AnimAtlas

CSC-475: Seminar in Computer Science
Fall 2023

Sophie Ngo
Department of Computer Science
Furman University
sngo@furman.edu

Abstract

Paleontology is a field that the general public is not commonly familiar with. Efforts to improve education of paleontology is therefore important for improving the public's knowledge and respect for the field. Paleontology has notable impacts in understanding climate change, reducing the spread of misinformation, and raising awareness to issues in the field such as the commercialization of fossil selling. AnimAtlas is a website that aims to make paleontology education more accessible to the general public. It is a website application featuring a map that can be used to interactively explore fossil information sourced from the Paleobiology Database, a well-known collection of recorded fossil information. Similar map visualization applications are often inaccessible to the layperson as they are often intended to be used by professionals that are familiar with complex concepts, so AnimAtlas also aims to simplify data exploration while still making learning interesting. The website, accessible at https://animatlas.me/, features an interactive map and sidebar that prioritizes a smooth user experience via a simple menu and concise communication of complex information. Users can also access different visualizations of the data, such as a heatmap showing fossil abundance, to expand the options of learning about fossils.

1 Introduction

The goal of this project was to create an educational website called AnimAtlas (animal + atlas) that features an interactive map of Earth, and is accessible at https://animatlas.me/. This map features the locations of fossil occurrences, and when clicked on, the user can learn facts including but not limited to its name, diet, time period it lived during, and journal articles that describe it. The user can explore which organisms lived during which periods of time by using a slider that filters by geologic time period.

AnimAtlas' goal is to educate a general audience about paleontology. Similar data visualizations exist, such as the one for the Paleobiology Database [1], however these are often tailored to be used by other experts who have a deep familiarity with the field already. As a result, they are not easily accessible to the general public. The project's goal is to offer easily digestible material to anyone with a mild interest or to those who want to gain a deeper understanding.

Education of paleontology is important and relevant to ongoing issues. Misinformation is rampant–it is a common misconception that even some adults believe that dinosaurs existed at the same time as humans [1]. Understanding the history of Earth's wildlife is important to put the ongoing climate crisis into context. The fossil record is used to understand how mass extinctions and drastic changes in climate have occurred in the past and how they affected life on Earth. Comparing current extinction rates to historic extinction rates show that Earth is experiencing drastically higher extinction than normal, which has led to the conclusion that Earth is experiencing its sixth mass extinction [2].

[1] https://paleobiodb.org/

The data source of this project is the Paleobiology Database (PBDB) [3]. It is a massive database, consisting of dozens of tables containing various types of data. Fossil occurrences are the core data concept around which the PBDB is built, and is the data that will be displayed on AnimAtlas' interactive map. Occurrences are not exactly analogous to individual fossil specimens, but rather they are a collection of one or more fossil specimens that coexisted at the same place and time. Simply put, a fossil occurrence is a singular piece of evidence that a specific type of organism existed in a certain location and time.

The project queries the PBDB API to acquire data and allows the user to manually search and filter through the data. Each data point on the map is selectable and will display a sidebar that shows information about the selected occurrence. Once these core functionalities were implemented, more advanced features were implemented, such as a fossil abundance visualization and an AI chatbot. The preliminary results show a decent foundation of basic functionality, including selectable data points and a time range interval filter.

2 Related Work

Paleobiology is rich in resources used to study and examine current knowledge and findings. In this aspect, the Paleobiology Database is simply one of many, but certainly without a lack of major influence. As of 2023, the PBDB has given the opportunity for more than 468 official publications to be made using its data [3] covering a diverse set of topics, including paleogeography, taxonomic diversity, and even computer science and machine learning [4]. Its free API resource allows for convenient and up-to-date access to the data [4], and so will be useful to this project in providing rich, accurate data to be used in the visualizations.

Paleobiology is not a commonly taught subject yet its importance in today's affairs remains strong. There is a current necessity to understand ongoing issues for the sake of the field's ability to progress. Commercialized fossil selling may be harming the field's ability to make new findings. It is a common occurrence that important fossils are privately owned, sold, and housed in private collections that are inaccessible to scientists [5]. Additionally, understanding the Earth's history is important to put its current health in context. Barnosky et al. state that humanity is currently experiencing the Earth's sixth-ever mass extinction (an event that the Earth has only experienced 5 other times in its entire 4.5-billion-year history), which has major implications for how our survival, well-being, and capacity for advancement will be negatively affected in the future [2]. The paper argues that a mass extinction is underway because the fossil record indicates higher levels of extinction than expected–specifically, that the modern extinction rate is orders of magnitude larger than the background extinction rate (which is the extinction rate expected based on non-human factors) [2]. As a side note–this paper utilizes the PBDB to calculate the background extinction rate. Paleontology can be used to find patterns in climate that may contribute to major climate change events by studying which ecosystems and species were the most vulnerable in these events [6].

Additionally, there are issues of misinformation and a lack of education about evolution due to the spread of ideologies such as creationism [7] that inherently hinder the field of paleobiology's capacity to be taken seriously. There is a large fraction of Americans who don't accept evolution, and even those who do often fail to have a deep understanding of how evolution is relevant to human affairs [8]. Since a massive point of paleobiology is to study how life changes over time, a lack of understanding and deep appreciation for evolution directly translates into a lack of understanding and respect for paleobiology.

