

# “Kinda like The Sims... But with ghosts?”: A Qualitative Analysis of Video Game Re-finding Requests on Reddit

Ida Kathrine Hammeleff Jørgensen, Center for Computer Games Research,  
IT University of Copenhagen, Copenhagen, Denmark, ihjo@itu.dk

Toine Bogers Department of Communication & Psychology,  
Aalborg University Copenhagen, Copenhagen, Denmark, toine@hum.aau.dk

## ABSTRACT

With the advent of the retro-gaming movement there is an in-creasing interest in rediscovering games once played. ‘Tip of my Joystick’ is a Reddit community dedicated to the re-finding of for-gotten games. In this subreddit, users describe games they wish to re-find so that other users may help them identify the game’s title. This community thus offers a unique opportunity for studying how players recall and describe games and play experiences of the past. This paper presents the results of an analysis of a random sample of 250 posts from this subreddit. The posts were analyzed in terms of what aspects of games they describe. For the purpose of this analysis we developed a coding scheme consisting of 38 individual codes belonging to 9 different main categories. Our findings may contribute to research on game archiving and collection as they may help inform the design of better game search engines.

**KEYWORDS:** video games, game description, game archives, re-finding, known-item search, complex information needs

## 1 INTRODUCTION

With a history spanning over 40 years, video games are not only novel artifacts, but also things of the past that may be forgotten and rediscovered at a later point in time. In the Reddit community [/r/tipofmyjoystick](#), users attempt to re-find games they once played, but cannot remember the title of. Typically, a Reddit user will start a thread by posting a description of the game they wish to re-find, after which other users, based on this description, then try to help by suggesting the title of the game. As such, this community offers a unique opportunity for studying game re-finding. Such knowledge may prove valuable to the design and development of public game archives. To our knowledge, little consideration has been given to the question of how archived games should be made available for users to find. When users know the title of a game, re-finding that game may be straightforward, but when they remember only few aspects of the game not including the title and when many years have passed since last playing it, re-finding it can prove problematic. Another issue further complicate this scenario: games are highly complex, dynamic media objects and players do not necessarily engage with the same aspects of a game. Therefore, memories and descriptions of games may vary greatly from player to player.

While such issues cannot be solved within the scope of a single paper, we aim to take a first step in the process. In this paper we analyze so-called game-oriented known-item needs, i.e., limited but (mostly) complete descriptions of an existing object. In this study, the known items are games that users know on a first- or second-hand basis. Our analysis limits itself to the initial request in which the user describes their game information need, rather than the subsequent

comments by other users that form a collective problem-solving process. We code these game requests in terms of the aspects of the target game they describe and analyze our sample by the relative frequency of these individual aspects. While the Reddit community [/r/tipofmyjoystick](#) is not a game archive in itself, it does function as a user-driven instrument for re-finding games. As such, insights into how end-users — rather than archivists or collectors—describe games may potentially offer an alternative to existing game classification and retrieval principles

## 2 RELATED WORK

### 2.1 Game archiving

Several scholars have discussed the problems associated with the archiving of games, often focusing on the issues pertaining to collection and preservation [4,13,16,17,20,21,24]. A commonly discussed issue is which game aspects, such as their code or hardware, to archive, and which to ignore. By necessity, answering such questions includes a value judgement about which aspects of games are deemed essential—and therefore important to preserve.

Accessibility is another important issue in game archiving. The logic by which archivists classify games may not necessarily be logical to laymen users, and even among users there may be great variation in how they describe information items [12]. It is therefore necessary to consider how users are supposed to find the archived games they are looking for. Game classification and typologies [1,10,11,25] may inform the construction of search algorithms. However, previous work has shown that a mismatch between classification terms assigned by professional indexers and the search terms entered by regular users can impede search success [7]. It therefore makes sense to study the users' needs and search behavior in more detail and players' game descriptions provide an interesting avenue for this. Similar to how existing research [19] shows that fan-generated and -curated game archives [21] can be used in in more institutional setting, the [/r/tipofmyjoystick](#) Reddit community represents a participatory culture that we argue should be taken into account when designing tools for archival game search.

