ViSCitR: Visual Summarization and Comparison of Hotel Reviews

Franziska Huth*
University of Stuttgart

Fabian Beck†
University of Bamberg

Johannes Knittel
Harvard University

Shahid Latif
University of Bamberg

Steffen Koch
University of Stuttgart

Thomas Ertl
University of Stuttgart

Abstract

Despite the availability of hotel booking platforms with customer reviews, selecting the best hotel or accommodation can still be difficult. This is partly because it is hard to directly contrast multiple shortlisted hotels. We propose a new visual interface, called ViSCitR, that goes beyond existing solutions in visually comparing ratings and personalizing the evaluation based on individual priorities. Targeted at a broad audience, it integrates several dimensions of comparison into a highly interactive document. The document contrasts the hotels visually and textually from different perspectives, including (i) a geographic perspective with points of interest, (ii) a comparison of ratings across different categories, (iii) rating changes over time, and (iv) summaries of the most noted positive and negative points from the review texts. The interface is accompanied by a sidebar for personalization, which supports selecting relevant points of interest and setting priorities. Changing these adapts the visual comparison across all interface sections to match the user's preferences. A qualitative evaluation with potential hotel customers showed that the personalized visual comparison is valuable for all users, while the advanced features of the interface are leveraged to different extents.

1 Introduction

Probably everybody who has booked a business travel, organized a family vacation, or planned a city trip knows the issue: Selecting the most suitable hotel or accommodation is difficult. This is despite hotel booking platforms making it easier to find available rooms within a certain geographic region for a selected price range. Moreover, such platforms also provide numerous customer reviews and ratings for each place, providing rich word-of-mouth evaluation that was hard to get only two decades ago. However, despite this tool support, one can easily spend hours comparing different available options. A reason for this is often that differences between a shortlist of accommodation options are subtle and cannot be sufficiently explained through numeric ratings alone. Comparing options in multiple browser windows side by side does not scale well beyond comparing two hotels. Moreover, one would have to go down to individual reviews and figure out qualitative differences that are relevant to oneself personally. Personalized rankings are important as individual priorities might differ quite significantly. A family has partly opposite needs than a business traveler.

In this paper, we specifically study the second phase of the hotel booking process, namely, contrasting a shortlist of comparable hotel options (from two up to ten in our prototype). The first phase of finding accommodations that satisfy search criteria—such as matching location, availability of the desired room type, and acceptable price range—is already well covered by popular booking platforms, such as *booking.com* [6], *Tripadvisor* [34], or hotel search on *Google Maps* [12]. In contrast, the support for analyzing and comparing reviews considering personal preferences is usually limited on such platforms. Although these platforms recently integrated better qualitative summaries of reviews (e.g., adding a summary phrase like *good breakfast*), we still clearly lack two aspects: (i) the comparative evaluation of reviews, explicitly pointing out differences between selected accommodations, and (ii) an option to personalize the comparative evaluation based on individual priorities.

In our work, we address these gaps by providing a new visual interface to compare hotel reviews, called *ViSCitR* (pronounced like *visitor*). As implied by the name, it targets anybody who books a hotel. Our approach, as showcased in Figure 1, is designed as an interactive, web-based document that structures the comparison into different sections. An always visible panel for setting personal preferences is shown on the left (I)—adapting the listed priorities immediately updates the comparison in the main part across all sections. At the top of the linear document, below the title image, a map not only provides personalized geographic context for the hotel; it also allows refining the hotel selection (II). The main comparison (III) then covers aspects such as contrasting numeric ratings in six different categories, rating trends over time, and characteristics mentioned in the reviews. To provide a fully self-explanatory interface, our solution mixes information that is visually encoded in simple diagrams with summarized content in short text statements. While the sequence of presented information guides the users through the comparison, they can also choose to explore the data on more detailed levels or to jump between sections. Typically, users would first select the main points of interest and prioritize the rating categories before selecting the best matching hotels on the map on that basis. Then, the selected hotel can be compared in detail across different comparison dimensions, exploring the data interactively and checking relevant details. Occasionally, users might adapt their personalization in the side panel or refine the hotel selection in the map, before finally being able to conclude their comparison.

With this application study and the *ViSCitR* interface, we hence contribute a novel visual comparison solution for textual and numeric review data. The focus on visual comparison and personalization in an interface for a broader audience is unique among such solutions (see Sect. 2). Specifically, we consider as our main contribution creating personalized and comparative visual summaries of hotel ratings and reviews, embedded into an easy-to-use interactive document-like interface. We worked in iterations with internal feedback to refine the interface, and finally evaluated it in a qualitative user study. For this study, we recruited six potential users with diverse levels of experience regarding hotel booking and using advanced visual interfaces. The results, on the one hand, confirm that all participants understood the interface and were able to make use of the personalization and visual comparison features. On the other hand, they showed relatively clear differences in usage profiles. Only some people intensely used more advanced features, like studying clusters of statements in the context of the original reviews.

The supplemental material [26] contains the source code and code for our data preprocessing pipeline, as well as a video demonstrating our interface prototype, evaluation videos, and results.

*e-mail: franziska.huth@vis.uni-stuttgart.de
†e-mail: fabian.beck@uni-bamberg.de

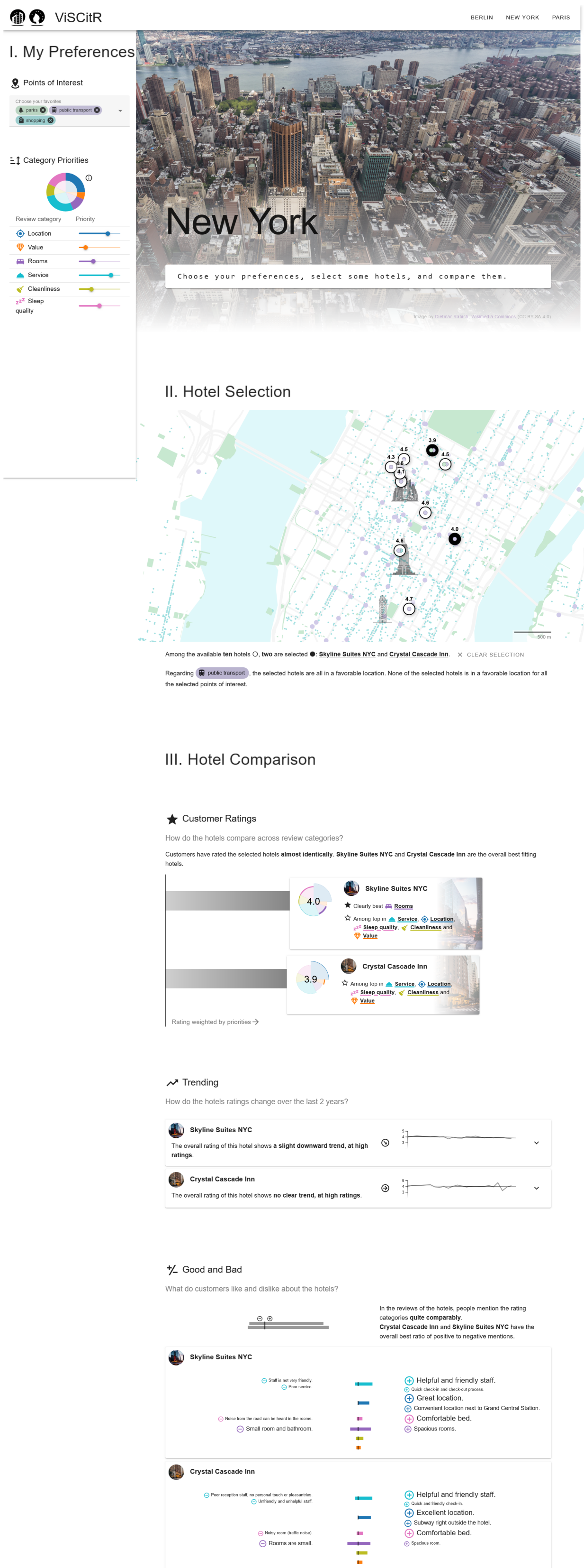


Figure 1: *ViSCitR* (full interface) contrasting two selected hotels in the city of New York under certain personalization preferences.

