

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

MARINA FORTES REY

**A Visualization-based Approach for the
TaxonomyBrowser Interface**

Work presented in partial fulfillment
of the requirements for the degree of
Bachelor in Computer Science

Advisor: Prof. Dr. Carla M.D.S. Freitas

Porto Alegre
December 2016

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Rui Vicente Oppermann

Vice-Reitora: Prof^a. Jane Fraga Tutikian

Pró-Reitor de Graduação: Prof. Vladimir Pinheiro do Nascimento

Diretora do Instituto de Informática: Prof^a. Carla Maria Dal Sasso Freitas

Coordenador do Curso de Ciência de Computação: Prof. Sérgio Luis Cechin

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

ABSTRACT

The process of collecting biological data is a continuous activity, specially in field work for university projects and courses. Data from these collections needs to be properly stored in order to be readily available for future analysis. This is particularly important since a significant amount of data tends to be lost when the students responsible for collection leave their undergraduate or graduate programs and do not share their contributions for future projects. These students are frequently worried about storing their work on a public database. *TaxonomyBrowser* is a biodiversity information system developed to manage data collected by biologists during field work, as well as specimens registered in museum collections. This work describes a new approach on the *TaxonomyBrowser's* user interface, focusing on providing an easier and more intuitive method of visualizing and managing the database information. The interface layout was based on Shneiderman's Visual Information Seeking Mantra: overview first, zoom and filter, then details on demand. We limited the access to certain specimen's information depending on the user's authentication by altering the original database structure in order to associate each specimen created to accounts or groups of users.

Keywords: Biodiversity information systems. information visualization. taxonomy.

Uma Abordagem Baseada em Visualização para a Interface do *TaxonomyBrowser*

RESUMO

O processo de coletar dados biológicos é uma atividade contínua, especialmente em trabalhos de campo realizados para projetos e disciplinas na universidade. Dados dessas coleções precisam ser guardados apropriadamente para que estejam disponíveis em caso de futuras análises. Isso é particularmente importante já que uma quantidade significativa de dados tende a ser perdido quando os alunos responsáveis pela coleção desligam-se da universidade sem compartilhar suas contribuições para trabalhos futuros. Esses estudantes estão frequentemente apreensivos em registrar o seu trabalho em um banco de dados de acesso público. *TaxonomyBrowser* é um sistema de informação de biodiversidade desenvolvido para o gerenciamento de dados coletados por biólogos em trabalhos de campo, como também de espécimes cadastrados em museus. Esse trabalho descreve uma nova abordagem à interface do *TaxonomyBrowser*, focando em prover um método mais fácil de visualizar e gerenciar informação do banco de dados. O layout da interface foi baseado no mantra de visualização de informação de Shneiderman: primeiramente visão geral, zoom e filtragem, e então detalhes por demanda. O acesso da informação de certos espécimes foi limitado dependendo do nível de autenticação do usuário acessando o sistema por meio de alterações na estrutura do banco de dados para associar cada espécime com contas e grupos de usuário.

Palavras-chave: sistema de informação de biodiversidade, visualização de informação, taxonomia.

LIST OF FIGURES

Figure 2.1 Example of taxonomy tree for the order Carnivora.....	14
Figure 2.2 Structure of a phylogenetic tree.....	15
Figure 2.3 Notebook page used in field work to record information about collected specimens or samples.....	15
Figure 3.1 SinBiota map visualization interface showing collection points and information about a selected specimen.	16
Figure 3.2 Types of visualization provided by the Interactive Tree of Life. a) normal (standard tree visualization); b) circular; c) unrooted (radial)	18
Figure 3.3 An example of taxonomic abundance, colored by classification confidence, built using Krona's visualization tool	19
Figure 3.4 Four types of visualization implemented by Dendroscope. a) circular cladogram; b) radial phylogram; c) rectangular phylogram; d) slanted cladogram.....	20
Figure 4.1 Interface developed for the TaxonomyBrowser. The buttons on the top refer to the visualizations below them.	22
Figure 4.2 Simplified entity-relationship diagram of the database model	24
Figure 4.3 Simplified diagram of the system's data flow. Data about specimens from all species are loaded to the Sunburst visualization and, upon selection by the user, part of the dataset feeds the other visualizations with the chosen specimens.....	25
Figure 4.4 Simplified Use Case Diagram for the three possible types of users. Standard login interaction not shown.....	27
Figure 4.5 Filtering pop-up with four filters specified. The first and second line indicate that all entries should have the value '2N' and 'CAIXA' recorded. The third and fourth line limit the value of the attribute 'PESO' between 150 and 250.	29
Figure 4.6 Characteristic manager pop-up. Characteristics can be added, edited or removed by an administrator user. Recorded characteristics cannot have their type changed in order to avoid invalid data in the database.....	30
Figure 4.7 The Sunburst visualization with the current database information in some of its shapes: 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.	33
Figure 4.8 Sunburst visualization with taxa inserted to demonstrate how the visualization will look like when the database is properly populated. Nodes labeled 'C', 'N', 'minusus', 'lami' and 'sp.' have been selected.....	34
Figure 4.9 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 this area of the interface.....	35
Figure 4.10 Specimens' View with different numbers of selected specimens: A) 34 specimens selected from one species; B) 165 specimens selected from two species; C) 1600 specimens selected from 7 different species.	36

Figure 4.11 Scatterplot visualization allows selecting characteristics for each axis and circle 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.....	37
Figure 4.12 Scatterplot with the Dynamic Lenses feature activated, produce a fish-eye effect on the axis relative to the mouse position. In this case the mouse was centered between 35 and 40 of the X axis and 200 and 250 of the Y axis.....	38
Figure 4.13 Map visualization using Google Maps API. Circles represent specimens and, when hovered, a tooltip with its main information is displayed.....	39
Figure 4.14 Parallel coordinates visualization with columns representing characteristics 'M1', 'M2', 'M3', 'M4' and 'M5'. Each line represents a specimen.....	40
Figure 4.15 Parallel coordinates with brushing applied on its axis. 'M1' is limited between 5 and 6, 'M2' is between 14 and 16, 'M3' is between 13 and 16 and 'M4' is limited between 14 and 16.	41
Figure 5.1 Taxonomy tree information in the previous version of Taxonomy-Browser. Only the direct children of each taxon are shown at a time. To check each taxon's children the user must click on their parent, loading a new web page each time.....	45
Figure 5.2 Alphabetical list of all analysis saved in the old version of the interface. It does not link to the full information of the specimens analysed, only displaying how many of them matched the filter.	47
Figure 5.3 Alphabetical list of all specimens recorded on the database available in the previous version of the interface. This is the only alternative users had to check a specimen's complete information.	48
Figure 6.1 Summary of success rate results from the tasks with objective answers asked on the questionnaire	58
Figure 6.2 Summary of results from questions on the user satisfaction about interface decisions	60
Figure 6.3 Summary of results from the SUS section.....	61
Figure 6.4 Summary of results from the SUS section.....	62
Figure 6.5 Previous style on the left compared to the new proposed style for selection on the Sunburst visualization on the right	62
Figure 6.6 Breadcrumbs and the Sunburst visualization, showing all parent nodes of the zoomed-in taxa.....	63
Figure A.1 Pop-up with information from a taxon. When clicking on the edit button, information such as the taxon's name, information and characteristics can be altered.....	68
Figure A.2 Pop-up with information from a specimen. Each tab represents a group of characteristics that are populated for this particular specimen. Data can be altered or edited with the edit button on the upper right of the pop-up.	69
Figure A.3 Pop-up with information from the selection made, such as the number of specimens selected for each species and an analysis on a certain characteristic selected by the user.....	70
Figure A.4 User login pop-up. Here you can also register a new account.....	70

LIST OF TABLES

Table 5.1 Summary of the comparison between the previous version of the system with the one implemented in this work based on Shneiderman's seven tasks for information visualization.	42
Table 6.1 Summary of success rate results from the tasks with objective answers asked on the questionnaire	57
Table 6.2 Summary of results from questions on the user satisfaction about interface decisions	59
Table 6.3 Summary of results from the SUS section	61
Table B.1 Personal information questions from the survey	71
Table B.2 Practical Tasks from the survey	71
Table B.3 User satisfaction questions from the survey	72
Table B.4 SUS questions from the survey.....	73

LIST OF ABBREVIATIONS AND ACRONYMS

FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
iTOL	Interactive Tree of Life
HTML	Hypertext Markup Language
SVG	Scalable Vector Graphics
JS	JavaScript
D3	Data Driven Documents
JSON	JavaScript Object Notation
RSF	Radial, Space Filling
SUS	System Usability Scale

CONTENTS

1 INTRODUCTION	11
1.1 Objectives.....	12
1.2 Structure of the text.....	12
2 BACKGROUND	13
2.1 Taxonomy Tree.....	13
2.2 Phylogenetic Tree.....	13
2.3 Biodiversity Information Systems.....	14
3 RELATED WORK	16
3.1 SinBiota.....	16
3.2 Interactive Tree of Life (iTOL).....	17
3.3 Krona.....	18
3.4 Dendroscope.....	19
3.5 Final Remarks.....	20
4 VISUALIZATION-BASED INTERFACE	22
4.1 Data Model.....	23
4.2 Application Architecture and Dataflow.....	24
4.3 User Authentication.....	26
4.3.1 Anonymous User.....	26
4.3.2 Standard User.....	27
4.3.3 Administrator User.....	28
4.3.4 Research Groups.....	28
4.4 Filtering.....	29
4.5 Characteristics Manager.....	30
4.6 Exporting.....	31
4.7 Visualization Techniques.....	31
4.7.1 Sunburst.....	31
4.7.2 Specimens' View.....	34
4.7.3 Scatterplot.....	35
4.7.4 Map.....	38
4.7.5 Parallel Coordinates.....	39
4.8 Final remarks.....	41
5 DISCUSSION	42
5.1 Assessing the Interface regarding the Shneiderman's Seven Tasks for In-formation Visualization	42
5.1.1 Overview.....	44
5.1.2 Zoom.....	45
5.1.3 Filter.....	46
5.1.4 Details-on-Demand.....	46
5.1.5 Relate.....	47
5.1.6 Extract.....	48
5.2 Essential Operations for Interaction with Hierarchically Constructed Data ... 49	49
5.2.1 Selection.....	49
5.2.2 Reconfiguration.....	49
5.2.3 Drill-down/Roll-up.....	49
5.2.4 Pan, Zoom, Rotation.....	50
5.2.5 Distortion.....	50
6 EVALUATION	51
6.1 Participants.....	51

6.2 Tasks and initial setup	51
6.3 Evaluation procedure	54
6.4 Results	56
7 CONCLUSIONS AND FUTURE WORK	64
REFERENCES.....	65
APPENDIX A — INTERFACE POP-UPS	68
APPENDIX B — SURVEY	71

1 INTRODUCTION

Computational tools have become essential in all scientific areas. This is also true in biology and ecology where they help the management, exploration, discovery, analysis and presentation of biological information (BARVE; OTEGUI, 2016). This process of collecting biological data takes place constantly, specially in field work for university projects and classes. It is also a very manual and tedious work, which demands that the researcher collects basic data from the specimen on the field, catalogues and stores them for further visualization. This entry can also be later modified when new laboratory studies are made on the collected samples, where it is fundamental to have a system capable of easily finding and altering previously stored specimens.

Still a vast number of these collections are stored inefficiently, such as in local spreadsheets, convoluting the process of sharing and updating this information. In order to preserve these data it is important to unify their storage in a biodiversity information system. This enables a variety of new studies to be produced by facilitating exchange of data between students and researchers.

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 from mammals collected by biologists all over the state (SILVA; FREITAS, 2007; HENKIN, 2010; TAVARES; CANETE, 2011; CAÑETE, 2011; HENKIN, 2013). Another important goal of *TaxonomyBrowser* was to catalogue specimens found in museums collections (TAVARES; CANETE, 2011). *TaxonomyBrowser* was conceived as a web application, and it has been available on-line since its first version. Users are able to add, edit and remove specimen and taxa only if they are authenticated by the system. Anonymous users can only view the information recorded.

The tool allows to filter specimens based on certain properties and show them by their geographic location as markers using the Google Maps' API. The user can draw lines and polygons on top of the map to select a group of specimens, although information about the specimens shown cannot be altered or deleted. This selection can be manipulated via R scripts, very commonly used among biologists, embedded on the web application.

The main motivation for the present work is to provide an intuitive way of accessing and querying data from the database. To accomplish this we have two main focuses: simplify the number of interactions users need to perform in order to perform certain actions and display a complete overview of the system constantly, while also showing

specific information if requested by the user.

1.1 Objectives

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 data in a clear and intuitive way, aiming at reducing the number of interactions required by the user to perform certain actions on the database.

The objectives pursued in this work are summarized below:

- Develop an intuitive system with a clear overview of all stored data using information visualization techniques for user interaction;
- Minimize the number of interactions necessary for managing information;
- Create a consistent and responsive user interface;
- Restrain the access to certain data stored in the database depending on the user's authentication;
- Facilitate selecting and filtering data without necessarily being a specialist on the field.

1.2 Structure of the text

The remainder of this dissertation is organized as follows. In Chapter 2 we present some important biological concepts for understanding the work. Next, in Chapter 3, we describe some of the related applications and on-line portals. Chapter 3 shows the implementation and features of the application. In Chapter 4 we go through the features of the work and the implementation decisions. Next, Chapter 5 provides a discussion and comparison, using a number of consolidated parameters, between the old and the new interface of the *TaxonomyBrowser*. Chapter 6 describes the user test applied and its results.

