

“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 increasing interest in rediscovering games once played. ‘Tip of my Joystick’ is a Reddit community dedicated to the re-finding of forgotten games. On this subreddit, users describe the game they wish to re-find as best they can remember it, 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, such as hardware, mechanics, or characters. 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 search engines enabling users to better retrieve archived games or information about games.

KEYWORDS

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

ACM Reference Format:

Ida Kathrine Hammeleff Jørgensen and Toine Bogers. 2020. “Kinda like The Sims... But with ghosts?”: A Qualitative Analysis of Video Game Re-finding Requests on Reddit. In *Proceedings of ACM Conference (Conference’17)*. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

With a history now spanning over 40 years, video games exist not only as novel artifacts associated with the latest technological advancements, but also as things of the past. As such, they may be forgotten and rediscovered at a later point in time. In the Reddit community [/r/tipofmyjoystick](https://www.reddit.com/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 and, based on this description, other commenters may then help solve the user’s problem by suggesting the title of the game. As such, this community offers a unique opportunity for studying how players describe games of the past. Such knowledge may prove valuable to the design and development of game

archives—collections, whose information is or should be made accessible to the public. Scholars have already studied a variety of professional or fan-based game archives, and suggested principles for how to classify games within these archives, but to our knowledge relatively little consideration has been given to the question of how these games should be made available for the users to find. When users know the title of a game, re-finding that game may be relatively straightforward, but this is not always the case. Many years may have passed since a user last played the game they are looking for, and as a result they may remember only few aspects of the game, which may not include the game’s title or perhaps a false memory of it.

Other issues further complicate this scenario. First of all, as many games are highly complex media objects with a great number of different properties, players do not necessarily engage with the same aspects of a game. Aarseth [3], for example, argues that games exist as both objects and processes and, furthermore, that the game object consists of both a mechanical layer—the internal structure of the game and the rules for how it functions—as well as a presentational layer, which represents how the game is communicated to the player, for example by means of text and images as well as characters, objects, and locations. The game as process on the other hands is about the events (narrative as well as actual) that unfolds during play. A natural consequence of this is that, due to the complex nature of games, how one remembers and describes a game may vary greatly from player to player. Furthermore, the characterization of an individual game is often relative to the sum of available games at a given point in time. In other words, when we describe a game, we often do that by making comparisons with other games that we know of. However, if a player describes a game as similar to Doom, for instance, this description does not in itself necessarily offer any information about the character of this similarity.

Such issues cannot be solved within the scope of a single paper. However, we do aim to take a first step in the process. The work described in this paper attempts to shed light on the complexities and logic underlying the ways in which users try to describe these so-called game-oriented *known-item* needs, i.e., limited but (mostly) complete descriptions of an existing object. In the case of this study, the known items are games that users know on a first- or second-hand basis, either because they have played these games themselves or because they have watched other people play them. The current analysis limits itself to the initial request in which the user describes the searched for game, 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. We then analyze our sample of annotated requests both by looking at how (un)common the individual codes are as well as how they relate to other codes. While the Reddit community [/r/tipofmyjoystick](#) is not a game archive in itself, it does function as a user-driven instrument for re-finding existing games. As such, insights into the principles with which end-users—rather than archivists or collectors—describe games may potentially offer an alternative to existing game classification and retrieval principles.

2 RELATED WORK

We argue that the analysis of game re-finding requests presented in this paper offers valuable insights into the archiving of games, as it specifically sheds light on the myriad ways in which users describe the games they are searching for. This is of particular importance if game archives are to be made available to the general public. However, the game re-finding requests also act as straightforward player-made analyses of games in the sense that they break down and describe the different elements of games, highlight some and ignore other. As such, they also represent examples of how players conceptualize and discuss games. This leads us to three different areas of related work in this section. First, we discuss the most important work on analytical game description and classification. Second, we offer an overview of research related to game archiving. Finally, we review relevant work on information seeking behavior related to video games and other related domains.

2.1 Game description & classification

In the field of game studies, a great deal of effort has been devoted to the classification and description of games, often in the form of game ontologies. A number of analytical frameworks have been proposed over the years. One such framework is Konzack's framework for game criticism [26], which distinguishes between hardware, program code, functionality, gameplay, meaning, referentiality and socio-culture. Other examples include the multidimensional typology of games developed by Aarseth et al. [2] and Elverdam and Aarseth [19], which considers a number of aspects including time, space, and player relations, and the Game Ontology Project by Zagal et al. [47], which considers four overall categories of interface, rules, entity manipulation and goals. In 2018, Aarseth and Grabarczyk presented a meta-ontology of games consisting of a physical layer, a structural layer, a communicational layer and a behavioral layer [1]. Finally, Debus [16] proposed a unified game ontology made up of seven game elements: mechanics, time, goals, space, randomness, entities, as well as a category of so-called 'unattached facets'. Although all of these frameworks differ in terms of their level of detail and their overall purpose, they are all centered around describing the game object in itself, and are therefore either constructed or harnessed from the formal analysis of games.

In contrast, there are also several studies that are concerned with a more empirical investigation of how games are described by users in various discursive contexts. Zagal and Tomuro [49], for example, employed natural language processing (NLP) techniques in an analysis of an extensive corpus of user reviews of games. The

purpose of their study was to analyze the ways in which the notion of 'gameplay' was described. They sampled all the sentences in which this term appeared and thereafter extracted all the adjectives associated with the word 'gameplay' and in which context it appeared. Interestingly, they found that when describing gameplay, the most important elements users describe refer to the game's story, characters, controls and in-game objects, whereas other elements, such as the country (or region) of origin were noticeably absent. Grace [23] analyzed descriptions of games on Apple's App Store in terms of the most frequent verbs appearing in those descriptions. For the author, verbs are an interesting site of inquiry, as she argues game descriptions often focus on how the games are supposed to be played, which is seldom the case with other media. Ryan et al. [36] collected game descriptions from their respective Wikipedia articles and used NLP and latent semantic analysis to cluster related games together. The authors developed a tool on top of this called GameNet, which allow users to look up a given game and see an ordered list of the 50 most related games to that specific seed game. The authors also discuss how this tool can be used as an alternative to genre typologies, as it shows how games can be related to one another, based on any notable aspect that they share instead of simply being similar in terms of specific game aspects, such as gameplay mechanics. However, as acknowledged by the authors, GameNet is not able to explain or make explicit in what respect two games are related.