If the goal of the project is to educate a large audience of laymen, it is important to understand that adults are notoriously hard to convince of anything–whether it's trying to motivate them to believe in something or to call to action. An issue presents itself: How can AnimAtlas cater to this audience in a way that can not only capture their attention span enough for them to learn something new, but to care enough to make an effort to understand on a deeper level? Luckily there are findings that show that free-choice learning is an important force for adult learning [9]. Two random phone surveys were conducted in 1997 and 2000 on California residents, and their results showed that the general public is interested in science [10]. In both surveys, people indicated that a large fraction of their knowledge in science was gained from sources outside of structured education (mandatory schooling or secondary education). Some specific sources differed: in the 1997 survey, radio and cable were common sources of extracurricular education, while in 2000, the Internet far surpassed the usage of the radio. While the percentages of specific sources used by people varied, the conclusion that free-choice learning plays a more significant role in lifelong learning remains consistent in both surveys. Falk and Needham in their 2013 paper note that such findings are preliminary insights into how the general public gains scientific knowledge and that future studies will build a more complete picture of how and why the public learns science [9]. However, it remains significant that using the Internet was found to be one of the strongest predictors of self-motivated knowledge of science and technology [9].

An effective way to communicate information in an attention-grabbing manner is to visualize data, which is where dashboards can be utilized to their fullest. Sarikaya et al. showcase the diverse ways dashboards can be utilized, including those for individuals and the general public, and provide insight as to what goals each dashboard type is trying to achieve, what methods they use to do so [11]. For example, a dashboard designed to be used individually may integrate personal customization features to increase the relevance of information to the individual [11]. Another article by Bach et al. finds that among many studies, scholars can agree on a general set of guidelines to follow when designing a dashboard [12]. This includes but is not limited to ensuring the dashboard is not overwhelming to users, not showing too much data or other clutter, and organizing charts symmetrically.

A major subject in visualizing paleobiological information is the phylogenetic tree, which is a diagram depicting the evolutionary ancestry of organisms. Since AnimAtlas may also display phylogenetic information, understanding how other professionals present such information will be useful in understanding what is important to visualize and how to do so in a clear and effective manner (likely in conjunction with dashboard design guidelines mentioned previously) [13]. The PBDB offers this information, yet it will require effort to ensure the data are comprehensible to members of the general public, and not too complex to be understood only by professionals. Kaya et al. document the creation of their phylogenetic tree visualization software that has shown promise in its ability to increase accuracy in categorizing newly discovered species and its source code is freely available [13].

It is evident that the PBDB is extremely versatile and is rich in information to be used for various project types. However, the open-collaboration nature of its data collection is vulnerable to a set of inconsistencies in the data therefore the database will need to be cleaned. There exists an R package that accomplishes this called CoordinateCleaner that is freely available to use [14]. The article describing this package identifies 6.5% of records in the database as potentially problematic as of 2018 [14].

With these issues addressed, AnimAtlas will be a great resource for educating the public in a riveting and approachable manner, exposing them to the exciting field of paleobiology. The website will model ideal design practices to maximize the effectiveness of its visual communication. It will take advantage of any attention it garners to educate its users on important issues relating to paleobiology such as climate change, commercialized fossil selling, and the spread of misinformation. By spreading awareness, AnimAtlas will breathe new life into the extinct creatures that once called Earth its home.

3 Data

The Paleobiology Database [3] is a vastly influential collection of paleontological information. It started during the Phanerozoic Marine Paleofaunal Database initiative, funded by the National Center for Ecological Analysis and Synthesis (NCEAS). It is currently maintained by an international non-governmental group of paleontologists and is made public for anyone to use. It is updated as contributors acquire new information and does not conform to regularly scheduled intervals for updates. Contributors have three levels of clearance: Authorizers, Enterers, and Students. Authorizers typically have doctorate degrees and must be approved by the Executive Committee. Enterers are typically graduate students working with an Authorizer, and Students are also given their status by Authorizers, and are typically using the PBDB for class or research purposes [15].

The database is relational and consists of dozens of tables. However, data is generally categorized into four different classes: bibliographic, taxonomic, geologic/geographic, and occurrences [15]. Occurrences are the data type that will be focused on in this project. An occurrence represents an individual specimen or set of specimens of a fossil species at a specific place and geologic time. It helps to understand that one occurrence belongs to one collection, which is loosely defined as a set of fossil occurrences that are located in the same place and geologic time.

As of November 2023, the Paleobiology Database has 1,597,192 fossil occurrences and 410 contributors.

The PBDB records a large number of attributes for each occurrence. In general, they may fall into the following categories (note that the listed attributes in each category are merely a few examples, intended to give an idea of what information is generally stored):

- **Taxonomic Information** - Levels of classification, e.g. Kingdom, Phylum, Order, Genus, Species
- **Location of Discovery** - Latitude, longitude, collection, state, county
- **Time Range** - Time when this organism may have been alive, given by MYA intervals and named intervals (e.g. Jurassic, Campanian)
- **Fossil Characteristics** - Common body parts, preservation quality, level of fragmentation
- **Organism Characteristics** - Life habit, reproduction, vision, motility
- **Geography of Discovery Location** - Environment, tectonic setting, lithology
- **Attribution** - Reference author(s) and publication year
- **Data Entry Information** - Names of Enterer, Authorizer, Modifier, creation date/time, modification date/time

It should be kept in mind that many attributes may be missing information due to the fact that fossil quality varies greatly and that some attributes may have niche situations in which they are applicable. However, this will not be much of an issue because basic information (such as name, location, and time range) will always be present and will be all that is needed for AnimAtlas. Extra information will be used when it is available.