2 BACKGROUND

Certain biology concepts need to be clarified in order to understand the terminology of this work. Such concepts include the notion of taxonomy tree, which our tool uses in order to organize data from the collected specimens. The difference between taxonomic and phylogenetic trees is often confused by the general public and will be clarified in the next sections. The definition and usage of a biological information system is also presented.

2.1 Taxonomy Tree

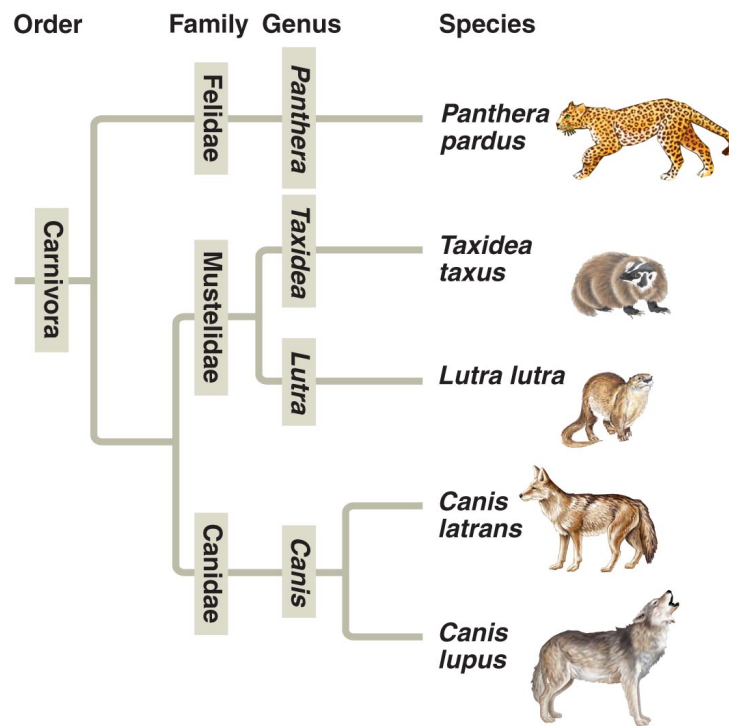
Enghoff's paper, 'What is Taxonomy? - An overview with myriapodological examples' (ENGHOFF, 2009) helps us understand the concept of a taxonomy tree. Etymologically, the word taxonomy is derived from Greek taxis, meaning 'arrangement or division', and nomos, meaning 'law'. According to Enghoff, taxonomy can thus be understood as meaning '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, as can be seen in Figure 2.1. These taxa 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) (LEE et al., 2004). The taxon's attributes are inherited by its 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.

2.2 Phylogenetic Tree

Formally a phylogenetic tree is a construction that attempts to form the ancestors and descendants relationships for a set of entities (TAHIR; AFZAL, 2014). They have 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.

Figure 2.1 – Example of taxonomy tree for the order Carnivora.



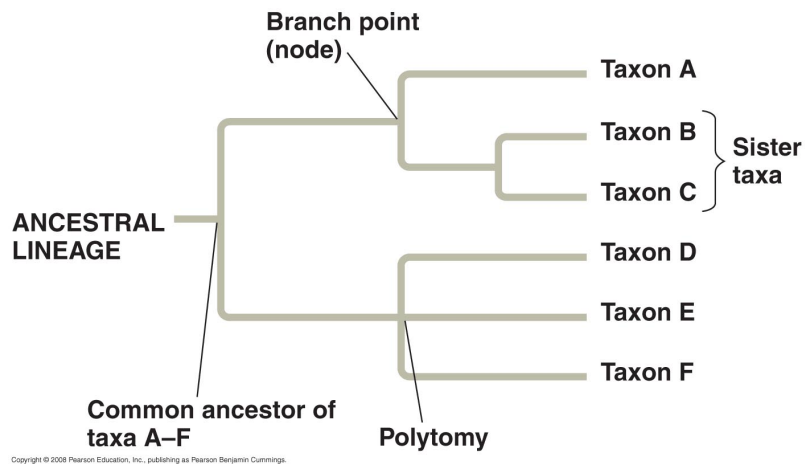
Source: <http://ksuweb.kennesaw.edu/~jdirnber/Bio2108/Lecture/LecPhylogeny/LecPhylogeny.html>

For this reason, internal nodes are rarely changed, while leaf nodes vary more constantly. Figure 2.2 illustrates this concept.

2.3 Biodiversity Information Systems

Biodiversity Informatics includes the application of information technologies to the management, algorithmic exploration, analysis and interpretation of primary data regarding life, particularly at the species level of organization (SOBERÓN; PETERSON, 2004). Biodiversity information systems consist of a database to store 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 specimen is written down in notebook pages as the one seen in Figure 2.3. 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 animal body.

Figure 2.2 – Structure of a phylogenetic tree.



Source: <http://ksuweb.kennesaw.edu/~jdimer/Bio2108/Lecture/LecPhylogeny/LecPhylogeny.html>

Figure 2.3 – Notebook page used in field work to record information about collected specimens or samples.

DEPARTAMENTO DE GENÉTICA UFRGS		tab. 51
Nº 689	SEXO ♀	PESO 245g
LOCAL	DATA	
HORA 13:20h	TEMPERATURA	
C. CORPO 291	LOCALIZ. A	③ (com filhote ↓ me toca) mas não pegamos ele!
C. CAUDA 89	B	
C. ps/cu 44	GPS	
C. ps/ss 37	OBS. Oculhada (amarelada)	
INCISIVO 2,65		
DEPARTAMENTO DE GENÉTICA UFRGS		tab. 52
Nº 60483	SEXO ♀	PESO 270g
LOCAL	DATA	
HORA 14:20h	TEMPERATURA	
C. CORPO 289	LOCALIZ. A	④
C. CAUDA 92	B	
C. ps/cu 44	GPS	
C. ps/ss 37	OBS. CICATRIZADA	
INCISIVO 2,83		
DEPARTAMENTO DE GENÉTICA UFRGS		tab. 52
Nº 690	SEXO ♀	PESO 130g
LOCAL	DATA	
HORA 15:05h	TEMPERATURA	
C. CORPO 236	LOCALIZ. A	⑤
C. CAUDA 76	B	
C. ps/cu 36	GPS	
C. ps/ss 33	OBS. IMPERFORADA	
INCISIVO 2,03		

Source: (HENKIN, 2013). Courtesy of T.R.O.Freitas, Dept. of Genetics, UFRGS.

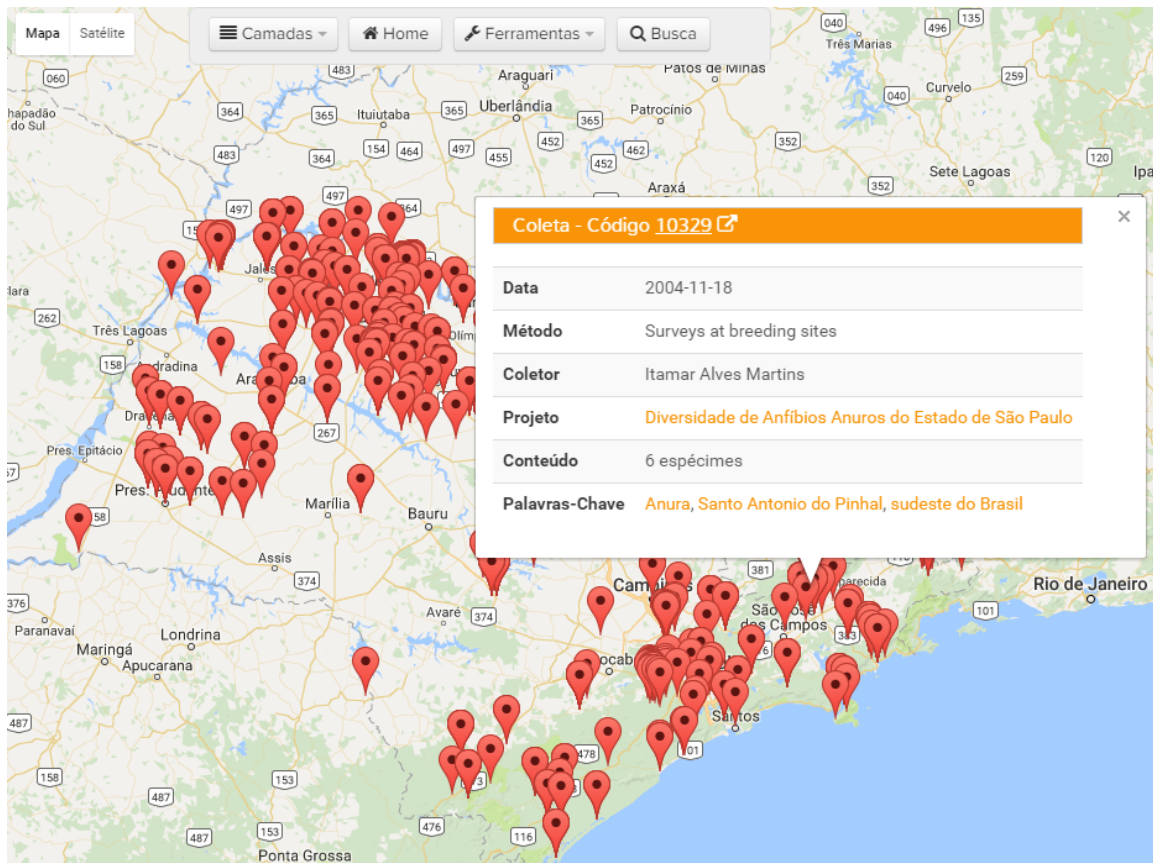
3 RELATED WORK

In this chapter we will review similar visualization applications, most using phylogenetic or taxonomic trees to represent information about species.

3.1 SinBiota

SinBiota ("Sistema de Informação Ambiental do Biota") was created by the Biota project (FAPESP) 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 (CANHOS, 2005). The database is very similar to the one used in this work, containing for each collected specimen its location, date and the name of researchers responsible for the capture.

Figure 3.1 – SinBiota map visualization interface showing collection points and information about a selected specimen.



Source: available at <http://sinbiota.biota.org.br/atlas>

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. It also shows information about regional ecosystems in São Paulo. To visualize the collected specimens the user must first select a certain area of interest in the map. Only in this specific area markers representing each collected sample will be shown. Another way to visualize the specimens in the system is to filter by a certain restriction. However, only one restriction can be made for each search. When left clicking a marker, information from that collected specimen such as date of collection and its related project can be seen, as shown in Figure 3.1.

As for the visualization of the (partial) taxonomy tree related to the specimens entered on the database, the system lacks any way to overview the data, presenting only a list with taxonomy nodes that, when clicked, lists its children, and so on.

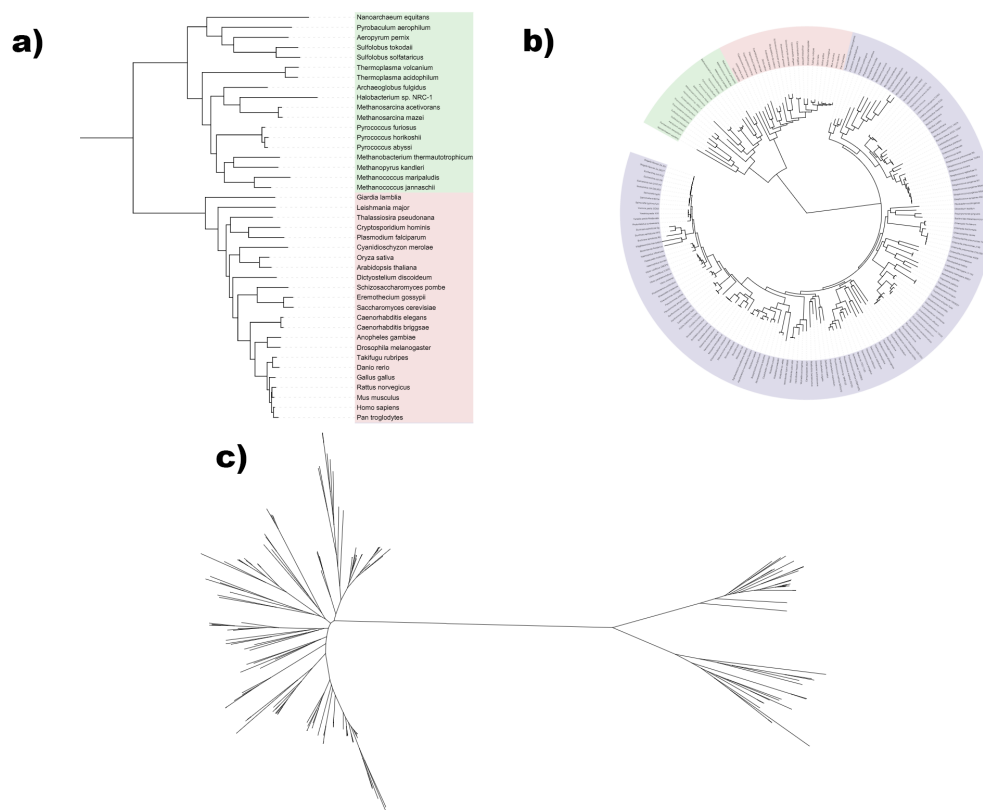
3.2 Interactive Tree of Life (iTOL)

The Interactive Tree of Life is an on-line tool for the display, annotation and management of phylogenetic trees (LETUNIC; BORK, 2007). Data can be entered by users using a plain text file on any of the supported formats. Users can manage multiple trees and share their workspace with other researchers, all available on an on-line interface developed using Javascript and HTML5 (ITOL. . . ,).

The system has 3 types of tree visualizations available, as shown in 3.2: normal (standard indented tree visualization), circular and unrooted (radial). The visualization can be zoomed by scrolling or using the assigned buttons on the interface, can be moved by dragging the visualization, and can display information by left clicking on any node. An important feature of this system is the creation of a pruned tree by selecting each node manually from the original tree. Other visualization details can be changed, such as the size and color of the font used to display leaf node's name and stroke used to link nodes, the visualization's rotation and size, and information of each internal node.