2.2 Game archiving

Several scholars have discussed the problems associated with the archiving of games, often focusing on the issues pertaining to collection strategies and how to preserve the game objects [6, 21, 31, 37, 39, 46]. An added difficulty, as argued by Grabarczyk and Aarseth [22], is that many game titles come in multiple versions, which requires that the archivist or collector makes sense of their differences and makes a decision about which version to preserve.

Common to all of this research is, that it discusses the basic problem of which aspects of games to include in the archives and which to ignore. Lowood [30], for example, asks whether it would be sufficient to preserve only a game's content, or whether it necessary to also preserve the gaming technology on which the game is played. Answering such a question must by necessity include a value judgement about which aspects of games are deemed essential—and therefore important to preserve—and is always relative to some purpose.

However, the collection, classification, and preservation of games is only one aspect of game archiving. If we are to make meaningful game archives, we should also consider how to make the archived materials available to the archive's intended users. This raises different questions, such as how users are supposed to find the games they are looking for in the archives. The logic by which archivists may classify games may not be as logical to their users, and even among users there may be great variation in how they describe information items [20]. While the game classifications and typologies discussed above could prove to be valuable tools for game archivists as detailed and thorough classification and annotation frameworks, the logic on which they are built may not necessarily be apparent to the archive's intended users, who are not necessarily librarians

or game scholars by training. Previous work has shown that this mismatch between classification terms assigned by professional indexers and the search terms entered by regular users can impede search success [13]. It therefore makes sense to look into the users' needs in more detail. This was exactly the aim of Lee et al. [29], who created a metadata schema for a video game collection in a museum context. They gathered existing metadata from popular game websites as well as Wikipedia, and assessed their relative importance to a set of personas, that they had defined for the purpose of their study. Compared to this study, the present paper studies actual re-finding requests of users of [/r/tipofmyjoystick](#) in order to analyze how players themselves describe games in the context of re-finding.

Stuckey et al. [39] argue for the necessity of considering the role of game fans in the documentation and preservation of games. Indeed, the authors characterize game fans as being at the forefront of video game preservation. These archival practices are expressed in the creation, maintenance, and curating of fan-generated archives, whereas Nicoll et al. [35] also discuss how these practices can be used in institutional settings, such as museums. In contrast, our study is not concerned with dedicated fans, but specifically with users of the [/r/tipofmyjoystick](#) Reddit community. Still, these users—both those who post the requests and those who help find the answers—represent a significant participatory culture that we argue should be taken into account when designing tools for archival game search.

2.3 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 [4] and Zhang and Smith [50] presented work on how users describe their information needs with respect to retrieving specific moments and events in video 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 has been studied in and across many different contexts, applications [8], and information types, such paper documents [32], websites [41], emails [18], and social media posts [34]. Re-finding is centered around satisfying so-called *known-item needs*. Dahlström and Gunnarsson [15] define a known-item need as a situation where a user has a limited but (mostly) correct description of an existing object. Typically, the user is convinced the object exists and would like to find it. Its actual existence is not necessarily required though and can even be doubted by the user. Studies have shown that users can often remember multiple aspects about the known item they are trying to re-find, such as how, when and where they encountered it the first time [7, 8, 17].

The closest to our work is the study by Bogers et al. [11], who collected and annotated a set of 521 threads from three gaming-focused Reddit subreddits ([/r/gamingsuggestions](#), [/r/gamesuggestions](#), and [/r/tipofmyjoystick](#)). In their analysis they focused specifically on how their annotations could inform the design of better game

search algorithms. Their coding scheme contained five top-level categories of the aspects that make the known-item relevant to the user—content, metadata, experience, context, and interactivity. In addition, they annotated the type of information need, one of which were known-item needs, which made up 50.8% consisted of their total sample. Our work overlaps with theirs in that we used one of the same subreddits ([/r/tipofmyjoystick](#)) to collect data from, but our analysis is informed more by the current game classification literature than by information retrieval concerns. There are also a handful of similar studies of known-item needs in different domains, such as book requests on the LibraryThing fora [10, 27], and movies on the IMDB fora [9, 10]. Another related study is the work by Hagen et al. [24], who analyzed a set of known-item questions (and answers) posted to social question-answering website Yahoo! Answers¹ from three different domains: movies, music, and websites. They specifically focused on the false memory aspect of these re-finding needs.

3 METHODOLOGY

3.1 Data collection

In order to perform a detailed analysis of the aspects that users think they need to describe when attempting to re-find a video game from their past, we collected and analyzed a representative sample of information need descriptions specifically focused on re-finding old video games. 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, such as books [10, 27], movies [9, 10], and games [12]. Similar to Bogers et al. [12], we turned to Reddit², a popular discussion and social news website. Reddit has many dedicated discussion groups—so-called 'subreddits'—focused on video games, and at least three of them are focused specifically on solving video game-related information needs on Reddit. Two of these subreddits, [/r/gamingsuggestions](#) and [/r/gamesuggestions](#), contained a wide variety of requests for new games to play, while the third ([/r/tipofmyjoystick](#)) is dedicated to known-item (i.e., re-finding) requests for video games. Another, more general subreddit called [/r/tipofmytongue](#) exists that contains re-finding requests for all types of media, including video games. However, we focused only on [/r/tipofmyjoystick](#) because of its singular focus on re-finding video games, which made data collection more straightforward.

The fact that Reddit has subreddits dedicated specifically to different types of video game discovery and re-finding shows that such requests are common enough to warrant these dedicated subreddits. It also suggests the threshold for users to post such requests is low and that traditional search engines are not able to satisfy them. It is possible that Reddit users are not representative of all players, but we have no reason to believe their requests differ in any significant way from those of others.

We adapted an existing Reddit crawler³ to collect the game re-finding requests from the [/r/tipofmyjoystick](#) subreddit. We set it to crawl all threads and comments posted to this subreddit from

¹ Available at <https://answers.yahoo.com/>, last accessed May 17, 2020

² Available at <http://reddit.com>, last visited May 17, 2020.

