

# GameGoogle

## A Search Engine for Mechanics in Video Games

Brooke Szajda<sup>1</sup>, M Charity<sup>1</sup>

<sup>1</sup>University of Richmond  
brooke.szajda@richmond.edu, mcharit2@richmond.edu

### Abstract

This paper introduces a new search engine for finding game mechanics internal to a game. The system utilizes a custom made database containing extracted gameplay information from openly available game walkthroughs, manuals, and guides along with general information about the game. This information is displayed as an interactive web application that displays relevant games to the user based on keyword queries. We conduct a preliminary user study to evaluate the relevancy of the results for the system and discuss future directions and feature implementations.

### Introduction

When searching for a game, whether through Google, ChatGPT, online game databases, or even by asking a friend, often the results lead to games matching the genre or overall gameplay of the query. For example, searching for “basketball” in video games will return many if not all of the NBA series games and other sports games such as Mario Hoops 3-on-3, Backyard Basketball, or Wii Sports Resorts. However, if someone such as a game developer, a games scholar, or a games educator (Anderson and Smith 2019) were looking for smaller and simpler examples of a basketball mechanic in games that were not necessarily the focus of the gameplay, the search then becomes more challenging. Currently, there are no search engines for games that can search within the content of the gameplay. Games that feature a “basketball” minigame mechanic or easter egg – such as in games like LittleBigPlanet, Schedule 1, or Thief: The Dark Project – may not come up as a result in traditional search engines but are nonetheless relevant.

This paper introduces a prototype search engine for video games called GameGoogle (GG for short.) GG searches through a game’s entire gameplay space based on a keyword query and returns any relevant game matches. The system uses video game guides, walkthroughs, and basic information about the games that are tokenized from various online sources to retrieve the most relevant games. GG is intended to be used by game scholars, developers, and educators as a quick and easy to use reference site for game content. We conducted a small user study to evaluate the relevancy of the

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.



Figure 1: Top row – basketball video games: Mario Hoops 3-on-3, Street Hoops. Bottom row – basketball *in* video games: LittleBigPlanet2, Thief: The Dark Project

results returned and discuss future feature implementations for the system.

## Background and Related Works

### Game Data

Games as a multimedia entity can be represented or vectorized in a variety of ways – either for searching, comparison, or content generation.

Games databases such as MobyGames<sup>1</sup>, Internet Game Database (IGDB)<sup>2</sup>, and the Launchbox Games Database<sup>3</sup> all have a comprehensive database of games and their general information including release year, the system it was released on, and summaries of the game. While these databases and their websites often have search engines themselves, they do not include in-depth details about the gameplay and offer only general information and similarities to other games.

Websites, community groups, and social media accounts

<sup>1</sup><https://www.mobygames.com/>

<sup>2</sup><https://www.igdb.com/>

<sup>3</sup><https://gamesdb.launchbox-app.com/>

can also leverage user submitted data about a game to identify specific aspects of gameplay. Howlongtobeat.com<sup>4</sup> provides information about a estimated game’s length of play-time and sorts based on main story, extras, and completionist categories. ‘The Woke Ranking’<sup>5</sup> – a search engine that leverages data collected from a Steam community – uses user submitted data to identify ‘wokeness’ – typically content that includes diversity, equity, and inclusion – in games. The Twitter account ‘@CanYouPetTheDog’<sup>6</sup> identifies whether the player can or cannot pet a dog NPC in game. These communities offer a wide range of games in the search corpus but focus on only a specific game mechanic.

## Game Search Engines

In the research community, there have been many systems made with the intent to search within games for a gameplay moment or for a specific genre or tag via a text query.

The game tools developed by Ryan et. al. – GameGlobs, GameTree, GameNet, GameSage, and GameSpace – all visualize related games using an input query based on genre and content, but focus more on exploration of game similarity at a overall, summative perspective rather than at a detailed level within the game. (Ryan et al. 2015a,b, 2016, 2017) These systems miss out on smaller mechanics within the game that may still be relevant to a search query.

Anderson, Zhang, Zhan, and Smith, all utilized search engine techniques of crawling, indexing, and searching to retrieve video game “moments” – vectorized representations of game states processed for querying and retrieval. These moments are gathered from emulator states, user playthroughs (either through speedruns or 15-minute sessions,) and Youtube gameplay videos. These systems used computer vision to analyze and annotate frames of the games along with cosine similarity to compare similar moments to an input image query. However, these experiments were limited to searching only within a few games (Super Mario, Super Metroid, Life is Strange, and the Last of Us.) (Anderson 2020; Zhang and Smith 2019; Zhan and Smith 2018)

As far as we know, there currently exist no search engines or user-facing applications that can search within the *entire* range of gameplay across multiple games for specific gameplay mechanics or content.