### 2.2 Game search & re-finding

Despite their popularity, there are only a handful approaches dedicated to information seeking behavior related to video games. In 2019, Anderson and Smith [2] and Zhang and Smith [26] presented work on how users describe their information needs with respect to retrieving specific moments in games. However, to the best of our knowledge there is no work on how the more general question of how people search for and decide which video games they want to play. In our paper, we focus on another type of information seeking behavior: *re-finding*. Re-finding is the act of attempting to locate and return to previously encountered information and is centered around satisfying so-called *known-item needs* — situations where a user has a limited but (mostly) correct description of an existing object [9]. Typically, the user is convinced the object exists and would like to find it. The closest to our work is the study by Bogers et al. [6], who collected and annotated a set of 521 threads from three gaming-focused Reddit subreddits. In their analysis they focused specifically on how their

annotations could inform the design of better game search algorithms. Their coding scheme contained five main categories of the aspects that make the known-item relevant to the user — content, metadata, experience, context, and interactivity. There are also a handful of similar studies of known-item needs in different domains, such as books and movies [5, 14].

### 3 METHODOLOGY

#### 3.1 Data collection

Earlier work has shown that online discussion forums are a fruitful source of descriptions of complex information needs for a variety of casual leisure domains [5,6,14]. We turned to Reddit, a popular discussion and social news website. Reddit has many dedicated discussion groups—so-called ‘subreddits’—focused on video games. There are a handful of subreddits focused on video game-related information needs, but only [/r/tipofmyjoystick](#) is dedicated to videogame re-finding requests. We adapted an existing Reddit crawler<sup>ii</sup> to collect the game re-finding requests from [/r/tipofmyjoystick](#). We set it to crawl all threads and comments posted to this subreddit from June 2-22, 2018, resulting in 1,131 threads. On average, these Reddit threads contained 4.4 comments for a total of 4,969 comments on all 1,131 threads. The length of these requests (as represented by the first post in a thread) varied greatly: from 1 to 1,137 words with an average of 115.5 words per request.

#### 3.2 Coding

We used *open coding* to develop our coding scheme for game re-finding requests on a random sample of 65 threads as our development set. Both authors developed their own individual coding schemes on this set. Coders were shown the title and the full text of the first post for each thread. Many subreddits, especially those aimed at solving other Reddit users’ problems, often includes guides or templates for how to phrase a request for help in a Reddit post. The [/r/tipofmyjoystick](#) community also offers such a template with a set of recommended categories to include: “Platform(s)”, “Genre”, “Estimated year of release”, “Graphics/Art style”, “Notable characters”, “Notable gameplay mechanics” and “Other details”. We were aware of the influence this template would have on our coding process, but our goal was to capture real-life information needs on Reddit, so we did not remove these headers before coding.

Our open-coding phase resulted in two different coding schemes with a combined total of 74 different codes. Many codes were proposed by both annotators. In the *axial coding phase*, we used affinity mapping to produce a single, unified coding scheme by merging similar codes and identifying hierarchical relationships between codes. The resulting coding scheme was then discussed by both authors until consensus was reached. In general, we made all axial coding decisions with the aim of both taking into account the needs of designers of game archivists as well as better supporting the search for new and old games by the general public. Our final coding scheme is described in the Section 4.

In the final coding phase, both authors annotated their own random sample of 150 game re-finding threads. None of the top-level or low-level codes are mutually exclusive. Top-level codes were never directly annotated, but instead inferred from the low-level annotations. In order to be able to examine the reliability of our content analysis, we made sure that 50 posts overlapped between both authors, resulting in a total of 250 unique coded posts. Agreement on the overlapping set of 50 posts was calculated using Cohen's kappa and was  $\kappa=0.46$  averaged over all 38 low-level codes, which can be classified as moderate agreement [18].

