A Visualization-based Approach for the TaxonomyBrowser Interface

Marina Fortes Rey
Institute of Informatics
Federal University of Rio Grande do Sul
mfrey@inf.ufrgs.br

Carla Maria Dal Sasso Freitas
Institute of Informatics
Federal University of Rio Grande do Sul
carla@inf.ufrgs.br

ABSTRACT

The process of collecting biological data is a continuous activity, specially in field work for research projects or teaching activities. Data obtained from these collecting activities need to be properly stored in order to be readily available for future analysis. *TaxonomyBrowser* is a biodiversity information system mainly developed to manage data collected by biologists during field work. It stores such data on a database organized according the taxonomic tree of leaving species. This work describes a new approach for the *TaxonomyBrowser*'s user interface, focusing on providing an easier and more intuitive method of managing and visualizing the information stored in such databases. The approach was assessed by means of a survey conducted with users from the field of study as well as users with no knowledge on biodiversity information systems. The participants considered the application to be good and mostly provided positive feedback.

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces; J.3. Life and Medical Sciences

Author Keywords

Biodiversity information systems; information visualization; taxonomy.

INTRODUCTION

The process of collecting biological data takes place constantly and is also a very manual and systematic work, which demands that the researcher collects basic data from the specimen on the field, catalogues and stores them for further visualization. Although usually such data are recorded in separate files, there are adequate systems for that, the so-called biodiversity information systems. These data records can be later modified when new laboratory studies are performed on the collected samples. So, it is important to have an unified biodiversity information system capable of providing easy access to previously stored specimens' data either for retrieving or updating the related information.

Most solutions for biodiversity information systems have unnecessarily complex user interfaces [4, 22]. Such systems usually do not provide any overview of the data or any simplified method of browsing their entries, only showing a list of all specimens recorded on the database. There are applications that focus on the visualization of taxonomic and phylogenetic trees especially built for biological data analysis, yet they normally do not comprise own databases for storing the information, requiring that users import their data on every use.

In a previous research project, a team composed of computer scientists and biologists developed *TaxonomyBrowser*, a biodiversity data management system with the purpose of maintaining data about mammals collected by biologists all over the state [17, 7, 22, 3, 8]. That version does not provide an overview of all data stored in the database nor an easy way of accessing a specific specimen. Moreover, all specimens and specimens' information are displayed as lists.

The present work describes the development of a completely new interface for the *TaxonomyBrowser*, using information visualization techniques as the primary means for user interaction. We based our work on an overview of the system's taxonomic hierarchy for displaying and interacting with the data in a clear and intuitive way, aiming at reducing the number (and complexity) of interactions required by the user to perform certain actions on the database.

The solution was validated by means of a remote survey with users from all levels of understanding about the application domain, i.e., with different levels of knowledge about biodiversity. The results showed the tool was well understood and praised by most of the participants.

The remainder of this paper is organized as follows. First, we present the biological concepts that are important for understanding our work, and briefly describe applications and on-line portals related to TaxonomyBrowser at a certain extent. Then, we present the system, its features and implementation details, as well as the results from user tests we conducted for assessing the new version of TaxonomyBrowser. Finally, we conclude the paper discussing open issues and future work.

BACKGROUND

Some concepts are needed for making easier the understanding of the terminology we use in this work. Such concepts include the notion of taxonomy and phylogenetic trees, and the definition and usage of a biological information system.

IHC'17, Proceedings of the 16th Brazilian symposium on human factors in computing systems. October 23-27, 2017, Joinville, SC, Brazil. ISBN 978-85-7669-405-2 (online).

Taxonomy Tree

Enghoff's paper, 'What is Taxonomy? - An overview with myriapodological examples' [5] helps us to understand the concept of a taxonomy tree. Etymologically, the word taxonomy is derived from Greek *taxis*, meaning 'arrangement or division', and *nomos*, meaning 'law'. According to Enghoff, taxonomy can thus be understood as 'laws of arrangement and division'. Such taxonomies are composed of taxonomic units known as *taxa* (singular: *taxon*), frequently arranged in a hierarchical structure and related to one another by 'parent-child' relationships. The taxa for living beings are distributed in a Linnaean classification, where groupings receive a rank, such as Kingdom, Phylum, Class, Order, Family, Genus, and Species (in decreasing order of inclusiveness) [12]. The taxon's attributes are inherited by the children nodes. Taxonomy also consists of the interpretation of names and the way we believe that the taxa are phylogenetically related to each other, being able to evolve as taxa are discovered or altered.

Phylogenetic Tree

Formally a phylogenetic tree is a construction that attempts to form the ancestors and descendants relationships for a set of entities [21]. They represent a clear notion of evolution from ancestors to current-day entities. An important characteristic of phylogenetic trees is that the descendants (leaf nodes of the tree) represent present-day entities, while common ancestors represent parents that existed in the past. For this reason, internal nodes are rarely changed, while leaf nodes vary more constantly.

Biodiversity Information Systems

Biodiversity Informatics includes the application of information technologies to the management, exploration, analysis and interpretation of primary data regarding life, particularly at the species level of organization [18]. Biodiversity information systems are built around a database that stores taxonomic information from a particular area or group of living organisms, mainly storing individual specimen's and species' information. The collection of these materials is performed during field work, when information about the captured specimens is usually written down in a conventional (paper) notebook. Samples from collected specimens are often stored physically, having a description of their location, for example, which box in which room it is kept, and attributes saved in the database. These samples can be tissue, blood, bones, DNA, organs and even the entire fluid-preserved or taxidermied animal body.