³ Available at <https://github.com/lucas-tulio/simple-reddit-crawler>, last visited May 17, 2020.

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 word in length to 1,137 words with an average of 115.5 words per request.

3.2 Coding

3.2.1 Open coding. We used an open coding approach to develop our coding scheme for game re-finding requests. We selected a random sample of 65 threads from our Reddit crawl to serve as our development set. Both authors developed their own individual coding schemes on this development set. We settled on a sample size of 65 threads because earlier work has shown that it provides a good balance between effort and recall, allowing even the more infrequent but meaningful code to have a chance of being identified [9, 10, 12]. 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 to its users with the aim of increasing the success rate of the requests posted to the subreddit. This template contains a set of categories that users are recommended to include when posting their requests: "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 template headers before coding.

3.2.2 Axial coding. Our open coding phase resulted in two different coding schemes with a combined total of 74 different codes. We then used affinity mapping in the axial coding phase to produce a single, unified coding scheme and identify relationships between codes and arrange them into higher-level codes. Many codes were proposed by both annotators. After the merging phase, we grouped related categories into top-level categories. The resulting coding scheme was then discussed by both authors until consensus was reached about all codes and their descriptions. 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. We added textual descriptions of each code to our code book along with prototypical examples to aid the final annotation process. Our final coding scheme is described in the Section 4.

3.2.3 Final coding. In the final coding phase, both authors annotated their own random sample of 150 game re-finding threads. Posts from the development set were not re-used in this final sample. 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 its results are covered in Section 4.2. None of the top-level or low-level codes are mutually exclusive, so a request could be annotated with multiple low-level codes at once. Top-level codes

were never directly annotated, but instead inferred from the low-level annotations: if at least one low-level code has been assigned to a specific post, then its parent top-level code is automatically applied as well. Some posts included links to screenshots or videos. we never followed these links to make sure we avoided implicit annotations not present in the original description.

4 RESULTS

4.1 Coding scheme

Figures 1 and 2 show our final coding scheme, along with the descriptions of the different codes. Our coding scheme 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. None of these codes are mutually exclusive and therefore a game re-finding request can be annotated with more than one code as different parts of a game description can represent different aspects of a game.

Coding was not always straightforward, as the same statement in a game's description could represent different aspects at once. For example, **Availability** and **Hardware & platform** often overlapped as the aspects these codes describe may in some cases be very closely connected. The following quote from one of the posts in our data set nicely illustrates this point: *"I know it was from a website with a large amount of other games that weren't choose your own adventure."* This quote can be coded as **Availability** as it identifies the distribution model of the game as a website, but it can also describe the platform on which the game is played. The same applies to the two low-level codes under **Audience & purpose**. For example, a game could be described as a *"Children's educational game"*, and therefore be coded as both **Audience** and **Extrinsic purpose**. However, it was important to keep the distinction, as not all educational games are for children, and not every post coded as **Audience** necessarily described games aimed at children. One post, for example, characterized a game as a *"Mature rated game"*, in which case it was coded as **Audience**, but not as **Extrinsic purpose**.

Sometimes, even specific terms used in the requests could have different meanings. For instance, the term 'arcade', was sometimes coded as **Hardware & platform** and sometimes as **Genre**. The following two quotes illustrates the difference. In one post, the title of the thread described the game as a *"...Tank-like two person arcade game"* and the post itself further clarified that *"I used to play this with my friends in the arcade"*. In this case, it was relatively clear from the post that the term 'arcade' should be coded as **Hardware & platform** as the user described having played the game on an actual arcade machine. However, in another post 'arcade' was used more as a genre tag: *"It was a very bright colorful cartoonish style arcade game"*. Since this user had already specified that the game was for PC, it was clear that the term 'arcade' was not used to describe the physical machine, but instead used as a genre label. While coding is in itself an interpretive practice, some posts proved particularly ambiguous and difficult to code. For instance the term 'dark' can be applied to many aspects of a game, but this was not always clearly specified by the request. One post, for example, simply described the game as a *"Dark FPS"*. In this case, 'dark' could be interpreted as a graphics aspect as well as the overall mood the game evoked.

Finally, it should be mentioned that the code **Similarity** was used for descriptions of similarity as well as 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 opposite concepts, both concepts are used to characterize the requested game by making a comparison to another game. For this reason, we use the same code for both usages, even though these were originally split into codes called “More like this” and “Not this one” during the development stage.

4.2 Inter-annotator agreement

As mentioned earlier in Section 3.2.3, we arranged for 50 overlapping posts to be coded by both authors. We used Cohen’s kappa on a code-by-code basis, because we had exactly two coders for each of the 50 posts⁴. Agreement averaged over all 38 low-level codes is $\kappa = 0.46$, which can be classified as moderate agreement according to Cohen [33], indicating that coding game re-finding requests for the properties of games users remember and express is far from a straightforward task. Some codes occur relatively infrequently in our overlapping set of 50 posts, so care should be taken in interpreting their κ values. In accordance with Sim and Wright [38], we only consider those low-level codes in our overlap set with a relative frequency greater than 5%.

Looking at the nine top-level categories, we can see considerable variation in agreement scores: **Metadata** has the highest agreement at $\kappa = 0.85$, whereas **Structure** has the lowest agreement at $\kappa = 0.18$. On the other top-level categories, agreement is $\kappa = 0.64$ for **Similarity**, $\kappa = 0.54$ for **Audience & purpose**, $\kappa = 0.48$ for **Design**, $\kappa = 0.38$ for **Narrative**, $\kappa = 0.34$ for **Experience**, $\kappa = 0.31$ for **Genre**, and $\kappa = 0.24$ for **Exposure**. Among the low-level codes with the highest agreement are **Setting** and **Characters** (from **Narrative**) with κ values of 0.68 and 0.57 respectively, **Release data** and **Hardware & platform** (from **Metadata**) with $\kappa = 0.65$ and $\kappa = 0.51$ respectively, and for **Design** the low-level codes **Graphics** ($\kappa = 0.55$) and **Perspective** ($\kappa = 0.53$). **Temporal structure** has a negligible agreement of $\kappa = 0.07$, signaling it is difficult to pick out of a game re-finding description. Finally, given the inclusion of gameplay mechanics in the subreddit template proposed by [/r/tipofmyjoystick](#), it is perhaps surprising that **Gameplay mechanics** shows almost no meaningful agreement with a *kappa* of 0.14. The code also elicited a fair amount of discussion in the consolidation phase, not in the least because Reddit users themselves also had difficulties describing these game aspects in their requests. Many users for instance, blended descriptions of genre and gameplay mechanics.