## System Design

The GameGoogle system is made up of a custom-made game database and a front-facing web search engine showing the details of resulting games from a query. The following subsections detail the development of the system.

### Creating the Dataset

While game databases such as IGDB and GameNet exist and provide substantive details on a wide corpus of games, they were all missing the internal gameplay details needed to make the search for the GameGoogle system more effective.

<sup>4</sup><https://howlongtobeat.com/>

<sup>5</sup><https://wokedetector.cirnoslab.me/full-list>

<sup>6</sup><https://x.com/canyoupetthedog>

Therefore, we decided to build our database from scratch; combining the game information from other openly available online databases and sources. As a proof of concept, we limited the database to the first 50 top rated games from the last 50 years (1975 - 2025.) The list of games was extracted from the user rated game website GlitchWave.com<sup>7</sup>. We used this list of games to search for the guides and information about the game across other databases and websites.

To extract the overall gameplay data for a game and find the range of mechanics available, we processed game manuals as well as user-made game guides and walkthroughs from the following websites: Vimm’s Lair<sup>8</sup>, Steam Community pages<sup>9</sup>, IGN<sup>10</sup>, and GameFAQ<sup>11</sup>. Many of these guides were already displayed as text on the website page, while other guides such as from Vimm’s Lair required image-to-text processing from the manuals. We tokenized the manual, guide, and walkthrough text data using the NLTK and SpaCy libraries. Since many game mechanics are described based on a combination of nouns and verbs and because we expected users would want to query a game based on game elements, we extracted only the nouns and verbs for each game (Sicart 2008). We also retrieved information from game databases such as IGDB, the Launchbox Games Database, and Steam to provide more general details about the games shown (e.g. the game’s system, game summary, and cover image.)

Unfortunately, we were not able to consistently retrieve data for all of the games from our list due to inconsistencies with naming across the different websites. Examples of these include the use of roman numerals vs. numbers, original games vs. re-releases or ports of the same game, game releases on multiple systems that may have different mechanics or gameplay, and games with the same names but different publishers and release times. There were also a number of games where entries were found at one source but not another – particularly for the older games in the dataset. As such, our system’s dataset is incomplete and has some minor inaccuracies that we hope to remedy in future iterations.

### Games Database

Unlike most search engines, GG does not retrieve game information through crawling. This system only uses the indexing and searching steps of search engine design (Seymour et al. 2011). The data to query is stored in a MySQL database on an Amazon Web Services cloud instance. The website is also hosted from this same instance to allow faster querying and retrieval from the database.

Two tables in the database are used for quick querying: a table containing all of the information about the games, *gameData*, and an inverse word lookup database, *inverseGameWords*. The columns for the *gameData* table are listed in Table 1 while the *inverseGameWords* table acts as an inverted index table. For *inverseGameWords*, the table

<sup>7</sup><https://glitchwave.com/charts/top/game/>

<sup>8</sup><https://vimm.net/>

<sup>9</sup><https://steamcommunity.com/>

<sup>10</sup><https://www.ign.com/wikis>

<sup>11</sup><https://gamefaqs.gamespot.com/>

Column	Description
<i>ID</i>	Unique identifier for the game
<i>Game Title</i>	Title of the game (without punctuation and upper-cased)
<i>Game System</i>	Console system the game was released on
<i>Game Year</i>	Year the game was released
<i>Description</i>	General summary of the game
<i>Cover Image</i>	Base64 representation of the game’s cover image
<i>Rating</i>	User voted rating for the game
<i>Genres</i>	List of genres associated with the game
<i>Tags</i>	List of tags associated with the game
<i>Word Count</i>	Combined frequency of the tokenized words that occur in the walkthroughs, guides, and manuals of the game
<i>Sources</i>	Databases and websites that the game information was extracted from

Table 1: Table for the game entries of the GameGoogle system. The data for each game is populated from extracted text and databases

keys are the individual words while the values are the list of games containing the word in its ‘word count’ column value. This allows the system to quickly query for a keyword and find associated games. For example, if the user were to type in “apple” in the search bar of the website, the server would first look up the word apple as a key in the *inverseGameWords* table. If the entry exists, it will return the full list of associated games (e.g. Cooking Mama, Animal Crossing, Minecraft) stored with that entry. Using those games, the server will retrieve the game’s information from the gameData table and rank the resulting games according to the user specified filter (see the following subsection.) Once they have been sorted, the server will return the sorted list of games and the necessary information (release year, system, cover image, title) to the client’s web page. In total the database has nearly 4500 game entries and over 120,000 keywords available to query.

