Systems*, 64:44–58.

Cheng, M. and Jin, X. (2019). What do Airbnb users care about? An analysis of online review comments. *International Journal of Hospitality Management*, 76:58 – 70.

Chu, W.-T. and Tsai, Y.-L. (2017). A hybrid recommendation system considering visual information for predicting favorite restaurants. In *Proceedings of the 26th Int. Conf. on World Wide Web*, WWW '17, page pages 1313–1331, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

Coelho, G., Hanel, P. H. P., and Wolf, L. J. (2020). The very efficient assessment of need for cognition: Developing a six-item version. *Assessment*, 27(8):1870–1885.

Conati, C., Barral, O., Putnam, V., and Rieger, L. (2021). Toward personalized XAI: A case study in intelligent tutoring systems. *Artificial Intelligence*, 298:103503.

Deldjoo, Y., Nazary, F., Ramisa, A., Mcauley, J., Pellegrini, G., Bellogin, A., and Di Noia, T. (2022). A review of modern fashion recommender systems.

Di Sciascio, C., Brusilovsky, P., Trattner, C., and Veas, E. (2019). A roadmap to user-controllable social exploratory search. *ACM Transaction on Interactive Intelligent Systems*, 10(1).

Di Sciascio, C., Sabol, V., and Veas, E. E. (2016). Rank as you go: User-driven exploration of search results. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, IUI '16, page 118–129, New York, NY, USA. Association for Computing Machinery.

Dong, R. and Smyth, B. (2017). User-based opinion-based recommendation. In *Proceedings 26th IJCAI*, pages 4821–4825, Melbourne, Australia.

Ekstrand, M., Harper, F., Willemsen, M., and Konstan, J. (2014). User perception of differences in recommender algorithms. In *Proceedings of the*

8th ACM Conference on Recommender Systems (RecSys'14), October 6–10, 2014, Silicon Valley, California, pages 161–168, United States. Association for Computing Machinery, Inc. 8th ACM Conference on Recommender systems (RecSys 2014), RecSys 2014 ; Conference date: 06-10-2014 Through 10-10-2014.

Fischer, G. (2000). User modeling in human-computer interaction. *User Modeling and User-Adapted Interaction*, 11.

Geninatti Cossatin, A., Mauro, N., and Ardissono, L. (2023). Enriching recommender systems results with data about sustainability and ethical standards of brands. In *2023 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 238–242, New York, NY, USA. IEEE.

Geninatti Cossatin, A., Mauro, N., and Ardissono, L. (2024). Promoting green fashion consumption through digital nudges in recommender systems. *IEEE Access*.

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., and Kagal, L. (2018). Explaining explanations: An approach to evaluating interpretability of machine learning. *CoRR*, abs/1806.00069.

Golbeck, J. and Hendler, J. (2006). Filmtrust: movie recommendations using trust in web-based social networks. In *CCNC 2006. 2006 3rd IEEE Consumer Communications and Networking Conference, 2006.*, volume 1, pages 282–286.

Gosling, S., Rentfrow, P., and Swann, W. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality*, 37(6):504–528.

Gou, L., You, F., Guo, J., Wu, L., and Zhang, X. L. (2011). Sfviz: Interest-based friends exploration and recommendation in social networks. In *Proceedings of the 2011 Visual Information Communication - International Symposium*, VINCI '11, New York, NY, USA. Association for Computing Machinery.

Guan, X., Cheng, Z., He, X., Zhang, Y., Zhu, Z., Peng, Q., and Chua, T.-S. (2019). Attentive aspect modeling for review-aware recommendation. *ACM Trans. Inf. Syst.*, 37(3).

Hearst, M., Elliott, A., English, J., Sinha, R., Swearingen, K., and Yee, K.-P. (2002). Finding the flow in web site search. *Communications of the ACM*, 45(9):42–49.

Hearst, M. A. (2006). Design recommendations for hierarchical faceted search interfaces. In *Proceedings of SIGIR 2006, Workshop on Faceted Search*, pages 26–30.