5 ANALYSIS

5.1 Complexity

The game re-finding requests posted to [/r/tipofmyjoystick](#) are often quite complex. In our sample of 250 posts, 7.1 different codes were assigned to a post on average, reflecting the richness that is also present in the game classification ontologies previously discussed

in Section 2.1 and contrasting with the more singular focus of web search behavior. In addition, [10] report a similar level of complexity for other domains, such as 7.4 different codes on average for books and 7.7 for movie requests. Figure 3 shows the distribution of the number of assigned codes per post, revealing an approximately normal distribution centered around the average of 7.1 low-level codes per post. Only six posts were assigned two low-level codes or less and only six posts were assigned more than 11 low-level codes. This shows that although the coding scheme we developed, with its 38 different codes, describes a relatively large set of possible game aspects, for the majority of the actual requests we coded, less than a fifth of the available codes were used in the coding of the individual requests. The highest number of low-level codes assigned to a single post was 16, and two posts were not assigned any codes at all, because they only contained a link to an image, which we excluded from our coding process as described in Section 3.2.3.

5.2 Frequency

Figure 4 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, with a handful of codes dominating the frequency distribution. Our analysis of code frequency is organized as follows: we first analyze our results in terms of the most frequent top-level codes (and their low-level codes), followed by the top-level codes that occur in less than half of our sample. In Section 5.2.3, we discuss the possible influence of that the suggested posting template of [/r/tipofmyjoystick](#) may have had on our results.

5.2.1 Most frequent codes. As Figure 4 shows, the four most frequently applied top-level codes were **Metadata** (89.2%), **Narrative** (83.2%), **Structure** (81.6%), and **Design** (70.8%). It is perhaps unsurprising to see **Structure** occurring in 81.6% of our coded posts if we compare our coding scheme with the typologies mentioned in Section 2.1. Structural aspects—such as game mechanics, temporal structure, number of players, and ludic space—are common elements in many of those typologies. It seems that these structural aspects are as important to game researchers as they are to players attempting to re-find lost games.

In contrast to the structure of video games, narrative aspects appear to be less prominent in the typologies by Debus [16], Elverdam and Aarseth [19], and Zagal et al. [48]. Nevertheless, **Narrative** is a common code category in our data set, which is understandable. Events that occur in a game and its characters in particular are explicit and easily identifiable aspects of game content and easily remembered by players. These aspects are also commonly remembered in other domains, such as books and movies [10, 13], 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. Furthermore, because of the re-finding context in which these game descriptions appear, users retrospectively describe the games they search for from the point of view of a finite game session in the past. As such, the descriptions of mechanics (e.g., player input and the game’s response) may often

⁴Agreement scores for all codes are available at ANONYMIZED URL.

Top-level code	Low-level code	Description
Design	Graphics	Descriptions relating to the visual appearance of particular graphical elements, as well as the more overall stylistic tradition they refer to.
	Interface	Descriptions relating to the visual interface elements that are not part of the game's diegetic universe, but rather provide the player with information about the game. It can also refer to the overall way in which the game is presented, such as 'point-and-click' or 'text-based'.
	Perspective	Descriptions relating to the perspective of the game space that the game affords the player (e.g., first-person view, third-person view, isometric, top-down perspective). It can also refer to so-called virtual cameras.
	Sound design	Descriptions that describe specific auditive qualities of the game, such as in-game music, sound effects, voice acting, or the way sound is used to create atmosphere.
	Soundtrack	Descriptions relating to a game's soundtrack. Examples include requests with information about soundtracks such as (named) songs, artist names or similar music.
Structure	Connectivity	Descriptions of a game that reference a specific type of connectivity, such as offline, online or LAN.
	Controls	Descriptions of a game by referencing the desired input devices or control settings (e.g., joystick, mouse, keyboard, VR, Kinect sensor) or level of difficulty with respect to controls.
	Expandability	Descriptions that address a games expandability, for instance through downloadable content (DLC) or map editors.
	Game mode	Descriptions relating to the game mode (e.g., single-player, multiplayer, cooperative (aka 'co-op'), turn-based, same-device (split-screen) vs. multiple devices, AI, ghost mode in racing games, etc.).
	Gameplay mechanics	Descriptions relating to the way(s) in which the game structures and govern the actions of the player. More colloquially speaking, it describes what the player can or should do in the game as well as how this is done.
	Ludic space	Descriptions relating to the spatial layout in which the game is placed. This code is used strictly for descriptions of the spatial structure in this space.
	Temporal structure	Descriptions that refer to the temporal structure of a game. Examples could include describing a game's episodic content, whether it features flashbacks, or whether the game deviates from a linear story structure.
Narrative	Characters	Descriptions relating to the NPCs as well as PC. Characters are everything that is described as having some sense of 'personhood', which in a fictional universe can apply to both humans and non-humans, even objects.
	Cutscene(s)	Descriptions relating to cut-scenes in the game, i.e., sequences in a video game that are not interactive and break up the gameplay. This code could also apply to memorable intro and title screens or tutorials.
	Dialogue	Descriptions of a game by referring to a particular line or style of dialogue, such as specific phrases or references to conversations or conversation styles within the game.
	Events	Descriptions that refer to one or more elements of a game's plot or events that take place during the game.
	Setting	Descriptions of games by referencing that it takes place in a specific location or near geographical landmarks as well as during a particular time period or around a specific historical event.
	World building	Descriptions of the backstory, fictional universe, a particular environment design or style of world building in the game. Examples include descriptions or requests for specific worlds/maps, world building styles or functions.
Genre	Genre	Descriptions of a game that reference which genre(s) it falls in (e.g., action-adventure, ego-shooter, racing, sport, simulation, strategy, etc.).

Figure 1: Overview of part one of our coding scheme along with descriptions of each of the 38 low-level codes.

be told in the form of a narrative event, and are therefore also more likely to be coded as such.