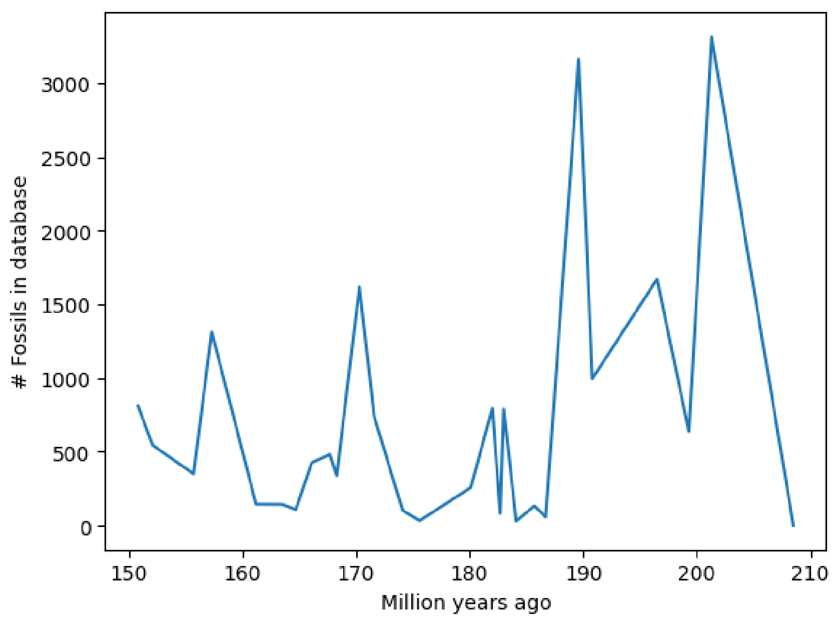

Figure 1: Distribution of dinosaur fossil occurences found over their maximum MYA interval

The size of the database makes data exploration difficult. To simplify the task, I only downloaded fossil occurrences within the clade Dinosauria (all Dinosaurs). This resulted in 19,074 records. Some basic data exploration (Figure 1 shows that most dinosaurs recorded in the database, at their earliest estimation, may have existed between 190 and 205 MYA–the late Triassic and early Jurassic.

Due to the scale of the database, data integrity is a concern. Data entry errors such as transcription errors are not unheard of. When spotted, anyone can report the error to the Authorizer who authorized the data entry. However, unspotted errors may still exist and affect accuracy of the data. Another major limitation of this database is it is vastly incomplete. There are tons of potential data that exist in the world that have not been entered in the database (such as undiscovered fossils, privately owned fossils, or fossils that have been discovered but no one has entered yet). As a result, AnimAtlas will not represent a complete picture of all of humanity's current paleontological discoveries.

4 Methods

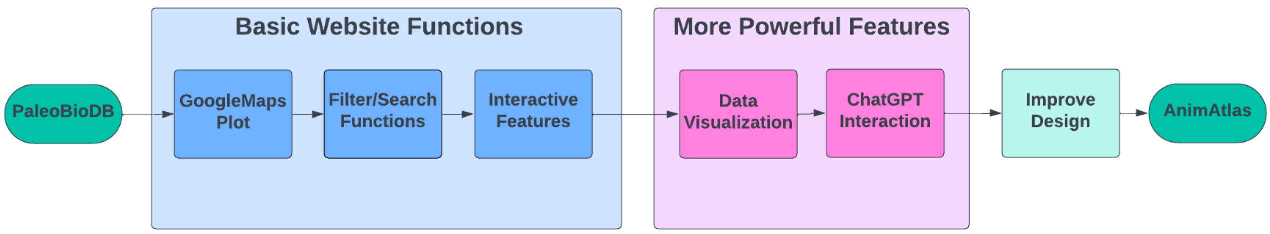

Figure 2: Pipeline of the project

The pipeline in Figure 2 lists the general process in which AnimAtlas was created. The map itself was initially created using Bokeh (a Python library for data visualization on websites) [16] with an underlying Google Map using the GoogleMaps API [17]. The reason to use Bokeh and not solely the GoogleMaps API was to allow Bokeh's useful features to be taken advantage of, such as widgets for searching/filtering the data and more efficient glyphs. JavaScript is used to enable user interactivity, such as displaying a sidebar that displays information about an occurrence when it is clicked on the map. While working the project, it was found that Bokeh is inefficient at creating heatmaps, a desired feature described later, so it was decided that AnimAtlas only needed the GoogleMaps API.

To access the data, the PBDB has an API that fetches data from the database given certain criteria used to filter it (e.g. taxonomic name or time range). It is not worth attempting to download the entirety of the PBDB because it updates frequently, making old versions obsolete, and because its large size would be an inefficient use of storage space. Using the API to query the entire database takes an uncomfortably long time (timed at over a minute) due to its large size. Additionally, when a user wishes to search or filter the data, a new query to the PBDB API can be easily constructed, meaning the user only sees the subset of the data that they wish to see.

A map offers a valuable source of opportunity for interactivity, but by itself it can only tell a limited part of a story. The sidebar complements the map by giving the user a "control panel" by making this list of features available to the user:

- A help/introduction page that explains to the user what a fossil occurrence is, what the user can learn, and how to use AnimAtlas.
- Detail view (e.g. name, time range, diet, description, reference) of a data point selected via the map. The AI chatbot is accessible here.
- Search area that allows the user to fill a form to change the data that is displayed on the map. Inputs include but are not limited to taxonomic name, time range, and record limit.
- A menu with radio buttons to allow the user to choose how the data is visualized on the map. These visualizations include a heatmap or clustering of data points

Each of these four features are accessible via a navigation bar at the top of the sidebar. The separation of these features into their own pages simplifies tasks for the user as it allows them to easily focus on only what they wish to accomplish at a point in time.