Herlocker, J. L., Konstan, J. A., and Riedl, J. (2000). Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, CSCW '00, page 241–250, New York, NY, USA. Association for Computing Machinery.

Hu, Z. F. (2022). Service-aware recommendation and justification of results. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, pages 341–345.

Hu, Z. F., Kuflik, T., Mocanu, I. G., Najafian, S., and Shulner-Tal, A. (2021). Recent studies of XAI - review. In Masthoff, J., Herder, E., Tintarev, N., and Tkalcic, M., editors, *Adjunct Publication of the 29th ACM Conference* *on User Modeling, Adaptation and Personalization, UMAP 2021, Utrecht, The Netherlands, June 21-25, 2021*, pages 421–431. ACM.

Hu, Z. F., Mauro, N., and Ardissono, L. (2023a). Image-based information filtering to compare and select items. In *2023 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 1–8, New York, NY, USA. IEEE.

Hu, Z. F., Mauro, N., Petrone, G., and Ardissono, L. (2023b). Service-based presentation of multimodal information for the justification of recommender systems results. In *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '23, page 46–53, New York, NY, USA. Association for Computing Machinery.

Hutto, C. and Eric, G. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*, pages 216–225, New York, NY, USA. AAAI.

Iovine, A., Narducci, F., and Semeraro, G. (2020). Conversational recommender systems and natural language:: A study through the converse framework. *Decision Support Systems*, 131:113250.

Jannach, D., Jugovac, M., and Nunes, I. (2019). Explanations and user control in recommender systems. In *Proceedings of the 23rd International Workshop on Personalization and Recommendation on the Web and Beyond*, ABIS '19, page 31, New York, NY, USA. Association for Computing Machinery.

Jocher, G. (2022). YOLOv5. `https://github.com/ultralytics/yolov5`.

Kashdan, T., Gallagher, M., Silvia, P., Winterstein, B., Breen, W., Terhar, D., and Steger, M. (2009). The curiosity and exploration inventory-II:

Development, factor structure, and psychometrics. *Journal of research in personality*, 43:987–998.

Kawano, Y., Sato, T., Maruyama, T., and Yanai, K. (2013). [demo paper] mirurecipe: A mobile cooking recipe recommendation system with food ingredient recognition. In *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pages 1–2.

Kirk, R. E. (2013). *Experimental design: Procedures for the behavioral sciences*. SAGE Publications, Inc.

Kitamura, R. and Itoh, T. (2018). Tourist spot recommendation applying generic object recognition with travel photos. In *2018 22nd International Conference Information Visualisation (IV)*, pages 1–5.

Kobyshev, K., Voinov, N., and Nikiforov, I. (2021). Hybrid image recommendation algorithm combining content and collaborative filtering approaches. *Procedia Computer Science*, 193:200–209. 10th International Young Scientists Conference in Computational Science, YSC2021, 28 June – 2 July, 2021.

Koren, Y. (2008). Factorization meets the neighborhood: A multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, pages 426–434, New York, NY, USA. ACM.

Koren, Y. (2009). Collaborative filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, page 447–456, New York, NY, USA. Association for Computing Machinery.

Koren, Y. and Bell, R. (2015). *Advances in Collaborative Filtering*, pages 77–118. Springer US, Boston, MA.

Kouki, P., Schaffer, J., Pujara, J., O'Donovan, J., and Getoor, L. (2017). User preferences for hybrid explanations. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, RecSys '17, page 84–88, New York, NY, USA. Association for Computing Machinery.

Kouki, P., Schaffer, J., Pujara, J., O'Donovan, J., and Getoor, L. (2019). Personalized explanations for hybrid recommender systems. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, IUI '19, page 379–390, New York, NY, USA. Association for Computing Machinery.

Kouki, P., Schaffer, J., Pujara, J., O'Donovan, J., and Getoor, L. (2020). Generating and understanding personalized explanations in hybrid recommender systems. *ACM Transactions on Interactive Intelligent Systems*, 10(4).