2 Related Work

Tourism research has studied hotel review data as an important factor that influences travelers' decisions and as a source that reflects customer opinions. Complementing this, visualization research has addressed the problem of visually helping to understand large text corpora like (hotel) review datasets. Moreover, personalization, comparison, and reporting are relevant concepts in our work.

Hotel Review Studies Sometimes referred to as *electronic word-of-mouth (eWOM)* [14], online reviews extend the set of opinions consumers can easily access when making product decisions. For hotel reviews, Vermeulen and Seegers [35] demonstrate the effect positive reviews have on booking decisions, especially for hotels (of brands) unfamiliar to the customer. Sparks and Browning [33] found that negative information in hotel reviews tends to be more salient than positive information. However, this effect can be mitigated if recent reviews (shown first) are positively framing the other reviews and if numerical ratings are good, which is why our approach shows textual, numerical, and visual summaries as well as comparisons. Easy access to information seems to play a major role, inasmuch as using categories or numerical ratings can aid in the efficient processing of information and potential decision-making. The perceived credibility of reviews might be higher for negative reviews, reviews written by experts, and if a review is consistent with others [24, 39]. Another factor for the high credibility of reviews seems to be that they show both positive and negative aspects of a hotel [7], and this credibility is important for people's booking decisions. On the other hand, reading conflicting opinions in reviews can lead to discomfort, and people defer the booking choice [1].

Review Visualization There are many visualization approaches on text and sentiment visualization, as described in respective literature surveys [17, 18]. We are most interested in techniques for visualizing product review data. For instance, Chen et al. [9] extract the sentiment of hotel reviews, which they then process into a basic color-coded visualization by topic. *OpinionSeer* [36] shows a projection of opinions from hotel reviews in the center, circle rings to encode numerical information, and on-demand tag clouds to encode term frequencies. Both approaches are designed as single interactive visualizations for sentiment or opinion exploring, in contrast to our interface, which is a combination of text and visualizations, targeted at a broad audience. Bjørkelund et al. [5] discuss how sentiment analysis can be used to analyze travel review sites to help people find hotels and areas to stay in. They color encode the areas on a map, which, however, gives people only an impression of the quality of the geographical area, but no information on average hotel ratings or sentiment. Other approaches [8, 15, 25, 37] focus on the comparison of sentiment or opinions mined from (product) reviews. In contrast, our approach has a more holistic focus on the reviews and additional contextual information, and tries to adapt to varying needs of level of detail. Another visual analytics approach [40] aims to reduce customers' self-selection bias when making decisions based on hotel reviews. It shows reviewers' experience, the degree of emotion, and emotional aspects. Their study found that the proposed design significantly increases people's satisfaction with decision-making. Similar to our approach, one of the proposed views depicts the distribution of reviews by category (e.g., food), broken down by rating. This correlates with our identification of discussed categories in hotel reviews (e.g., cleanliness).

Visual recommender systems support information retrieval tasks by proposing results, sometimes personalized, based on online reviews [23, 38]. The approach described in this work complements these, as it is applied after an initial retrieval of suitable hotels.

Personalized and Comparative Visual Reporting Whereas most visualization interfaces are adaptable through selections, filters, and settings, these interactions reflect personal interest and background only indirectly. In contrast, we directly ask people to express their interests through the interface, which can be considered as an aspect of *personalization*. In such a way, we make the personalization explicit and do not implicitly try to detect personal preferences from usage data (e.g., like a search engine personalizes search results based on previous searches and user location). This is linked to *personal visualization* [16], especially in the case *personal context* is considered for the visualization; it should not be confused with visualizing *personal data* (e.g., personal activities [4]), which is also a common form of personal visualization. We also do not go as far as *personalized interfaces* (e.g., [10]), where the input elements and layout can be individualized, but restrict the effects of personalization on the selection and weighting of the visualized data. Another concept we apply is *visual comparison*, where typical comparison patterns can be classified into *juxtaposition*, *superposition*, and *explicit encoding* [11]. We mostly leverage juxtaposition (i.e., placing review visualizations of the different hotels next to each other), but occasionally also use explicit encoding (e.g., computing and visualizing minimum ratings or textually summarizing differences between hotels). The selection of what is being compared can be regarded as another element of personalizing.

We present an interactive document, which is a *magazine style* type of *narrative visualization* [30] and *data-driven storytelling* [29]. Our case can be classified as a geographic story [21]. However, since typical embedding texts that connect the data-driven texts and visualizations to a full-fledged story are lacking, we rather consider it a reporting solution. Whereas most comparable reporting and storytelling approaches are rather static regarding different user input and data, some approaches focus on adapting the reporting [20, 31]. For instance, Calliope [31] very flexibly generates data stories from tabular data of various sorts. VIS Author Profiles [20], in contrast, is more confined to generating reports on specific data (i.e., bibliographic information); its linear document structure with an interactive integration of textual and visual elements inspired our approach. Interactive Map Reports [19] also have an element of comparative reporting included, where a comparative data-driven textual description of two selected regions is generated. However, overall, we are not aware of a visual reporting solution that, as much as ours, adapts to personalization and comparison.

3 Data Preparation

To compare hotels, we needed customer reviews that contain both numeric ratings and textual content. Moreover, certain interpretation steps of the review texts were necessary to abstract the content into comparable statements. Finally, the data was contextualized with geographic information to personalize the hotel selection better.

Dataset We used a subset of HotelRec [2], a dataset with reviews from the popular booking platform *TripAdvisor* [34]. As sample cities with many available reviews, we selected Berlin, New York, and Paris. For diversity, we selected cities from different countries; for familiarity, we selected popular tourist destinations; otherwise, the selection was arbitrary and our approach should work for any city with hotels having a sufficient number of reviews. We focused on the ten most reviewed hotels in each city to replace a preceding retrieval step. This makes sure a more than sufficient number of reviews is available. Each review contains the hotel name, a timestamp, a headline, the review text, a numerical overall rating, and numerical ratings for the pre-defined categories *location*, *value*, *rooms*, *service*, *cleanliness*, *sleep quality*, *business service*, and *check in/front desk*. However, we omitted the categories *business service* and *check in/front desk* because they were only present in very few reviews.

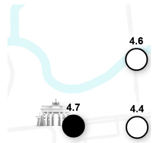
Review Text Summarization To summarize positive and negative review statements, we used the *GPT 3.5 turbo* [13] model through OpenAI's API to extract them from a random sample of 1 000 reviews per hotel. We evaluated the output of the summarization with 100 random samples manually, to test if (i) the positive and negative items are classified correctly, and (ii) the item texts accurately reflect the review text (no hallucinations, nothing important omitted). The overall quality of the summary points was above expectations. A total of 709 positive and negative sentiment items were extracted. Sentiment was assigned incorrectly in only four cases. Even with awkwardly formulated or grammatically incorrect reviews, the summary worked surprisingly well. In the case of very long reviews and enumerations, individual points were sometimes not explicitly included in the summary. Please refer to the supplemental material [26] for more details. Some reviews digress very much, or report on personal events, which have nothing to do with the hotel itself. Such points also end up in the summaries from time to time. However, the subsequently described category assignment step likely filters out the latter.