## 4 RESULTS

Our coding scheme<sup>iii</sup> includes nine top-level categories: **Design**, **Structure**, **Narrative**, **Genre**, **Metadata**, **Audience & purpose**, **Experience**, **Exposure**, and **Similarity**. These nine main categories are further subdivided into 38 low-level codes. **Figure 1** shows the structure of our coding scheme with top-level codes in bold and low-level codes in a lighter color variant.

Coding was not always straightforward, as the same statement in a game's description can represent multiple aspects. For example, **Availability** and **Hardware & platform** often overlapped, as illustrated by the following quote: *"I know it was from a website with a large amount of other games that weren't choose your own adventure."* This quote could be coded as **Availability** as it identifies the game's distribution model as a website, but it could also describe the platform on which the game is played. Sometimes, even specific terms used in the requests could have multiple meanings. For instance, the term 'arcade' was sometimes coded as **Genre**, as when described a game for PC with *"It was a very bright colorful cartoonish style arcade game"*. In other cases, it was used as **Hardware & platform**, as exemplified by the quote *"Tank-like two person arcade game. I used to play this with my friends in the arcade"*. In this case, the term 'arcade' should be coded as **Hardware & platform**, as the user described playing the game on an actual arcade machine. While coding in itself is an interpretive practice, some posts proved particularly difficult to code. For instance the term 'dark' can be applied to many game aspects, but this is not always clearly specified, such as in a post describing a game as a *"Dark FPS"*.

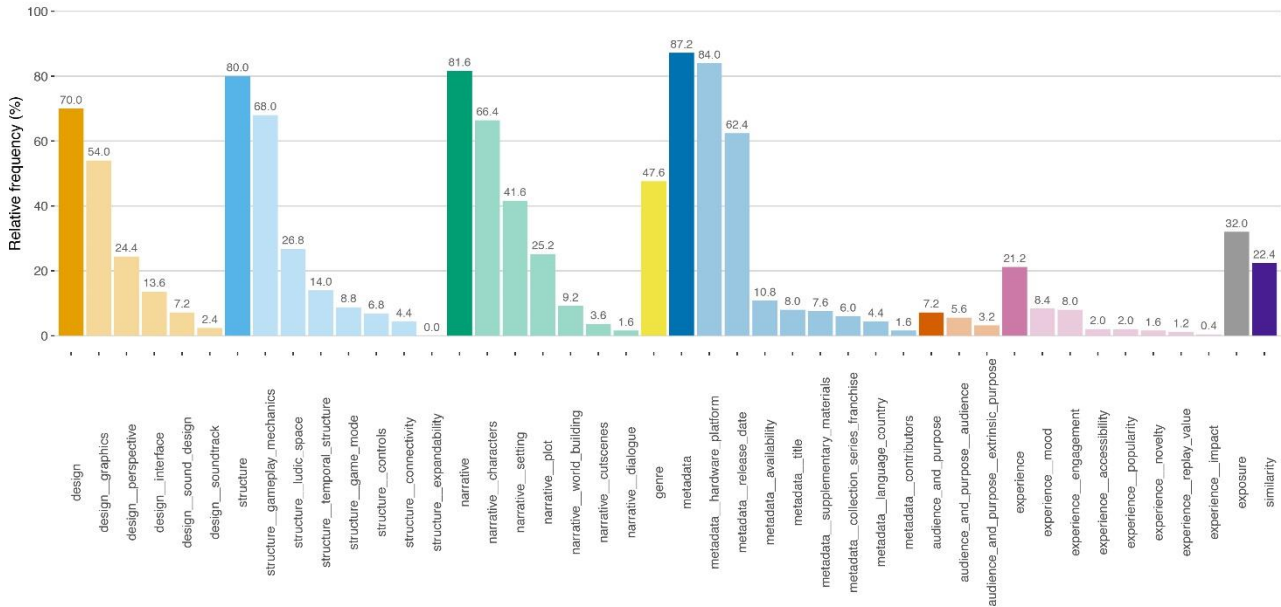
Finally, it should be mentioned that the code **Similarity** was used for descriptions of both similarity and dissimilarity. The following quotes demonstrate both uses: *"I already looked up Menace and Blood Money and most of Psygnosis's games that came out around the same time. They're not it"* and *"...it was a JRPG style turn based RPG much like Final Fantasy but was 100% not final fantasy"*. Even though one could argue that similarity and dissimilarity are opposites, both concepts are used to characterize a game by making comparisons to other games, which is why we combined them into a single code.