Krening, S., Harrison, B., Feigh, K. M., Isbell, C. L., Riedl, M., and Thomaz, A. (2017). Learning from explanations using sentiment and advice in rl. *IEEE Transactions on Cognitive and Developmental Systems*, 9(1):44–55.

Lakkaraju, H., Bach, S. H., and Leskovec, J. (2016). Interpretable decision sets: A joint framework for description and prediction. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 1675–1684, New York, NY, USA. Association for Computing Machinery.

Lee, C. K. H. (2022). How guest-host interactions affect consumer experiences in the sharing economy: New evidence from a configurational analysis based on consumer reviews. *Decision Support Systems*, 152:113634.

Lei, T., Barzilay, R., and Jaakkola, T. S. (2016). Rationalizing neural predictions. *CoRR*, abs/1606.04155.

Lewis, J. R. and Sauro, J. (2009). The factor structure of the System Usability Scale. In Kurosu, M., editor, *Human Centered Design*, pages 94–103, Berlin, Heidelberg. Springer Berlin Heidelberg.

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P. (2015). Microsoft COCO: Common objects in context.

Linden, G., Smith, B., and York, J. (2003). Amazon.com recommendations - item-to-item collaborative filtering. *IEEE Internet Computing*, January-February:76–80.

Liu, X., Li, J., Wang, J., and Liu, Z. (2021). MMFashion: An open-source toolbox for visual fashion analysis. In *Proceedings of the 29th ACM International Conference on Multimedia*, MM '21, page 3755–3758, New York, NY, USA. Association for Computing Machinery.

Loepp, B., Herrmanny, K., and Ziegler, J. (2015). Blended recommending: integrating interactive information filtering and algorithmic recommender techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 975–984, New York, NY, USA. ACM.

Lops, P., de Gemmis, M., and Semeraro, G. (2011). *Content-based recommender systems: state of the art and trends*, pages 73–105. Springer US, Boston, MA.

Loria, S. (2020). TextBlob: Simplified text processing. `https://textblob.readthedocs.io/en/dev/index.html`.

Lu, Y., Dong, R., and Smyth, B. (2018). Why i like it: Multi-task learning for recommendation and explanation. In *Proceedings of the 12th ACM Conference on Recommender Systems*, RecSys '18, page 4–12, New York, NY, USA. Association for Computing Machinery.

Margaris, D., Vassilakis, C., and Spiliotopoulos, D. (2020). What makes a review a reliable rating in recommender systems? *Information Processing & Management*, 57(6):102304.

Mauro, N., Ardissono, L., Capecchi, S., and Galioto, R. (2020a). Service-aware interactive presentation of items for decision-making. *Applied Sciences, Special Issue Implicit and Explicit Human-Computer Interaction*, 10(16):5599.

Mauro, N., Ardissono, L., and Lucenteforte, M. (2020b). Faceted search of heterogeneous geographic information for dynamic map projection. *Information Processing & Management*, 57(4):102257.

Mauro, N., Hu, Z. F., and Ardissono, L. (2022a). Justification of recommender systems results: a service-based approach. *User Modeling and User-Adapted Interaction*, 33(3):643–685.

Mauro, N., Hu, Z. F., and Ardissono, L. (2022b). Service-aware personalized item recommendation. *IEEE Access*, 10:26715–26729.

Mauro, N., Hu, Z. F., Ardissono, L., and Izzi, G. (2021). A service-oriented perspective on the summarization of recommendations: Preliminary experiment. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, page 213–219, New York, NY, USA. Association for Computing Machinery.

McAuley, J. and Leskovec, J. (2013). Hidden factors and hidden topics: Understanding rating dimensions with review text. In *Proceedings of the*

7th ACM Conference on Recommender Systems, RecSys '13, page 165–172, New York, NY, USA. Association for Computing Machinery.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444.

Millecamp, M., Htun, N. N., Conati, C., and Verbert, K. (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*, IUI '19, page 397–407, New York, NY, USA. Association for Computing Machinery.

Millecamp, M., Htun, N. N., Conati, C., and Verbert, K. (2020). What's in a user? Towards personalising transparency for music recommender interfaces. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '20, page 173–182, New York, NY, USA. Association for Computing Machinery.

