

What is That? Crowdsourcing Questions to a Virtual Exhibition

Johannes Kiesel
johannes.kiesel@uni-weimar.de
Bauhaus-Universität Weimar
Weimar, Germany

Volker Bernhard
volker.bernhard@uni-weimar.de
Bauhaus-Universität Weimar
Weimar, Germany

Marcel Gohsen
marcel.gohsen@uni-weimar.de
Bauhaus-Universität Weimar
Weimar, Germany

Josef Roth
josef.roth@uni-weimar.de
Bauhaus-Universität Weimar
Weimar, Germany

Benno Stein
benno.stein@uni-weimar.de
Bauhaus-Universität Weimar
Weimar, Germany

ABSTRACT

Virtual environments with an ambient natural interface that allows to retrieve information for learning about the environment are a promising combination for implementing engaging virtual exhibitions. As a step towards better understanding search behavior in such exhibitions, this paper contributes the data from an exploratory study with participants asking questions on a real-world historical room, the Gropiuszimmer at the Weimar Bauhaus, while being on an “online virtual tour” through the room. The dataset comprises 849 manually categorized questions (557 in English, 292 in German) from 63 participants combined with a detailed interaction log, which allows replaying each session (29 hours total). The presented dataset and analyses aim to provide researchers and practitioners with a starting point to develop in-depth studies and prototypical systems.

CCS CONCEPTS

• **Information systems** → **Multimedia information systems**; **Specialized information retrieval**; • **Human-centered computing** → **Interaction paradigms**; • **Applied computing** → *Education*.

KEYWORDS

answer retrieval, conversational search, crowdsourcing, dataset, exhibition, question answering, user study, virtual environment

ACM Reference Format:

Johannes Kiesel, Volker Bernhard, Marcel Gohsen, Josef Roth, and Benno Stein. 2022. What is That? Crowdsourcing Questions to a Virtual Exhibition. In *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '22)*, March 14–18, 2022, Regensburg, Germany. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3498366.3505836>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHIIR '22, March 14–18, 2022, Regensburg, Germany

© 2022 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9186-3/22/03.

<https://doi.org/10.1145/3498366.3505836>

1 INTRODUCTION

Recent advancements in interface technology enable immersive virtual environments, thereby allowing everyone to experience locations that are far away,¹ long gone, or otherwise inaccessible. Coupled with an ambient natural interface for retrieving information, virtual environments provide a promising opportunity to “bring history to life.” But also without virtual reality equipment, on-screen virtual places with conversational search interfaces will allow engaging with historic sites in intuitive ways: asking questions right away instead of skimming through information boards.

Due to the aforementioned advancements, such virtual exhibitions are now feasible, yet little is known on the information needs visitors to such exhibitions would have. What kind of questions do they have? How do they connect visual and auditive information? How do they interact with the environment?

As a step towards answering these questions, the paper at hand contributes the data from an exploratory study in which participants asked questions on a real-world historical room, the Gropiuszimmer at the Weimar Bauhaus, while being on a “virtual tour” in an online version of the room. Participants were able to look around in a 360° panorama of the room, listen to an audio clip with background information, and set markers in the panorama to represent them pointing there. As the tour progresses, the camera position and audio clips change. For the study, 63 people were hired from a crowdsourcing platform, spending more than 29 hours total in the virtual environment and asking 849 questions. After a brief review of the historical and scientific background (Section 2), Section 3 details the study setup and execution. Section 4 then provides an overview of the collected data, which is freely available online.

2 BACKGROUND

This study takes place in the context of interactions in museums. Though social aspects are a key factor of the museum experience in many regular kinds of museums [5], the work at hand assumes the solitary visitor, which we expect to be the case for first generations of virtual exhibitions. Previous research on interactive interfaces in museums mainly focused on interactive museum guides, especially voice-based ones. Such guides can be broadly categorized into being mainly reactive to the visitor’s questions (e.g., [6, 8]) or proactive (i.e., questioning the visitor, e.g., [2, 7]). The study at hand mixes both behaviors, proactively providing information but waiting for the visitor’s question.