## Website Design

For the website design, we display search results with a variety of query ranking options. Users can sort by exact keyword matches, title, release date, game rating, and a default ranking defined by a custom equation. The website also has placeholders for future implementations such as an image gallery to show the game element queried in context of the game. Figure 2 shows a screenshot of the GG website with the search results for “basketball.”

For all filters in the system (except “exact match”) games that contain any keyword from the query are included in the



Figure 2: Capture of GameGoogle search results for the keyword ‘basketball’.

output results shown to the user. With the “Exact match” filter, only games that have all of the keywords from the query are used. The default ranking score for a game  $i$  is calculated with the following equation:

$$R_i = 0.5(K_i) + 0.2(T_i) + 0.15(G_i) + 0.15(A_i) \quad (1)$$

where  $R_i$  is the game rank value.  $K_i$  is the frequency total of each keyword from the query over the total frequency of all of the words associated with the game.  $T_i$ ,  $G_i$ , and  $A_i$  are the percentage of words in the game’s title, genre list, and tag list respectively that match the keywords.

The GameGoogle system is available at the following link: <http://18.233.50.163:3000/>. Each returned result has a grid of images associated that currently shows ‘No Image found.’ For future work, we hope to incorporate examples and screenshots of the associated query keyword as it appears in the game.

## System Evaluation

While the intention of the system is to discover mechanics in games that would be overlooked in regular search engine results, it is important that the system is still able to identify games that are extremely relevant to a particular mechanic. For example, it would be confusing to the user if

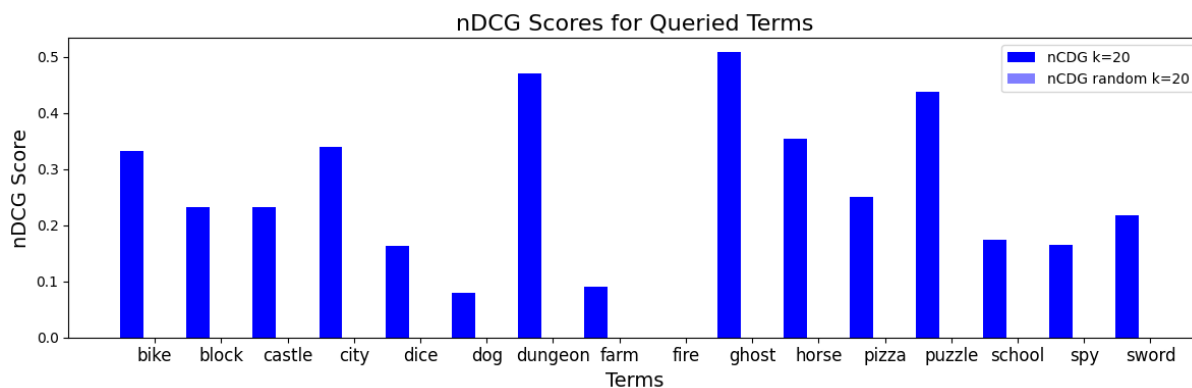


Figure 3: Results of the user study games vs. the returned games from GameGoogle using nDCG

they searched "basketball" and no basketball video games were shown. We conducted a user study asking participants to list the first game they thought of when they heard a specific word. This study was collected using Google Forms and spread via social media posts, university newsletters, and flyers. We collected a total of 60 responses from participants.

We asked users to name games for 16 different words: *bike*, *block*, *castle*, *city*, *dice*, *dog*, *dungeon*, *farm*, *fire*, *ghost*, *horse*, *pizza*, *puzzle*, *school*, *spy*, and *sword*. Like Zhang and Smith (Zhang and Smith 2019) we used the Normalized Discounted Cumulative Gain (nDCG) to evaluate the relevance and ranked quality of the returned results from the GG system. Our true relevance scores were based on the frequency of a game being mentioned for a word in our user study. Before calculating, we cleaned the user responses by editing spelling mistakes and combining games from a series into one term. For example, if participants said "The Legend of Zelda: Ocarina of Time", "Twilight Princess", or "zelda games" for their answer to *sword*, all 3 responses would count towards "zelda". If any game from the The Legend of Zelda series appeared in the system results (e.g. Breath of the Wild), it would be counted for "zelda". A random baseline was also used as comparison for each word which returned a randomly chosen subset of games from the entire GG dataset.

Figure 3 shows the nDCG scores for the 16 words based on the user results based on a  $k$  value of 20 (i.e. the first page of results from the system.) Of note, none of the randomly selected games from the over 4000 possible games in the list appeared, resulting in a nDCG score of 0. The word *fire* returned no matching games – most likely because the verb and noun forms of the word "fire" were included for the games dataset. Many older games (such as arcade games and retro console games) included the verb form in their walkthroughs and manuals (e.g. "press the trigger button to fire"), making these games take precedence in the search results. Dungeon and ghost returned the highest number of matches out of the words, most likely because of the "Zelda" game series and the "Ghosts n Goblins" game series being popular responses from the user studies for their respective terms.