Category Matching and Clustering To prepare for comparability and personalization, we assigned each of the extracted items to the rating categories of the HotelRec [2] dataset (e.g., *'central location'* → *location*, *'unfriendly staff'* → *service*). We assign an item to a category if its BERT-based embedding [28] is similar to the embedding of category keywords we curated based on the terms in the reviews. To generate a summary of the sentiment items, we consider the most frequently recurring items. We clustered the items within each category and hotel, based on their sentence embeddings, with spherical *k-means*, and determined the best *k* with the elbow method on the silhouette score. As a representative of each cluster to be displayed in the interface, we selected the item with the highest centrality score. We also want to distinguish the relevance of each item cluster for the sentiment summary, and thus use the size of the cluster to determine how prevalent the sentiment item cluster is in the reviews. For the interface, we then sort the clusters and, for example, display them in decreasing order of relevance. With this categorization, we count positive and negative mentions of categories, and augment the numerical ratings. The goal was not to keep every data point, but to extract re-occurring patterns. Of the 208 142 sentiment items extracted from the reviews for our selected hotels, 66 253 remained and were grouped into 1 212 clusters with representative statements; both can be explored in the sentiment section of the interface.

Geographic Context and Metadata We enrich the data with the geographic location of the hotels, their geographic context, and points of interest (POIs) within the respective areas of the cities. As listed in Table 1, this includes general geographic features like water bodies, main roads, and landmarks, as well as POIs such as parks, public transport, restaurants, sightseeing, and shopping. All geographic data was extracted from *Open Street Map* (OSM) through *Overpass Turbo*, partly simplified with *Mapshaper*; landmark icons were AI-generated through *Midjourney*.

To allow hotel comparison without biases of familiarity and previous experience, we replaced the original names of the hotels with fictional ones. We used AI-generated images (through Midjourney) of hotels roughly resembling the style of the hotel. Through this obfuscation, we also wanted to rule out any intervention with existing businesses. As providing a detailed visual impression of each hotel is out of the scope of our approach, only one image per hotel was sufficient, which could act as a visual identifier for the hotel in the interface. To rate each hotel regarding each POI type, we inspected the extracted location data in a radius of 500–1 000 m, in addition to occasionally using *Google Maps*, and formulated a statement. The statements are manually written for each hotel and type of points of interest and refer to information that can be judged relatively objectively, such as the restaurant density in the neighborhood or the access to main public transport lines. Each statement is classified as either *positive* (+), neutral (±), or negative (−) (for examples, see Fig. 3, cutout).

Table 1: Different layers of the map shown by default and on demand.

Type	*Map Encoding*	*Example*
General features (by default)		
Water bodies	Thick light blue lines and filled polygons	
Main roads	Light gray lines	
Landmarks	Icon images (semi-transparent)	
Hotels	Large black/white (selected/non-selected) circles with rating scores	
Points of interest (on demand)		
parks	Filled polygons	
public transport	Medium-sized circles	
restaurants	Dots	
sightseeing	Small circles and filled polygons	
shopping	Dots	

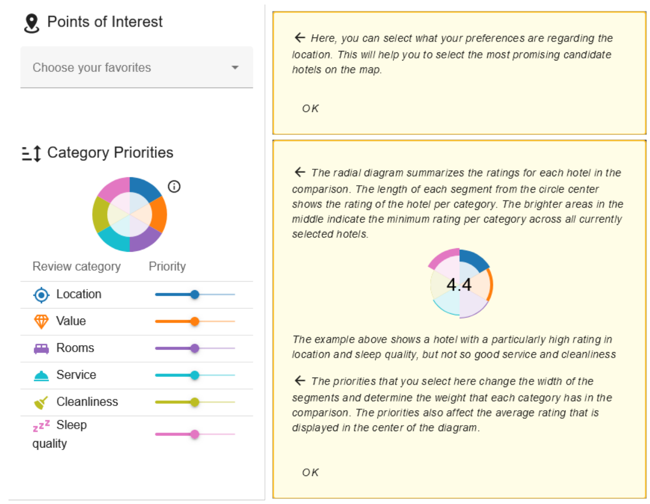


Figure 2: The personalization panel in its default configuration, with tutorial boxes introducing the features to the user.

4 Visual Interface

In the design process of *ViSCitR*, we followed clear design concepts (DCs), which stayed consistent throughout the development process:

- **DC1 – Divide and Conquer:** Break down the comparison into aspects that are simple to analyze.
- **DC2 – Guidance and Exploration:** Guide the user through the interface, but always offer options for data exploration.
- **DC3 – Visualization and Text:** Present data visually, but complemented with textual data summaries.
- **DC4 – Comparison and Personalization:** Support comparison and react to the personalization in each data representation.

With **DC1**, we address the complexity of review data. Targeted users of our system are laypeople, which motivates **DC2**, to provide sufficient high-level information and interaction prompts, and enable deeper discovery if desired. With **DC3** especially, the interface supports users with low visualization literacy. One of the main goals of the interface is to highlight differences between hotels, with respect to personal preferences, which we aim at with **DC4**. Aside from the DCs, of course, we followed established guidelines for user interface and visualization design, such as, using colors and other visual encodings consistently, providing details on demand, applying animation for smooth transitions, etc. Furthermore, we have implemented different levels of importance that can be discovered step by step (referenced as ① – *highly important*, ② – *moderately important*, and ③ – *detail*), which we matched with the visual saliency and discoverability of the respective information. While some people might be satisfied by viewing only the most important information ①, others want to go deeper regarding certain aspects of comparison (levels ② and ③). Hence, we do not expect each user to discover and use all features, but anticipate seeing different usage profiles.

The resulting visual interface is an interactive, linear document (Fig. 1). We chose this linear structure because it allows breaking down the analysis into sections (**DC1**) and provides the intended guidance (**DC2**) to the targeted, broad audience. To first draw the visitors' attention and introduce the purpose of the interface, the document only starts with the name and a representative image of the currently selected city and the introductory text: *'Choose your preferences, select some hotels, and compare them.'* To provide context on the research project, the system redirects to a landing page when first visiting the interface. The landing page is an overlay that offers a tutorial for the interface features that guides through the hotel selection and personalization with overlay messages that appear next to the respective features (partly illustrated in Fig. 2). Although we designed the basic functionality to be intuitive and visually self-explanatory, we learned through initial user testing that minimal extra explanations are still considered helpful.

We implemented the interface in JavaScript and vue.js, using D3.js for visualization. All data handling described in Sect. 3 is performed as preprocessing steps. The interface was developed for screen resolutions of 1280×1024 and above.

4.1 Personalization

On the left-hand side of the interface, we present an always-visible personalization panel, as shown in Fig. 1 in the context of the whole interface and enlarged with tutorial explanations in Fig. 2. To keep the usage effort of this primary feature of the interface ① limited, we restrict the personalization to two aspects that we deem highly important for hotel booking according to personal preferences. (I) People can select geographical **points of interest** (POIs), like parks or public transport, which the hotel should be close to (a full list of available POIs is included in Table 1). These POIs are then displayed on the map, and the hotel's proximity to the POIs is evaluated and summarized. (II) People can set **category priorities** for each of the rating categories *location*, *value*, *rooms*, *service*, *cleanliness*, and *sleep quality*, which influences how the ratings in the interface are displayed and weighted in comparison. When people hover the mouse on a category, a tooltip explains the category, and across the whole interface, references to the category are highlighted.

As a secondary purpose ②, the panel also introduces the two main color schemes of the interface. For POIs, we have selected non-saturated, light background colors that suit a map-based encoding (see Table 1). In contrast, for the rating categories, we use saturated, intense colors that serve as markers throughout the interface and should hardly be confused with the POI colors (see Fig. 2). While the color-coding can support understanding geographic context and link categories for hotel comparison quickly, we made sure not to rely merely on color for the correct interpretation of a data encoding; where necessary, we redundantly encoded the information textually, e.g., the name of the category (**DC3**).