¹E.g., the CHIIR’21 virtual treasure hunt through Canberra using Google Street View.



Figure 1: The web interface at the start of the study with instructions folded in. Participants were able to listen to information, look around, and set a marker. They were instructed to ask at least two questions for each of 6 tasks.

Datasets of museum interactions are rare. Barth et al. [1] report on a dataset of more than 5000 visits to a real modern art museum with a voice-based interactive smartphone app. Based on manual categorization, they find that nearly 60% of visitors’ questions are related to reasonings, with only minor differences between the seven artworks. In contrast, the participants in our study used fewer such questions, with significant differences between the topics (cf. Section 4), likely due to our study proactively providing information that evokes different kinds of questions. Moreover, whereas the participants of Barth et al. were regular visitors, we recruited from the demographically diverse pool on Mechanical Turk [9].

We selected the Gropiuszimmer (“Gropius room”) in Weimar, Germany, as exhibition object, which is considered the first holistic spatial concept of modernism [3, 4, 10, 11]. Bauhaus school founder Walter Gropius designed the room for the Bauhaus exhibition in 1923. The office marks the transition of the Bauhaus from a wild, Expressionist-Romantic early phase to the world-famous functional style that Gropius then proclaimed as “Art and Technology - a New Unity.” Gropius wanted to demonstrate with this office the collaboration of the Bauhaus workshops and make his idea of holistic environmental design tangible. The exhibition, including the office, was intended to justify the modern teaching concepts and anti-bourgeois habitus of the Bauhaus students. The exhibition’s international success (15,000 visitors), however, could not stop the criticism of conservative elites, which led to the school’s moving in 1925. The room was still unfinished for the 1923 exhibition and later dismantled. The study is based on a reconstruction from original concept drawings. The study’s audio clips connect these pieces of information, amongst others, to the room’s furniture.

3 STUDY

In order to investigate which questions visitors to a virtual exhibition may ask, we created an exhibition mockup that gathers questions and logs all interactions, though it provides no answers.² We employed Amazon’s Mechanical Turk to hire study participants for asking questions, which we then categorized manually. Though the participants may not be representative of visitors to a virtual exhibition—a population that either way likely differs between exhibitions—, we hope the collected questions and behavior allow others to kickstart their own respective investigations and systems.

3.1 Setup

Figure 1 shows the web interface of the study.³ The instructions tell the participant to imagine themselves in the room, look around in the 360° panorama by dragging the mouse, and ask at least two questions “like you would ask the [museum] guide” for each of six tasks. The camera changes its position in the room between tasks to focus on specific objects (tasks 2 to 6). Each task has a audio clip containing background information, either on the room as a whole (task 1) or on the specific object in focus (task 2 to 6). Participants were able to pause the clip by pressing a button or typing a question. The instructions tell to imagine the latter as politely interrupting the speaker. Moreover, participants were able to set a marker on the image as substitute for them pointing somewhere.

The virtual environment consists of 360° panorama views created from high-resolution photographs of the Gropiuszimmer in its

²Dataset, code, and replay: <https://data.webis.de/#webis-exhibition-questions-21>

³For participants, the interface is embedded in the crowdsourcing platform’s page.

current reconstructed form (cf. Section 2). We employ third-party JavaScript libraries for rendering the panorama in a browser. Mouse dragging moves the camera to mirror the interaction in Google’s Street View, which we assume to be the virtual environment most familiar to the participants. The camera is then repositioned automatically between tasks to put other objects into the focus.