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. Other visualizations are available, such as pie charts, when each node is associated with multiple numeric values. There are also text labels to describe a single text for multiple species (a taxonomic group's name, for example), colored gradients and strips for grouping, con-

Figure 3.2 – Types of visualization provided by the Interactive Tree of Life. a) normal (standard tree visualization); b) circular; c) unrooted (radial)



Source: <http://itol.embl.de/help.cgi>

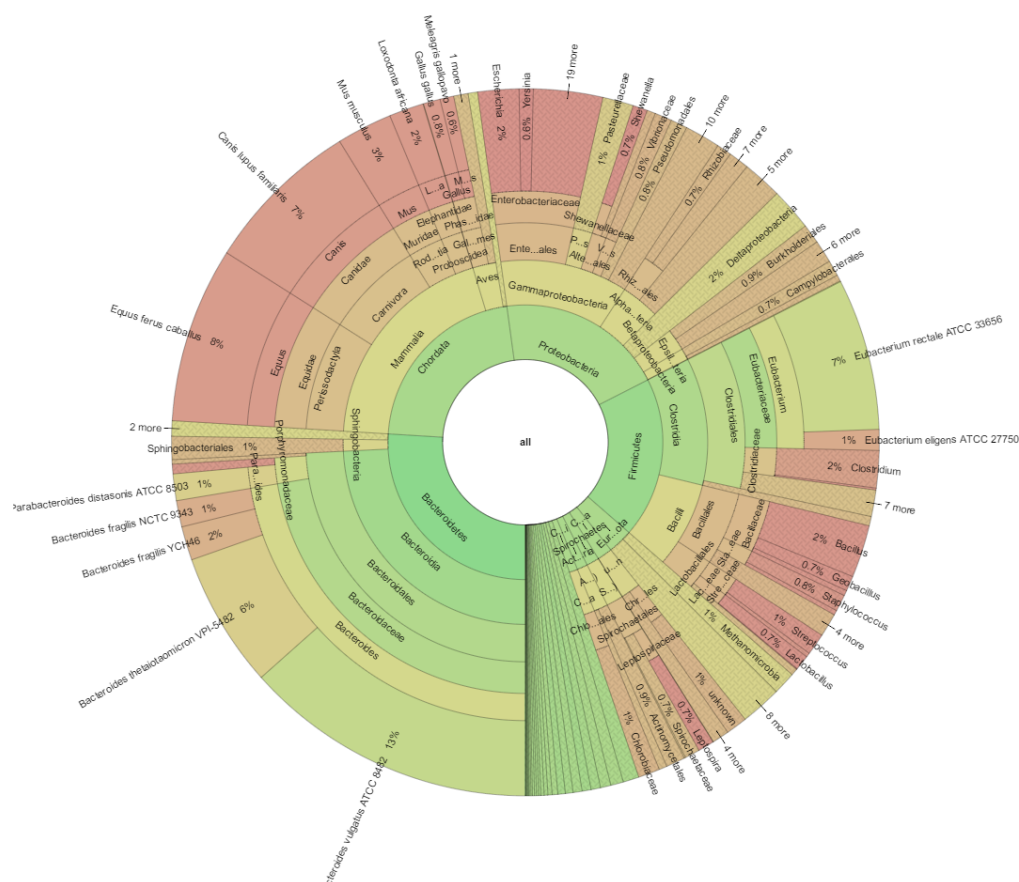
nections describing interactions between nodes, and heat maps for associating each node with several numeric values.

3.3 Krona

Krona is a visualization tool that allows intuitive exploration of relative abundances and confidences within the complex hierarchies of meta-genomic classifications (ONDOV; BERGMAN; PHILLIPPY, 2011). The system was developed using a hybrid approach in terms of application code and data storage, having data stored locally but the system available on-line. Krona uses a Radial, Space Filling (RSF) visualization (Figure 3.3), 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 on 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. This solution, despite being useful for removing extra interactions required in order to discover the names of smaller sectors, can confuse the overall visualization.

Figure 3.3 – An example of taxonomic abundance, colored by classification confidence, built using Krona's visualization tool



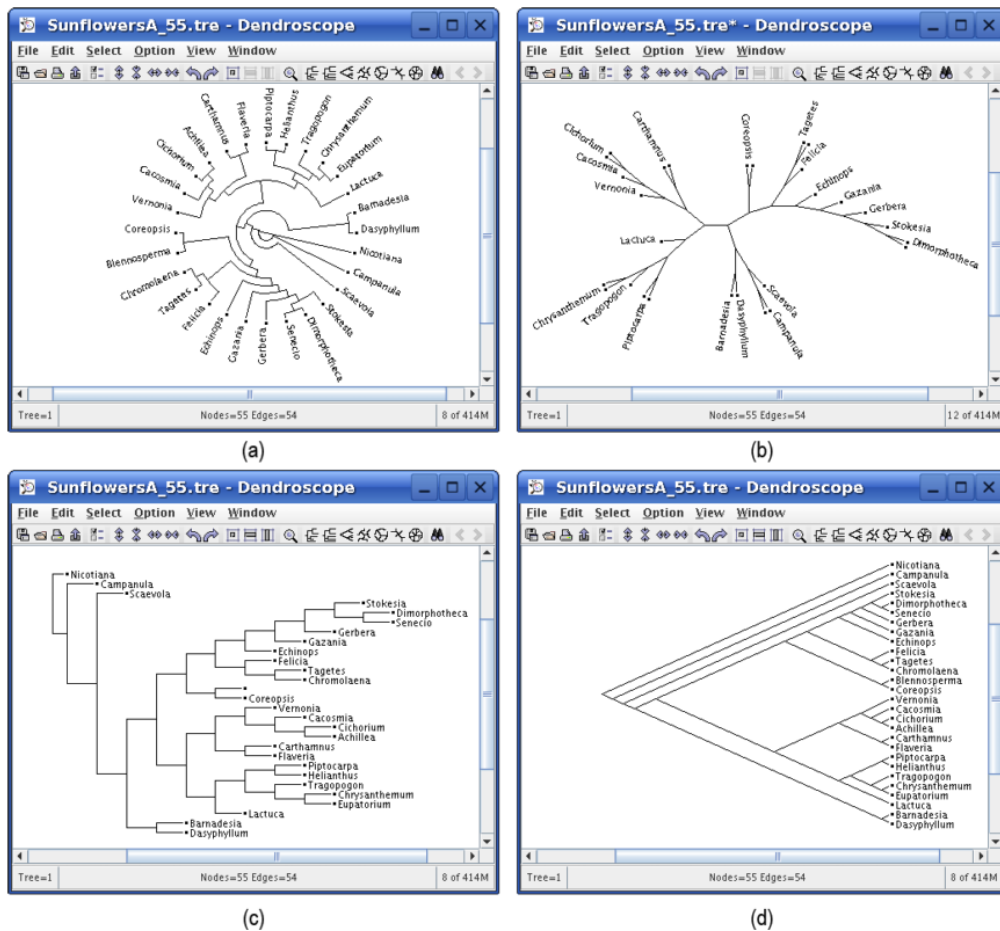
Source: <http://marbl.github.io/Krona/examples/rdp.krona.html>

3.4 Dendroscope

Dendroscope is designed as an all-round tree visualization tool that can handle trees with hundred thousands of taxa (HUSON et al., 2007). The tree can be displayed in seven different ways, including: circular cladogram, radial phylogram, rectangular phylogram and slanted cladogram. The four types listed can be seen in Figure 3.4. Nodes and edges can be customized, the user being able to select a color, size and shape for each node or node group.

The system can also handle the display and correlation of multiple phylogenetic

Figure 3.4 – Four types of visualization implemented by Dendroscope. a) circular cladogram; b) radial phylogram; c) rectangular phylogram; d) slanted cladogram.



Source: (HUSON et al., 2007)

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 its corresponding nodes (HUSON; SCORNAVACCA, 2012). For the correlation, a *tanglegram* is used, displaying 2 trees or networks opposite to each other and connecting the corresponding taxa by lines. For better understanding the visualization, the structure of the tree is changed so that the minimum amount of lines need to be crossed.

3.5 Final Remarks

The works mentioned above generally aim at the visualization of taxonomic or phylogenetic trees or just storing information. iTOL and Dendroscope are specially efficient in providing many types of hierarchical visualizations while also allowing compar-

isons between several trees. iTOL, Krona and Dendroscope are focused on presenting an overview of the data, lacking more complex functions for analysis and management of information. SinBiota is able to store and handle a substantial amount of data, even when displaying points on a map visualization. However, this system lacks an overview of the recorded specimens. In this work we focus on providing tools that are required for the management of information integrated with the display of an overview of the data as well as other visualization methods.

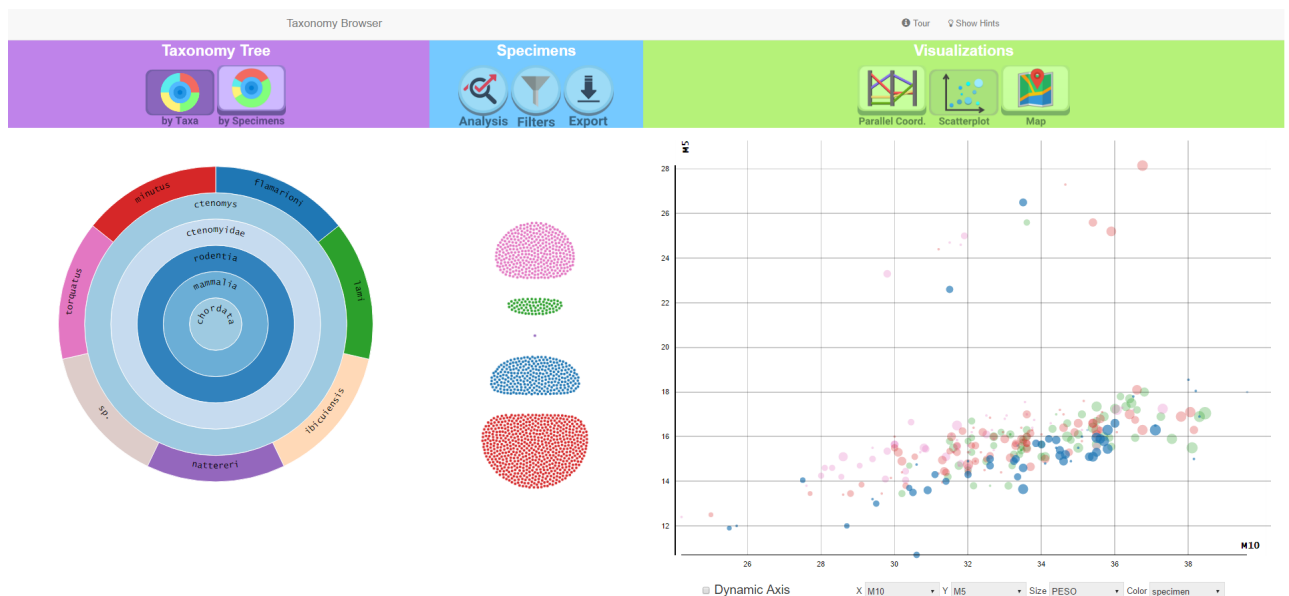
4 VISUALIZATION-BASED INTERFACE

We started this work by comparing the new system design with the previous version (see Chapter 5) following Shneiderman's work "The eyes have it: A task by data type taxonomy for information visualizations" (SHNEIDERMAN, 1996).

In that article, 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 multi-dimensional 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 based on a Sunburst (STASKO; ZHANG, 2000) visualization, which is used to represent and manage the taxonomic tree. As each species is selected (4.1) in the Sunburst visualization, their specimens are exhibited as small circles in the center of the layout and three types of visualizations can be created with their measures.

Figure 4.1 – Interface developed for the TaxonomyBrowser. The buttons on the top refer to the visualizations below them.



Source: the author

The *TaxonomyBrowser* database is accessed through a web based approach 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 interface of the system. Other libraries were used for specific parts of the interface, such as the Intro.js library (INTRO...), only needed for the tours and hints available, and the Toolbar.js plugin (A...), responsible only for the tool-tip used on the Sunburst visualization, described on Section 4.7.1.

4.1 Data Model

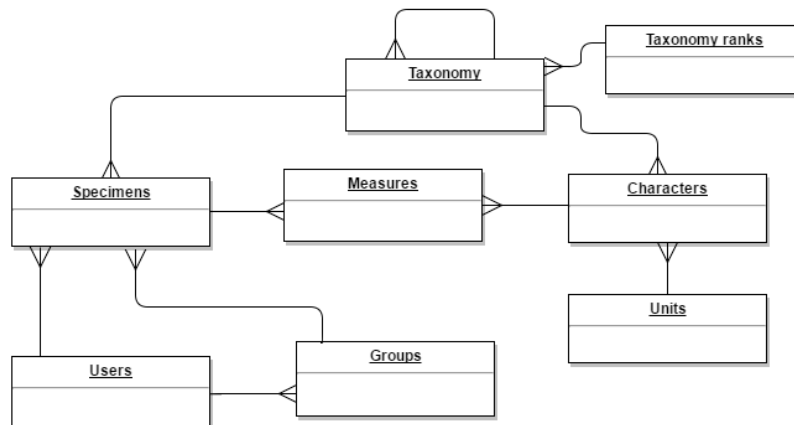
The purpose of the database developed for the *TaxonomyBrowser* is to store data about specimens collected or observed by biologists in order to support the queries they need for data analysis purposes as well as browsing and specific searches performed by anonymous users, either for entertaining and educational purposes (TAVARES; CANETE, 2011).