Moreover, the category priorities section explains the glyph-based visualization of the weighted rating used for quantitatively comparing the hotels in the main document. We have decided to use a circular metaphor to represent that the encoded weights define fractions

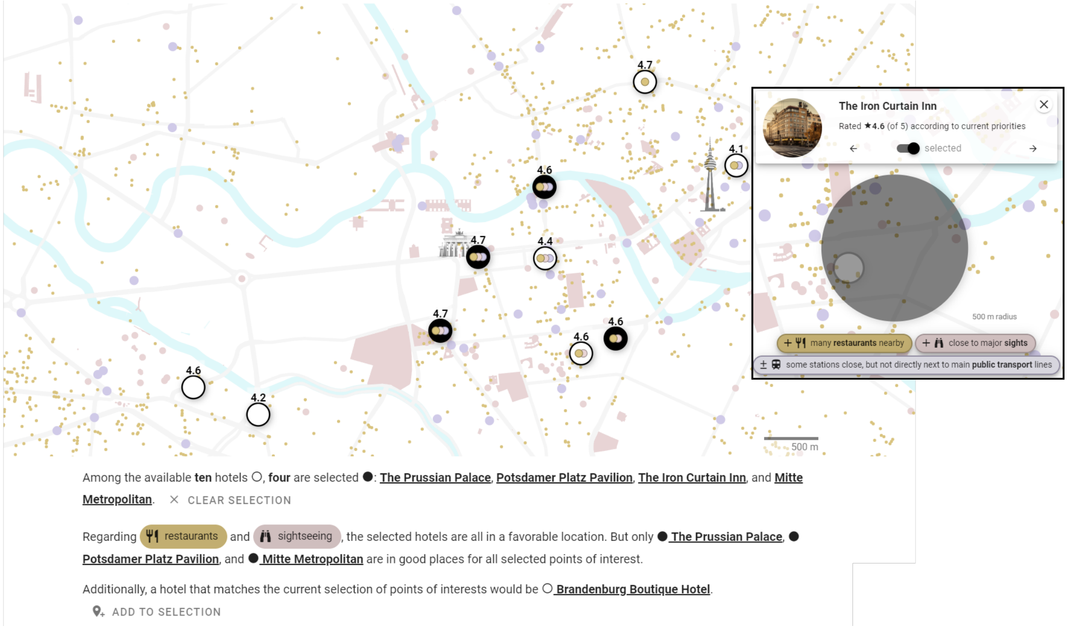


Figure 3: Map view showing all available Berlin hotels (selected: filled circles; unselected: unfilled circles) in their geographic context with textual descriptions below; the shown state corresponds to three types of points of interest are activated: *restaurants*, *sightseeing*, and *public transport*. Overlayed on the right is a zoomed-in view of a focused hotel.

(sectors) of a whole (circle). When changing the priority weights, the respective circle sectors change in size (i.e., the angle it covers, see Fig. 1); higher priority categories have a larger angle. This is to represent the importance that people assigned to the categories visually—more important categories take up more space in the glyph. The aim of the glyph is to act as a compact and reusable interface element that helps users grasp the numerical ratings. We allow for a better comparison of the category ratings between hotels by introducing a baseline, which is the minimum category rating across all selected hotels ③. In the glyph, the part of the slice below the minimum rating is displayed in a pale shade. The emphasis is on ratings above the minimum, which are displayed in full color. The tutorial text (see Fig. 2), similar to the tooltip text that appears when hovering over the attached info icon, explains the encoding further. While we tried to choose an intuitive encoding here, the data and encoding might still be too complex for some people with lower visualization literacy. Hence, the middle of the glyph depicts the weighted (personalized) rating average as a simple number for a very broad overview ①. We consider understanding how the weighted average is computed as a secondary goal ②, while contrasting each category's value to the respective minimum value is a tertiary option ③.

4.2 Map-based Hotel Selection

As the first section of the document, a map provides an overview ① of available hotels (Fig. 3). We have decided to take this geographic perspective as an entry point, as the specific locations might be a key criterion for choosing a hotel and can act as the first filter. At the same time, geography provides an important context for interpreting the ratings and reviews that are presented in the subsequent sections. For instance, reviews of hotels in the historic downtown area might need to be read differently than those of a hotel next to the airport, as expectations of customers would be different.

The background of the map shows important city features such as main roads, water bodies, and some landmarks to provide sufficient geographic context. Additionally, we blend in favored points of interest (**DC4**, as selected in the personalization panel), for instance, marking restaurants as small beige circles or sights as rose polygons and circles ②. Positions of all available hotels are highlighted with contrast-rich circular markers. Clicking one of the markers selects the hotel for comparison, but users can also zoom in on the hotel ③, which enlarges the marker (Fig. 3, right), thus offering an entry point for more detailed exploration (**DC2**). Overlays appear that provide additional information: at the top, hotel picture, name, and current rating according to selected priorities, as well as selection and navigation controls. Additionally, statements are shown regarding how well the location matches the selected points of interest.

Below the map, in accordance with **DC3**, a short text summarizes ① the current hotel selection in three paragraphs (Fig. 3). The first paragraph lists the selected hotels, with the option to clear the selection. The second one provides a concise textual rating of the selected hotels regarding their proximity to active points of interest. The last paragraph lists hotels that would also fit the active points of interest but have not been selected yet. Initially, none of the hotels are selected. To guide the user (**DC2**), adaptive instructions hint at how to select a hotel and, if not specified already, prompt people to set their points of interest.

4.3 Comparison of Numeric Ratings

The next section in our interactive document provides a quantitative comparison of the selected hotels condensed from the individual reviews (Fig. 4). Here, the prioritization of different rating categories as shown in the personalization panel plays a key role (**DC4**). The section starts with a textual summary of the numerical ratings of the selected hotels ①. It states whether *customers have rated the selected hotels* **almost identically**, **quite comparably**, **somewhat differently**, **quite variably**, or **with clear differences**, depending on the variety of minimum and maximum category ratings. A list of the top-rated hotels follows, cut off when the difference to the highest-rated hotel is larger than .2.

The list of hotels is sorted by their average rating, beginning with the highest rated hotel. The main visual representation of the ratings is a horizontal bar chart and the glyph. The bar visually supports the average rating of each hotel, which can also be found

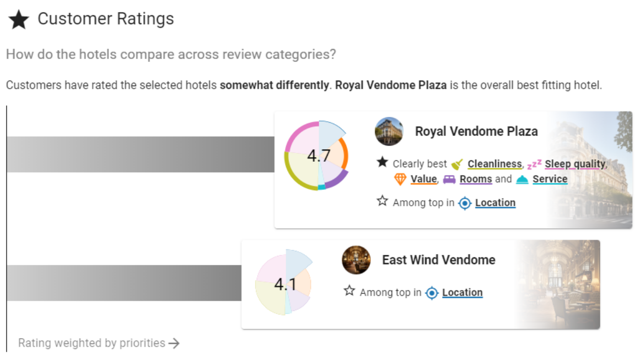


Figure 4: Overview of the numeric ratings of the currently selected hotels. A textual summary helps to compare the hotels. The glyph represents the ratings in each category, with the average rating in the center. The angles of glyph slices are unequal because of category priorities; for example, the *service* category was given a low priority by that user, whereas *cleanliness* and *sleep quality* have a high priority.

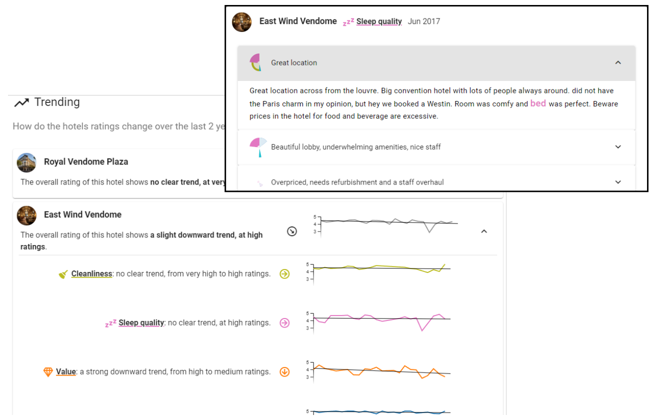


Figure 5: Ratings trend over time, with trend per category and descriptive text. On the top right, a details-on-demand dialog shows the reviews that make up an average monthly rating, with a ratings glyph and highlighted category text.

written out as a number in the center of the glyph ①. As already introduced in the personalization panel, the glyph contains further information, like illustrating the weighted average rating ② and individual ratings per category compared to the minimum among all selected hotels ③ (**DC4**).