All 6 tasks ask for at least two questions related to the current audio clip, what people see, or whatever else comes to their mind. The 6 tasks focus on the following topics, in this order, positioning the camera to focus on the respective object at the start of the task: (1) the room in its entirety, with the audio clip providing historical context; (2) the large carpet/rug, with the audio clip detailing the imaginary inner room of representation and exchange delineated by carpet, panel on the left, and lights above; (3) the desk (behind the armchair in Figure 1), with the audio clip detailing how the desk stayed with Gropius his entire life. (4) the Bauhaus lamp (on the desk), with the audio clip detailing the collaborative design process at the Bauhaus; (5) the wall hanging (on the right hand side), with the audio clip detailing the life of its creator Else M \ddot{o} gelin; and (6) the antechamber (the camera’s position in Figure 1), with the audio clip detailing the antechamber’s detachment to make the main room equilateral.

The audio clips differed between participants in terms of language, length, and perspective. We conducted the study in both *German* and *English*, used audio clips with text lengths about 280 characters (*short*, Tweet size, ~15 seconds), 560 characters (*medium*, twice as long), and 840 characters (*long*, three times as long as short), and changed the perspective of the audio clip’s voice between being a *narrator* (like a museum guide) and *personal* (like the object talking itself).⁴ Only the lengths vary within subject, for which we employ three length patterns that ensure a equal distribution of lengths across topics and participants. The audio clips are written by a Humanities scholar with expertise on the room and assured to be consistent with each other across languages (being translations of each other), length (longer clips containing all parts of shorter clips), and perspective (containing the same information). The written text was voiced by Amazon’s Alexa voice or different Amazon Polly voices in the “personal” perspective. The source code for this setup and employed resources are shared alongside the dataset. However, usage of the panorama images other than viewing them in our interaction replay service requires consent from the respective copyright owners.

3.2 Crowdsourcing

We used Amazon’s Mechanical Turk to hire participants for the study. Each participant received \$3.00 (USD) for compensation, which amounts to a hourly wage of \$7.50 for the median study time of ~24 minutes, thus more than the US federal minimum wage. As an incentive to ask more questions, we announced a small bonus payment of \$0.10 for participants who asked three questions for at least one task. 27 participants gained this bonus payment. We rejected payment in 7 cases in which no questions were asked, and sorted out 14 participants who seem to have misunderstood the task (e.g., asking to re-decorate the room), but filled up these positions with

⁴Except for the first task, which has no assigned object but serves as introduction.

new participants until we gathered 7 participants for each configuration (language, length, perspective). We restricted participation for English to US workers and for German to Germany, Austria, and Switzerland. However, due to a shortage of participants for German, even after doubling the monetary compensation, we had to omit the configurations with German language and personal perspective. In total, the study comprises 849 questions from 63 participants. This amounts to 12% more questions than we demanded, showing overall a mild interest beyond the task’s minimum requirements.

3.3 Post-processing

To enable further analyses, we manually categorize each question based on the context it likely originated from (4 labels), what it is about (8 labels), what kind of answer it expects (16 labels), and whether it directly addresses the speaking object (perspective: personal; 1 label). The dataset’s README details each label and Section 4 provides for an analysis of the questions using these labels. We consider the collected questions as they are entered, even though 38 “questions” actually contain more than one question in the usual sense. However we assume that the participant intended these to form a single question, as in obvious cases like “Was this ever a working office? or just a display?” or “What is the table made out of exactly? Is that wood?”

Moreover, we link interface events (playing/stopping audio, looking around, setting the marker, or writing a comment) to the question the participant is asking at that time. We define this time period(s) of a participant asking a question q as starting directly after the last modification to some other question or the start of q ’s task until (including) the last consecutive modification of q .

4 DATASET OVERVIEW

For practitioners, the main use case of the dataset is to study in detail the questions the participants asked, so that one can make reasonable first decisions when designing and implementing a virtual exhibition. For a first approach to the dataset, this Section provides an overview of the dataset’s contents. Though the mockup did not provide answers, it provided piecewise new information with each task to provoke new questions.