RELATED WORK

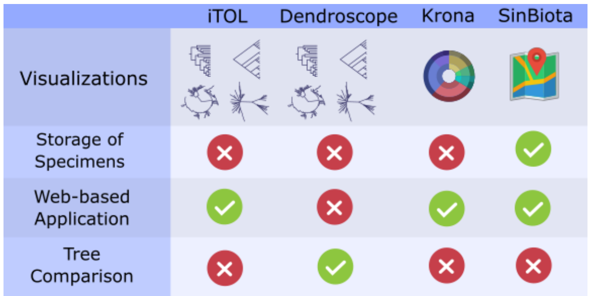
There is a considerable number of similar visualization applications, most using phylogenetic or taxonomic trees to represent information about species. SinBiota ("Sistema de Informação Ambiental do Biota") was created within the Biota project (FAPESP - Fundação do Amparo Ãă Pesquisa do Estado de São Paulo) to provide a infrastructure for consolidating data obtained by researchers from their projects, and make biodiversity information from the region of São Paulo, Brazil, readily available [4]. SinBiota is able to store and handle a substantial amount of data. The system's main feature is to provide a map populated with the location of collected specimens, showing all provided information when selecting each marker displayed. To visualize the collected specimens the user must first either select a certain area of interest or filter the specimens by a certain parameter.

The Interactive Tree of Life (iTOL) is an on-line tool for the display, annotation and management of phylogenetic trees [13]. Users can manage multiple trees and share their workspace with other researchers, all available on an on-line interface developed using Javascript and HTML5. The system has 3 types of tree visualizations available: standard indented tree visualization, circular, and unrooted (radial) layout. Another important feature is the possibility of creating a pruned tree by selecting each node manually from the original tree. Besides the standard taxonomic information, specific measures can be displayed for each species entered in the database. The values are displayed in a bar linked to each leaf node, thus creating a bar chart that uses the tree visualization as one of its axis.

Dendroscope is designed as an all-round tree visualization tool that can handle trees with hundred thousands of taxa [9]. The tree can be displayed in seven different ways, including: circular cladogram, radial phylogram, rectangular phylogram and slanted cladogram. The system can handle the display and correlation of multiple phylogenetic trees, including being able to zoom certain parts of the tree, reshape, re-root, reorder, extract a sub-tree or network and even attach images to be displayed next to corresponding nodes.

Krona is a visualization tool that allows intuitive exploration of relative abundances and confidences within the complex hierarchies of meta-genomic classifications [15]. It uses a radial, space filling visualization, which subdivides classes into sectors and places them depending on their biological lineage. The sectors are labeled with the scientific name of each taxon and, even though most would not fit in the space given for each partition, the system has an algorithm to increase textual information by orienting leaf node labels along the radial configuration and internal ones along the tangent of the partition.

We compared the four systems, and Figure 1 shows a summary of their features.



	iTOL	Dendroscope	Krona	SinBiota
Visualizations				
Storage of Specimens	✗	✗	✗	✓
Web-based Application	✓	✗	✓	✓
Tree Comparison	✗	✓	✗	✗

Figure 1. Comparison between iTOL, Dendroscope, Krona and SinBiota systems regarding their features.

We have found that iTOL, Krona and Dendroscope are focused on presenting an overview of the data, lacking more complex functions for analysis and management of information. SinBiota, on the other hand, targets mainly the storage of data, missing a tool for an overall observation of the collected specimens. Krona, iTOL and SinBiota have very clean and modern user interfaces, allowing users to access it on-line. Dendroscope is quite the opposite, with a common interface for Windows programs and requiring to be downloaded and installed.

In this work we focus on providing every feature shown in Figure 1 by means of tools for managing the information, with the additional characteristic that we integrate them with visualization techniques that allows for an overview of the data as well as details when needed, all available through a web-based application.

VISUALIZATION-BASED INTERFACE

As mentioned before, *TaxonomyBrowser* was developed in a previous research project in order to store biological data using a taxonomy tree [17, 7, 22, 3, 8]. So, we started this work by comparing the new system design with the previous version, based on Shneiderman's tasks [16]. In that seminal paper, Shneiderman describes seven tasks for information visualization (overview, zoom, filter, details-on-demand, relate, history and extract) and seven data types (1-, 2-, 3-dimensional data, temporal and multidimensional data, and tree and network data). He also defines as the basic principle of visual design the Visual Information Seeking Mantra: "overview first, zoom and filter, then details on demand".

These principles were followed in the design of the new interface, which is based on a Sunburst [20] visualization that allows for representing and managing the taxonomic tree. As each species is selected in the Sunburst visualization, the specimens classified as belonging to that species are exhibited as small circles in the center of the layout and three types of visualizations can be created with the specimens' measures, as can be seen in Figure 2. Following the Visual Information Seeking Mantra, the Sunburst represents the overview of the data, while the specimens' visualization serve as zooming and filtering, and finally the parallel coordinates, scatterplot and geospatial visualization present details on demand.

TaxonomyBrowser is a web-base application for providing wide availability and readiness of use. The visualizations and main features of the application were implemented using JavaScript, while the connection to the database was developed using PHP. Widely known libraries such as jQuery and Bootstrap were used, mainly for developing the graphical user interface. Other libraries were used for specific parts of the interface, such as the Intro.js library [11], only needed for the tours and hints available, and the Toolbar.js plugin [23], used only for the tooltips provided within the Sunburst visualization.

Data Model

The database structure is based on the taxonomic tree, which ranks each node based on the Linnaean classification: Kingdom, Phylum, Class, Order, Family, Genus, and Species (in decreasing order of inclusiveness) [12]. Each collected specimen is recorded in the database as children of its species node, which is children of a higher rank and so on, forming the taxonomic hierarchy. Depending on its taxonomic level, different characteristics can be recorded for each specimen. The values for the characteristics can be either numerical measurements or textual attributes. If a taxon has a certain characteristic, all its descendants (other taxon or specimens) will inevitably inherit it.

The design of the database was not within the scope of this work since it was already developed in the previous version of the *TaxonomyBrowser*. However, some modifications had to be made in order to add new features, such as the authentication needed to access specimens' data.

Application Architecture and Dataflow

In order to display information from recorded specimens, data is obtained from the database, and then encoded in JSON, a lightweight data-interchange format that is syntactically identical to JavaScript objects. Due to this, data in JSON can be imported to a JavaScript module with standard JS functions, which leads to a better performance compared to other approaches.