## 5 ANALYSIS

### 5.1 Complexity

The game re-finding requests posted to [/r/tipofmyjoystick](#) are of-ten quite complex: in our sample of 250 posts, 7.1 different codes were assigned to a post on average, reflecting the richness of these descriptions and

contrasting with the more singular focus of web search behavior. [5] report a similar level of complexity at 7.4 different codes on average for books and 7.7 for movie requests.



**Figure 1: Relative frequencies of the 38 codes in our sample of annotated Reddit threads (N= 250). Dark bars in each color group represents the top-level code and lighter-colored bars to their right represent the low-level codes.**

## 5.2 Frequency

Figure 1 shows the relative frequencies of the nine top-level codes as well as our 38 low-level codes. Despite the richness of the game re-finding requests in terms of number of codes applied per post, not all codes occur in equal measure in our sample. As **Figure 1** shows, the four most frequently applied top-level codes were **Meta-data** (89.2%), **Narrative** (83.2%), **Structure** (81.6%), and **Design** (70.8%). Each of these top-level codes were dominated by a few low-level codes. As shown in **Figure 1**, the codes **Characters**, **Setting** and **Events** — low-level codes of **Narrative** — occur with a relative frequency of 68.0%, 41.6% and 25.2% respectively. In contrast, **World building**, **Cutscene(s)** and **Dialogue** are much more infrequent (<10%). It seems that in particular characters and settings are more easily remembered by players, and also easy to communicate in the relative short format of the Reddit requests. These aspects are also commonly remembered in other domains, such as books and movies [5,7], and can be easily described without the need for a specialized vocabulary as well as understood by the other members of the [/r/tipofmyjoystick](#) community. A similar pattern emerges with **Metadata**, **Structure** and **Design**. For each of the three top-level codes only two low-level codes occur with a frequency of over 20%. This distribution is unsurprising since most frequently used codes also coincided with the template often used in [/r/tipofmyjoystick](#). Nevertheless, our analysis sheds light on the many aspects not already part of the template, but which appeared consistently, albeit less frequently, in the game descriptions.

Among the less frequently used top-level codes are **Genre** (47.6%), **Exposure** (32.4%), **Similarity** (22.4%), **Experience** (21.2%), and **Audience and & purpose** (7.6%). Of these, only **Experience** contains more than two low-level codes, although they are all fairly infrequent at 0.4% to 8.4% of all posts.

With regard to **Genre** (47.6%) — albeit a contested concept among game researchers [3,8,15,23] — it is unsurprising that it is a relatively frequently described aspect, as it remains a widely-used shorthand for describing games in popular discourse. What may seem more remarkable, is the frequency of **Exposure** (32,4%). Afterall, this code describes a somewhat idiosyncratic aspect of the game, that at first sight may not seem immediately useful in a community-based search. However, we also frequently encountered situations where the user did not know when the game was released, but instead offered an estimate of when they themselves played the game. In the absence of more accurate release date information, this could still help other users narrow down the search by eliminating other games by filtering out ‘impossible’ release dates.

Another interesting and complex top-level code is **Similarity**, which we encountered in 22.4% of all posts. It is evidence of the comparative manner in which games tend to be described, as players point out other, known game titles that are, in different ways, (dis)similar to the game that the user wishes to re-find. It should be noted, though, that it was not always easy to deduce from the description in what way(s) that other game was similar to the one the user was trying to re-find. Therefore, while it is evident from the frequency of **Similarity**, that this is an important aspect of game re-finding request, it would be far from trivial to operationalize as part of a retrieval algorithm for a game search engine.