The first priority was to implement the ability to select a data point and display its information on the sidebar. The displayed information also includes reference information, which requires another API call as the PBDB stores reference information in a separate table and uses the reference number as a foreign key in the fossil occurrence table. Then, the search function was implemented. The current accepted fields for searching are taxonomic name, time range, and maximum number of records. The help/introduction page only needed static text, so was filled with a placeholder description.

A feature that was implemented to improve the user's experience are tutorials. The reason for this is not many laymen are likely going to be familiar with most of the names of the fossils present on the map, so tutorials may be useful for easing them into AnimAtlas and convincing them to learn how to use the site. Popular extinct animals such as Tyrannosaurus Rex are featured in these tutorials, giving new users a way to connect what they may already know to what new information AnimAtlas may teach them. These tutorials are accessible via the help page. The Tyrannosaurus Rex tutorial teaches the user how to use the search function to filter the dataset to see only T. Rex fossils by offering the user two choices: a button that, when clicked, performs the operation automatically, or instructions to follow to perform the operation manually. In this case, the user has the freedom of choice and can take learning at their own pace while still being able to get immediate results from data exploration.

Each record has a minimum MYA (minimum million years ago) and maximum MYA (maximum million years ago) which are estimates of the time range in which the organism may have been alive. These values can be used to filter based on a specified time range via the slider at the bottom of the page. To decide which records should appear on the map based on the specified time range, the algorithm chooses records that have one or both of their time range bounds within (inclusive) the specified time range.

Once basic functionality was completed, more advanced features were implemented. To give the user more insightful information as opposed to the basic, surface-level information available in the database's attributes, AnimAtlas can be used to display other spatial information such as the number of fossils over a distance (or fossil abundance). This was done by creating the ability to visualize the data as a heat map, where more concentrated areas correspond to areas with high fossil abundance. It is also useful to display the data points as clusters to improve readability (as it is easier to see how many data points are in an area) and efficiency. These options were made available in the sidebar's visualization menu.

The GPT Chatbot was then implemented. This was done using Chatbase [18], a service that creates chatbots using OpenAI's GPT models [19]. AnimAtlas' chatbot uses the GPT-3.5 Turbo model because it is currently the most cost-effective model made available by OpenAI. Chatbase trains the chatbots by requesting input data, however it does not seem to accept tabular data (only .pdf, .doc, .docx, and .txt are accepted), so the PBDB User Guide [15] was used to train the chatbot. As a result, the chatbot currently available on AnimAtlas cannot be trusted to not hallucinate facts about the organisms if asked about them, but it will be useful to a user who has questions regarding how to interpret the data or how the data are organized. However, if AnimAtlas one day gains traffic, a severe limitation is that the free plan of Chatbase permits only 20 messages per month.

The GitHub repository for AnimAtlas can be found at https://github.com/Sophie-Ngo/Animatlas.

5 Results

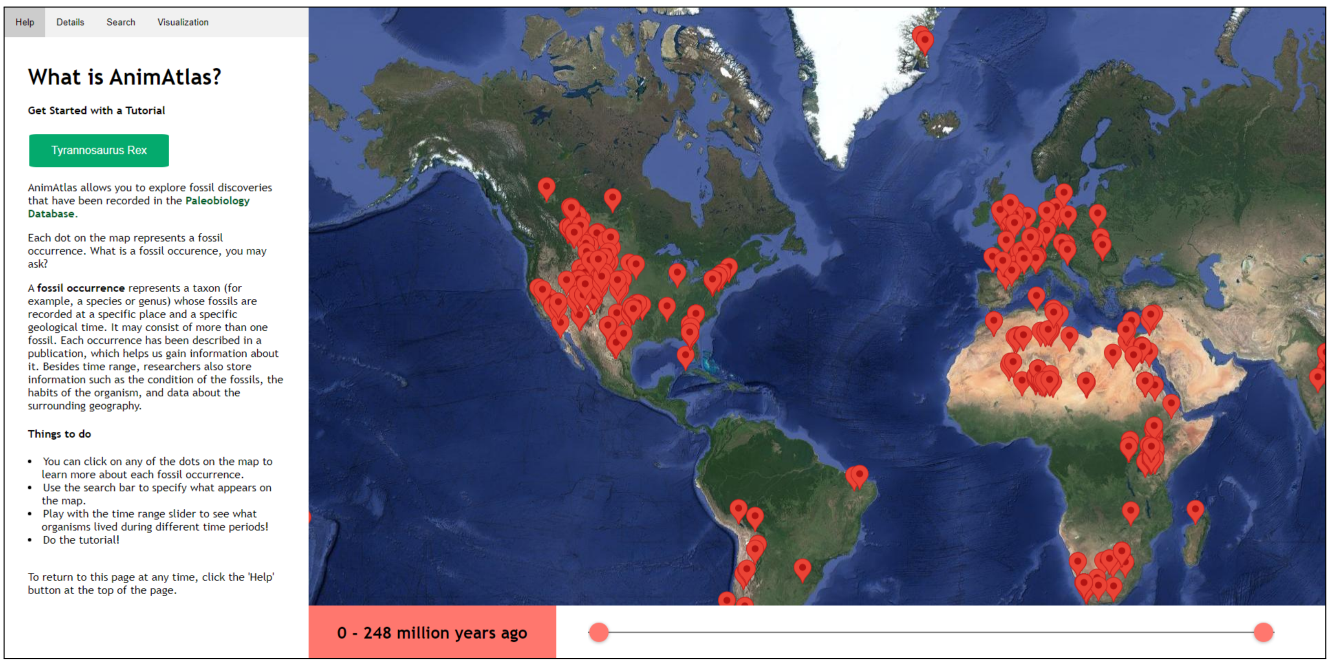

Figure 3: Appearance of AnimAtlas upon first loading the page.