When the page loads, the data acquired from the database is displayed as the (main) Sunburst visualization, where the user is able to freely interact with all taxa currently recorded in the database. This provides the overview of the whole dataset. From the Sunburst visualization it is possible to select species for further inspection. All specimens from the selected ones are added to a new list of objects and displayed in the Specimen's View (in the central area of the interface). Moreover, this selection can be filtered by its characteristics and visibility. Visibility means that a specimen can be available publicly, or can be private to a certain user or visible only to all members of a certain research group). The specimens contained in this new list will be used for the visualizations shown in the right area of the screen. Each of the features mentioned will be further explained in the next subsections.

Visualization Techniques

From the complete overview of the database to the details of each specimen, the visualizations presented are a fundamental part of this work. The main tool used for implementing all the visualizations is Data-Driven Documents (D3) [1], a JavaScript library created for manipulating documents based on data used mainly for creating interactive visualizations.

Sunburst

Due to the organization of the provided data, a hierarchical visualization capable of displaying a full overview of the database was required. Even though the standard tree visualization would be the safest choice since it is widely used for viewing such datasets, this approach would not enable a complete overview of all data in a constrained area. Radial, Space Filling (RSF) techniques for hierarchy visualization have several advantages over traditional nodeâĂŞlink diagrams, including the ability to use the display space efficiently while effectively representing the hierarchy structure[24] and allowing to analyze and alter in detail a variety of regions simultaneously without loosing the overview of the dataset. The RSF Sunburst uses a radial configuration where the inner circle represents the root of the hierarchy and deeper levels are layered around this central node. We chose Sunburst also because an analysis reported by Stasko et al.[19] suggested that participants strongly

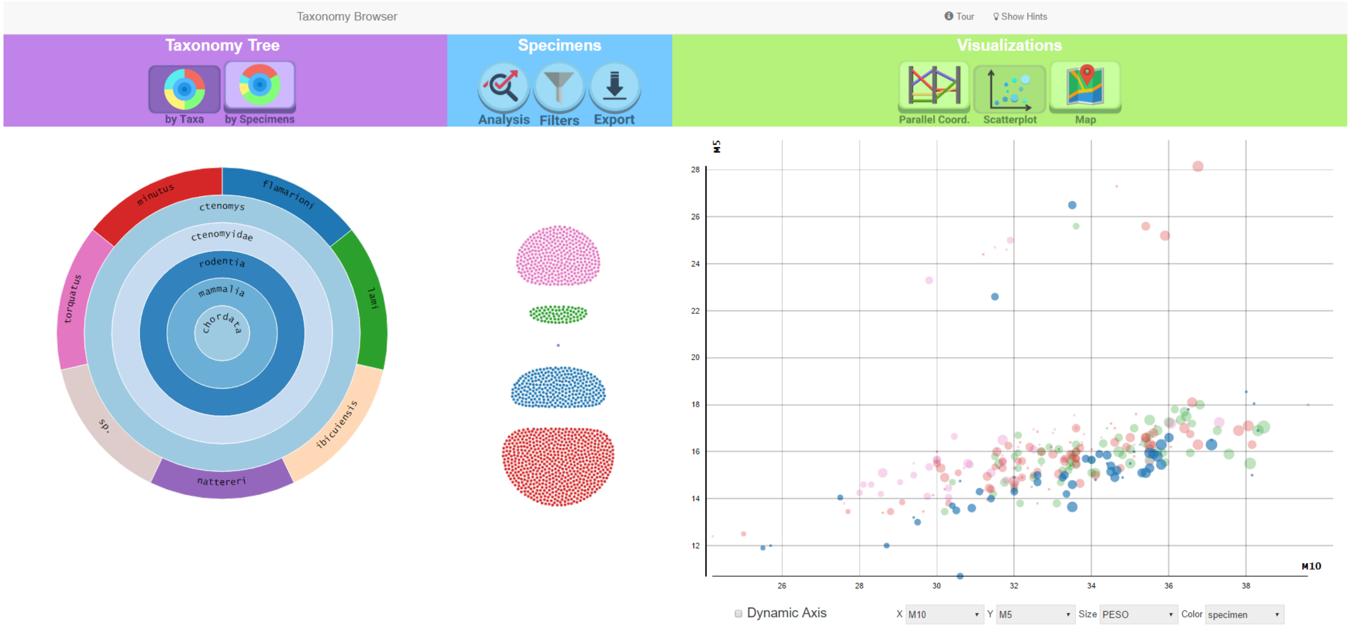


Figure 2. The interface developed for the TaxonomyBrowser. Left: Sunburst visualization showing all species represented in the database starting from the highest level (*Chordata*, in this case), with species *minutus*, *torquatus*, *flamarioni*, *lami* and *natteteri* selected. Center: all specimens selected in the Sunburst visualization, with or without further filtering. Right: example of a visualization available for displaying data about selected specimens. Top: options for tree visualization, specimens manipulation, and data attributes visualization (parallel coordinates, scatterplot, and map.

preferred this tool, citing better ability to convey structure and hierarchy. They also were more successful and presented a faster performance in the tasks they were asked to complete.

The Sunburst visualization implemented allows the following interactions:

- *Hover*: Hovering on each node triggers a toolbar-styled tooltip, with the node's name at the top and icons representing options that allow the user to show information, add children nodes and edit the hovered node (Figure 3A).
- *Change Partition Size*: By clicking the buttons on top of the Sunburst visualization the user can define if the partitions' sizes are defined by the number of species of each taxon or by the number of specimens recorded (in the database) for each taxon (Figures 3A and 3B).
- *Zoom*: When right-clicking a node, it becomes the main node displayed in the Sunburst view, showing only its children and therefore presenting them in more detail (Figure 3C).
- *Select*: In order to select specimens for further inspection or visualization, the user must use the left mouse button on the desired node. If a node of higher rank is selected, all of its children will also be selected. The result will be shown to the right of the Sunburst as small circles, each one representing one specimen, as can be seen in Figure 2 (center) and Figure 5.

The current database still does not contain many taxa, with only 7 species of the same taxonomic group. Figure 4 demonstrates how the visualization would look like with many taxa stored. It also shows how color is distributed in the inner nodes of the hierarchy: if a taxon has more than one child, a new color is assigned to it in order to simplify inspection.