Again, we follow **DC3** and show the visualizations accompanied by text. For each category, the best-performing hotels are highlighted with a short textual description ② next to the respective glyph. For each of the categories, it is computed from the average rating scores whether there is a single hotel that is *clearly best* (at least .3 better rating than the other selected hotels) or whether there is at least a set of hotels that can be considered *among top* (at least .3 better than the worst rating of the other selected hotels, but at most .3 worse rating than the best hotel). These computed sets are then used to mark the respective hotels (category name below the hotel name, grouped by ★ *clearly best* and ☆ *among top* as shown in Fig. 4). If all ratings are similar across hotels, none of the hotels is marked for the respective category. If multiple categories are listed, they are in order of their user-assigned priority. Furthermore, categories with a priority value below 10% are not considered.

4.4 Temporal Trend Analysis

Hotel ratings can change when, for example, renovations led to upgrades in room quality, the service personnel received better training, or customer expectations changed. The temporal development of hotel ratings provides a clearer picture of how an overall rating relates to recent developments. Hospitality research [33] shows that customers value insights on the temporal development of ratings, and more recent ratings are regarded as more important. To reflect this, our interface shows a line chart of the ratings for each hotel over time ① (Fig. 5). Even though the dataset contains ratings dated as far back as seven years, we decided only to show ratings from the last two years because these are the most relevant. The ratings are binned by month and calculated as a weighted average of the category ratings, to reflect the personalization (**DC4**). If there are less than five ratings in a bin, we omit the average of that bin as not being representative enough. For better comparability between hotels and their differences, the y-axis starts not at zero, but is faded out at a baseline of 3 (we did not observe averaged hotel ratings below this value). We used linear regression to compute a trend line, generated a descriptive text with the trend and value (**DC3**), and depicted an arrow icon to allow for quickly grasping the trend.

With a click on the trend panel for a hotel, it expands to show the rating trends in each category ②. The structure mirrors that of the overall trend: For each category, there is a line chart with the average ratings per month, a trend line, and a description with a trend arrow. Hovering over the line charts reveals the date and value of the data points ②. To explore more (**DC2**), users can click on a data point to retrieve relevant reviews within that month, enriched with glyphs that visualize the personalized category ratings ③ (Fig. 5, top right).

4.5 Summary of Review Sentiments

Finally, the interface shows a comparison of review sentiments (Fig. 6, top). Reading both positive and negative reviews increases their credibility [7] and allows for a deeper understanding of the benefits and drawbacks of a hotel. We show positive and negative aspects summarized from the reviews as sentiment items ①. The items we extracted from each hotel's review texts are categorized and clustered within a category (described in Sect. 3), and shown as bullet points. To guide people's attention, we vary the font size of the items ② depending on how salient the aspect is in the reviews (cluster size). A tooltip on mouse-over ② shows the percentages of the positive and negative mentions, in relation to the total number of reviews. The number of visible sentiment items depends on the category priorities set by the user. The higher the priority, the more items are shown for that category ③. Sentiment items in each category rating are juxtaposed: negative sentiment on the left and positive sentiment on the right. In line with **DC4**, this allows people to compare the hotel features directly, and if there are contradicting sentiment items, evaluate the different saliency of each of them.

Between the positive and negative sentiment items, we show a bar chart of the percentage of reviews with positive and negative mentions of each category. At the top, there is a bar chart with the overall number of positive and negative mentions in the review texts. On mouse-over of a category, the percentage of the rating of that category is shown ② (Fig. 6, top left). Next to the overall sentiment chart, combining text with visualization (**DC3**), we show a generated textual summary ①, similar to that in the numeric ratings (Sect. 4.3). Based on the computed variance of the positive and negative mentions, it states whether the rating categories have been mentioned (very) similarly or differently in the reviews. It also lists the hotels with the best positive-to-negative sentiment ratio.

People can retrieve the detailed reviews of specific items in the summary (**DC2**). When clicking on a sentiment item, an overlay dialog ③ opens (Fig. 6, bottom). It shows the reviews that contributed to the cluster, with the sentiment from each review as the title and the review text as an expandable panel. In the text, keywords from the respective category are highlighted in category colors.

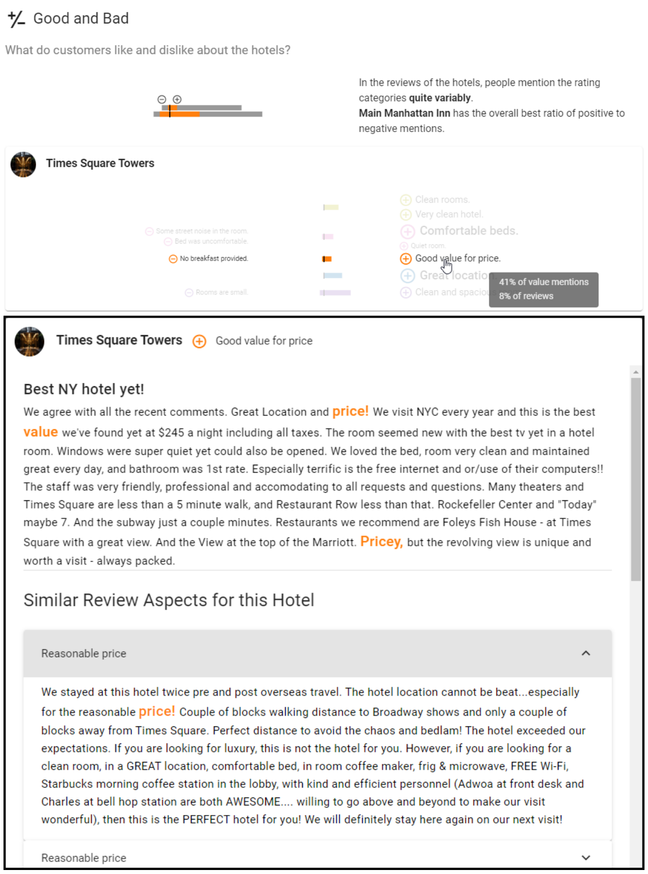


Figure 6: Positive and negative sentiment from the review texts per category with mouse-over highlight of a category (top), and an overlay dialog with reviews of the selected category and sentiment (bottom).

5 Evaluation

In a formative, informal evaluation process, we already tested the interface during its design with different experts for visualization and web development (mostly members of the participating research groups). This process started when the first interface components were implemented. We demonstrated the interface and generally asked for any feedback. Many usability issues could be resolved quickly through this, but the process also led to higher-level insights and modifications. For instance, it became clear that—despite the efforts to make all interface parts self-explanatory—a short tutorial would be beneficial at the beginning. Moreover, we learned that not all features and representations might have the same relevance, which led to the introduction of explicitly designed levels of importance.

As a formal evaluation, we performed a qualitative user study to evaluate our interface as presented in its final version above. The goal of the evaluation was to analyze how well the interface supports a diverse set of users in summarizing and comparing reviews, and how that might help with the decision-making process of booking a hotel. We wanted to find which of the features were valuable and could be recommended for designing similar interfaces. Moreover, the study should show whether the features are self-explanatory and easily discoverable.

We conducted interviews with six participants with diverse backgrounds and traveling profiles—including solo travelers, business travelers, and people who travel with their families. We used *Prolific* [27] to recruit four participants and paid them 12£ for their participation, which was up to an hour long and held over a video conferencing software. Because *Prolific* participants were rather inexperienced travelers or visualization users, we additionally recruited two participants with more traveling experience from our lab. They were given a non-monetary compensation for their participation.