The database structure (see Figure 4.2) is based on the taxonomy tree, which ranks each node based on the Linnaean classification: Kingdom, Phylum, Class, Order, Family, Genus, and Species (in decreasing order of inclusiveness) (LEE et al., 2004). 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 a taxonomic hierarchy. For each registered specimen, the user can store information such as its collection ID, the names of the biologists responsible for its collection, the latitude and longitude of the collection's site and its description. If the specimen is made private or member of a group, it will also store the identification of the user that inserted it in order to limit its display depending on who is accessing the database. More about the authentication logic implemented is reported in Section 4.3. Depending on its taxonomic hierarchy different characteristics can be recorded for each specimen. The values for the characteristics are called measures and can be either numerical or textual.

The possible characteristics to be associated to each taxonomic level are defined globally: there is only a table in the database relating them with each taxon. 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 in the scope of this work since it was already a well-developed topic in previous versions of the *TaxonomyBrowser* (SILVA; FREITAS, 2007; HENKIN, 2010; TAVARES; CANETE, 2011; CAÑETE, 2011; HENKIN, 2013). However, some modifications were made in order to add new features, such as the au-

Figure 4.2 – Simplified entity-relationship diagram of the database model



Source: the author

thentication needed to access certain specimens. Two new tables were added, one to store the research groups and another to link the users to their groups. Also, the table used for storing specimens was slightly altered, adding two new columns. One column is used to define the user ID of who registered the specimen and the other is used for the group ID in case one was assigned to it.

4.2 Application Architecture and Dataflow

Figure 4.3 illustrates the application providing an overview of the different modules and the data flow between them. In order to display information from recorded specimens, data is obtained from the database using a PHP function and then encoded in JSON, a lightweight data-interchange format that is syntactically identical to JavaScript objects. For this reason data in JSON can be imported to a JavaScript code with standard JS functions, which leads to a better performance compared to other formats.

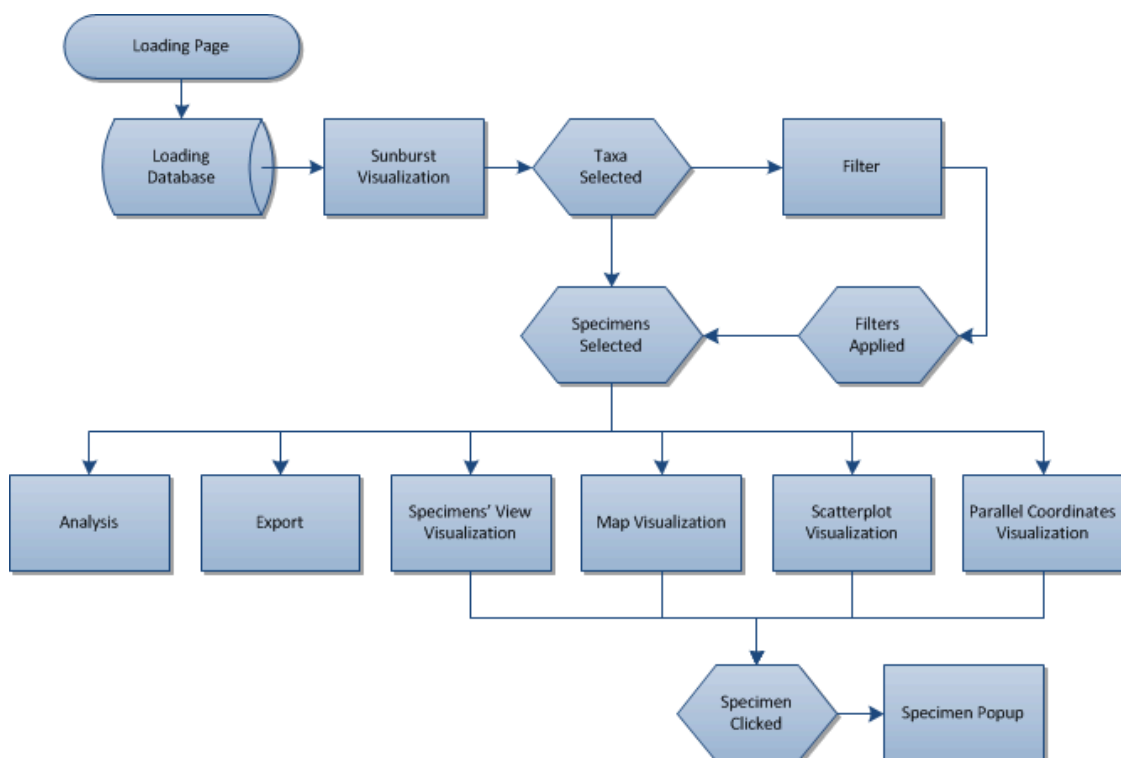
The main JSON dataset contains a single object, the root of the tree, which contains the next node as its child, and so on. Each of these objects contains standard information about each taxon and also a list with all its associated characteristics and inherited ones. The characteristics are saved using two different lists in the JSON structure so that one can specify which characteristics cannot be changed (inherited) and which ones can (defined for this taxon and consequently inherited by its children). Finally, a species taxon can have specimens linked as its children, with information such as their collection ID, the location where the specimen was collected, and the value for each of its inherited

characteristics. Specimens cannot define a new characteristic for itself, all characteristics being inherited from its species.

Each characteristic has an ID, a name, a group, a type and a description. For storing this information another JSON is imported and used as a list for correlating the characteristic's ID stored in each taxon or specimen node with its full information. Doing so, there is no need of storing all data each time a characteristic is added to a certain taxonomic level, avoiding duplicating information. Also, this approach does not cause any loss in performance, since all lists are implemented using associative arrays (hash) and use the characteristics' ID as key, hence each search has a complexity of $O(1)$.

When the page loads, the data acquired from the database is displayed as a Sunburst visualization, where the user is able to freely interact with all taxa currently recorded in the database. Specimen entries are not displayed directly in Sunburst, since it would reduce the performance of the visualization drastically when interacting with a large number of specimens. It would also be difficult for the user to select any specimen since each partition would be too small to be easily clicked on. In Section 4.7.1 we describe all interactions available to users.

Figure 4.3 – Simplified diagram of the system's data flow. Data about specimens from all species are loaded to the Sunburst visualization and, upon selection by the user, part of the dataset feeds the other visualizations with the chosen specimens.



Source: the author

From the Sunburst visualization it is possible to select certain species for further inspection. All specimens from these species are added to a new list of objects and displayed on the Specimen's View, described in Section 4.7.2. The specimens contained in this new list, named 'selected', will be used for the visualizations shown in the right are of the screen (or bottom part, depending on the resolution). These visualizations are described in Sections 4.7.3, 4.7.4 and 4.7.5. In Figure 4.2 we can see how the database information flows in the system. When taxa are selected on the Sunburst visualization it can be filtered (extensively described in Section 4.4) limiting the number of specimens displayed. This filtering generates a new list, but this one does not contain a copy of the objects but the index of objects from the 'selected' list that passed the user defined filters. This approach saves memory and also facilitates the removal of filters when required, since it keeps both the filtered and unfiltered list. If the filtered list is empty, the visualizations will use the 'selected' list. If not, data will be acquired by indexing the 'selected' list using the filtered list.

4.3 User Authentication

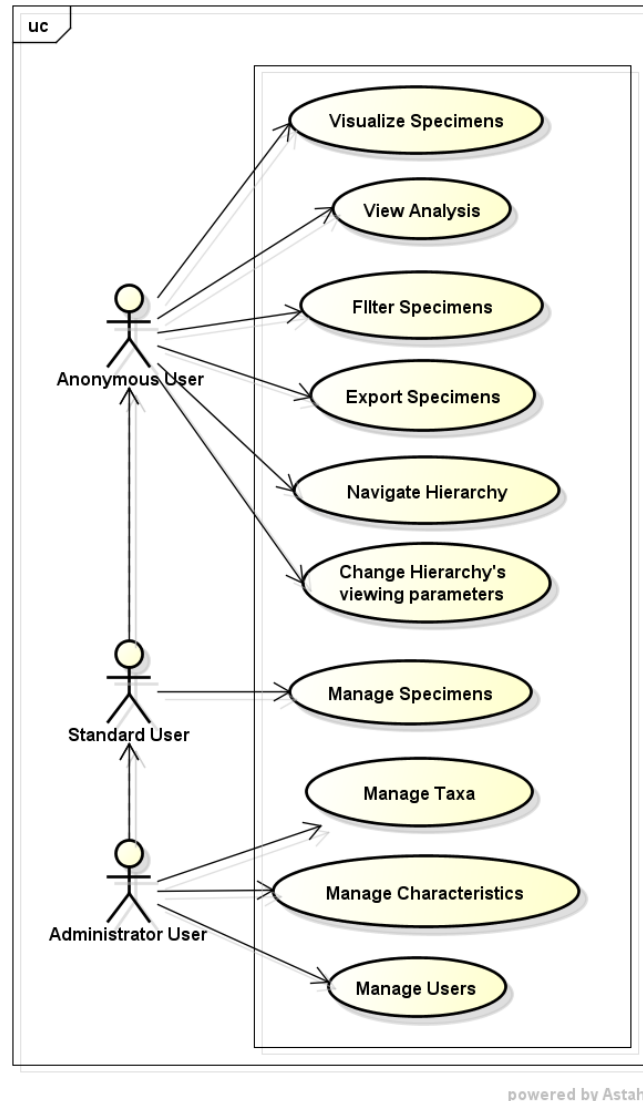
An authentication system was created to limit the access to the management functions of the database and to the viewing of private information. Since this system can be accessed by anyone with an Internet connection, it is important to limit the access of certain data so that only authorized personnel can use it on their researches and publications.

4.3.1 Anonymous User

An anonymous user (a user that has not yet logged in the system) can only interact with or view specimens that are made public. This user can see the complete taxonomy tree, but will not be able to alter its information in any way. Also, all managing options are disabled, since it would be imprudent to let an anonymous user alter (or delete) information in the database. Any other interaction that does not involve altering information or displaying private information can be done by any type of user, including visualizing specimens' data, displaying plots built from the data, filtering specimens, exporting data, navigating the taxonomy tree or changing its partitioning parameters. A simplified use case diagram (Figure 4.4) shows an overview of all interactions, illustrating which type

of user has access to which action.

Figure 4.4 – Simplified Use Case Diagram for the three possible types of users. Standard login interaction not shown.



Source: the author

4.3.2 Standard User

In order to access all possible features the user must first log in or create a new account. There are two types of accounts: standard and administrator. A standard account can view all the specimens it has recorded, the ones made public by other users, and the ones from the any of the groups this user belongs to (groups are explained in Section 4.3.4). Also, this type of account cannot manage any information on the taxa and

characteristics, i.e., cannot alter taxa or characteristics descriptions, only being able to manage specimens and their measures. This limitation was created to prevent this user from changing the structure of the database, creating inconsistencies, and compromising other users' data without realizing it. This type of account was envisioned as a student's account, able to store information related to his/her individual research in an on-line system, while keeping the information private from other users, still being able to share the data with the research groups if so desired. Of course, any user, when creating a new specimen entry, can decide if the data being recorded will be made public, private or shared in a group. After entered, it is possible to change this decision unless the specimen was marked as public, since it would make possible for users to limit public information only to themselves.

4.3.3 Administrator User

An administrator account can view all specimens recorded in the database and can fully manage the taxonomy tree, characteristics, specimens and users. This type of account was created to be used by professors to keep track and manage their students' research, also being able to keep the data after their project is completed for further use.

Regarding user management, this type of account can change the level of authorization of users as well as the groups they belong to. They can also completely delete any user from the system, except themselves.

4.3.4 Research Groups

Research groups were envisioned as means to make it possible for a user to share his/her collection with the research team. A specimen can be assigned to only one group, yet a user can belong to as many groups as she/he wants. Any user belonging to the group assigned to the specimen is able to see it and, when clicking on the specimen to display its information, the name of who recorded it and the associated group will also be displayed.

4.4 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 (SHNEIDERMAN, 1996). 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 (described in Section 4.7.2), a pop-up with all filtering options is displayed (Figure 4.5). There is no limit in the number of filters that can be used simultaneously.

Figure 4.5 – Filtering pop-up with four filters specified. The first and second line indicate that all entries should have the value '2N' and 'CAIXA' recorded. The third and fourth line limit the value of the attribute 'PESO' between 150 and 250.

Attribute	Filter Type	Value	Delete
2N	exists		X
CAIXA	exists		X
PESO	is bigger than	150	X
PESO	is smaller than	250	X

Source: the author

For each filter a new line is added to the pop-up. For each line there are two Combo Box (a commonly used graphical user interface widget for selecting between multiple options) elements, one graphical element for entering text and a red button for deleting the filter. The first Combo Box is used for selecting a characteristic for filtering. This can be any information such as the specimens' collection ID, name of its collector or any of their measures. The next Combo Box defines the type of filtering applied. The options vary depending if the characteristic's type selected on the previous Combo Box is numeric or textual. Textual options can only be compared to other texts or checked for their existence, while numerical options can receive additional operations, such as if a measure is larger or smaller than a certain value. The value used for comparison can be defined in the text entry line right next to the second Combo Box, and a consistency test

is performed to check if the value entered matches the characteristic's type, warning the user if otherwise).

4.5 Characteristics Manager

The characteristics of specimens can be added, edited and removed by the user using the characteristic manager interface. This interface shows all characteristics recorded in the database separated in tabs by groups which are designated by administrator users, when they are created and associated to some taxonomic level. Characteristics are described by name, unit, type and description (see Figure 4.6 and refer to Section 4.1). The characteristic's name and description can be altered freely, but once registered in the database its type and unit cannot be changed. This restriction is made so that measures of that characteristic already recorded for a specimen will not be of an invalid type.