The template provided by the [/r/tipofmyjoystick](#) admins overlaps with our coding scheme. However, with its 10 top-level codes and 38 low-level codes, our coding scheme is considerably richer and more nuanced, reflecting that users do not necessarily limit themselves to the template — or even follow it altogether. It would therefore be of particular interest to determine which codes are relatively frequent in our data set despite not being a part of the template. Of these, the most noteworthy ones are probably **Setting**, **Events** and **Similarity**. None of these can easily be contained within the categories of the template, but they nevertheless occur in between 20 and 40% of all posts.

## 6 DISCUSSION & CONCLUSIONS

Our findings open up many interesting directions for future work. For instance, we did not analyze the demographic characteristics of [/r/tipofmyjoystick](#) users, but assumed that they are representative of the entire population of video game players. However, many attempts have been made to design games targeted at specific demographics, such as female players. If these games differ significantly from others, re-finding requests for them may differ as well. A better understanding of the (lack of) homogeneity of [/r/tipofmyjoystick](#) could help us frame our findings and assess their generalizability. Drawing conclusions based on a more homogeneous user group could affect the results in certain ways, such as by promoting some aspects of games at the expense of others, which runs counter to the idea of building inclusive game search engines.

In the present study, we also did not annotate the historical period (e.g., the decade) in which the game was released. However, design conventions, distribution and business models, as well as available technologies change over time. For example, due to CD-ROM technology, game developers increased the number of cut-scenes in their games [22]. One could hypothesize that this might also be observable from the frequency of the low-level **Cutscene(s)** code when comparing requests decade by decade.

These issues reinforce the complexity of building search engines for (re-finding) games and require a detailed understanding of general game search behavior. To achieve this, we should not only take into account the complexities of current games, but also past design conventions and game distribution patterns. Our study represents a first step in this direction. We may also need to look beyond the request itself and examine the collaborative problem-solving process taking place in the [/r/tipofmyjoystick](#) community after a request has been posted. Investigating the success factors of game re-finding requests would be valuable for optimizing the game re-finding process specifically and game search in general.

## REFERENCES

- [1] Espen Aarseth, Solveig Marie Smedstad, and Lise Sunnanå. 2003. A multidimensional typology of games. In DiGRA Conference.
- [2] Barrett R Anderson and Adam M Smith. 2019. Understanding User Needs in Videogame Moment Retrieval. In Proceedings of FDG '19. 1–10.
- [3] Thomas H. Apperley. 2006. Genre and game studies: Toward a critical approach to video game genres. *Simulation and Gaming*37, 1 (2006), 6–23.
- [4] Alasdair Bachell and Matthew Barr. 2014. Video game preservation in the UK: A survey of records management practices. (2014).
- [5] Toine Bogers, Maria Gäde, Marijn Koolen, Vivien Petras, and Mette Skov. 2018. “What was this Movie About this Chick?” A Comparative Study of Relevance Aspects in Book and Movie Discovery. In Proceedings of iConference 2018. Vol. 10766. Springer, 323–334.
- [6] Toine Bogers, Maria Gäde, Marijn Koolen, Vivien Petras, and Mette Skov. 2019. “Looking for an Amazing Game I Can Relax and Sink Hours into...” A Study of Relevance Aspects in Video Game Discovery. In Proceedings of iConference 2019. Vol. 11420. Springer, 503–515.
- [7] Toine Bogers and Vivien Petras. 2015. Tagging vs. Controlled Vocabulary: Which is More Helpful for Book Search?. In Proceedings of iConference 2015. iDEALS.
- [8] Rachel Ivy Clarke, Jin Ha Lee, and Neils Clark. 2017. Why video game genres fail: A classificatory analysis. *Games and Culture*12, 5 (2017), 445–465.
- [9] Mats Dahlström and Mikael Gunnarsson. 2000. Document Architecture Draws a Circle: On Document Architecture and its Relation to Library and Information Science Education and Research. *Information Research*5, 2 (2000).
- [10] Michael Sebastian Debus. 2019. Unifying Game Ontology: A Faceted Classification of Game Elements. Ph.D. Dissertation. The IT University, Copenhagen, Denmark.
- [11] Christian Elverdam and Espen Aarseth. 2007. Game classification and game design: Construction through critical analysis. *Games & Culture*2, 1 (2007), 3–22.
- [12] George W. Furnas, Thomas K. Landauer, Louis M. Gomez, and Susan T. Dumais. 1987. The Vocabulary Problem in Human-System Communication. *Communications of the ACM*30, 11 (1987), 964–971.
- [13] Paul Gooding and Melissa Terras. 2008. ‘Grand theft archive’: a quantitative analysis of the state of computer game preservation. (2008).
- [14] Marijn Koolen, Toine Bogers, Jaap Kamps, and Antal van den Bosch. 2015. Looking for Books in Social Media: An Analysis of Complex Search Requests. In ECIR '15: Proceedings of the 37th European Conference on Information Retrieval (Lecture Notes in Computer Science, Vol. 9022). Springer, 184–196.