Language. The example questions we provide in this Section are of participants for English language only. The questions of participants for German tend to be mostly similar, though the historic background of the Bauhaus school, which originated in Germany, seems to be more known in Germany: considerably fewer questions ask for basic background information in German than in English. This difference in knowledge might have also been caused by the public 100 years anniversary celebrations of the Bauhaus two years ago (2019), which was broadcasted in German media. Figure 2 exemplifies the general similarity across languages for the ratio of questions expecting a reason (compare “German, narrator” and “English, narrator”). The one task with a clear difference is that with the topic of the desk, which focuses on the historic developments after the Bauhaus exhibition, which are likely known to German participants who thus asked more questions on specific reasons than on what was going on in general.

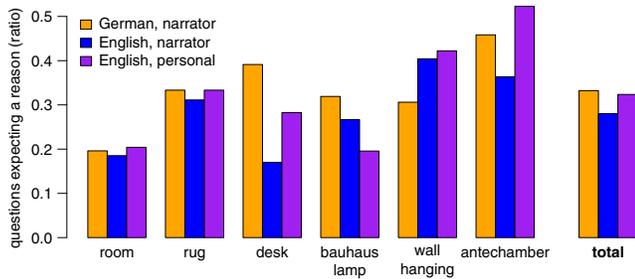


Figure 2: Ratio of questions that ask for reasons, by task, language, and perspective.

Perspective. For one half of the English participants the audio clips were written from the perspective of an ambient museums guide providing information (*narrator*), whereas for the other half the perspective was that of the respective object in the task’s focus providing the same information to the participant (*personal*). Surprisingly, this difference had very little effect on the questions asked. Only 13 of the 275 question for the personal perspective (~5%) directly address the speaking object (e.g., “Where were you made?”). Figure 2 further indicates similarity in questions that ask for reasons (no significant differences across tasks according to a χ^2 -test, p -values are all above 0.29). We thus conclude that changing the audio clip alone is not sufficient if the goal is to get visitors engaged in a dialogue with the objects. Possibly, an additional animation or highlighting of the object may provoke the desired effect.

Expects: Reasoning. A total of 265 of the 849 questions (31%) ask why something happened, why some object is as it is, or what the function of something is. As Figure 2 shows, the ratio of these questions varies between tasks (significantly as per χ^2 -test with p -value of 0.0002), showing that the information provided in the audio clip has a strong effect on how many of such questions are asked. Indeed, the questions in this category focus on design considerations and the historical context that the audio clips introduce. Most questions in this category are either on some object (115, e.g., “That table seems a little out of place in this room isn’t?”, “Why did Gropius choose that wall hanging?”, or “Is there any specific explanation for the art?”) or on the room itself (106, e.g., “Why did Gropius want to create a cube within a cube?”, “What makes this room unique?”, and “Isn’t the desk a little small to be used for every day work?”). About half of questions in both groups are also on design choices (53 and 51, respectively). The questions neither on an object nor a room are mostly on people (30, e.g., “How did Gropius feel about waste byproducts in his design? Was he a strict perfectionist?” and “Why did Gropius move to Massachusetts?”). Though nearly a third of questions might seem much, it is only half of the percentage reported by Barth et al. [1] for a modern art museum (Section 2). Modern art, however, might also be especially provocative of “why”-questions.

Expects: Description. A total of 121 of the 849 questions (14%) ask for descriptions or definitions, for which a suitable answer might be similar to the first paragraph of the corresponding Wikipedia article (if it exists). Most of these are either on an object (51, e.g., “What is that?” and “Is the design based on design diagrams? It seems

almost architectural.”), a concept (31, e.g., “What is modernism?” and “What are the key parts of Bauhaus style?”), or a person (20, e.g., “Who is Elza Mogulin?” and “What set Gropius’s designs apart from other designers of his time?”).