Figure 4.6 – Characteristic manager pop-up. Characteristics can be added, edited or removed by an administrator user. Recorded characteristics cannot have their type changed in order to avoid invalid data in the database.

Characteristics

add, edit and delete characteristics below:

LOCAL DE DEPOSITO DE ESPECIMES
PROCEDENCIA
BIOMETRIA
ETARIOS
REPRODUCAO

CITOGENETICOS
AMOSTRAS COLETADAS
RECAPTURAS
OUTROS
OBSERVAOES

MORFOMETRICOS

CRANIO	unit undefined	text	colecão onde o crânio do espécime está depositado/ou caixa	✖
PELE	unit undefined	text	colecão onde a pele do espécime está depositada/ou caixa	✖
TECIDO	unit undefined	text	caixa onde amostra de tecido está depositada/	✖
CAIXA	unit undefined	text	caixa onde material está depositado	✖
NROPUC	unit undefined	text	NROPUC	✖

+

Close
Submit

Source: the author

In order to remove a certain characteristic the user must only click on the 'remove' button on the right of each line. This will not instantaneously delete the characteristic, only mark it for deletion. For every group of characteristics there is an 'add' button for appending a new one on the category. Any changes, be it from editing, deleting or adding characteristics, will only be applied when the user clicks on the 'submit' button.

4.6 Exporting

The selected specimens can be exported to CSV format, using semicolons for separation, with just one click. This feature was implemented so that the user can export data resulting from a filtered selection, and use them in other tools, such as the R language, for analysis or other visualizations.

4.7 Visualization Techniques

From the complete overview of the database to intricate 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) (BOSTOCK; OGIEVETSKY; HEER, 2011), a JavaScript library created for manipulating documents based on data.

D3 (<https://d3js.org/>) is a novel representation-transparent approach to visualization for the web that enables direct inspection and manipulation of a native representation: the standard document object model (DOM). This solution delivers improved performance over other options on the web, such as Protovis (BOSTOCK; HEER, 2009) and Flash. Since D3 can be a very powerful library for visualization, its learning curve can be much steeper than with other solutions.

4.7.1 Sunburst

Due to the organization of the provided data, a hierarchical visualization capable of displaying a full overview of the database was required. Three techniques were elected as possible candidates: a standard tree visualization, a TreeMap visualization and a Radial, Space-Filling (RSF) visualization. Even though the standard tree visualization would be

the safest choice since it is widely used for viewing such databases, this approach would not enable a complete overview of all data in a restricted area. The TreeMap visualization, developed by Johnson and Shneiderman (JOHNSON; SHNEIDERMAN, 1991), makes full use of the available display space, mapping the entire hierarchy onto a rectangular region in a space-filling manner. The RSF Sunburst approach is an alternative to the TreeMap visualization, that uses a radial configuration. The inner circle represents the beginning of the hierarchy and deeper levels are layered around this central node.

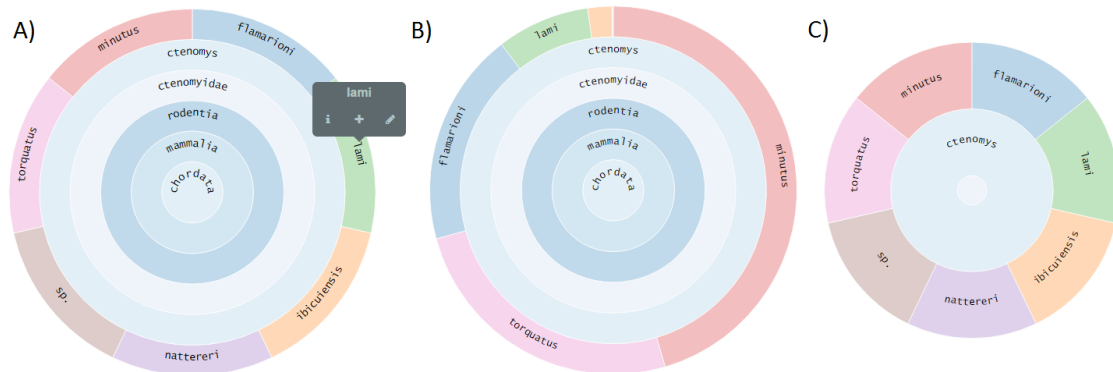
In order to decide between these techniques, a research was taken into consideration, evaluating their performance in an empirical study. The analysis made by Stasko et al. (STASKO et al., 2000) suggests that participants strongly preferred the Sunburst tool, citing better ability to convey structure and hierarchy. They also were more successful and presented a faster performance in their tasks, reporting the tool to be better for organizing and finding files due to its global view of the information.

RSF techniques for hierarchy visualization have several advantages over traditional node-link diagrams, including the ability to use the display space efficiently while effectively conveying the hierarchy structure (YANG et al., 2003). Other common problems solved by RSF visualizations include being able to analyse and alter in detail a variety of regions simultaneously without losing the overview of the dataset. Taken these points into account, the Sunburst technique was selected for representing the overview visualization in this work.

The Sunburst visualization implemented allows the following interactions:

- *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 4.9. More on the specimens' visualization will be described in the next section.
- *Zoom*: When right-clicking a node it becomes the main node displayed on the Sunburst, showing only its children and therefore presenting them in more detail (Figure 4.7). In order to go back for showing nodes higher in the hierarchy the central circle on the centre of Sunburst must be clicked.
- *Hover*: Hovering on each node shows a toolbar-styled tooltip, with the node's name on the top and icons representing options that allow get information, add children nodes and edit the hovered node.

Figure 4.7 – The Sunburst visualization with the current database information in some of its shapes: 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.

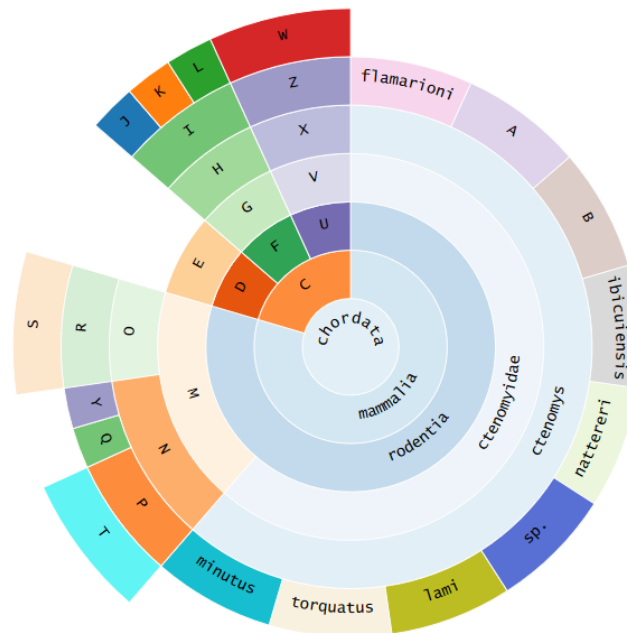


Source: the author

- *Information*: When clicked, this option opens a pop-up with the taxon's information and separates its characteristics in tabs by their groups, to simplify search. From this pop-up it is possible to edit its information and values of any selected characteristic, as long as it is not inherited from its parents.
- *Add*: When the node is a species, a pop-up for populating a new specimen will be shown. If the selected node is of any other rank, a similar pop-up will appear, only to add a new taxon below the one hovered.
- *Edit*: When clicked, this option opens the same pop-up as *Information*, with editing functions already enabled.
- *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. When viewed by the number of species, the user can have a better overview of all taxa recorded in the database. When viewed by the number of specimens, the user can have a general idea of which of those taxa are most populated. Both alternatives are displayed in Figure 4.7.

The current database still does not contain many taxa, with only 7 species of the same taxonomic group. Figure 4.8 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, where if a taxon has more than one child, a new color is assigned to it in order to simplify inspection.

Figure 4.8 – 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.



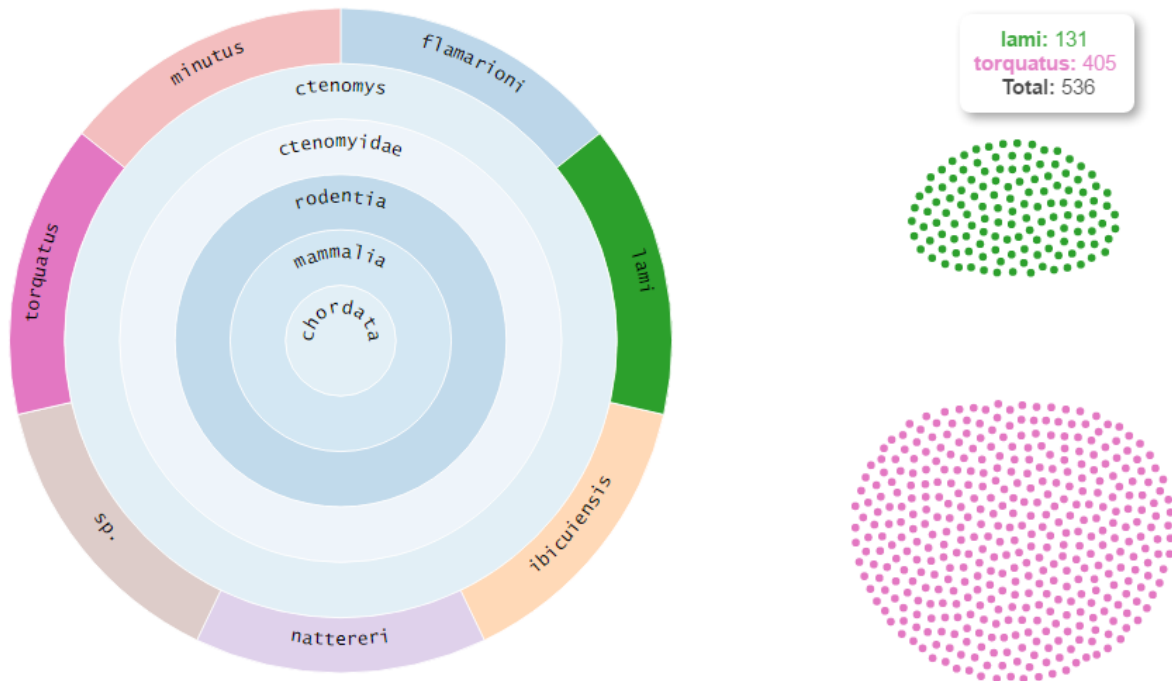
Source: the author

4.7.2 Specimens' View

This visualization was designed for showing all selected specimens. When a filter is applied, this visualization will only show, from the species selected in the Sunburst view, the specimens that fit all filters' criteria. The implementation is based on a force layout, circles of the same color attracting each other and forming groups (Figure 4.9). 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 interact by moving the circles (although only for entertaining reasons), clicking on them, which results in opening a pop-up with all the specimen's information (from this window it is also possible to edit the information and delete the specimen), or hovering the screen, which 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, as can be seen in Figure 4.10. 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.

Figure 4.9 – 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 this area of the interface.

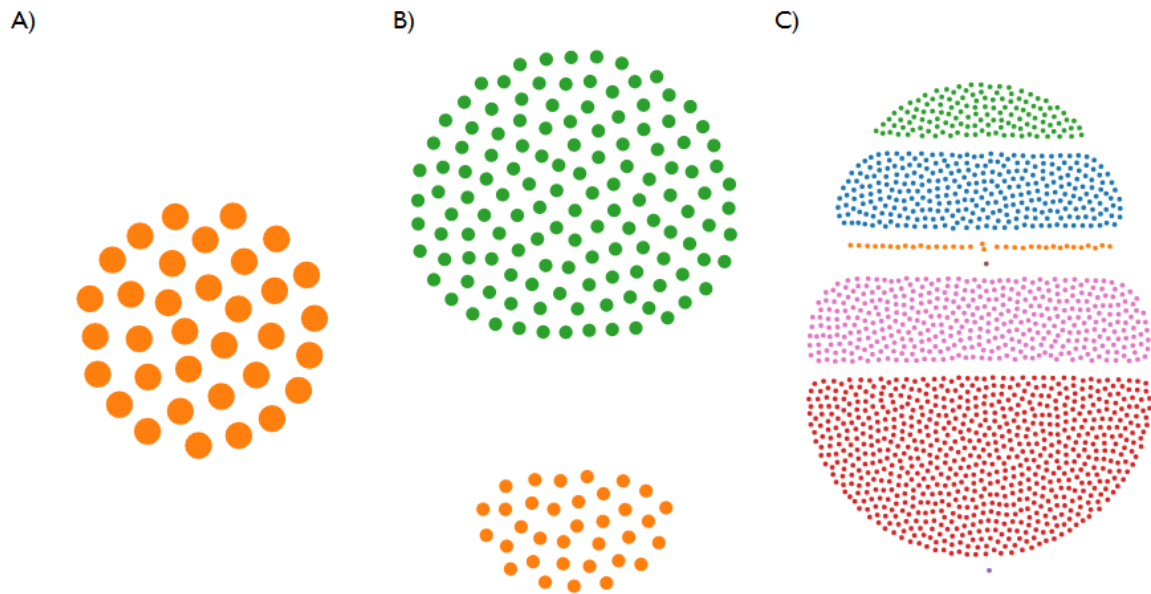


Source: the author

4.7.3 Scatterplot

According to Friendly and Denis (FRIENDLY; DENIS, 2005), "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". This technique is a plot of two variables, usually indicated as x and y , measured independently to produce bi-variate pairs (x_i, y_i) , 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 .