Muhammad, K. I., Lawlor, A., and Smyth, B. (2016). A live-user study of opinionated explanations for recommender systems. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, IUI '16, page 256–260, New York, NY, USA. Association for Computing Machinery.

Musto, C., de Gemmis, M., Lops, P., Narducci, F., and Semeraro, G. (2022). *Semantics and Content-Based Recommendations*, pages 251–298. Springer US, New York, NY.

Musto, C., de Gemmis, M., Lops, P., and Semeraro, G. (2021). Generating post hoc review-based natural language justifications for recommender systems. *User-Modeling and User-Adapted Interaction*, 31:629–673.

Musto, C., Narducci, F., Lops, P., de Gemmis, M., and Semeraro, G. (2019). Linked open data-based explanations for transparent recommender systems. *International Journal of Human-Computer Studies*, 121:93 – 107.

Nelson, P. (1974). Advertising as information. *Journal of Political Economy*, 82(4):729–754.

Ni, J., Li, J., and McAuley, J. (2019). Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 188–197, Hong Kong, China. Association for Computational Linguistics.

Ning, X., Desrosiers, C., and Karypis, G. (2015). *A Comprehensive Survey of Neighborhood-Based Recommendation Methods*, pages 37–76. Springer US, Boston, MA.

Nunes, I. and Jannach, D. (2017). A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, 27(3–5):393–444.

O'Donovan, J., Smyth, B., Gretarsson, B., Bostandjiev, S., and Höllerer, T. (2008). Peerchooser: visual interactive recommendation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1085–1088.

Parra, D. and Brusilovsky, P. (2015). User-controllable personalization: a case study with SetFusion. *International Journal of Human-Computer Studies*, 78:43 – 67.

Parra, D., Brusilovsky, P., and Trattner, C. (2014). See what you want to see: visual user-driven approach for hybrid recommendation. In *Proceedings of the 19th international conference on Intelligent User Interfaces*, pages 235–240, Haifa Israel. ACM.

Pazzani, M. J. and Billsus, D. (2007). *Content-Based Recommendation Systems*, pages 325–341. Springer Berlin Heidelberg, Berlin, Heidelberg.

Perano, M., Casali, G. L., Liu, Y., and Abbate, T. (2021). Professional reviews as service: A mix method approach to assess the value of recommender systems in the entertainment industry. *Technological Forecasting and Social Change*, 169(C).

Pu, P. and Chen, L. (2007). Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems*, 20(6):542 – 556.

Pu, P., Chen, L., and Hu, R. (2011). A user-centric evaluation framework for recommender systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys '11, page 157–164, New York, NY, USA. Association for Computing Machinery.

Rana, A., D'Addio, R. M., Manzato, M. G., and Bridge, D. (2022). Extended recommendation-by-explanation. *User-Modeling and User-Adapted Interaction*, 32:91–131.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, Los Alamitos, CA, USA. IEEE Computer Society.

Ren, L., Qiu, H., Wang, P., and Lin, P. M. (2016). Exploring customer experience with budget hotels: Dimensionality and satisfaction. *International Journal of Hospitality Management*, 52:13 – 23.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "why should I trust you?": Explaining the predictions of any classifier. *CoRR*, abs/1602.04938.

Ricci, F., Rokach, L., and Shapira, B. (2022). *Recommender Systems: Techniques, Applications, and Challenges*, pages 1–35. Springer US, New York, NY.

Richardson, A. (2010). Using Customer Journey Maps to improve customer experience.

Saia, R., Boratto, L., and Carta, S. (2016). A class-based strategy to user behavior modeling in recommender systems. In Chen, L., Kapoor, S., and Bhatia, R., editors, *Emerging Trends and Advanced Technologies for Computational Intelligence*. Springer International Publishing, Cham, Switzerland.

Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of 10th Int. World Wide Web Conference (WWW'2001)*, Hong Kong.

Schwartz, B. (2004). *The paradox of choice: Why more is less*. Harper Perennial.