Expects: Fact. A total of 226 of the 849 questions (23%) ask for single facts, for which thus knowledge graphs could provide suitable answers. The kinds of fact requested are, however, quite diverse. In descending frequency, questions asked for dates and time periods (51, e.g., “In what year was that desk made?” and “How long did it take the original weaver to create that?”), materials used (50, e.g., “What material is the rug made of?”), costs (42, e.g., “How much is the weaving worth now? How much did this whole room cost?”), sizes (10, e.g., “What are the dimensions of the desk?”), design/product names (9, e.g., “What is the name for the style of those lovely yellow chairs and couches?”), locations (5, e.g., “Where was the team from that helped design the lamp?”), number of people (5, e.g., “How many people attended the exhibition in 1923?”), electrical values (4, e.g., “How bright is that lamp? In watts?”), and even color names (2, e.g., “What would you call the shade of brown used in the art hanging?”).

Expects: Interaction. A total of 9 of the 849 questions (1%) could be answered by changing the visual environment. Though this seems few, the interface does not encourage such interactions. For the few questions that still asked for interactions, possible responses include displaying pictures (“Are there pictures of the unfinished room?”), activating objects (“Can we see the light turned on?”), changing camera position (“Can we sit on the couch?”), and highlighting object (“Where are the soffit lamps you are referring to [...]?”). A separate study is needed to shed more light on voice controls for virtual exhibitions.

Interrupting the Audio Clip. Participants frequently paused the audio clips. For each clip length participants solved 126 tasks, pausing 107 times for short clips (0.85 times on average), 174 times for medium clips (1.38), and 186 times for long clips (1.48). Interestingly, participants interrupted even the short clips (~15 seconds). With minor differences between task lengths, the participants paused about half of the time by using the pause button and half of the time by typing a question. This balance might indicate different preferences of participants on how to interrupt a speaker, which would be interesting trying to reproduce with voice-based systems.

Question Context. For each of the 849 questions we also annotated the context from which it likely originated. 549 questions (65%) are likely from something the participant heard and 336 (40%) from something they saw, with some overlap. As Figure 3 shows, these questions are not evenly distributed across tasks, with more questions from something heard in the later tasks (probably because questions on eye-catching object had already been asked) and the most questions from something seen for the desk (especially on its materials and small size, but also on hidden compartments). Of the questions from something seen, 76 (9% of all, 23% of something seen) are about what the participant marked in the interface, indicating a clear need for virtual exhibitions to allow for pointing. Moreover, 389 questions (46%) are specifically on the object (or antechamber) in the respective tasks’ focus, indicating an interest in both the object itself and the larger story around it.

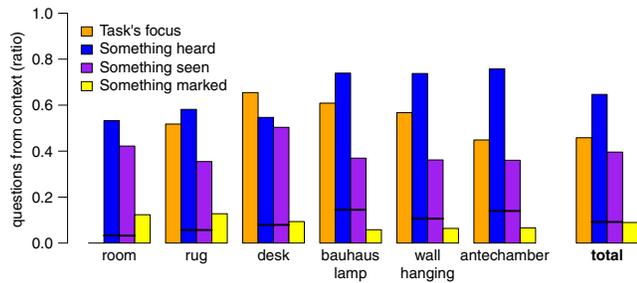


Figure 3: Ratio of questions from each context, by task. Horizontal lines indicate questions from both heard and seen.

5 CONCLUSION

This paper presents a first step towards understanding how visitors would interact with a virtual exhibition. To this end, we gathered a dataset of 849 manually categorized questions that provides insights on typical information needs visitors express within a virtual environment. These insights can help developers of virtual exhibitions to bootstrap the interface, especially the retrieval part.

With future advancements in virtual reality technology—and a wider dissemination of the corresponding devices, especially in the consumer sector—, implementations of such virtual exhibitions will become ever more immersive. We expect conversational interactions to present themselves as the preferred way to retrieve information with little disruption to the immersive experience. Whether future exhibitions may be able to also employ force fields to provide haptic feedback like in Star Trek’s Holodeck, remains to be seen.