Specimens' View

This view was designed for showing all specimens selected in the Sunburst visualization. When a filter is applied, only the specimens that fit all filters' criteria are displayed. The visualization is based on a force layout, circles of the same color attracting each other and forming groups (as shown in Figure 5). Each circle represents a single specimen, with its color representing its species (the same color code used in the Sunburst partition for representing that taxon).

The user can click on each of them in order to display, edit or delete all the specimen's information. Hovering over the view also shows the number of specimens from that species currently selected and the species' name. Depending on the number of specimens selected, the size of each circle and their proximity with each other is altered. This visualization is important to give the user a notion of the number of selected elements, particularly for checking how many remain selected after applying a filter.

Parallel Coordinates

During the mid-1980s and early 1990s, Inselberg and Dimsdale [10] introduced a technique for visualization of multidimensional data they called Parallel Coordinates. In this approach, each dimension, representing an attribute, is drawn as a vertical (or horizontal) line, and each multidimensional point is visualized as a polyline that crosses each axis at the appropriate position (depending on the attribute's value) to reflect the nD position [14].

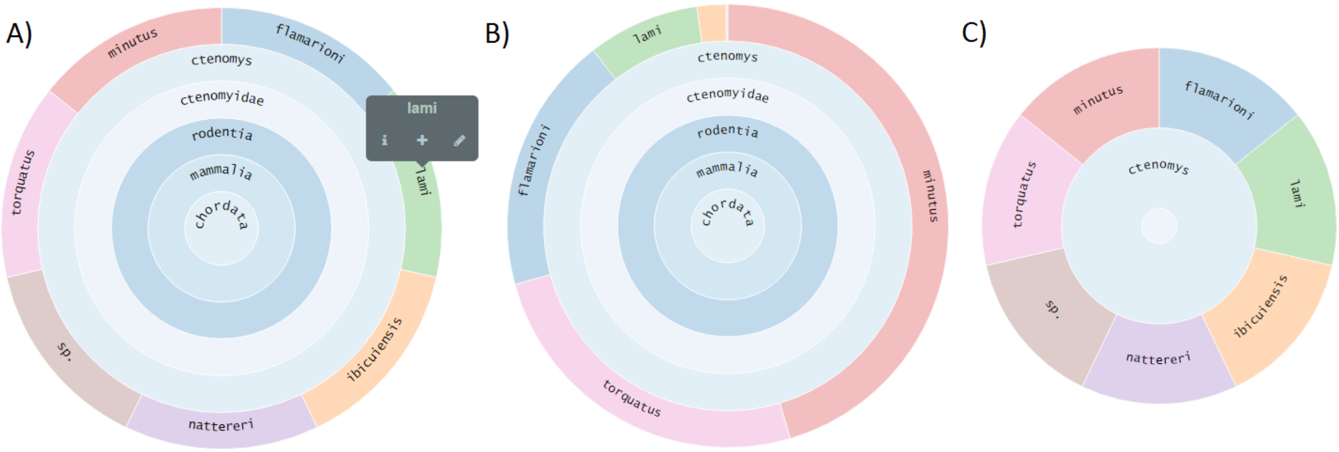


Figure 3. Examples of Sunburst visualization with the current database information: A) initial Sunburst layout, partitioned by the number of taxa and hovered on the 'lami' partition, showing its tooltip; B) Sunburst layout when partitioned by specimens; C) Sunburst partitioned by taxa when zoomed on the *Ctenomys* taxon.

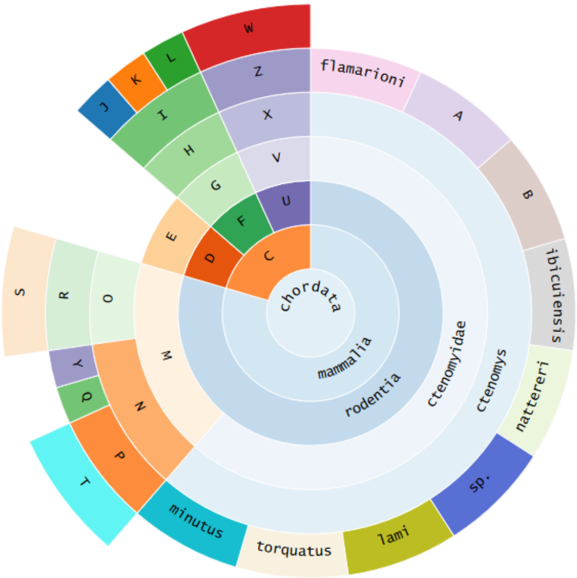


Figure 4. Sunburst visualization with taxa inserted to demonstrate how the visualization will look like when the database is properly populated. Nodes labeled 'C', 'N', 'minutus', 'lami' and 'sp' have been selected.

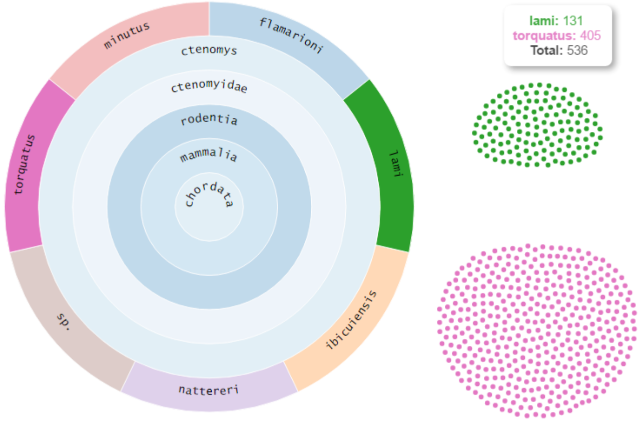


Figure 5. Sunburst visualization with the *torquatus* and *lami* species selected. Circles representing the specimens associated to these species are displayed on the right with a force-based layout. The tooltip displays the number of selected specimens when hovering over this area of the interface.