The distribution of sub-level codes confirms this pattern. As shown in Figure 4, 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, the codes **World building**, **Cutscene(s)** and **Dialogue** are much more infrequent (<10%). It seems that in particular characters and settings are aspects that are

easily remembered by players, and also easy to include in the relative short format of the Reddit requests. With regard to **Structure**, its low-level codes are even more rare. Here, the low-level codes **Gameplay mechanics** and **Ludic space** appear in 69.6% and 26.8% of our posts respectively, whereas the remaining low-level codes under **Structure** are much less common.

Metadata is another top-level code with an unexpectedly high frequency in our sample, given its absence from the aforementioned formal game typologies. It could be considered a relatively

Top-level code	Low-level code	Description
Metadata	Availability	Descriptions of a game by referring to its purchase cost or distribution model (e.g., shareware, free demo). This also covers free-to-play games that support in-app-purchases.
	Collection, series and/or franchise	Descriptions relating to a game collection, series or franchise which the particular title was a part of. It could also be used to describe games that are part of a specific collection, requests with explicit links to content in another domain, as well as descriptions relating to a specific version of the game, such as a port or a reboot.
	Contributor(s)	Descriptions of a game that refer to one or more of the contributors to that game. Examples include referring to a publisher, director, designer or developer, to voice actors or to specific creative contributions by other companies.
	Hardware & platform	Descriptions relating to physical or digital specifications of the system the game was played on. Examples include mentioning hardware or a specific device, operating system, platform.
	Language & country	Descriptions that refer to the language of the game or country in which it was produced Other examples could include referring to the fictional language(s) spoken in the game or whether the game allowed the user to switch to another language.
	Release date	Descriptions relating to the time (period) the game in question was released.
	Supplementary materials	Descriptions that reference related materials external to the game itself (e.g., booklets, jewel cases, game manuals, physical maps, packaging) as well as descriptions of promotion materials for the game.
	Title	Descriptions that mention (parts of) a title or memories of title aspects.
Audience & purpose	Audience	Descriptions that refer the specific audience the game was aimed at (e.g., suitable for kids, hardcore gamers).
	Extrinsic purpose	Descriptions relating to the extrinsic purpose of the game, typically used for educational games.
Experience	Accessibility	Descriptions that reference the skill level or coordination required to play the game, such as different levels of difficulty or its learning curve.
	Engagement	Descriptions that refer to the player's engagement with the game, either first-hand or second-hand. This could cover descriptions of how long/often they played the game themselves as well as watching a friend or Youtuber play a game.
	Impact	Descriptions referring to the impact the game had on the requester or how it motivated them. Examples include requests for games that change the searcher's view or behavior on certain issues like gender, environment, economy, games that motivate the searcher to learn something new or be more active in certain areas.
	Mood	Descriptions of a game that refer to the mood, tone or gaming experience it evoked.
	Novelty	Descriptions that mention a game's quirky, unusual or novel nature. Examples include requests for a game that is out of the ordinary or that promoted perspectives on issues that would be novel to the requester.
	Popularity	Descriptions that refer to a game's (lack of) popularity or well-known it is.
	(Re-)play value	Descriptions that reference a games (re-)play value, duration, longevity, complexity, and/or variety.
Exposure	Context of exposure	Descriptions that refer to the place or context in which the game was played or first encountered. It is also used in cases where instead of the release date the time when it was (first) played is described instead.
Similarity	Similarity	Descriptions that explicitly refer to other games this game was (not) similar to in some respect (such as genre, plot, experience, mechanics etc.).

Figure 2: Overview of part two of our coding scheme along with descriptions of each of the 38 low-level codes.

insignificant element of game descriptions, as it does not describe the actual content of a game itself. However, much like with the low-level codes **Characters** and **Setting**, **Metadata** includes aspects that may simply be more easily remembered by players, such as the platform they played the game on. In addition, game metadata, such as developer or release date, may also be especially useful in the re-finding process, as users may use them, for instance, to look up all the games from a particular year or developer. If we look at distribution of the low-level codes in Figure 4, we see that the two most-frequently applied codes are **Hardware & platform**

(86.0%) and **Release date** (63.2%), with **Hardware & platform** in fact being the single-most frequently used low-level code in our entire data set. The remainder of low-level **Metadata** codes were applied relatively infrequently.

When looking at the distribution of low-level **Design** codes, it was the **Graphics** of a game that were most commonly described by users, with a relative frequency of 54.4%, followed by **Perspective** (24.4%) and **Interface** (14.0%). The least-used codes in **Design** were **Sound design** (7.2%) and **Soundtrack** (2.4%). It is important to note here that if a game was described as being in 3D or 2D, this was

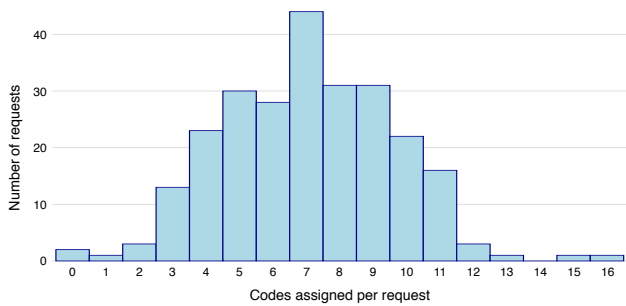


Figure 3: Histogram of code distribution in our sample of annotated Reddit threads ($N = 250$). The average number of codes assigned to a game re-finding requests was 7.1.

coded as **Graphics** and if a description mentioned a game’s first-person or third-person perspective, this was coded as **Perspective**. This could explain why **Design** occurs as often as it does, due to popular game-tags or genre descriptions, such as “*First-person shooter*” or “*2D platformer*”.

5.2.2 Less frequent codes. 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, as shown in Figures 1 and 2, although they are all fairly infrequent, occurring in only 0.4% to 8.4% of all posts.