ACKNOWLEDGMENTS

We would like to thank Thomas Gebhardt and Jan Frederick Eick for providing us with the 3D imagery of the room and the respective copyright holders for allowing us to use the image material for this research: Axel Kufus (furniture ante room by Axel Kufus), Bodelschwingsche Stiftungen Bethel (carpet by Benita Koch-Otte), Roswitha Campbell and Frank Schnellert (wall hanging by Else Mögelin), and VG Bild-Kunst (overall architecture and interior design by Walter Gropius; Bauhaus lamp by Wilhelm Wagenfeld). We would also like to thank the anonymous referees for their valuable comments and helpful suggestions. This work is supported by the Thüringer Ministerium für Wirtschaft, Wissenschaft und Digitale Gesellschaft (TMWWDG) under Grant 5575/10-5 (MetaReal).

REFERENCES

- [1] Fabricio Barth, Heloisa Candello, Paulo R. Cavalin, and Claudio S. Pinhanez. 2020. Intentions, Meanings, and Whys: Designing Content for Voice-based Conversational Museum Guides. In *2nd Conference on Conversational User Interfaces (CUI 2020)*, Maria Inés Torres, Stephan Schlögl, Leigh Clark, and Martin Porcheron (Eds.). ACM, 8:1–8:8. <https://doi.org/10.1145/3405755.3406128>
- [2] Roberta Beccaceci, Francesca Fallucchi, Cristina Giannone, Francesca Spagnoulo, and Fabio Massimo Zanzotto. 2009. Education with "Living Artworks" in Museums. In *1st International Conference on Computer Supported Education (CSEDU 2009)*, José A. Moinhos Cordeiro, Boris Shishkov, Alexander Verbraeck, and Markus Helfert (Eds.). INSTICC Press, 346–349.
- [3] Magdalena Droste. 2006. *Bauhaus 1919–1933*. Köln.
- [4] Walter Gropius. 1923. *Staatliches Bauhaus in Weimar 1919–1923*. Weimar.
- [5] Eva Hornecker and Luigina Ciolfi. 2019. *Human-Computer Interactions in Museums*. Morgan & Claypool Publishers. <https://doi.org/10.2200/>

S00901ED1V01Y201902HCI042

- [6] Stefan Kopp, Lars Gesellensetter, Nicole C. Krämer, and Ipke Wachsmuth. 2005. A Conversational Agent as Museum Guide - Design and Evaluation of a Real-World Application. In *5th International Working Conference on Intelligent Virtual Agents (IVA 2005) (Lecture Notes in Computer Science, Vol. 3661)*, Themis Panayiotopoulos, Jonathan Gratch, Ruth Aylett, Daniel Ballin, Patrick Olivier, and Thomas Rist (Eds.). Springer, 329–343. https://doi.org/10.1007/11550617_28
- [7] Thies Pfeiffer, Christian Liguda, Ipke Wachsmuth, and Stefan Stein. 2011. Living with a Virtual Agent: Seven Years with an Embodied Conversational Agent at the Heinz Nixdorf MuseumsForum.
- [8] Susan Robinson, David Traum, Midhun Ittycheriah, and Joe Henderer. 2008. What would you ask a conversational agent? Observations of human-agent dialogues in a museum setting. In *Proceedings of the sixth international conference on language resources and evaluation (LREC'08)*.
- [9] Kim Bartel Sheehan. 2017. Crowdsourcing research: Data collection with Amazon’s Mechanical Turk. *Communication Monographs* 85, 1 (2017), 140–156. <https://doi.org/10.1080/03637751.2017.1342043>
- [10] Hans M. Wingler. 1968. *Das Bauhaus. 1913–1933 Weimar Dessau Berlin und die Nachfolge in Chicago seit 1937*. Köln.
- [11] Klaus-Jürgen Winkler and Gerhard Oschmann. 2008. *Das Gropius-Zimmer*. Weimar.