Shen, X. and Stamos, I. (2023). Unified Object Detector for different modalities based on vision transformers.

Sherchan, W., Loke, S. W., and Krishnaswamy, S. (2008). Explanation-aware service selection: rationale and reputation. *Service Oriented Computing and Applications*, 2(4):203–218.

Shneiderman, B. (1992). Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics*, 11(1):92–99.

Smith, B. and Linden, G. (2017). Two decades of recommender systems at amazon.com. *IEEE Internet Computing*, 21:12–18.

Song, S., Lichtenberg, S. P., and Xiao, J. (2015). SUN RGB-D: A RGB-D scene understanding benchmark suite. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

Stickdorn, M., Schneider, J., and Andrews, K. (2011). *This is service design thinking: Basics, tools, cases.* Wiley.

Terano, T., Suzuki, M., Onoda, T., Uenishi, K., and Matsuura, T. (1989). Cses: an approach to integrating graphic, music and voice information into a user-friendly interface. *In: International Workshop on Industrial Applications of Machine Intelligence and Vision*, pages 572–581.

Tintarev, N. and Masthoff, J. (2012). Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction*, 22(4-5):399–439.

Tintarev, N. and Masthoff, J. (2015). *Explaining recommendations: design and evaluation*, pages 353–382. Springer US, Boston, MA.

Tintarev, N. and Masthoff, J. (2022). *Beyond Explaining Single Item Recommendations*, pages 711–756. Springer US, New York, NY.

TripAdvisor (2017). Tripadvisor. `https://www.tripadvisor.it/`.

Tsai, C.-H. and Brusilovsky, P. (2019a). Exploring social recommendations with visual diversity-promoting interfaces. *ACM Transactions on Interactive Intelligent Systems*, 10(1):5:1–5:34.

Tsai, C.-H. and Brusilovsky, P. (2019b). Exploring social recommendations with visual diversity-promoting interfaces. *ACM Trans. Interact. Intell. Syst.*, 10(1).

Tsai, C.-H. and Brusilovsky, P. (2021). The effects of controllability and explainability in a social recommender system. *User Modeling and User-Adapted Interaction*, 31(3):591–627.

Ullman, J. B. and Bentler, P. M. (2012). Structural equation modeling. *Handbook of Psychology, Second Edition*, 2.

Verbert, K., Parra, D., and Brusilovsky, P. (2016). Agents vs. users: visual recommendation of research talks with multiple dimension of relevance. *ACM Transactions on Interactive Intelligent Systems*, 6(2).

Wang, F., Zheng, Z., Zhang, Y., Li, Y., Yang, K., and Zhu, C. (2023). To see further: Knowledge graph-aware deep graph convolutional network for recommender systems. *Information Sciences*, 647:119465.

Wang, H., Zhang, F., Wang, J., Zhao, M., Li, W., Xie, X., and Guo, M. (2018). Ripplenet: propagating user preferences on the knowledge graph for recommender systems. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, page 417–426, New York, NY, USA. Association for Computing Machinery.

Wyner, A., Olson, M., Bleich, J., and Mease, D. (2015). Explaining the success of adaboost and random forests as interpolating classifiers. *Journal of Machine Learning Research*, 18.

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., and Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3485–3492.

Yi, J., Yuan, G., and Yoo, C. (2020). The effect of the perceived risk on the adoption of the sharing economy in the tourism industry: The case of Airbnb. *Information Processing & Management*, 57(1):102108.

Zhang, Y., Wang, J., and Luo, J. (2020). Knowledge graph embedding based collaborative filtering. *IEEE Access*, 8:134553–134562.

Zheng, L., Noroozi, V., and Yu, P. S. (2017). Joint deep modeling of users and items using reviews for recommendation. *CoRR*, abs/1701.04783.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., and Torralba, A. (2017). Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhou, X. (2023). Mmrec: Simplifying multimodal recommendation. In *Proceedings of the 5th ACM International Conference on Multimedia in Asia Workshops*, MMAsia '23 Workshops, New York, NY, USA. Association for Computing Machinery.