Table 2: Participants' previous experience with booking hotels and using visualization.

	Low	*Medium*	*High*
Travel experience	P_1 P_2 P_3 P_4	L_2	L_1
Visualization experience	L_2	P_1 P_2 P_3 P_4	L_1

5.1 Study Design

After presenting the study description and consent form to the participants, we asked demographic questions. We inquired about English language skills and visualization expertise, whether they work or have worked in the hospitality industry, how often they have booked a hotel in the past twelve months, and their main reason for booking a hotel (solo vacation, family vacation, or business trip). Further, we interviewed participants on how they usually research hotels.

Next, participants opened *ViSCitR*, read the introduction, and selected one of the three available cities, which was randomly assigned. With the data for that city loaded, participants completed the tutorial and spent five minutes exploring the interface. We instructed participants to comment on their actions, motivations, and insights when using the interface (think-aloud method [22]). To test how self-explanatory the interface is, we were careful not to give hints, unless participants requested clarification (which rarely happened).

Afterward, the participants switched to a second randomly selected city. We asked them to put themselves in the position of someone who wants to book a hotel and has already selected a date and price range so that they now see ten available hotels for their selection. We gave the participants a travel reason matching their traveler profile (i.e., solo or family vacation or a work trip) and told them to evaluate the available hotels based on that motivation. They were instructed to give a short action plan of how they wanted to proceed with evaluating the hotels. Then, we gave participants the following small tasks to facilitate further exploration: We asked participants first to compare two, then five hotels, to re-consider their personalization settings, and to reflect how setting their preferences affected their opinion of the hotels. We also requested summarizing one of the hotels based on the trend and sentiment sections, respectively. After these tasks, we showed interface features they might have missed (if any), and asked how they would have used those.

Ultimately, we interviewed participants about what features and visualizations they deemed important for gaining insights. Further, we asked whether there were any difficulties in understanding. For structured data, the participants rated the interface features, by selecting how much they agree or disagree with statements about intuitiveness and usefulness of the features on a 1 - 7 Likert Scale. We also inquired about the balance of visualization and text.

Screenshots of the interview questionnaire, the full list of tasks, complete documentation of the interview sessions, and participants' answers can be found in the OSF repository [26]. The ethics board of the first author's affiliation approved the described study.

5.2 Results

In the following summary of results, we denote (and color-code) lab participants as L_1 and L_2, as well as *Prolific* participants P_1, P_2, P_3, and P_4. L_1 and P_4 were male, L_2, P_1, P_2, and P_3 were female; L_1, P_1, P_3, and P_4 were aged between 20 and 30, and L_2 and P_2 between 40 and 50. P_1 and P_2 were located in South Africa, P_3 and P_4 in Eastern Europe, and the lab participants in Germany. P_1 and P_2's native language was English, the others stated a high English language proficiency, which matches our impression in the interviews. The experience with visualizations or statistical graphics covered a

Table 3: Usage and ratings of the interface features. For each color-coded participant, a filled circle ● indicates high usage, a half-filled circle ◒ medium usage, and an empty circle ○ encodes low usage of an interface feature and the respective level (for the personalization, there are no clear levels).

Feature	①	②	③	*Intuitive*	*Useful*
POI	●◒○◒●◒	-	-		
Cat. priorit.	●◒○◒◒◒	-	-		
Map	◒●○◒◒◒	●●○○○○	◒◒●●●		
Ratings	●◒◒○◒●	●●○○●●	●◒ ○		
Trend	◒◒◒◒◒●	◒◒◒○◒●	◒ ◒		
Sentiment	◒○◒●●◒	◒○●●◒◒	◒○○		

wide range, and participants' hotel booking experience spread from none to eleven hotels booked in the past twelve months (Table 2). **L_1**, **P_1**, and **P_2** stated that they regularly use booking platforms, such as *booking.com* or *TripAdvisor*, to inform themselves about hotels, the other three participants said they use general-purpose search engines, such as *Google*, or ask friends. **L_2** additionally said they prefer to book hotels that are part of a franchise they already know.

In Table 3, we show the usage intensity of the interface features, which we manually coded from the sessions. Using a feature multiple times or for several minutes is considered high usage; otherwise, briefer use is considered medium or low usage, respectively.

Personalization All participants except **P_1** set their preferred points of interest and category priorities after the tutorial at the beginning. When prompted in the respective task, all participants adapted their personalization, with **L_1** and **P_3** exploring a little longer. Experienced travelers and those who are used to booking websites appeared to be more decisive about their personalization settings.

Usage Profiles We observed that participants used the features of the interface to different extents. Table 3 shows two main profiles emerge. The first profile includes people mainly interested in numerical data (**L_1**, **L_2**, **P_4**), for example, in the ratings section of the interface (Fig. 4). The second profile matches people reading the more detailed textual data (**P_1**, **P_2**, **P_3**), for example, in the sentiment section (Fig. 6). For instance, **L_1** did not discover the review details overlay of the sentiment section (Fig. 6), but when shown at the end, they said they found it interesting, but would not have used it much. In contrast, **P_1** used the sentiment section relatively soon and often in the interview, and frequently opened the details overlay feature of both the trending and the sentiment section. The trending section was the least used by participants overall, but still rated as useful by most participants. Level ① and ② interface elements were discovered by all participants. In the map, five of six participants zoomed in to level ③ information, whereas the level ③ features of the ratings, trending, and sentiment sections were not discovered or used by about half of the participants (different subsets each, see Table 3). Even without using level ③ information, however, participants were able to solve the tasks well and without any major problems. The participants provided varying degrees of detail in their responses, but none of them had problems finding relevant information. The time to solve a task generally varied between one and six minutes, and the level of detail of the answers correlates with it. For the task of comparing hotels, only **P_4** used all the interface sections, although least frequently the sentiment section. The other participants focused on either the ratings section or the sentiment section for the hotel comparison tasks, and **L_2** also frequently employed the map. Participants who had not previously used the trending or sentiment sections gave satisfactory answers in the tasks that prompted their use, but the responses were less detailed, and the participants spent less time than those who had explored these sections previously.

Preferences and Overall Satisfaction Participants found all but one interface element easy to understand (Table 3, *Intuitive*). Only the category priorities were rated less intuitive, likely because participants also included the glyph in the rating. When asked about what participants thought were the most important features, three participants mentioned the sentiment section (**L_1**: "I think it's nice because it gives you a little bit more information about why exactly people rated it the way they did."), and the map was also mentioned as most important by three participants (**P_2**: "I definitely like the map because the location is something that's very important, and the location relative to my point of interest."). Two participants remarked that the points of interest were among the most important features to them. **P_3**, however, said the map and the points of interest were the least important. The trending section was considered least important by two participants (**L_1**: "I think the least useful for me is probably the trend, because I don't really care if the rating was bad, like two years ago."). **P_4** stated that the glyphs are of minor importance, especially if the hotel ratings (and thus the glyphs) are similar, and **P_2** found the ratings section least important. These mentions are also reflected in the usefulness ratings of the features (Table 3, *Useful*).

Participants' overall satisfaction with the interface was high, and they found the interface intuitive. Two participants stated that they found the amount of visualization too much. For instance, **P_1** stated "sentiment [section] too much [visualization]", and **P_3** said "In the map the circles get confusing and the pie chart at some appearances seemed overcomplicated" (they were irritated by partially overlapping hotel markers on the map). When asked to rate the amount of text, **P_3** stated "When hovering over words [there is too much text], at the hotel rating summary a bit too"; for the other participants, the amount of visualization and text was just right.