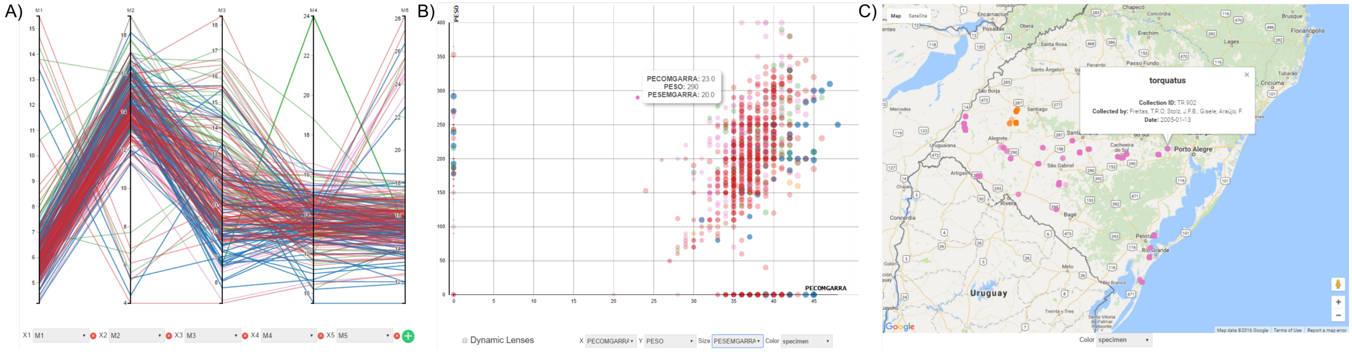
Accordingly, in this work, each vertical axis represents a single characteristic. The user can select the characteristic for each axis from the checkboxes at the bottom. Axes can be added dynamically by the user and viewed simultaneously, as can be seen in Figure 6A. Every line corresponds to one of the selected specimens that possess all the selected characteristics.

As with other implemented visualizations, the color represents the species of the specimen as defined on the Sunburst view and, when clicked, more information and possible actions can be seen on a pop-up window.

Scatterplot

According to Friendly and Denis [6], "of all the graphic forms used today, the scatterplot is arguably the most versatile, polymorphic, and generally useful invention in the history of statistical graphics". The most used scatterplot is a plot of two variables, usually indicated as x and y, measured independently to produce bi-variate pairs (xi, yi), and displayed as individual points on a coordinate grid, typically defined by Cartesian axes, where there is no necessary functional relation between x and y.

This visualization was implemented to be used as a ready-to-use solution for quickly analyzing characteristics of the selected specimens. Users can choose three characteristics to be displayed, one for each axis and another for defining the size of each marker. Each circle represents a specimen and can be hovered for displaying the exact values of the selected characteristics, as can be seen in Figure 6B. Users can also click on a circle to show a pop-up with all the specimen's information. From this pop-up it is also possible to edit or completely delete the specimen. The colors of each point, as it occurs in all implemented visualizations, represent the



PECOMGARRA: 23.0
PESO: 290
PESEMGARRA: 20.0

torquatus

Collection ID: TR.902
Collected by: Freitas, T.R.O; Stolz, J.F.B.; Gisele; Araújo, F.
Date: 2005-01-13

Figure 6. A) Parallel coordinates visualization with vertical axes representing characteristics 'M1', 'M2', 'M3', 'M4' and 'M5'. Each polyline crossing the axes represents a specimen. B) Scatterplot visualization allows selecting characteristics for each axis and circular marker attributes. In the example, the X axis encodes the 'PECOMGARRA' characteristic, the Y axis contains the 'PESO' characteristic and the size of each circle represents the 'PESEMGARRA' characteristic. Hovering on a circle shows a tooltip with the values of the three selected characteristics. C) Map visualization using Google Maps API. Circles represent specimens and, when hovering over a circle, a tooltip with the main information regarding that specimen is displayed.

color of the specimen's species, the same used in the Sunburst visualization.

Geospatial Visualization

A Geospatial visualization is very important for biologists in order to show the exact location of each captured specimen. For the implementation of this view we chose the Google Maps API, a widely used and simple solution for displaying maps on the web. If a specimen is selected and has longitude and latitude values, a point with the same color as its species in the Sunburst view is displayed at the corresponding location. When hovered, these points show the specimen's basic information: its collection ID, who collected it and the date of collection. When clicking on the marker, the same pop-up used in all visualizations for displaying specimen's information is shown, also allowing the user to change information and delete the specimen record from the database. This visualization is shown in Figure 6C.

Filtering

Users have highly varied needs for filtering features. By allowing users to control the contents of the display, they can quickly focus on their interests by eliminating unwanted items [16]. In this work, filtering is applied on specimens selected in the Sunburst visualization. When clicking the filtering button, located above the selected specimen's visualization, a pop-up with all filtering options is displayed. A filter can be defined for a specific characteristic of the specimens or their visibility (showing only specimens that are public, private or belonging to a certain research group). There is no limit in the number of filters that can be used simultaneously.

Final remarks

In this section we described the new interface for the TaxonomyBrowser. An overview of the whole system is available as a video posted at https://youtu.be/eYmcUOPDr50.

EVALUATION

In order to assess the new visualization-based interface, a remote survey was conducted. This evaluation aimed at measuring the users' understanding of the system and their ability to perform tasks unassisted.

The assessment involved 40 participants, 75% male and 25% female, with age between 19 and 58 years old. 32.5% of these participants are from the field of Biology, 52.5% are from Computer Science, and 15% are from other fields, such as Engineering, Health and Social Sciences. The participant's education levels were very varied: 47.5% undergraduate students, 22.5% graduated, 20% with a MSc degree, and 10% with a PhD degree. All participants had experience with web-based systems, while 40% had some experience with biological information systems. 85% of the participants believed they knew what a taxonomy tree is.

The first part of the survey had practical tasks to be accomplished by the subjects using the system in the form of questions to be answered. These tasks were as simple as finding which species had the largest number of specimens, selecting and deselecting species for visualization, and the name of the specimen with the higher or lower value in some measure. It was also asked how the users reached their results. Then, several sentences about the user's satisfaction with the interface features and visualizations were presented to be rated.

The final part of the survey was envisioned to measure the usability of the system by means of the System Usability Scale (SUS) [2]. This scale consists of 10 questions whose purpose is to provide an overview of subjective assessments of usability.