The current state of AnimAtlas is shown in Figure 3. It features the sidebar defaulted to the 'Help' page, which has basic instructions to get the user started. The tutorial is available on this page. The dataset that is initially loaded is 1,000 fossil occurrences identified to be in the clade Dinosauria (all Dinosaurs). A larger dataset could be used, however this increases loading time.

The data is initially visualized as points. The user can use the dual slider at the bottom of the page to filter fossil occurrences based on time range. The label to the left of the slider represents the current values of the filter and change dynamically when the sliders are changed by the user.

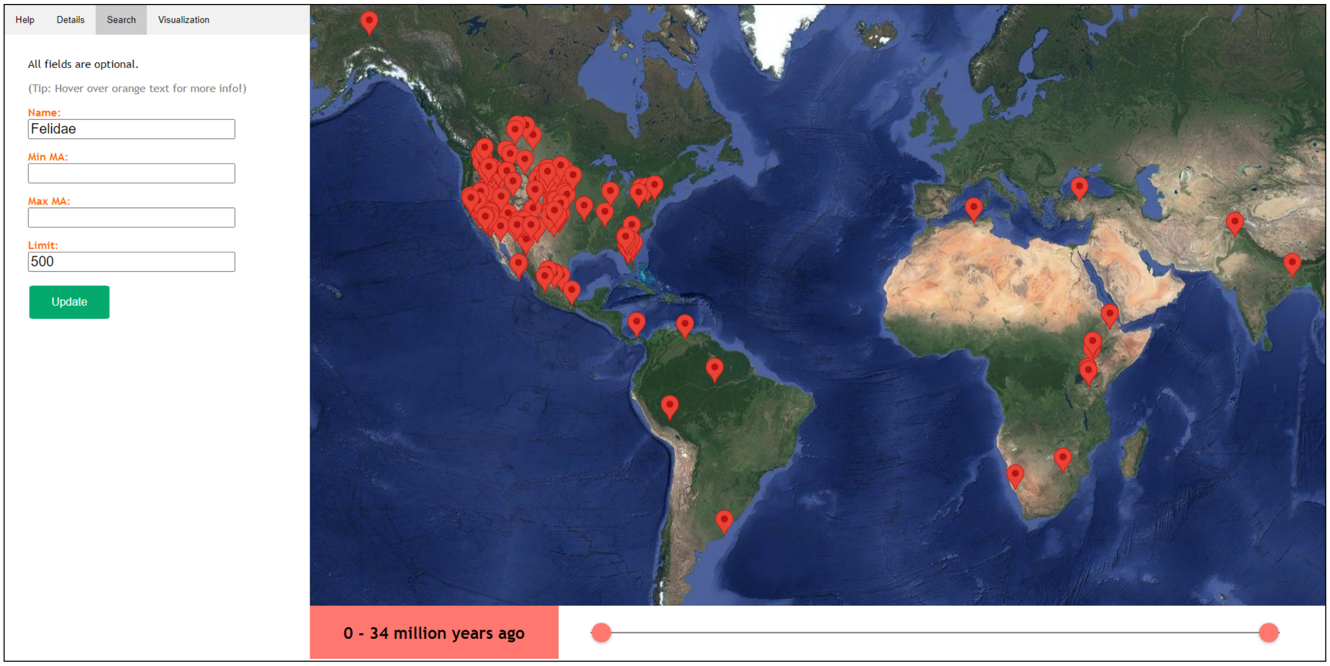

Figure 4: Example of using the search feature to select fossils identified to be in the Felidae (cat) family.

The user can search by certain criteria in the Search page. The taxonomic name can be any name of any rank, and the results will include data that are found to be identified with that taxa or any taxa lower in hierarchy to the given taxa. For example, 'Dinosauria' includes all dinosaurs and birds, while 'Aves' only includes birds. An example is shown in Figure 4, in which the resulting data is 500 fossil occurences identified as cats (in the family Felidae). The minimum and maximum MYA values are to limit the time range in which the resulting fossil occurrences lived.