Additional Comments Participants appreciated the additional information in tooltips, for example, in the personalization panel about what the categories mean, or in the sentiment section about how many of the reviews the respective statement items were mentioned. Some participants suggested more tooltips, e.g., showing exact category ratings when hovering on a glyph slice. **P_3**, however, stated they were a bit overwhelmed and wished for the elements in the interface to be more static and less interactive on mouse-hover, but rather on mouse-click. Some participants had a small screen size and, thus, only saw the title image of the city in the interface panel initially and had to be told to scroll down to see the map and the other sections. **L_2** and **P_4** suggested that the tutorial should not just include the personalization, but also the hotel comparison sections of the interface, so that they are aware of level ② and ③ features earlier (**P_4**: "It would be nice too if there was [a more in-depth tutorial] like when you first visit the website").

5.3 Study Limitations

We did not directly compare our prototype to any other interface, mostly due to the different approach we used, different levels of detail in the available data, and the limited duration of the interview sessions. We targeted the scenario that people have already narrowed down their selection of hotels, which is not directly supported in a specific view in most booking platforms. It would also be challenging to avoid biases related to familiarity, comparing our interface to well-known ones. Furthermore, we focused on qualitative evaluation with a small number of participants, trying to explore a realistic setting in depth. Hence, we cannot yet quantify the usage performance of our interface and cannot contrast it to another tool. To get a rather comprehensive understanding of how the interface can be used, we tried to recruit a diverse sample of participants (two participants from our lab, four via *Prolific*). However, we were unable to fully diversify, for instance, cultural background or age, and might not have yet reached a complete description of different usage profiles and relevant cases. We had limited control over the environment of the *Prolific* participants, for instance, used screen size and resolution, but clarified in the interviews that the interface was displayed reasonably in their setting. Reflecting on motivational factors, *Prolific* participants are paid a set sum for participating in an online study, and thus have an incentive to complete the study quickly, which might result in less detailed exploration. However, possibly because of the format of an interview study, we found that the participants stayed engaged throughout the interview session. For lab participants, ending the study in a short time could also have been desirable; however, we rather assume that—being researchers themselves—they better understand the value of detailed findings and might have been more committed to exploring all features. Further, participants in our evaluation saw the interface for the first time; repeated visits were not covered in the study.

6 Discussion

Study Results Participants were able to use the interface without any major problems, despite having only a minimal introduction via the tutorial. Different interface sections catered to different user profiles; for example, some people focused more on the sentiment than on the ratings section. The trending section was the least used one but was rated only slightly less useful than the rest of the interface. From this result, and from tourism research findings [33], it still makes sense to keep it in the interface, but making it a less prominent level ② feature. The glyph appeared to be the least intuitive interface element. However, the most important information it encodes is redundant with textual statements of summaries and comparisons next to it, and a full understanding of the glyph was only necessary to retrieve details (e.g., more precise differences in ratings regarding certain categories). We presume that most people can learn how to read the glyph effectively with repeated use of the interface. Still, a larger user study is necessary to optimize the level of complexity of the visualizations, catering to a broad range of users, from first-time to frequent users. In conclusion, we recommend offering interface features that provide contextual information even if not used extensively, to cater to different user preferences. Some participants did not use some interface features because they did not discover them. Participants solved the tasks well regardless, potentially only with limited depth. We went beyond what people are used to from typical hotel booking websites, but still, the study participants understood and appreciated it. As a similar unfamiliar element for most users, we successfully applied personalization that is not implicitly inferred from user behavior, but explicitly set by users and therefore transparent.

Practical Applicability On the one hand, our approach scales to larger datasets (more reviews, more hotels) with reasonable adaptions. The review processing is fully automatic, and manual or semi-automatic steps only relate to the contextual data per city or hotel. For instance, the *points of interest* proximity (Fig. 3) is independent of the number of reviews. A larger number of hotels on the map leads to overlapping location markers, requiring more elaborate filtering and zooming options. On the other hand, the approach is inherently limited to a few different hotels to be selected and compared at the same time; in our tests, we used up to six hotels and received satisfying results. Still, more empirical work is needed to investigate the number of candidate hotels customers would typically like to compare, and what type of comparison is most appropriate for what number of hotels. Moreover, restricting a flexible use, the current interface prototype is too wide for screens smaller than a laptop, such as mobile phones. However, the linear structure of the interface makes it easy to adapt to mobile devices, and the layout of interface elements can be changed to fit a narrow form factor. On touch devices, mover-over interactions are lost and should be transformed to on-click events.

Natural Language Processing With large language models, we transformed unstructured review texts into semi-structured sentiment items and confirmed that the results were reliable. After clustering these items, we condensed them to the set of cluster representatives. Some open ends are worth pursuing further. For instance, it might be interesting to compare the performance of our *GPT*-based sentiment extraction to other approaches [3, 32]. Another question is whether clusters of items are similar for different hotels, or whether some hotels have unique clusters. Those could then be visually highlighted as a noteworthy feature of a hotel. Finally, we could go even further and employ large language models to generate a personalized textual report by incorporating peoples' preference settings in the prompt. However, this would pose challenges regarding the accuracy of the resulting report.

Integration, Extension, and Generalization. *ViSCitR* is not a full-fledged hotel booking platform, but we envision that its interface (or large parts thereof) can be integrated into such platforms. It could specifically act as a detailed comparison page when customers have already bookmarked a pre-selection of hotels—somewhat similar to a comparative feature matrix of products available in some online shops. It can also be tailored to use cases such as business travel, where the personalization aspect might elaborate, for example, by selecting more specifically the location of a conference venue and offering business-related rating categories or keywords. Generally, personalization can be explored in greater depth. Our approach has primarily focused on two aspects (points of interest and category priorities), but others might be relevant, too (hotel services and facilities of personal interest, interior style of a hotel, spoken languages, etc.). Moreover, recommendations and data insights could be informed by the previous booking history of a customer. Extending the target audience to professionals from the hospitality sector also, it would be interesting to add more advanced options, for instance, to explore outliers such as reviews that the text processing could not categorize, or reviews with statements that do not match the rest of the reviews for a hotel. We decided to include categories and used the ones from the dataset since they seemed reasonable. However, the approach can be easily changed to include other categories and respective sentiment. Finally, the interface can be extended to reporting on other kinds of products. For geo-referenced businesses like restaurants, the required changes might be limited to identifying new relevant rating categories and related points of interest. For other products, changes would necessarily be more substantial. For instance, when comparing consumer electronics, the products would not have a location attribute, and personalization would need to work differently. However, we assume that a similar processing of reviews and a similar style of reporting is still applicable, as well as that consumers would profit from personalization and comparison features.

7 Conclusion

In this paper, we showcased how the information needs of hotel customers—like comparing summarized ratings or sentiments of hotel reviews—can be satisfied through an easy-to-understand, document-like visual interface called *ViSCitR*. For review summarization and comparison, we exploited AI-supported text processing and integrated contextual data. A unique feature of the interface is that all visual comparison components adapt to personalization, which the user can adjust at any time. As demonstrated through a qualitative study, users with diverse backgrounds were able to leverage the personalization and extract relevant information from the review comparison. The predefined exploration levels reflected different usage profiles regarding the main scope and depth of usage. Whereas this application study focuses on hotel reviews only, our ideas can be transferred to analyzing other review or opinion- and rating-based datasets. Moreover, *ViSCitR* can be regarded as an example of a novel type of personalized and comparative reporting interface that targets broad audiences.

Acknowledgments

We thank all participants of the study. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – projects 314647693 (VA4VGI) and 424960846 (vgiReports) as part of the Priority Program *VGIscience* (SPP 1894), and project 517412225 (CoCo).