Results

Participants provided correct answers to most of the tasks/questions. As questions become more complex, the success rate tends to drop, as expected. The main results from the practical section of the questionnaire are summarized in Figure 7 and Table 1 in the Appendix. All users were able to find the name of the species with the largest number of specimens (T1), where 47.5% used the Sunburst's partitioning by specimens and 22.5% compared the number of circles from the Specimens' View after selecting all species. When describing how they managed to find how many specimens belong to a species (T2), 27.5% of users used the information button from the Sunburst's tooltip to check the number, 30% checked the value by hovering over the Specimens' View, and 20% used the analysis button.

Then, a filtering was asked to be performed by the users, and soon after that the total of the remaining specimens selected was asked (T4). 40% of the users hovered over the Specimens' View to check the number of specimens after the filter, and 20% used the analysis button. A small number of participants did not understand how to answer the question correctly. Only 0.75% of users had an erroneous idea of some of the application's features. One participant thought that zooming on a taxon provided the same effect as selecting, while another participant thought the answer was the number of different colors shown in the Specimens' View. One participant did not add any filters before answering the question.

The tasks labeled T6 and T7 asked for the participants the collection ID of the lightest and heaviest specimens. 80% checked on the Scatterplot which circles were on the extremities of the graph and clicked on them. One of them used the dynamic axis to make sure he wasn't selecting wrongly.

In the final task (T9), users had to count how many specimens had the measure 'PESO' between 150 and 250gr. 90% of the participants used the filters to answer the question. Considering all participants, 40% filtered and then opened the analysis pop-up to check the sum of all specimens selected, and only 10% hovered over the Specimens' View to check the value after filtering. Only 0.75% tried to count manually the number of circles on the Specimens' View, all failing to reach the correct answer. The remainder of participants did not provide a complete answer, only specifying that they used two filters and citing their parameters.

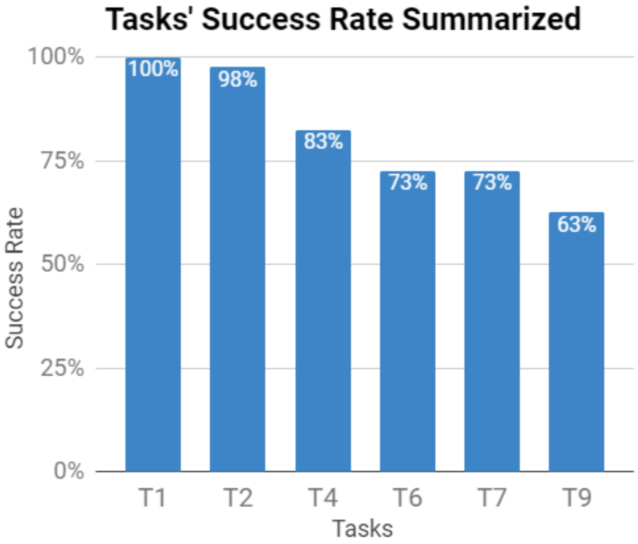


Figure 7. Summary of success rate in the practical tasks performed by experimental users. Answers were asked for tasks listed in the questionnaire shown in Table 1.

The results from questions about the user satisfaction with the interface design decisions (Section 3 of the questionnaire) are summarized in Figure 8 and Table 2 in the Appendix. In general, users liked and understood the proposed visualizations and the layout of the application. They particularly enjoyed the color scheme, which had 97.5% of approval by the participants. The only question slightly controversial was the system's response time, with 72.5% of participants satisfied, 15% neutral and 12.5% non-satisfied with the performance. This can be partially related to the variety of hardware possibly used for testing the tool, since it was a remote survey.

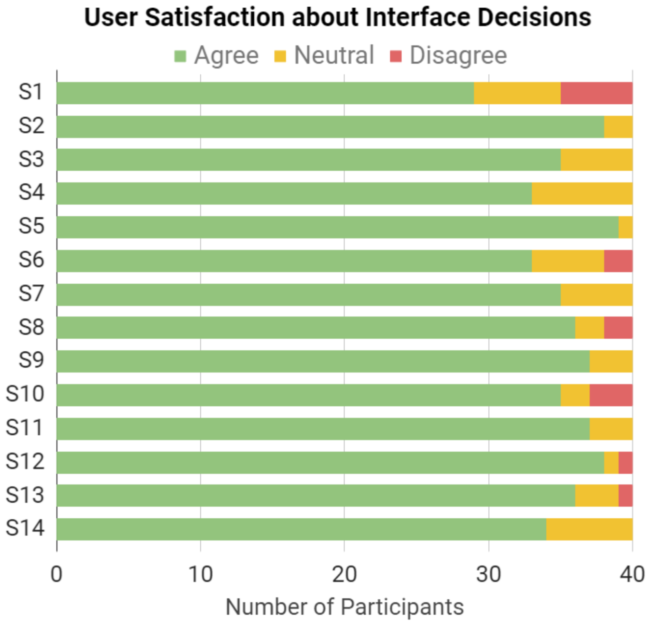


Figure 8. Summary of results from questions related to the user satisfaction about interface design decisions. The complete sentences and percentages can be seen in Table 2.

Regarding the System Usability Scale, the average SUS score was 78.3, which is above the average of 68 and close to an A grade score (above 80.3). The results for each SUS sentence can be observed in Figure 9. Feedback was mostly positive, specially regarding how well integrated the system's features were and its overall consistency. Also, 87.5% thought that they did not need to learn many things before they could operate with the system. The only truly controversial sentence was SUS.1: "I think that I would like to use this system frequently", with only 55% of agreement. This result can be explained since a significant number of participants were not biologists, and therefore would have no practical application for using *TaxonomyBrowser*. When analyzing only the answers by participants from Biology, 75% agreed they would like to use the system frequently, 12.5% were neutral and only 12.5% disagreed.

From these results we can conclude that the tool was generally visually interesting and understandable to all users, regardless of their field of study or age. Moreover, participants with practical applications for the tool were more interested in using the system frequently. The tasks results can be considered relevant, especially since it was the users' first experience with the application.