- [15] Jin Ha Lee, Natascha Karlova, Rachel Ivy Clarke, Katherine Thornton, and Andrew Perti. 2014. Facet analysis of video game genres. In Proceedings of iConference2014.
- [16] Henry Lowood. 2004. Playing History with Games steps towards historical archives of computer gaming. In Electronic Media Group Annual Meeting of the American Institute for Conservation of Historic and Artistic Works Portland, Oregon June, Vol. 14.
- [17] Henry Lowood. 2009. Game capture: The machinima archive and the history of digital games. Mediascape: UCLA's Journal of Cinema and Media Studies(2009).
- [18] Mary L McHugh. 2012. Interrater Reliability: The Kappa Statistic. *Biochemia Medica* 22, 3 (2012), 276–282.
- [19] Benjamin Nicoll, Melanie Swalwell, Ndalians Angela, and Helen Stuckey. 2017. Sega Saturn Fan Sites and the Vernacular Curation of Videogame History. 180.
- [20] Adrienne Shaw. 2017. What's next?: the LGBTQ video game archive. *Critical Studies in Media Communication* 34, 1 (2017), 88–94.
- [21] Helen Stuckey, Melanie Swalwell, Angela Ndalians, and Denise de Vries. 2013. Remembrance of Games Past: The Popular Memory Archive. In Proceedings of the 9th Australasian Conference on Interactive Entertainment. 1–7.
- [22] Carl Therrien. 2008. CD-Rom Games. Greenwood Press Westport, 121–134.
- [23] Zach Whalen. 2004. Game/genre: A critique of generic formulas in video games in the context of "the real". *Works and Days* 22, 43/44 (2004), 289–303.
- [24] Megan A. Winget and Caitlin Murray. 2008. Collecting and preserving videogames and their related materials: A review of current practice, game-related archives and research projects. *Proceedings of the American society for information science and technology* 45, 1 (2008), 1–9.
- [25] José P. Zagal, Michael Mateas, Clara Fernández-Vara, Brian Hochhalter, and Nolan Lichti. 2007. Towards an ontological language for game analysis. *Worlds in play: International perspectives on digital games research* 21 (2007), 21.
- [26] Xiaoxuan Zhang and Adam M Smith. 2019. Retrieving videogame moments with natural language queries. In Proceedings of the FDG '19 Workshop on Workshop on Games and Natural Language Processing. 1–7.

---

<sup>i</sup> The original 10-page version of this paper can be found at <http://www.toinebogers.com/content/publications/fdg2020-long.pdf>

<sup>ii</sup> Available at <https://github.com/lucas-tulio/simple-reddit-crawler>

<sup>iii</sup> Due to lack of space, the final coding scheme can be found at <https://bit.ly/fdq2020>.