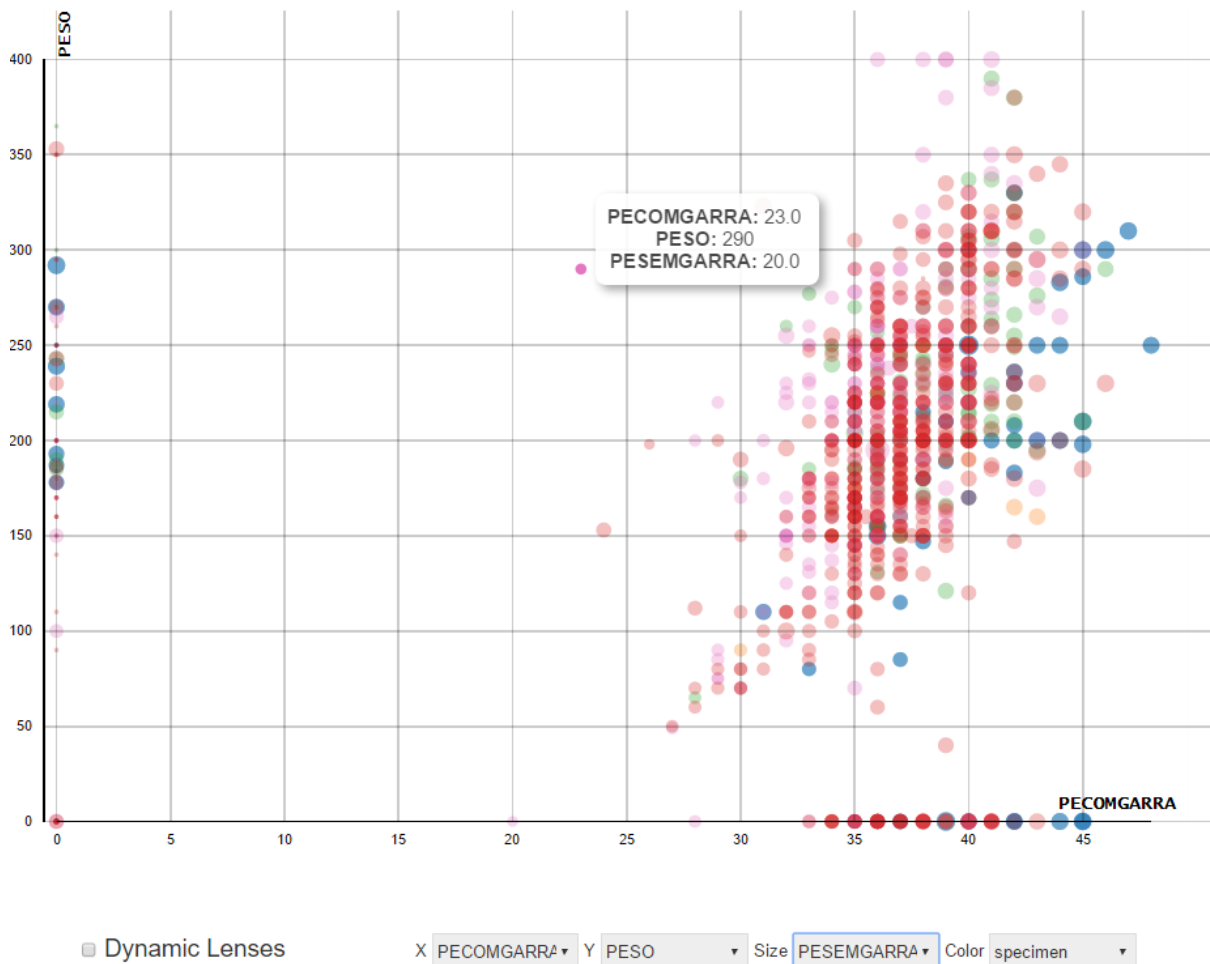
Figure 4.10 – Specimens' View with different numbers of selected specimens: A) 34 specimens selected from one species; B) 165 specimens selected from two species; C) 1600 specimens selected from 7 different species.



Source: the author

This visualization was implemented to be used as a ready-to-use solution for quickly analysing the selected data. 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 4.11. 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 color of the specimen's species, the same used in the Sunburst visualization.

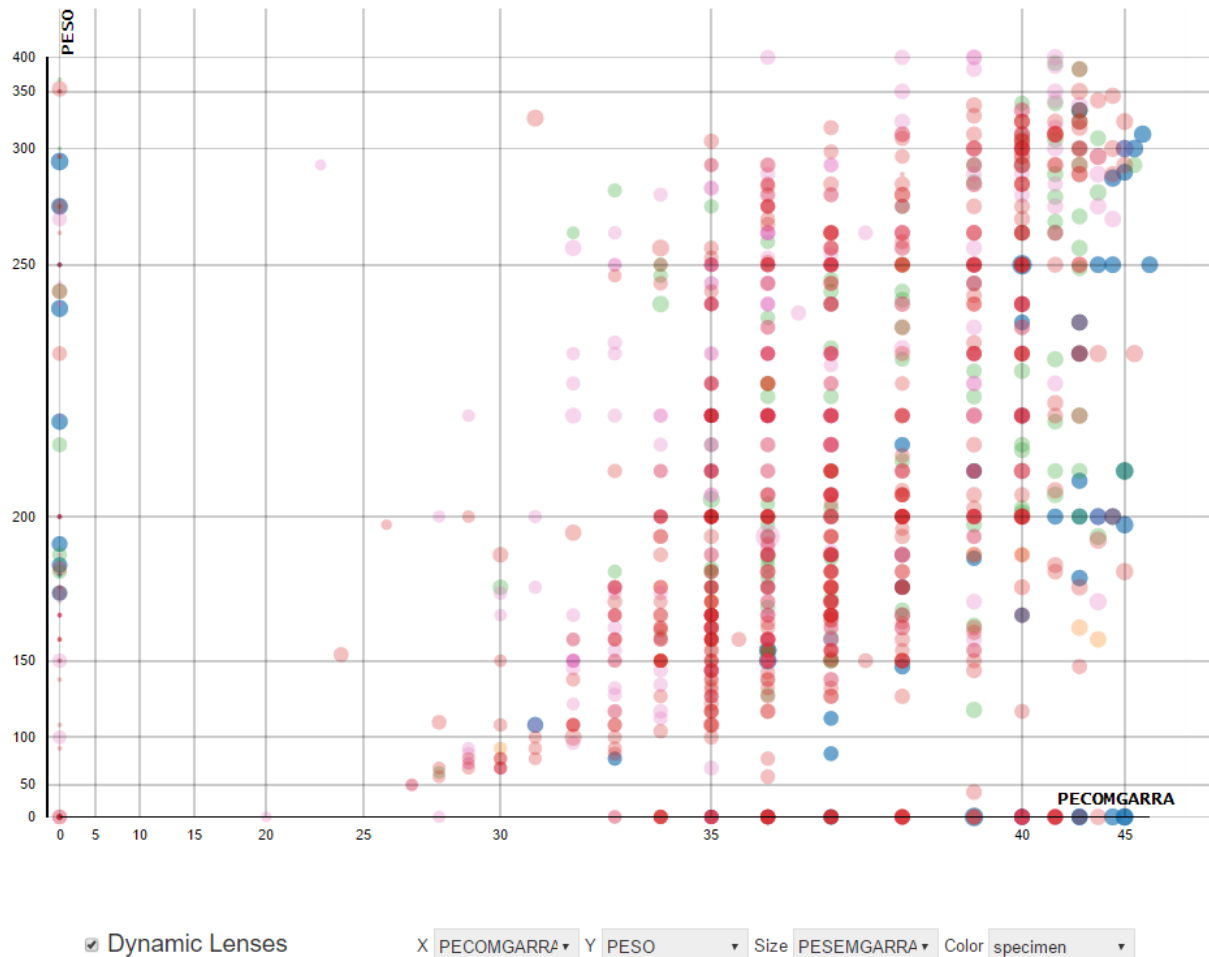
Figure 4.11 – Scatterplot visualization allows selecting characteristics for each axis and circle attributes. In the example, the X axis encodes the 'PECOMGARRA' characteristic, the Y axis contains the 'PESO' characteristic and the size of each circle represents the 'PESEMGARRA' characteristic. Hovering on a circle shows a tooltip with the values of the three selected characteristics.



Source: the author

In the lower left space of the visualization there is a Check Box named as 'Dynamic Lenses'. This enables a fish-eye representation of the axis in the Scatterplot. The axis proportion is changed dynamically depending on the user's mouse position, as can be seen in Figure 4.12.

Figure 4.12 – Scatterplot with the Dynamic Lenses feature activated, produce a fish-eye effect on the axis relative to the mouse position. In this case the mouse was centered between 35 and 40 of the X axis and 200 and 250 of the Y axis.



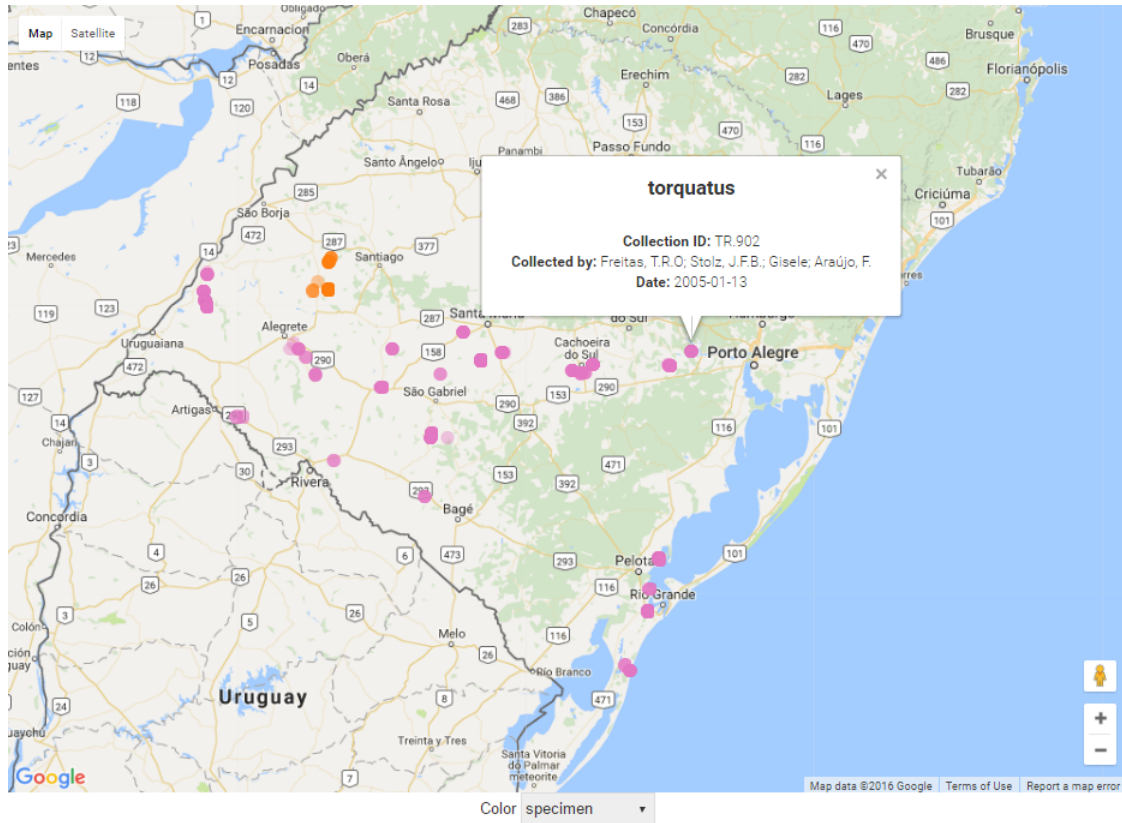
Source: the author

4.7.4 Map

A Map visualization is very important for biologists in order to show the exact location of each captured specimen. For the implementation of this view iwe used 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 its specified 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 information of the specimen is shown, also allowing the user to change information and delete the specimen from the database. This visualization is

shown in Figure 5.3.

Figure 4.13 – Map visualization using Google Maps API. Circles represent specimens and, when hovered, a tooltip with its main information is displayed.



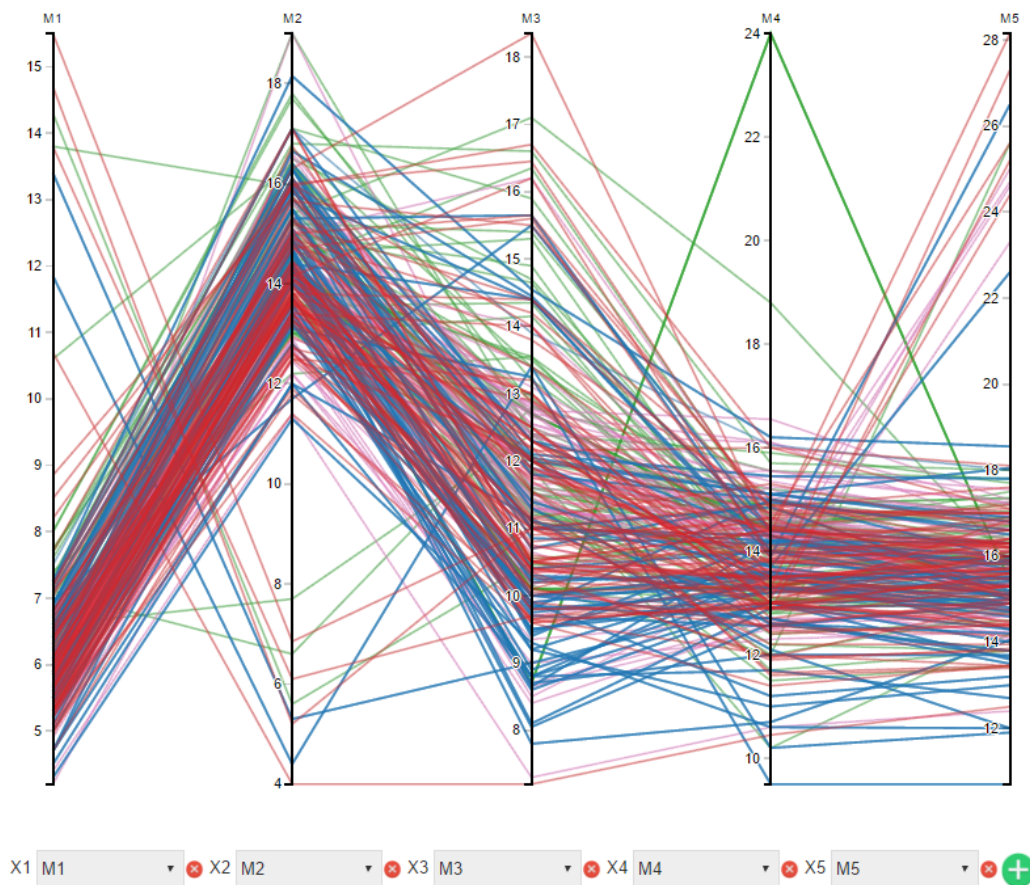
Source: the author

4.7.5 Parallel Coordinates

During the mid-1980s and early 1990s, Inselberg and Dimsdale (INSELBERG; DIMSDALE, 1990) introduced a technique for visualization of multidimensional data they called Parallel Coordinates. In this approach, each dimension 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 to reflect the N-D position (MCDONNELL; MUELLER, 2008).