With regard to **Genre**, it should be noted that genre is a somewhat contested concept among game studies researchers [5, 14, 28, 44] Nevertheless, it is unsurprising that **Genre** 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**. After all, this code describes a somewhat idiosyncratic aspect of the game, that at first sight may not seem immediately useful in a community-based search. For example, many examples of **Exposure** posts focused on the setting or context in which a game was played, such as “*An old PC game I played in elementary school...*” and “*...Played this at someone’s house as a kid and it was amazing...*”. 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, similar (or dissimilar) to the game that the user wishes to re-find. It should be noted, though, that it was not always straightforward to deduce from the description in what way(s) another mentioned game was similar to the one the user was trying to re-find. Sometimes the similarity was simply just stated, such as in “*...kinda like a mix of Digimon and Pokémon*”, whereas in other instances it was more clear in what way the games were similar, for example “*...i remember*

there was a merchant character that said things in a similar manner to the merchant from resident evil 4...”. The frequency of **Similarity** shows that it is an important aspect of game re-finding requests. However, as the notion of similarity in itself is very vague, and does not define in what sense two (or more) objects are similar, it would be far from trivial to operationalize as part of a retrieval algorithm for a game search engine.

Audience & purpose is the least-frequent code with a relative frequency of only 7.6 % for the top-level code, and 5.6% for its low-level code **Audience** and 3.6% for **Extrinsic purpose**. Most of the cases where **Audience** was applied, are descriptions of games as being a game for children. Only in a few cases was it used for descriptions of games rated ‘Mature’ and similar classifications. However, it is not surprising that users are attempting to re-find games they played in their childhood, given the main purpose of the [/r/tipofmyjoystick](#) subreddit.

5.2.3 Influence of the posting template. The categories specified in the template provided by the [/r/tipofmyjoystick](#) admins overlap with our coding scheme, although with its 10 top-level codes and 38 low-level codes, our coding scheme is considerably richer and more nuanced. However, even though users do not necessarily limit themselves to the template— or even follow it altogether—the relative frequency of our top-level and low-level codes does reflect that many users take the template as a starting point when formulating their re-finding requests. As such, some of the most-frequently used codes in our coding scheme, in one way or the other, overlap with this template. For example **Hardware & platform** and **Release date** coincide with the categories “Platform(s)” and “Estimated release date”, which is reflected in their high relative frequency. Similarly, **Gameplay mechanics**, **Characters**, **Genre**, and **Design** overlap with “Notable gameplay mechanics”, “Notable characters”, “Genre”, and “Graphics/Art” in the template. 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.

5.3 Co-occurrence

In addition to analyzing how often the individual codes occur in our data set, it could also be interesting to examine which codes *co-occur* with each other. To this end, we calculated Cramer’s V —a correlation coefficient for nominal data—on all pairs of top-level codes and all pairs of low-level codes. The resulting correlation matrix for the top-level codes is shown in Figure 5. Our hypothesis was that we might be able to detect strong correlations between certain codes, such as **Genre** and **Perspective**, **Interface** or **Gameplay mechanics**, due to the fact that many users employ popular genre tags such as “*first-person shooter*”, “*point-and-click adventure*”, and so forth. One could also expect to see a strong correlation between **Extrinsic purpose** and **Context of exposure**, for instance, as users could be expected to include the school context in which they played an educational game.

However, as Figure 5 shows, there are no strong correlations between the top-level codes given the maximum Cramer’s V value

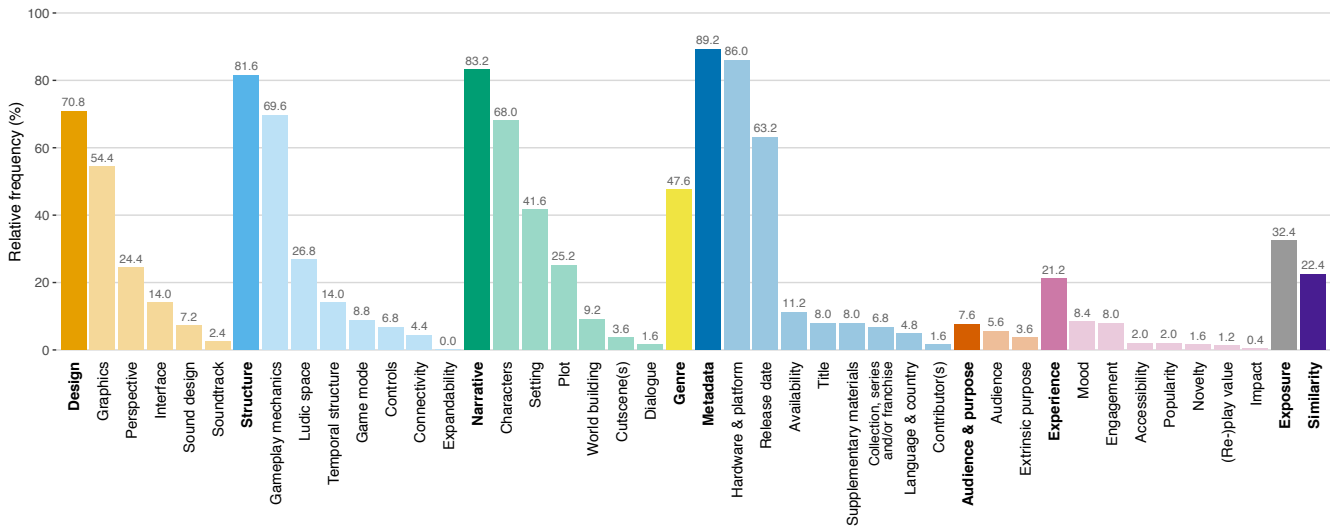


Figure 4: Relative frequencies of the 38 different codes in our sample of annotated Reddit threads ($N = 250$). The dark bars in each color group represents the top-level code (e.g., **Design**) and any lighter-colored bars to their right represent the low-level codes (e.g., **Graphics**, **Perspective**).

of 0.28, which is commonly interpreted as negligible. The same was true for the correlations between the low-level codes, which we do not visualize here. The code pairs with the highest correlation values include **Metadata–Narrative** ($V = 0.28$), **Genre–Design** ($V = 0.27$), **Structure–Design** ($V = 0.26$), and **Metadata–Genre** ($V = 0.24$). The fact that they rise above the other pairs is most likely due to the fact they are among the most frequent top-level codes due to their inclusion in the posting template and

by virtue of this are more likely to co-occur. The code pairs that do not appear to be correlated at all include **Audience & purpose–Narrative**, **Audience & purpose–Metadata**, and **Audience & purpose–Exposure**, which all had a V value of 0.01.