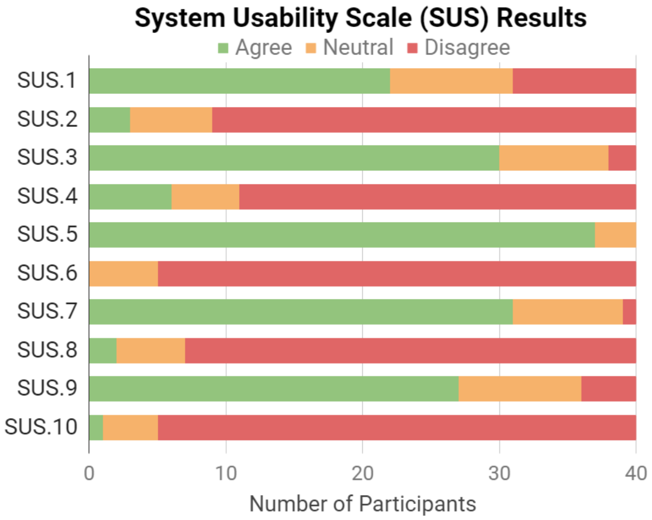


Figure 9. Summary of the answers obtained in the SUS questionnaire. In sentences 1, 3, 5, 7, and 9 it's considered a positive feedback when users agree to the phrase, while sentences 2, 4, 6, 8 and 10 it's considered a positive feedback when users disagree. The complete sentences and percentages can be seen in Table 3.

FINAL REMARKS

This work presented a new interface for *TaxonomyBrowser* based on Shneiderman's Visual Information Seeking Mantra: overview first, zoom and filter, then details on demand. A Sunburst visualization was implemented for displaying an overview of the taxonomy tree and for managing the database information, such as its taxon and recorded specimens. The user's selected and/or filtered information can be viewed on different visualizations, besides being displayed as small circles, for a better understanding of the size and characteristics of the set of filtered specimens.

The interface was assessed by means of a remote survey based both on specific tasks and a questionnaire. The evaluation yielded promising results, specially considering that most participants were unfamiliar with the tool and had no external assistance. The answers and suggestions provided by the participants have already allowed improvements in the application. Considering the wide range of hardware used by the participants, it was possible to receive feedback on performance issues and tweak the system accordingly.

These preliminary results demonstrate that even users without any experience with biological databases could use and obtain satisfactory results from the tool. This also indicates that the interface is overall intuitive to potential users.

In order to improve this work, more visualization techniques can be easily embedded in the application. Moreover, it would be interesting to build a mobile version of the interface, in order to facilitate the recording of new specimens during field work.

ACKNOWLEDGMENTS

We thank CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and FAPERGS (Fundação do Amparo à Pesquisa do Estado do Rio Grande do Sul), federal and state funding agencies, respectively, for financial support. We are grateful to the members of the team that built the first version of *TaxonomyBrowser*, which helped us to shape the new interface, and also to the subjects that assessed our application and provided insightful comments. Finally, we thank our lab colleagues for providing many suggestions that improved this paper.

REFERENCES

1. Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D3 data-driven documents. *IEEE transactions on Visualization and Computer Graphics* 17, 12 (2011), 2301–2309. DOI: `http://dx.doi.org/10.1109/TVCG.2011.185`
2. John Brooke. 1996. SUS - A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7. DOI:`http://dx.doi.org/10.1002/hbm.20701`
3. Samantha Castro Cañete. 2011. Interface de gerenciamento e consultas visuais em banco de dados de biodiversidade. PPGC da UFRGS (MSc Dissertation). (2011). `http://www.lume.ufrgs.br/handle/10183/29014`
4. Vanderlei Perez Canhos. 2005. Sistema de Informação Distribuído para Coleções Biológicas : A Integração do Species Analyst e SinBiota. Relatório Técnico Final do projeto Species Link. (October 2005). `http://splink.cria.org.br/docs/outubro2005.pdf`
5. Henrik Enghoff. 2009. What is taxonomy? - An overview with myriapodological examples. *Soil organisms* 81, 3 (2009), 441–451.
6. Michael Friendly and Daniel Denis. 2005. The early origins and development of the scatterplot. *Journal of the History of the Behavioral Sciences* 41, 2 (2005), 103–130. DOI:`http://dx.doi.org/10.1002/jhbs.20078`
7. Rafael Henkin. 2010. Interface de consultas analíticas para bases de dados de biodiversidade. INF/UFRGS (BSc Dissertation). (2010). `http://www.lume.ufrgs.br/bitstream/handle/10183/26354/000757817.pdf?sequence=1`
8. Rafael Henkin. 2013. A study on visual analysis of georeferenced haplotype networks. PPGC da UFRGS (MSc Dissertation). (2013).
9. Daniel H Huson, Daniel C Richter, Christian Rausch, Tobias Dezulian, Markus Franz, and Regula Rupp. 2007. Dendroscope: An interactive viewer for large phylogenetic trees. *BMC bioinformatics* 8, August 2016 (2007), 460. DOI: `http://dx.doi.org/10.1186/1471-2105-8-460`
10. Alfred Inselberg and Bernard Dimsdale. 1990. Parallel Coordinates: A Tool for Visualizing Multi-dimensional Geometry. In *Proceedings of the 1st Conference on Visualization '90 (VIS '90)*. IEEE Computer Society Press, 361–378. `http://dl.acm.org/citation.cfm?id=949531.949588`
11. Intro.js. 2016. Step-by-step Guide and Feature Introduction for your Website. `http://introjs.com/`. (2016). (Accessed on 18/11/2016).