## Future Work and Features

To build on GG, additional information like gameplay images and obscure game titles are features to consider. While GG works in the context of larger and more popular games, including games from evolving platforms like IOS and niche titles from sites like Itch, Newgrounds, and Roblox would enhance the current data set and provide more comprehensive results. Previous works have looked to catalog and search within indie games as well, however these have still only been done at the general level of gameplay and used to compare to triple A and more well-known titles (Anderson 2020; Vu 2020; Vu and Bezemer 2021). Many games of this nature don't release official manuals or have thorough walkthroughs. Including these titles would involve compiling game descriptions and player review data directly from the publishing site and from community forum sites like Fandom and GameFAQs if available.

Additionally, including screenshots of in-game moments would bridge the gap between the game title and the game moment. The goal is to be able to search "bike" and display a list of games including bikes and screenshots of those specific moments in the game. To source these images, we've collected screenshots from longplay videos on Youtube and intend to train a contrastive learning model and CNN model to interpret these images accurately – similar to the work done by Trivedi et. al and Breve (Trivedi, Liapis, and Yannakakis 2021; Breve 2025).

An additional future metric to enhance GG could involve testing its effectiveness in identifying games from the Reddit community 'Tip of my Joystick' (Jørgensen and Bogers 2020). This community helps identify forgotten games based on context provided by users. Utilizing GG to potentially aid in identifying these games could be a good metric to test the success of search results.

The prototype built thus far as well as the results of the user study has provided many insights into future improvements for the system overall. We look forward to adding more games and detailed information to the system to make game discovery much quicker and useful for designers, researchers, and educators.

## References

- Anderson, B. 2020. *Searching Within and Across Games*. University of California, Santa Cruz.
- Anderson, B. R.; and Smith, A. M. 2019. Understanding user needs in videogame moment retrieval. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*, 1–10.
- Breve, F. 2025. From Pixels to Titles: Video Game Identification by Screenshots using Convolutional Neural Networks. *IEEE Transactions on Games*.
- Jørgensen, I. K. H.; and Bogers, T. 2020. “Kinda like The Sims... But with ghosts?”: A qualitative analysis of video game re-finding requests on Reddit. In *Proceedings of the 15th International Conference on the Foundations of Digital Games*, 1–4.
- Ryan, J.; Kaltman, E.; Fisher, A. M.; Owen-Milner, T.; Mateas, M.; and WARDRIP-FRUIIN, N. 2017. GameSpace: An explorable visualization of the videogame medium. *Univ. California Santa Cruz, Santa Cruz, CA, USA, UCSC-SOE-17-14*.
- Ryan, J.; Kaltman, E.; Hong, T.; Isbister, K.; Mateas, M.; and Wardrip-Fruin, N. 2016. GameNet and GameSage: Videogame discovery as design insight. In *Proceedings of DiGRA/FDG 2016 Conference*.
- Ryan, J. O.; Kaltman, E.; Fisher, A. M.; Hong, T.; Owen-Milner, T.; Mateas, M.; and Wardrip-Fruin, N. 2015a. Large-scale interactive visualizations of nearly 12,000 digital games. *Proc. Foundations of Digital Games*.
- Ryan, J. O.; Kaltman, E.; Mateas, M.; and Wardrip-Fruin, N. 2015b. What we talk about when we talk about games: Bottom-up game studies using natural language processing. *Proc. FDG*.
- Seymour, T.; Frantsvog, D.; Kumar, S.; et al. 2011. History of search engines. *International Journal of Management & Information Systems (IJMIS)*, 15(4): 47–58.
- Sicart, M. 2008. Defining game mechanics. *Game studies*, 8(2): 1–14.
- Trivedi, C.; Liapis, A.; and Yannakakis, G. N. 2021. Contrastive learning of generalized game representations. In *2021 IEEE Conference on Games (CoG)*, 1–8. IEEE.
- Vu, N. Q. 2020. Leveraging Data From the Itch. io Online Game Distribution Platform to Help Indie Game Developers.
- Vu, N. Q.; and Bezemer, C.-P. 2021. Improving the Discoverability of Indie Games by Leveraging their Similarity to Top-Selling Games: Identifying Important Requirements of a Recommender System. In *Proceedings of the 16th International Conference on the Foundations of Digital Games*, 1–12.
- Zhan, Z.; and Smith, A. M. 2018. Retrieving Game States with Moment Vectors. In *AAAI Workshops*, 586–592.
- Zhang, X.; and Smith, A. M. 2019. Retrieving videogame moments with natural language queries. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*, 1–7.