6 DISCUSSION & CONCLUSIONS

In the work presented in this paper, we focused on the analysis of game re-finding requests posted to a dedicated Reddit community called [/r/tipofmyjoystick](#) with the aim of better understanding which aspects of games users remember and describe in their posts. As expected, we found that many requests were centered around describing the content of the game: their structure, narrative, and design. Even more frequent than content were descriptions of game metadata. Descriptions of the experience of playing the game were mentioned least frequently in these re-finding requests, which is unsurprising given the context of the posts and the personal and subjective nature of such experiences—and subsequent lack of value in identifying the game in question. Other codes occur relatively infrequently in our sample of 250 requests, but our aim was to capture as much variety as possible; a larger follow-up study could better determine the relative frequencies of the different codes.

Our findings open up for many interesting directions for future work. For instance, we did not analyze the demographic characteristics of the users of [/r/tipofmyjoystick](#), but instead assumed that they are representative of the entire population of video game players. However, there may be good reasons for testing this assumption in future work. As several scholars have pointed out, gender norms are an important factor in gaming practices [40, 43, 45], and it might therefore also be worthwhile to analyze the gender distribution among [/r/tipofmyjoystick](#) users. To be clear, we are not hypothesizing that gender straightforwardly determines how we relate to game objects, remember them and describe them to others. Nevertheless, throughout the history of video games there have been

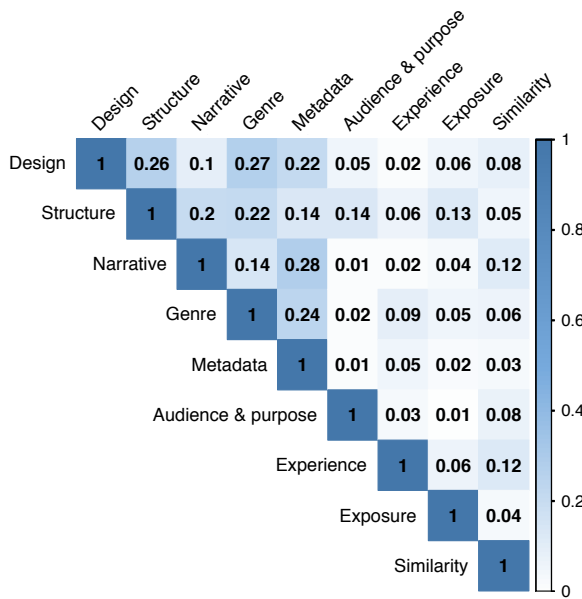


Figure 5: Correlation matrix of Cramer’s V correlations between the nine different top-level codes ($N = 250$).

various attempts at designing ‘girl games’, based on the assumption that female players have gaming preferences that were not being met by the video games available on the market [25]. It is therefore not unreasonable to speculate about whether requests for such games look different from requests for other games, both because there may be subtle differences in which game aspects are described as well differences between the games themselves. A better understanding of the (lack of) homogeneity of the [/r/tipofmyjoystick](#) community 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. In order to build inclusive search engines for games or for more specific archival purposes, such information could be valuable.

This raises an additional question. In the present study, we have not annotated the historical period (e.g., the decade) in which the games were released. However, design conventions, distribution and business models, as well as available technologies change over time. For example, with the emergence of CD-ROM technology, game developers increased the number of cut-scenes they included in their games [42]. One could, therefore hypothesize that this might also be observable from the frequency of the low-level code **Cutscene(s)** when comparing requests decade by decade. A similar thing could happen with the **Exposure** code: while this aspect is already relatively frequently mentioned in our data set, we might be able to detect changes in its frequency over time. For instance, streaming gameplay live on Twitch has become highly popular in recent years. This might result in more requests for re-finding games that the user has not played themselves, but instead watched other people play. This may make it necessary in the future to include a more fine-grained distinction at the level of **Exposure** and **Engagement** than what is currently covered by our coding scheme.

All of the above issues reinforce that building search engines for (re-finding) games is a highly complex task and requires a more detailed understanding of general search behavior with regard to games. To achieve this, we should not only take into account the complexities of current games, but also past design conventions and game distribution patterns. There may also be demographic patterns in the aspects of games that a user deems relevant in a re-finding request. This 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. An investigation of the factors that influence the success or failure of game re-finding requests would be just as valuable. After requests are posted to [/r/tipofmyjoystick](#), other Reddit users often attempt to help identify the game in question, after which the original poster can mark the problem as solved. Is describing certain (combinations of) aspects more conducive to successfully solving a game re-finding request? Does mentioning more aspects or simply writing down as many details as possible in the form of longer post texts increase the chances of success? The dynamics of this collaborative problem solving process, and the role that our codes play in this, needs to be studied in more detail so as to optimize the search process for game re-finding specifically and for game search in general.