Accordingly, in this work, each column represents a single characteristic. The user can select the characteristic for each vertical axis from the checkboxes at the bottom. Columns can be added dynamically by the user and viewed simultaneously, as can be seen in Figure 4.14. Every line corresponds to one of the selected specimens that possess

Figure 4.14 – Parallel coordinates visualization with columns representing characteristics 'M1', 'M2', 'M3', 'M4' and 'M5'. Each line represents a specimen.

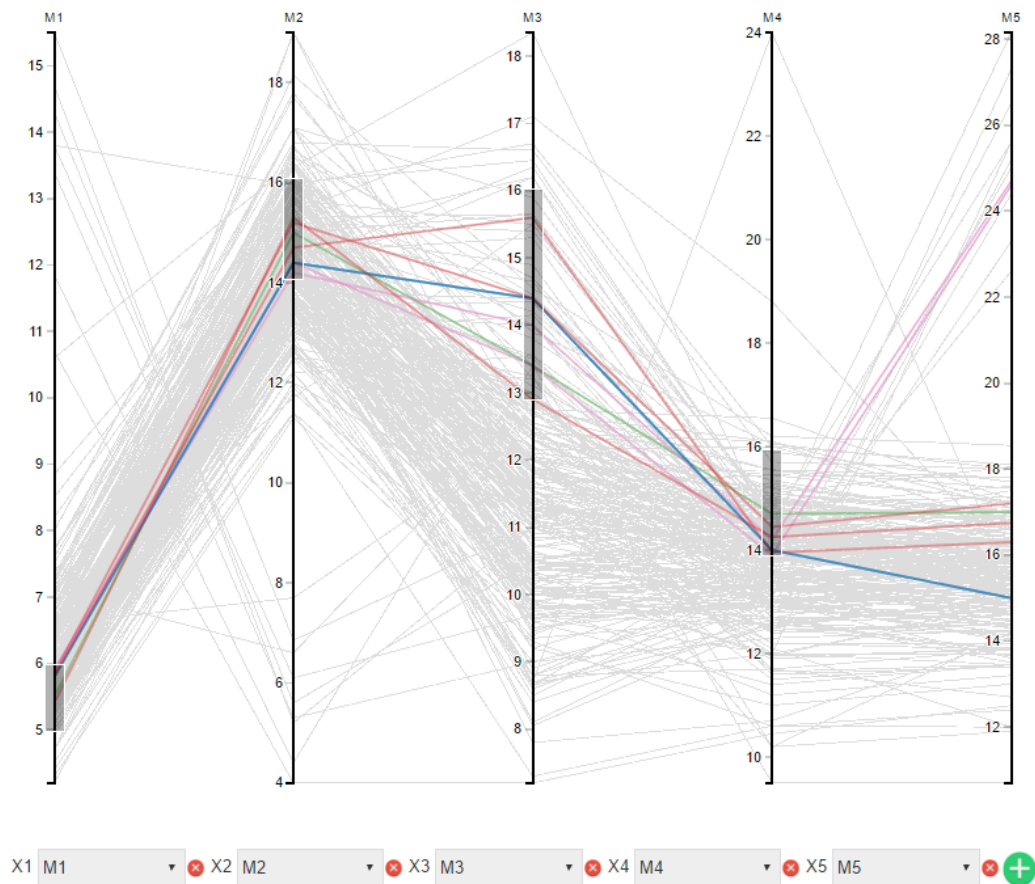


Source: the author

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

When visualizing many specimens at the same time the viewing can be overcrowded with lines. To counter this, a brushing can be made on each of the columns. This filtering is applied by holding and dragging on the axis, selecting a region. Only specimens with values ranging from the values inside this selection will be shown, changing the remaining lines to a faded gray color, as can be seen in Figure 4.15. This feature is useful for a more precise visualization of certain data without needing to modify the current specimen selection.

Figure 4.15 – Parallel coordinates with brushing applied on its axis. 'M1' is limited between 5 and 6, 'M2' is between 14 and 16, 'M3' is between 13 and 16 and 'M4' is limited between 14 and 16.



Source: the author

4.8 Final remarks

In this section we described the new interface for the Taxonomy-Browser. An overview of the whole system is available as a video posted at <https://youtu.be/eYmcUOPDr50>, while the application used for testing purposes can be accessed at <http://inf.fortesrey.com/taxonomybrowser/> and the final version at <http://inf.fortesrey.com/taxonomybrowserfinalversion/>.

5 DISCUSSION

5.1 Assessing the Interface regarding the Shneiderman's Seven Tasks for Information Visualization

As mentioned before, Shneiderman (SHNEIDERMAN, 1996) describes a list of essential tasks in visualization and the Visual Information Seeking Mantra: *overview first, zoom and filter, then details on demand*. Fitting the new interface described here regarding this mantra, the overview is provided by the Sunburst visualization, zoom is provided by the zooming mechanism applied to Sunburst, and the filtering method and details on demand are provided by interacting with the Specimen's View, Scatterplot, Map and Parallel Coordinates visualizations.

In this chapter we compare the present interface with the previous version of TaxonomyBrowser using Shneiderman's seven tasks for information visualization. Since we have not fully described the previous TaxonomyBrowser version here, the reader should refer to Tavares et al. (TAVARES; CANETE, 2011) for further information on that.

Table 5.1: Summary of the comparison between the previous version of the system with the one implemented in this work based on Shneiderman's seven tasks for information visualization.

Begin of Table 5.1			
Task	Description	Previous Version	Developed Version
Overview	Gain an overview of the entire collection. See Subsection 5.1.1 for a detailed comparison.	The interface does not provide an overview of all taxa and specimens stored in the database.	The Sunburst Visualization provides a complete overview of the data structure, being also able to show partitions' size based on the number of collected specimens.

Continuation of Table 5.1			
Task	Description	Previous Version	Developed Version
Zoom	Zoom in on items of interest. See Subsection 5.1.2 for a detailed comparison.	In order to access a certain species, the user must click on each of its parent taxa, loading a new web page each time.	In the Sunburst any taxon can be zoomed in with right-clicking its partition and with left-clicking its specimens are selected for a close view and management.
Filter	Filter out uninteresting items. See Subsection 5.1.3 for a detailed comparison.	Filtering can only be done when generating an analysis. When a filter is applied, the results only appear on the map, allowing only to access to specimen simple information.	Filtering can be done without the need of creating a new interface to store the filter's result and work on its analysis. Each selected specimen that passes the filter can be selected for displaying or managing all its information.
Details-on-Demand	Select an item or group and get details when needed. See Subsection 5.1.4 for a detailed comparison.	Details from each taxon and species recorded in the database can be viewed by selecting them from an alphabetical list.	Taxon details can be viewed on demand directly on the Sunburst Visualization and specimen details can be viewed in any other visualization.

Continuation of Table 5.1			
Task	Description	Previous Version	Developed Version
Relate	View Relationships among items. See Subsection 5.1.5 for a detailed comparison.	When browsing for specimens the interface does not make it clear when a certain specimen is from a similar hierarchy level than the others, being only displayed on an alphabetical list.	Relationship between items can be viewed by the organization of the Sunburst's hierarchy and also by the filtering of the specimens selected.
History	Keep a history of actions to support undo, replay and progressive refinement.	There is no history of previous actions stored by the system.	There is no history of previous actions stored by the system.
Extract	Allow extraction of sub-collections and of the query parameters. See Subsection 5.1.6 for a detailed comparison.	There is no option for exporting a certain sub-collection of the database selected by the user.	The selected and possibly filtered specimens can be exported to a CSV file format that can be used as input on other systems such as the R language.
End of Table 5.1			

In the next sections, we provide some details regarding the comparison summarized in Table 5.1.

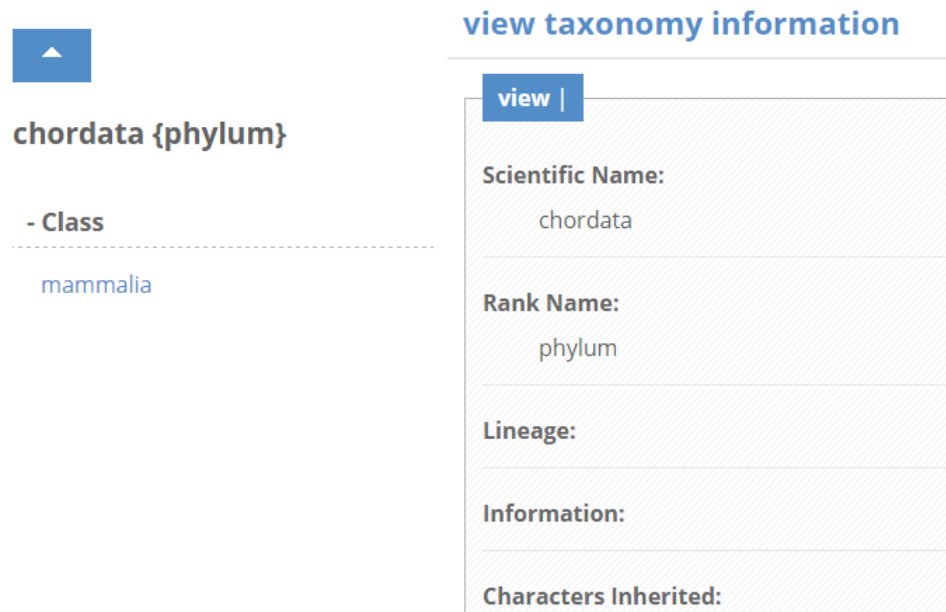
5.1.1 Overview

According to Shneiderman, it is essential to provide an overview of the entire collection. In the previous version of the system, to inspect the recorded data, the user had to navigate the entire collection level by level, if he/she wished to have a general idea of the elements stored in the database, as can be seen in Figure 5.1. If the user

wanted to inspect the specimens, the only interface provided is an alphabetical list with the first letter of the species name, which once clicked showed all specimens entries also in alphabetical order.

In the newly developed version a full overview of the recorded taxa can be visualized without the need of navigating between user interface panels. Each visible taxon can be easily inspected and modified by hovering. Moreover, an overview of the percentage of recorded specimens for each species can be seen with one click by changing the Sunburst's partitioning criteria for a better overview of each species' predominance.

Figure 5.1 – Taxonomy tree information in the previous version of TaxonomyBrowser. Only the direct children of each taxon are shown at a time. To check each taxon's children the user must click on their parent, loading a new web page each time.



Source: darwin.inf.ufrgs.br/taxonomybrowser

5.1.2 Zoom

The user's ability to zoom in terms of interest is an important aspect of information visualization. In the previous version of the system, for zooming in (accessing) specific taxa it was necessary to click on each of its taxonomical parents, beginning from the root of the taxonomic tree and loading a new web page each time. This is a time consuming task specially if the user is not familiar with the taxon's hierarchy.

In the new version, using the Sunburst, any node can be zoomed in or its speci-

mens selected for a closer view with just one click. For finding a certain species it is not necessary for the user to know its hierarchy in the tree, since a complete overview of the taxa recorded in the database is displayed with no effort using the Sunburst technique.

5.1.3 Filter

Filtering items is important for allowing users to control the contents of the display, eliminating unwanted items and focusing on what is really important to them. In the previous version of the system there is no tool for filtering data without consequently generating and saving a new analysis map. When a filter was applied, the results only appear on the map, being only possible to access simple specimen information from each point, preventing the user to edit or delete the filtered entries. For finding a particular specimen by its collection ID and be able to manipulate its contents, the user must search on a vast list organized alphabetically by its collection ID and species' name. Finding a specimen's entry with some particular measure value is not an intuitive task also.

In the implementation presented in this work, filtering can be done in a simple manner, not requiring the creation of a new page to store the filter's result and work on its analysis. Any information or characteristic of the specimens can be used for filtering. Also there is no fixed number of filters that can be applied at the same time. Each selected specimen that passes the filter can be chosen for viewing, editing or deleting, facilitating for users to look for invalid data recorded in the database and modify it.

5.1.4 Details-on-Demand

After filtering a collection, the few items left should have their details readily available for the user. The previous version of TaxonomyBrowser shows details from each taxon, species and specimen as an alphabetical list, where they can be clicked for retrieving any details. All attributes are listed on the right of the list.