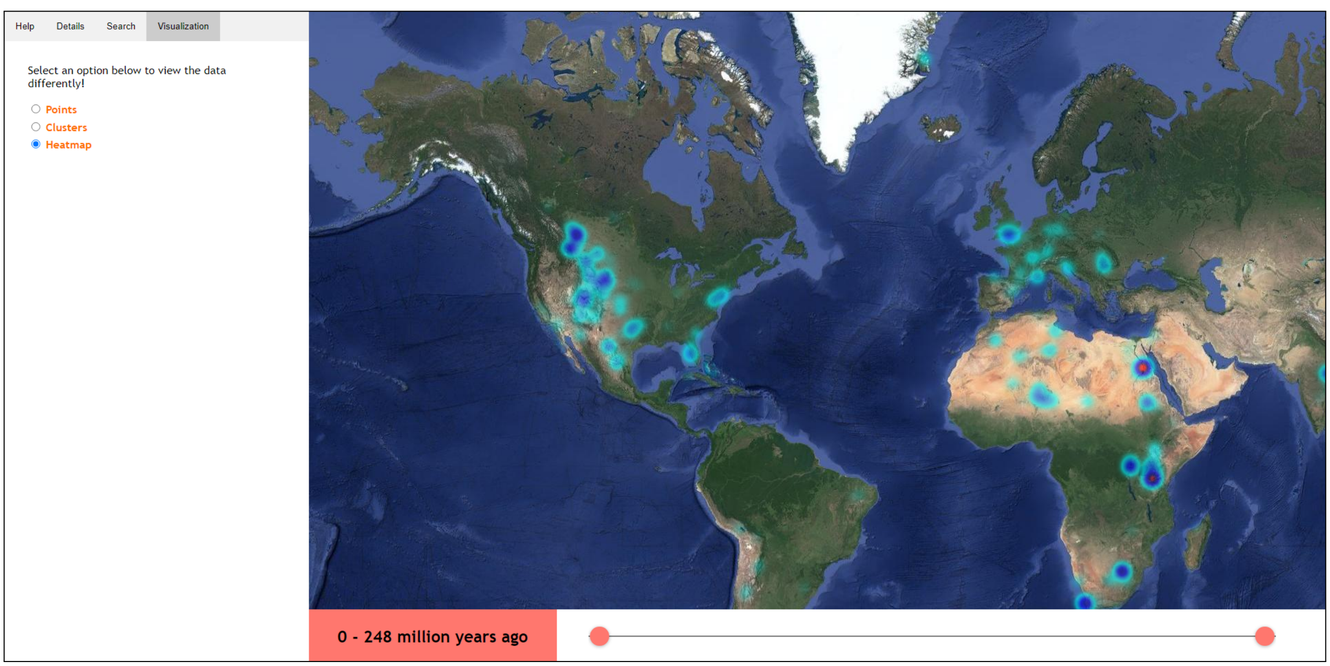

Figure 5: Heatmap visualization of the Dinosauria dataset (with 1,000 records), showing fossil abundance by location.

Figure 5 shows the user interface for choosing the visualization method. The heatmap allows users to explore where the most fossils have been discovered for the current dataset, where red represents the highest density and blue the lowest density.

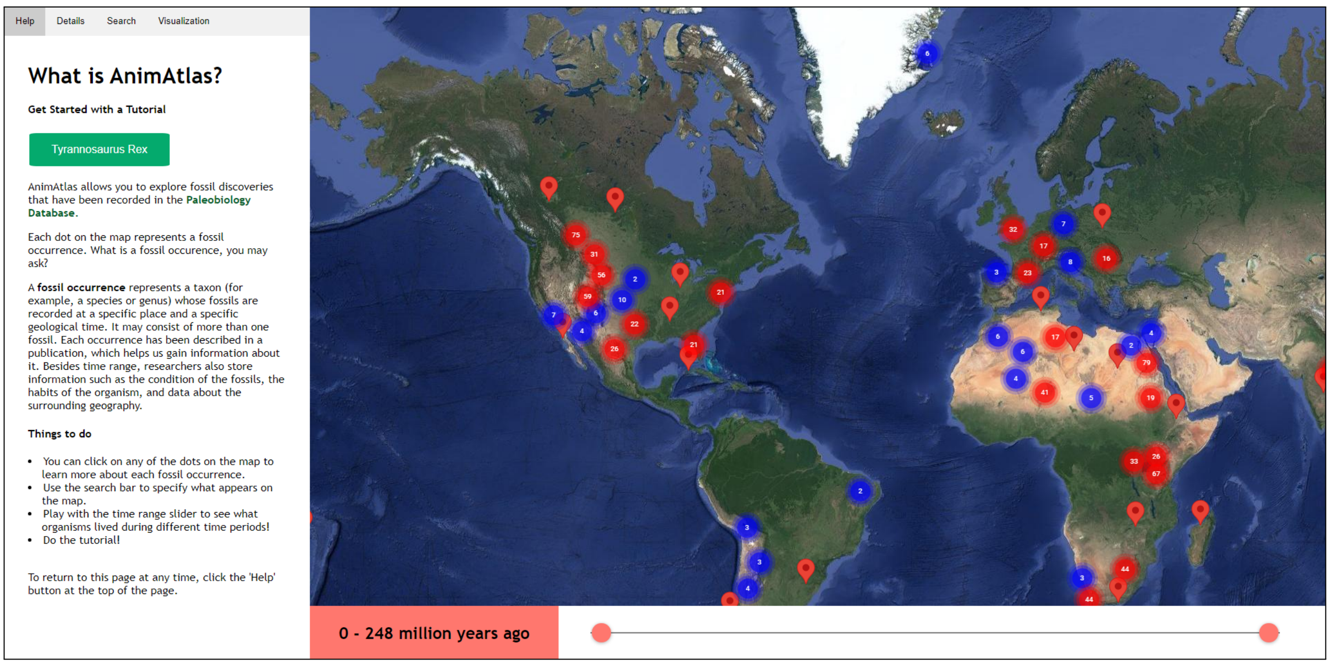

Figure 6: Clustering visualization of the same Dinosauria dataset, showing the location of each fossil occurrence and clusters of occurrences close in proximity.

Figure 6 is similar to the default 'points' visualization, however any fossil occurrences that are too close to each other and would appear to overlap when displayed on the map are now conglomerated into clusters. The number on each cluster represents how many points are in close proximity to each other in the area of the cluster.

Figure 7: Example detail page that appears in the sidebar when a record on the map is clicked.

Figure 7 features detailed information of a selected record. At the bottom of the page, the user has access to the AI chatbot where they have the ability to ask any questions.