REFERENCES

- [1] Espen Aarseth and Pawel Grabarczyk. 2018. An Ontological Meta-Model for Game Research. In *Proceedings of the Digital Games Research Association Conference*.
- [2] Espen Aarseth, Solveig Marie Smedstad, and Lise Sunnanå. 2003. A multidimensional typology of games.. In *DiGRA Conference*.
- [3] Espen J. Aarseth. 2011. *Define real, Moron!: Some remarks on game ontologies*. DIGAREC Series, Vol. Vol. 6. Universität Potsdam, 50–68.
- [4] Barrett R Anderson and Adam M Smith. 2019. Understanding User Needs in Videogame Moment Retrieval. In *FDG '19: Proceedings of the 14th International Conference on the Foundations of Digital Games*. 1–10.
- [5] Thomas H. Apperley. 2006. Genre and game studies: Toward a critical approach to video game genres. *Simulation Gaming* 37, 1 (2006), 6–23.
- [6] Alasdair Bachell and Matthew Barr. 2014. Video game preservation in the UK: A survey of records management practices. (2014).
- [7] D. Barreau and B. A. Nardi. 1995. Finding and Reminding: File Organization from the Desktop. *SIGCHI Bull.* 27, 3 (1995), 39–43.
- [8] R. Boardman and M. A. Sasse. 2004. "Stuff Goes into the Computer and Doesn't Come out": A Cross-tool Study of Personal Information Management. In *Proc. of CHI'04*. ACM, New York, NY, USA, 583–590.
- [9] Toine Bogers. 2015. Searching for Movies: An Exploratory Analysis of Movie-related Information Needs. In *Proceedings of iConference 2015*. iDEALS.
- [10] 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 *Proc iConf 2018*. Vol. 10766. Springer, 323–334.
- [11] 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.
- [12] 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 the 2019 iConference*. Springer, 503–515.
- [13] Toine Bogers and Vivien Petras. 2015. Tagging vs. Controlled Vocabulary: Which is More Helpful for Book Search?. In *Proceedings of iConference 2015*. iDEALS.
- [14] 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.
- [15] 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).
- [16] Michael Sebastian Debus. 2019. *Unifying Game Ontology: A Faceted Classification of Game Elements*. Ph.D. Dissertation. The IT University, Copenhagen, Denmark.
- [17] David Elswailer, Mark Baillie, and Ian Ruthven. 2008. Exploring memory in email re-finding. *TOIS* 26, 4 (2008), 21.
- [18] David Elswailer, Mark Baillie, and Ian Ruthven. 2011. What Makes Re-finding Information Difficult? A Study of Email Re-finding. In *Advances in Information Retrieval*, Paul Clough, Colum Foley, Cathal Gurrin, Gareth J. F. Jones, Wessel Kraaij, Hyowon Lee, and Vanessa Mudoch (Eds.). Springer Berlin Heidelberg, 568–579.
- [19] Christian Elverdam and Espen Aarseth. 2007. Game classification and game design: Construction through critical analysis. *Games and Culture* 2, 1 (2007), 3–22.
- [20] 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.
- [21] Paul Gooding and Melissa Terras. 2008. 'Grand theft archive': a quantitative analysis of the state of computer game preservation. (2008).
- [22] Pawel Grabarczyk and Espen J. Aarseth. 2019. Port or Conversion? An Ontological Framework for Classifying Game Versions.. In *DiGRA '19: Proceedings of the 2019 DiGRA International Conference: Game, Play and the Emerging Ludo-Mix*.
- [23] Lindsay D. Grace. 2014. A linguistic analysis of mobile games: Verbs and nouns for content estimation. *Proc. FDG* (2014).
- [24] Matthias Hagen, Daniel Wäger, and Benno Stein. 2015. A Corpus of Realistic Known-Item Topics with Associated Web Pages in the ClueWeb09. In *ECIR '15: Proceedings of the 37th European Conference on Information Retrieval*. Springer, 513–525.
- [25] Yasmin B. Kafai, Carrie Heeter, Jill Denner, and Jennifer Y. Sun. 2008. Preface: Pink, purple, casual, or mainstream games: Moving beyond the gender divide. *Beyond Barbie and Mortal Combat. New Perspectives on Gender and Gaming*. The MIT Press: London (2008).
- [26] Lars Konzack. 2002. Computer Game Criticism: A Method for Computer Game Analysis.. In *CGDC Conf*.
- [27] 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.

- [28] Jin Ha Lee, Natascha Karlova, Rachel Ivy Clarke, Katherine Thornton, and Andrew Perti. 2014. Facet analysis of video game genres. *ICConference 2014 Proceedings* (2014).
- [29] Jin Ha Lee, Joseph T. Tennis, Rachel Ivy Clarke, and Michael Carpenter. 2013. Developing a video game metadata schema for the Seattle Interactive Media Museum. *International journal on digital libraries* 13, 2 (2013), 105–117.
- [30] 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.
- [31] Henry Lowood. 2009. Game capture: The machinima archive and the history of digital games. *Mediascape: UCLA's Journal of Cinema and Media Studies* (2009).
- [32] Thomas W. Malone. 1983. How Do People Organize Their Desks?: Implications for the Design of Office Information Systems. *TOIS* 1, 1 (1983), 99–112.
- [33] Mary L McHugh. 2012. Interrater Reliability: The Kappa Statistic. *Biochemia Medica* 22, 3 (2012), 276–282.
- [34] Florian Meier and David Elsweller. 2016. Going Back in Time: An Investigation of Social Media Re-finding. In *Proc. of SIGIR'16*. ACM, New York, NY, USA, 355–364.
- [35] Benjamin Nicoll, Melanie Swalwell, Ndalians Angela, and Helen Stuckey. 2017. *Sega Saturn Fan Sites and the Vernacular Curation of Videogame History*. 180.
- [36] James Owen Ryan, Eric Kaltman, Michael Mateas, and Noah Wardrip-Fruin. 2015. What we talk about when we talk about games: Bottom-up game studies using natural language processing. *Proc. FDG* (2015).
- [37] Adrienne Shaw. 2017. What's next?: the LGBTQ video game archive. *Critical Studies in Media Communication* 34, 1 (2017), 88–94.
- [38] Julius Sim and Chris C. Wright. 2005. The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy* 85, 3 (2005), 257–268.
- [39] 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: Matters of Life and Death*. 1–7.
- [40] T. L. Taylor. 2009. The Assemblage of Play. *Games and Culture* 4, 4 (2009), 331–339.
- [41] J. Teevan, E. Adar, R. Jones, and M. A. S. Potts. 2007. Information Re-retrieval: Repeat Queries in Yahoo's Logs. In *Proceedings of SIGIR '07*. ACM, New York, NY, USA, 151–158.
- [42] Carl Therrien. 2008. *CD-Rom Games*. Greenwood Press Westport, 121–134.
- [43] Lotte Vermeulen, Jan Van Looy, Frederik De Grove, and Cédric Courtois. 2011. You are what you Play?: A Quantitative Study into Game Design Preferences across Gender and their Interaction with Gaming Habits. In *DiGRA 2011: Think, design, play*. Digital Games Research Association (DiGRA).
- [44] 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.
- [45] Dmitri Williams, Mia Consalvo, Scott Caplan, and Nick Yee. 2009. Looking for Gender: Gender Roles and Behaviors among Online Gamers. *Journal of Communication* 59, 4 (2009), 700–725.
- [46] 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.
- [47] 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.
- [48] 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.
- [49] Jose P. Zagal and Noriko Tomuro. 2010. The aesthetics of gameplay: a lexical approach. In *Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments*. 9–16.
- [50] 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.