References

[1] N. Akhtar, U. I. Siddiqi, M. N. Akhtar, M. Usman, and W. Ahmad. Modeling attitude ambivalence and behavioral outcomes from hotel reviews. *International Journal of Contemporary Hospitality Management*, 32(9):2831–2855, jul 2020. doi: 10.1108/ijchm-11-2019-0962

[2] D. Antognini and B. Faltings. HotelRec: a novel very large-scale hotel recommendation dataset. In *Proc. LREC*, pp. 4917–4923. European Language Resources Association, Paris, 2020. https://www.aclweb.org/anthology/2020.lrec-1.605.

[3] D. Antognini and B. Faltings. Fast multi-step critiquing for VAE-based recommender systems. In *Proc. RecSys*, p. 209–219. ACM, New York, 2021. doi: 10.1145/3460231.3474249

[4] B. A. Aseniero, C. Perin, W. Willett, A. Tang, and S. Carpendale. Activity River: Visualizing planned and logged personal activities for reflection. In *Proc. AVI*. ACM, New York, 2020. doi: 10.1145/3399715.3399921

[5] E. Bjørkelund, T. H. Burnett, and K. Nørvåg. A study of opinion mining and visualization of hotel reviews. In *Proc. IIWAS*. ACM, New York, 2012. doi: 10.1145/2428736.2428773

[6] booking.com. `https://booking.com`, 2023.

[7] U. Chakraborty. Perceived credibility of online hotel reviews and its impact on hotel booking intentions. *International Journal of Contemporary Hospitality Management*, 31(9):3465–3483, aug 2019. doi: 10.1108/ijchm-11-2018-0928

[8] C. Chen, F. Ibekwe-SanJuan, E. SanJuan, and C. Weaver. Visual analysis of conflicting opinions. In *Proc. VAST*, pp. 59–66. IEEE, New York, 2006. doi: 10.1109/VAST.2006.261431

[9] Y.-S. Chen, L.-H. Chen, and Y. Takama. Proposal of LDA-based sentiment visualization of hotel reviews. In *Proc. ICDMW*. IEEE, New York, 2015. doi: 10.1109/icdmw.2015.72

[10] K. Z. Gajos, D. S. Weld, and J. O. Wobbrock. Automatically generating personalized user interfaces with Supple. *Artificial Intelligence*, 174(12-13):910–950, may 2010. doi: 10.1016/j.artint.2010.05.005

[11] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, sep 2011. doi: 10.1177/1473871611416549

[12] Google Maps Hotel Search. `https://google.com/maps/search/Hotels`, 2023.

[13] GPT 3.5 turbo. `https://platform.openai.com/docs/models/gpt-3-5`, 2023.

[14] T. Hennig-Thurau, K. P. Gwinner, G. Walsh, and D. D. Gremler. Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*, 18(1):38–52, nov 2004. doi: 10.1002/dir.10073

[15] M. Hu and B. Liu. Mining and summarizing customer reviews. In *Proc. KDD*, pp. 168–177. ACM, New York, 2004. doi: 10.1145/1014052.1014073

[16] D. Huang, M. Tory, B. A. Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury. Personal visualization and personal visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 21(3):420–433, sep 2014. doi: 10.1109/TVCG.2014.2359887

[17] K. Kucher and A. Kerren. Text visualization techniques: Taxonomy, visual survey, and community insights. In *Proc. PacificVis*, pp. 117–121. IEEE, New York, 2015. doi: 10.1109/PACIFICVIS.2015.7156366

[18] K. Kucher, C. Paradis, and A. Kerren. The state of the art in sentiment visualization. *Computer Graphics Forum*, 37(1):71–96, jun 2017. doi: 10.1111/cgf.13217

[19] S. Latif and F. Beck. Interactive Map Reports summarizing bivariate geographic data. *Visual Informatics*, 3(1):27–37, apr 2019. doi: 10.1016/j.visinf.2019.03.004

[20] S. Latif and F. Beck. VIS Author Profiles: Interactive descriptions of publication records combining text and visualization. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):152–161, aug 2019. doi: 10.1109/tvcg.2018.2865022

[21] S. Latif, S. Chen, and F. Beck. A deeper understanding of visualization-text interplay in geographic data-driven stories. *Computer Graphics Forum*, 40(3):311–322, jun 2021. doi: 10.1111/cgf.14309

[22] C. Lewis and J. Rieman. Task-centered user interface design. *A practical introduction*, 1993.

[23] P. Li, Z. Wang, Z. Ren, L. Bing, and W. Lam. Neural rating regression with abstractive tips generation for recommendation. In *Proc. SIGIR*, pp. 345–354. ACM, New York, 2017. doi: 10.1145/3077136.3080822

[24] A. S. Lo and S. S. Yao. What makes hotel online reviews credible? An investigation of the roles of reviewer expertise, review rating consistency and review valence. *International Journal of Contemporary Hospitality Management*, 31(1):41–60, jan 2019. doi: 10.1108/ijchm-10-2017-0671

[25] D. Oelke, M. Hao, C. Rohrdantz, D. A. Keim, U. Dayal, L.-E. Haug, and H. Janetzko. Visual opinion analysis of customer feedback data. In *Proc. VAST*, pp. 187–194. IEEE, New York, 2009. doi: 10.1109/VAST.2009.5333919

[26] OSF—Personalized Visual Summarization and Comparison of Hotel Reviews. `https://osf.io/pu6gz`, 2023.

[27] Prolific. `https://app.prolific.co`, 2023.

[28] N. Reimers and I. Gurevych. Sentence-BERT: Sentence embeddings using siamese BERT-networks. In *Proc. EMNLP*. ACL, Kerrville, TX, USA, 2019. doi: 10.48550/arXiv.1908.10084

[29] N. H. Riche, C. Hurter, N. Diakopoulos, and S. Carpendale. *Data-Driven Storytelling*. CRC Press, 2018.

[30] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, oct 2010. doi: 10.1109/tvcg.2010.179

[31] D. Shi, X. Xu, F. Sun, Y. Shi, and N. Cao. Calliope: Automatic visual data story generation from a spreadsheet. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):453–463, oct 2021. doi: 10.1109/tvcg.2020.3030403

[32] Z. Shu, R. A. C. González, J. P. García-Miguel, and M. Sánchez-Montañés. Clustering using ordered weighted averaging operator and 2-tuple linguistic model for hotel segmentation: The case of TripAdvisor. *Expert Systems with Applications*, 213:118922, oct 2023. doi: 10.1016/j.eswa.2022.118922

[33] B. A. Sparks and V. Browning. The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management*, 32(6):1310–1323, jan 2011. doi: 10.1016/j.tourman.2010.12.011

[34] Tripadvisor. `https://tripadvisor.com`, 2023.

[35] I. E. Vermeulen and D. Seegers. Tried and tested: the impact of online hotel reviews on consumer consideration. *Tourism Management*, 30(1):123–127, jul 2008. doi: 10.1016/j.tourman.2008.04.008

[36] Y. Wu, F. Wei, S. Liu, N. Au, W. Cui, H. Zhou, and H. Qu. OpinionSeer: Interactive visualization of hotel customer feedback. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1109–1118, oct 2010. doi: 10.1109/tvcg.2010.183

[37] X. Zhang, J. Engel, S. Evensen, Y. Li, Ç. Demiralp, and W.-C. Tan. Teddy: A system for interactive review analysis. In *Proc. CHI*, pp. 1–13. ACM, New York, 2020. doi: 10.1145/3313831.3376235

[38] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In *Proc. SIGIR*, pp. 83–92. ACM, New York, 2014. doi: 10.1145/2600428.2609579

[39] X. Zhao, L. Wang, X. Guo, and R. Law. The influence of online reviews to online hotel booking intentions. *International Journal of Contemporary Hospitality Management*, 27(6):1343–1364, aug 2015. doi: 10.1108/ijchm-12-2013-0542

[40] Q. Zhu, L. Y.-H. Lo, M. Xia, Z. Chen, and X. Ma. Bias-aware design for informed decisions: Raising awareness of self-selection bias in user ratings and reviews. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):1–31, nov 2022. doi: 10.1145/3555597