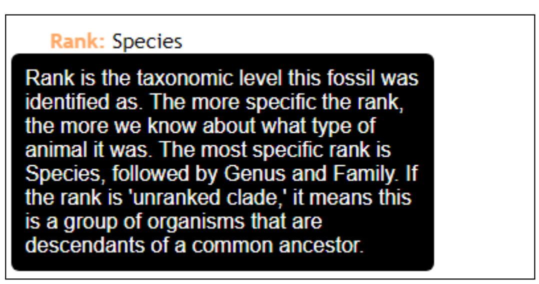

Figure 8: Tooltip description of the term 'Rank' that appears when hovered over.

To clarify what certain terms mean, each text in bold orange can be hovered over by the user to gain a more apt understanding, seen in Figure 8. Rank is an example of a term that may be unfamiliar to laymen, so a more in-depth description is especially required. The challenge in having tooltips for confusing terms is that they must communicate as much information as possible in the least amount of words to avoid confusing the reader. The great benefit of tooltips is they are the easiest and quickest way for a user to reference clarifying information. Alternatives that are not as effective include navigating to another page to read this information (which increases the amount of clicks done by the user), displaying the text below the term (clutters up valuable space in the sidebar), or opening a pop-up window (can be seen as invasive and annoying to some users).

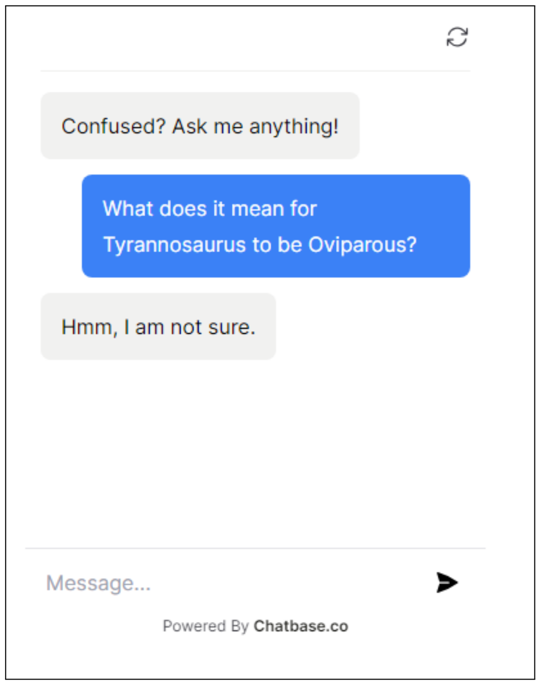

Figure 9: Example of asking a question to the chatbot that it cannot answer.

The AI chatbot has various limitations. By default, Chatbase primes the model with the following instructions:

I want you to act as a support agent. Your name is "AI Assistant". You will provide me with answers from the given info. If the answer is not included, say exactly "Hmm, I am not sure." and stop after that. Refuse to answer any question not about the info. Never break character.

As a result, any question the user asks that relates to the organism itself will be unanswerable by the model due to the fact it was trained on the user guide and not the data. An example of this can be seen in Figure 9.

These instructions are changeable, and it may be acceptable to remove the part of the instructions forcing the model to answer with "Hmm, I am not sure." upon receiving a question that is not answerable with the source material. However, this does not guarantee the model will not hallicinate, plus it may be more beneficial in the future to replace the Chatbase chatbot with an alternative method due to the fact that 1) Chatbase only allows 20 messages per month with the free plan and 2) it would be more ideal to train the model using the data, which Chatbase does not seem able to perform.

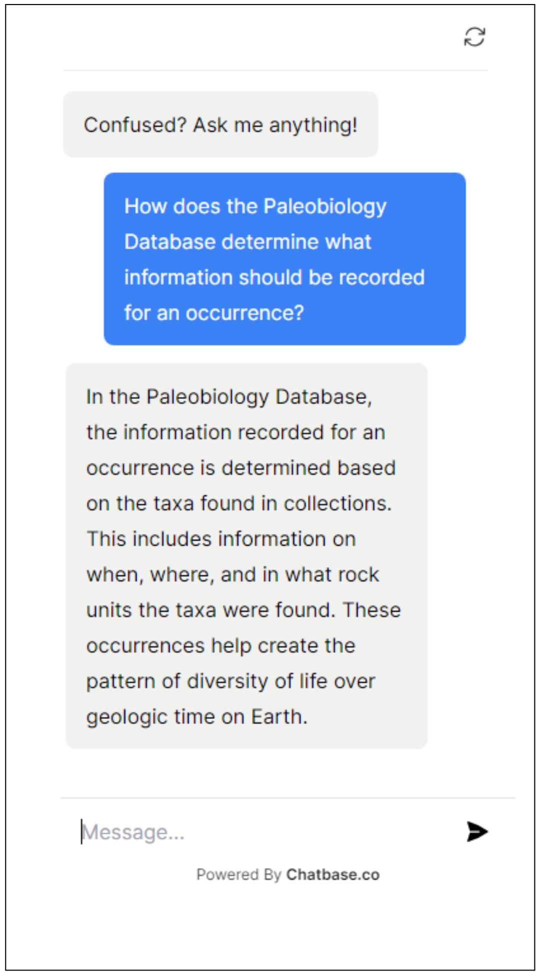

Figure 10: Example of asking a question to the chatbot that it can answer, using information from the Paleobiology Database User Guide [15].

The current model can be helpful in understanding the structure of the database itself, however. An example can be seen in Figure 10. However, this example, while decent, has the potential to be vague to the asker due to the use of commonly unfamiliar terms such as "taxa" and "collections."