12. Bongshin Lee, Cynthia Sims Parr, Dana Campbell, and Benjamin B. Bederson. 2004. How Users Interact with Biodiversity Information Using TaxonTree. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '04)*. ACM, 320–327. DOI: http://dx.doi.org/10.1145/989863.989918

13. Ivica Letunic and Peer Bork. 2007. Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* 23, 1 (2007), 127–128. DOI:http://dx.doi.org/10.1093/bioinformatics/btl529

14. K. T. McDonnell and K. Mueller. 2008. Illustrative parallel coordinates. *Computer Graphics Forum* 27, 3 (2008), 1031–1038. DOI: http://dx.doi.org/10.1111/j.1467-8659.2008.01239.x

15. Brian D Ondov, Nicholas H Bergman, and Adam M Phillippy. 2011. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 12, 1 (2011), 385. DOI:http://dx.doi.org/10.1186/1471-2105-12-385

16. B. Shneiderman. 1996. The eyes have it: a task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages* (1996), 336–343. DOI: http://dx.doi.org/10.1109/VL.1996.545307

17. Jerônimo Silva. 2007. Projeto e Desenvolvimento de Sistema Web para Armazenamento de Coleções de Espécimes. Porto Alegre, INF/UFRGS (BSc Dissertation). (2007).

18. Jorge Soberón and Townsend Peterson. 2004. Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 359, 1444 (2004), 689–698. DOI: http://dx.doi.org/10.1098/rstb.2003.1439

19. John Stasko, Richard Catrambone, Mark Guzdial, and Kevin McDonald. 2000. An evaluation of space-filling information visualizations for depicting hierarchical structures. *International Journal of Human-Computer Studies* 53, 5 (2000), 663–694. DOI: http://dx.doi.org/10.1006/ijhc.2000.0420

20. John Stasko and Eugene Zhang. 2000. Focus+Context Display and Navigation Techniques for Enhancing Radial, Space-Filling Hierarchy Visualizations. In *Proceedings of the IEEE Symposium on Information Vizualization 2000 (INFOVIS '00)*. IEEE Computer Society, 57–65. http://dl.acm.org/citation.cfm?id=857190.857683

21. Shahzaib Tahir and Muhammad Tanvir Afzal. 2014. A novel phylogenetic tree data visualization application for researchers. *Proceedings of 2014 Science and Information Conference, SAI 2014* August (2014), 93–99. DOI:http://dx.doi.org/10.1109/SAI.2014.6918177

22. D.L.M. Tavares, S.C. Canete, P.C. Estrela, R. Henkin, T.R.O. Freitas, R. Galante, and C.M.D.S. Freitas. 2011. TaxonomyBrowser: a biodiversity data management system. *Journal of Computational and Interdisciplinary Science* 2 (2011), 37–46. http://www.inf.ufrgs.br/

23. Toolbar.js. 2016. A jQuery plugin that creates tooltip style toolbars. http://paulkinzett.github.io/toolbar/. (2016). (Accessed on 18/11/2016).

24. Jing Yang, Matthew O Ward, Elke A Rundensteiner, and Anilkumar Patro. 2003. InterRing: A Visual Interface for Navigating and Manipulating Hierarchies. *Information Visualization* 2, 1 (2003), 16–30. DOI: http://dx.doi.org/10.1057/palgrave.ivs.9500035

APPENDIX

The following tables contain the success rate shown by subjects in performing the tasks (Table 1), the sentences included in our survey and respective results (Table 2), and the results for the SUS questionnaire (Table 3).

Task	*Summarized Question*	*Success Rate*
T1	What is the species with most specimens?	100%
T2	How many specimens this species has?	97.5%
T3	Describe briefly how you reached this answer	—
T4	Select only the species 'Ctenomys lami'. Add a new filter testing if the parameter 'Data' exists. How many specimens match the filter?	82.5%
T5	Describe briefly how you reached this answer	—
T6	Remove the filter. Select all species. Use the Scatterplot graph to view the parameter 'PESO' (weight). What is the collection ID of the lightest specimen (ignoring null and '0' values)?	72.5%
T7	What is the collection ID of the heaviest specimen?	72.5%
T8	Describe briefly how you reached this answer	—
T9	Out of all specimen registered in the database, how many have the parameter 'PESO' between 150 and 250 (including specimen with exactly 150 and 250)?	62.5%
T10	Describe briefly how you reached this answer	—

Table 1. Success rate in performing the tasks: results are the percentage of correct answers to the questions posed to subjects that participated in the assessment.

ID	*Sentence*	*Agree*	*Neutral*	*Disagree*
S1	I think the tool has good response time.	72.5%	15%	12.5%
S2	I think the chosen visualizations are adequate.	95%	5%	0%
S3	I think the proposed visualizations allow a good understanding of the data base	87.5%	12.5%	0%
S4	I think the menus for each visualization are adequate	82.5%	17.5%	0%
S5	I think the color scheme is pleasant.	97.5%	2.5%	0%
S6	I liked the taxonomy tree visualization (Sunburst).	82.5%	12.5%	5%
S7	I liked the visualization of the selected specimens (circles on the center of the screen).	87.5%	12.5%	0%
S8	I liked and understood the Scatterplot visualization.	90%	5%	5%
S9	I liked and understood the map visualization.	92.5%	7.5%	0%
S10	I found the filtering method intuitive.	87.5%	5%	7.5%
S11	I found the options of visualization, search and comparison of data adequate.	92.5%	7.5%	0%
S12	I think the layout of the system is appealing.	95%	2.5%	2.5%
S13	I found the tool interesting.	90%	7.5%	2.5%
S14	I found the tool effective in its features.	85%	15%	0%

Table 2. Results from the survey measuring user satisfaction in relation to interface decisions.

ID	*Sentence*	*Agree*	*Neutral*	*Disagree*
SUS.1	I think that I would like to use this system frequently.	55%	22.5%	22.5%
SUS.2	I found the system unnecessarily complex.	7.5%	15%	77.5%
SUS.3	I thought the system was easy to use.	75%	20%	5%
SUS.4	I think that I would need the support of a technical person to be able to use this system.	15%	12.5%	72.5%
SUS.5	I found the various functions in this system were well integrated.	92.5%	7.5%	0%
SUS.6	I thought there was too much inconsistency in this system.	0%	12.5%	87.5%
SUS.7	I would imagine that most people would learn to use this system very quickly.	77.5%	20%	2.5%
SUS.8	I found the system very cumbersome to use.	5%	12.5%	82.5%
SUS.9	I felt very confident using the system.	67.5%	22.5%	10%
SUS.10	I needed to learn a lot of things before I could get going with this system.	2.5%	10%	87.5%

Table 3. Results from the SUS questionnaire.