6 Future Work

AnimAtlas has a few limits. It is slow to execute user queries, however this is because the PBDB API takes a nontrivial amount of time to respond that scales with the size of the resulting dataset. It may be possible to overcome this by somehow downloading a snapshot of the PBDB and caching it at regular, automated intervals so that when a user visits the page, there will not be any loading time. A limitation to this, however, is that the database is massive and therefore will demand many computer resources to download at regular intervals.

Another limit is that it may overall lack many exciting features. Being able to explore a map, read about fossils, and visualize fossil abundance are interesting features, but it is plausible that, especially an adult, may quickly become bored once those avenues have been exhausted. With more time, AnimAtlas could host a number of more engaging features. This could perhaps entail a way to compare multiple organisms at once (e.g. a side-by-side view of the estimated size of the organism), take educational pop quizzes/games testing what users may have learned, or create accounts so that users can bookmark fossil occurrences to view at a later date or store personal high scores on the aforementioned quizzes/games.

There is also a lack of data on living organisms, which would have been very useful to shed light on the ongoing mass extinction. This is because it would be possible to compare modern organism data to the paleobiology data to see differences in biodiversity. To implement it, one would have to figure out how to compare this data with the data from the PBDB. Fossil "occurrences" are unique to the PBDB and therefore cannot blend seamlessly with other data. The step to take here would then be to interpret fossil occurences in a way that would be comparable to other data.

Lastly, if AnimAtlas is to be used regularly by a general audience, it needs testing, bug fixes, and further improvements. The chatbot AI could be replaced with an alternative to Chatbase[18], preferably with a model that can read tabular data from the Paleobiology Database. The visualizations could be executed faster with improvements to the code, as currently, switching visualizations re-queries the API, when all that is needed is to re-draw the map. Some ideas for future UI improvements include the slider handles themselves showing their current value, adding automatically scraped images to the Details page, and adding more search criteria that laymen will be able to understand. Another tutorial could be added to help the user learn how to interpret the different visualizations and compare them to each other.

References

[1] Mark Francek. A compilation and review of over 500 geoscience misconceptions. *International Journal of Science Education*, 35(1):31–64, 2013.

[2] Anthony D. Barnosky, Nicholas Matzke, Susumu Tomiya, Guinevere O. Wogan, Brian Swartz, Tiago B. Quental, Charles Marshall, Jenny L. McGuire, Emily L. Lindsey, Kaitlin C. Maguire, and et al. Has the earth's sixth mass extinction already arrived? *Nature*, 471(7336):51–57, 2011.

[3] Paleobiology database. `https://paleobiodb.org/`. Online, Accessed 10/6/2023.

[4] Shanan E. Peters and Michael McClennen. The paleobiology database application programming interface. *Paleobiology*, 42(1):1–7, 2015.

[5] K Shimada, P Currie, E Scott, and S Sumida. The greatest challenge to 21st century paleontology: When commercialization of fossils threatens the science. *Palaeontologia Electronica*, 2014.

[6] Wolfgang Kiessling, Jansen A. Smith, and Nussaïbah B. Raja. Improving the relevance of paleontology to climate change policy. *Proceedings of the National Academy of Sciences*, 120(7), 2023.

[7] Jason R. Wiles. Challenges to teaching evolution: What's a head? *Futures*, 43(8):787–796, 2011.

[8] David Sloan Wilson. Evolution for everyone: How to increase acceptance of, interest in, and knowledge about evolution. *PLoS Biology*, 3(12), 2005.

[9] John H. Falk and Mark D. Needham. Factors contributing to adult knowledge of science and technology. *Journal of Research in Science Teaching*, 50(4):431–452, 2013.

[10] John H. Falk, Martin Storksdieck, and Lynn D. Dierking. Investigating public science interest and understanding: Evidence for the importance of free-choice learning. *Public Understanding of Science*, 16(4):455–469, 2007.

[11] Alper Sarikaya, Michael Correll, Lyn Bartram, Melanie Tory, and Danyel Fisher. What do we talk about when we talk about dashboards? *IEEE Transactions on Visualization and Computer Graphics*, 25(1):682–692, 2019.

[12] Benjamin Bach, Euan Freeman, Alfie Abdul-Rahman, Cagatay Turkay, Saiful Khan, Yulei Fan, and Min Chen. Dashboard design patterns. *IEEE Transactions on Visualization and Computer Graphics*, page 1–11, 2022.

[13] Gizem Kaya, Chisom Ezekannagha, Dominik Heider, and Georges Hattab. Context-aware phylogenetic trees for phylogeny-based taxonomy visualization. *Frontiers in Genetics*, 13, 2022.

[14] Alexander Zizka, Daniele Silvestro, Tobias Andermann, Josué Azevedo, Camila Duarte Ritter, Daniel Edler, Harith Farooq, Andrei Herdean, María Ariza, Ruud Scharn, and et al. Coordinatecleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, 10(5):744–751, 2019.

[15] Mark D. Uhen, Bethany Allen, Noushin Behboudi, Matthew E. Clapham, Emma Dunne, Austin Hendy, Patricia A. Holroyd, Melanie Hopkins, Philip Mannion, Phil Novack-Gottshall, and et al. Paleobiology database user guide version 1.0. *PaleoBios*, 40(11), 2023.

[16] Bokeh Development Team. *Bokeh: Python library for interactive visualization*, 2018.

[17] Google maps javascript api. `https://developers.google.com/maps/documentation/javascript`. Online, Accessed 12/5/2023.

[18] Chatbase. `https://www.chatbase.co/`. Online, Accessed 12/7/2023.

[19] Openai. `https://openai.com/`. Online, Accessed 12/7/2023.