According to Shneiderman, the usual approach is to simply click on an item to get a pop-up window with values of each of the item's attributes. This is exactly the interaction method used in this work: in all visualizations with specimens representations, the user can click on the representation of the desired specimen and visualize all information about it. For example, in the Scatterplot visualization each circle can be clicked on to

Figure 5.2 – Alphabetical list of all analysis saved in the old version of the interface. It does not link to the full information of the specimens analysed, only displaying how many of them matched the filter.

The image shows a web interface with two main parts. On the left, under the heading "Analysis", there are two sections: "- B" and "- T". Under "- B", the text "busca-flamarioni" is displayed. Under "- T", the text "teste" is displayed. On the right, a "view analysis information" panel is open, showing details for the "teste" analysis. The panel includes a "view" button, a "Name:" field with the value "teste", a "Color:" field with a blue square and the hex code "#5C98FF", an "Information:" field with the text "teste de GPS com ctenomys torquatus." and "PESO > 200", a "Scientific Name:" field with the value "ctenomys torquatus", an "Analysis Filters:" section with a bullet point "• PESO: > 200", and a "Result:" section with the text "Specimens Count: 162".

Source: darwin.inf.ufrgs.br/taxonomybrowser

show all the data from that particular specimen. Details from each taxon can be viewed on demand by hovering and choosing to see the data directly on the Sunburst Visualization.

5.1.5 Relate

Shneiderman states that designing user interface actions to specify which relationship is to be manifested is still a challenge. The clearer relationship of the *Taxonomy-Browser's* database is the hierarchical relationship between taxa.

In the previous version of the system, the only relationship among items that could be easily viewed is that of sibling taxa in the hierarchy, since, in order to show a taxon's children, the user must click the taxon and, in doing so, loses all context from any parent

Figure 5.3 – Alphabetical list of all specimens recorded on the database available in the previous version of the interface. This is the only alternative users had to check a specimen's complete information.

The screenshot shows a web interface with two main sections. On the left, under the heading "Specimens", there is a sub-heading "- F" followed by a list of specimen entries: "flamarioni (0000038)", "flamarioni (0000039)", "flamarioni (0000045)", "flamarioni (0000045)", "flamarioni (0000045)", and "flamarioni (0000047)". On the right, a "view specimens information" panel is open, displaying details for a specimen. The panel has a "view" button at the top left. The details are organized into sections: "Scientific Name:" with the value "ctenomys flamarioni"; "Lineage:" with the value "chordata; mammalia; rodentia; ctenomyidae; ctenomys;"; "Extra Taxonomy:" which is currently empty; and "Collection ID:" with the value "0000038".

Source: darwin.inf.ufrgs.br/taxonomybrowser

taxa to see only the children of the taxon selected. When browsing for specimens the interface does not make it clear when a certain specimen is from a similar hierarchy than the others, since they are displayed on an alphabetical list with all recorded specimens. On the other hand, the newest version of the application provides an overview of all taxa relationships within the Sunburst visualization.

5.1.6 Extract

According to Shneiderman, once users have obtained the item or set of items they desire, it would be useful to be able to extract that set and save it to a file in a format that would facilitate other uses, such as sending by email, printing, creating new plots, or inserting into a statistical or presentation packages (SHNEIDERMAN, 1996). In the previous version there is no option for exporting a certain sub-collection of the database selected by the user. This was changed in the newest implementation, now being possible to export the selected and, if so wished by the user, filtered specimens as a CSV file format that can be used as input to most programs as well as scripts in the R language, facilitating the filtered information to be viewed by already commonly used visualizations created.

5.2 Essential Operations for Interaction with Hierarchically Constructed Data

The interactions provided in our application can also be discussed in the light of Yang's essential operations for interaction with hierarchically constructed data: *selection, reconfiguration, drill-down/roll-up, pan, zoom, rotation and distortion* (YANG et al., 2003).

5.2.1 Selection

Selection, according to Yang, is the process of selecting some nodes from the hierarchy for further operations, such as highlighting, masking, deleting and distorting. In this implementation, highlighting can be achieved by right clicking nodes, causing all selected species' specimens to be selected and appear in the next visualizations. Deletion can be achieved by hovering and opening the taxon's information and clicking on the 'trash' button at the upper right of the pop-up window. This interaction could be made directly on the tooltip from the Sunburst visualization to reduce user clicks and waiting period, but since this operation will hardly be used and is much more likely to be wrongly selected by users we decided to require further steps before providing this feature.

5.2.2 Reconfiguration

Reconfiguration, the process of interactively modifying the hierarchy structure, can be achieved by adding and removing nodes using the tooltip shown when hovering the partitions.

5.2.3 Drill-down/Roll-up

Drill-down/roll-up, the process of exposing/hiding sub-branches of the hierarchy, is comparable to the zoom features implemented in the Sunburst visualization. This type of distortion is used for a better view of subsections of the tree.

5.2.4 Pan, Zoom, Rotation

Pan, zoom, and rotation are the process of modifying the focus, scale, and orientation of the display. Pan and zoom can be achieved using the same technique described on the previous section, while the rotation was not perceived as a crucial feature and is left for future work.

5.2.5 Distortion

Distortion is the process of enlarging some objects in the display while maintaining the context of surrounding objects. This feature is not implemented in the Sunburst visualization, but can be done in the Scatterplot. Selecting the check box labeled 'Dynamic Lenses', the position of the mouse causes a fish-eye effect in the plot, having the scale of the region closer to the mouse exponentially larger than the more distant ones.

6 EVALUATION

In order to validate this work, a remote survey was conducted¹. This test aimed at measuring the users' understanding of the overall flow of the system and their ability to perform tasks and receive results without any specialist's assistance. Another goal was to obtain feedback from user satisfaction related to the provided features as well as about the overall usability of the application.

6.1 Participants

The test involved 40 participants, 75% male and 25% female between the age of 19 and 58 years old. 32.5% of these participants are from the field of biology, 52.5% are from the field of computer science and 15% are from other fields, such as engineering, health and social sciences. The participant's education levels were very varied, having 47.5% undergraduate students, 22.5% graduated, 20% with a MSc degree and 10% with a PhD degree. 100% of participants had some kind of experience with on-line systems, while 40% had some experience with biological information systems. 85% of the participants knew or thought they knew what a taxonomy tree is.

6.2 Tasks and initial setup

Users were initially presented with a brief description of the system's purposes and the definition of a taxonomy tree. Subsequently, the users were directed to the website and asked to perform the tour available on the upper left of the screen. There is no log in required for the tests, only public information is handled and no management of the database is needed, since it could compromise other tester's results. In light of this, some interface options concerning user login and data management were intentionally hidden.

The following questions (here translated to English, but originally written in Portuguese) were presented in order to measure the users' understanding of this work's features and also to ensure that they truly interacted with the system before answering the usability questions. These questions are based on the main basic tasks the user could do and receive a solid result. It was also considered the lack of patience users could have if

¹The survey is available at <https://goo.gl/forms/lo6zC0CdlmeVTUCy1>

questions took too much of their time to be completed.

- **T1: What is the species with most specimens?**

The correct answer is 'minutus'. The objective of this question was to verify if the user could get an overall notion of the database, either by changing the partitions to be shown by the number of specimens or by selecting all species and checking, using the number of circles of each color in the Selected Specimens' visualization, which has the most entries.

- **T2: How many specimens this species has?**

The correct answer is 727. This value can be found by hovering the 'minutus' partition on the Sunburst and choosing to view its complete information in the tooltip, selecting the species and checking with the analysis button how many were selected or hovering on the circles shown after the selection was made. This question was created to check if the user was able to seek information from a certain species or has the intuition to select the specimen and check the circles or analysis button.

- **T3: Describe briefly how you reached this answer.**

The purpose of this question was to check what method was used to reach the previous answers, particularly to make sure the user did not count each circle to find out how many specimens are recorded.

- **T4: Select only the species 'Ctenomys lami'. Add a new filter testing if the parameter 'Data' exists. How many specimens match the filter?**

The correct answer is 57. The aim of this question was to check if the user could find, create and apply filters. Also to verify if he/she understands that the counting made with the analysis button or by hovering on the circles shows the number of selected specimens that match the current filtering.

- **T5: Describe briefly how you reached this answer.**

In this question we hoped that the user answers that she/he did indeed look up this number by using the analysis button or in the circles' tooltip and did not count all circles shown in the interface.

- **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)?**

The correct answer for this question is '0000282' or 'TR.414A', although answers such as '282' and '414A' were also accepted. The reason for both answers being accepted is the fact that, since it was asked for the user to ignore values zeroed, it was almost impossible to check that the specimen with the collection ID 'TR.414A'

has the value 1 in this measure (it was even overlooked when creating the test's questions, only to be realised when many participants answered with this particular ID). This question was done in order to analyse if the Scatterplot visualization is clear to the user in terms of understanding of the graph and its input parameters. Also this question verifies if the user knows that clicking on a dot in the graph will open all information about the specimen and if he/she knows how to remove filters.

- **T7: What is the collection ID of the heaviest specimen?**

The correct answer for this question is 'TR.1408', although answers such as '1408' were also accepted. This answer serves to the same purpose as the previous one.

- **T8: Describe briefly how you reached this answer.**

In this question we hoped that the user answers that he/she did indeed use the Scatterplot graph or Parallel Coordinates for the analysis, changing or not the axis' parameters.

- **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)?**

The correct answer is 827. The aim of this question was to test if the user knows how to build more than one filter and limit a parameter by their values.

- **T10: Describe briefly how you reached this answer.**

In this question we hoped the user answers whether she/he used two filters and checked the number using the analysis button or hovering on the Specimen's View. This question is important since, if a wrong value is given, it is possible to verify if the user's intuition on how to solve the problem was correct.

6.3 Evaluation procedure

The survey was conducted in four steps, each represented as a section of the questionnaire. First, users were asked to answer some personal information in order to understand their profile. These questions include their age and gender, their professional profile and experience with interactive systems, websites and biological information systems. The evaluation was conducted remotely and with no instructions on how to operate the system other than the tours and hints provided in the website. After gathering the user's information, we provide the link to the system. It is asked that the user uses Chrome or

Firefox browsers because of compatibility issues. It is also strongly recommended that, before interacting with the TaxonomyBrowser, she/he uses the tour available (enabled by a button at the upper right of the system's interface). It is even asked, before continuing, if he/she really did use it, to make sure that the user is aware of the tour and also, if the he/she did not use it, to take into consideration when analysing the answers. No further guidance regarding the features of the system is given on the questionnaire, letting subjects have the same experience as any new user of the system. The only explanation added to the survey was regarding the biological information in the database and the meaning of the taxonomy tree, in order to enlighten doubts users with no experience in the field might have.

The next part of the survey had practical tasks to be accomplished by the user. These tasks are the questions described in section 6.2, which lead the user to perform basic interactions with the system, and then, request values that can only be obtained if the user comprehends the features provided and explained during the tour.

After the user has experienced the website, questions on the user's satisfaction with the interface features were asked. These questions are actually sentences, which have to be answered by choosing one out of 5 options based on a Likert scale of agreement (LIKERT, 1932), indicating whether the user "strongly agreed", "agreed", is "undecided", "disagrees", or "strongly disagrees" with each sentence. This section of the questionnaire consisted of 14 sentences, which are listed below:

- S1: I think the tool has good response time.
- S2: I think the visualizations are adequate.
- S3: I think the proposed visualizations allow a good understanding of the database
- S4: I think the menus for each visualization are adequate
- S5: I think the color scheme chosen is pleasant.
- S6: I liked the taxonomy tree visualization (Sunburst).
- S7: I liked the visualization of the selected specimens (circles in the center of the screen).
- S8: I liked and understood the Scatterplot visualization.
- S9: I liked and understood the map visualization.
- S10: I found the filtering method intuitive.
- S11: I found the options of visualization, search and comparison of data adequate.
- S12: I think the layout of the system is appealing.

- S13: I found the tool interesting.
- S14: I found the tool effective in its features.

The final part of the survey was envisioned to measure the usability of the system by means of the System Usability Scale (SUS), a reliable, low-cost usability scale that can be used for global assessments of system's usability (BROOKE, 1996). This scale consists of 10 questions whose purpose is to provide an overview of subjective assessments of usability. These 10 items were very carefully selected to have half of them with the most preferred response being "strongly agree" and the other half "strongly disagree", in order to recognize response biases caused by users not fully reading and comprehending each statement.

6.4 Results

Participants provided correct answers to most of the questions. As the 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 Table 6.1 and Figure 6.1.

In order to check which methods were used to reach the answers we analysed the answers provided on questions T3, T5, T8 and T10. In T3, 47.5% used the Sunburst's partitioning by specimens to check which species had the most specimens, while 22.5% compared the number of circles from the Specimens' View after selecting all species. 30% did not describe how this task was accomplished. When describing how they managed to find how many specimens belong to this species, 27.5% of users used the information button from the Sunburst's tooltip to check the number. 30% checked the value by hovering on the Specimens' View and 20% used the analysis button. The remaining participants had vague answers or only described how they found which was the most populated species, not mentioning how they acquired the number.

In T5 40% hovered the Specimens' View to check the number of specimens after the filter. 20% used the analysis button and 32.5% did not provide a complete answer to infer what method was used to retrieve the information. 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 tool'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 task labeled T8 asked for the participants to describe how they acquired 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 them. One of them used the dynamic axis to make sure he wasn't selecting wrongly.

In the final task, T10, 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. Only 10% hovered the Specimens' View to check the value after filtering and Only 0.75% tried to count manually the number of circles on the Specimens' View, all failing to reach the desired answer. The remainder of participants did not provide a complete answer, only specifying that they used two filters and citing their parameters.

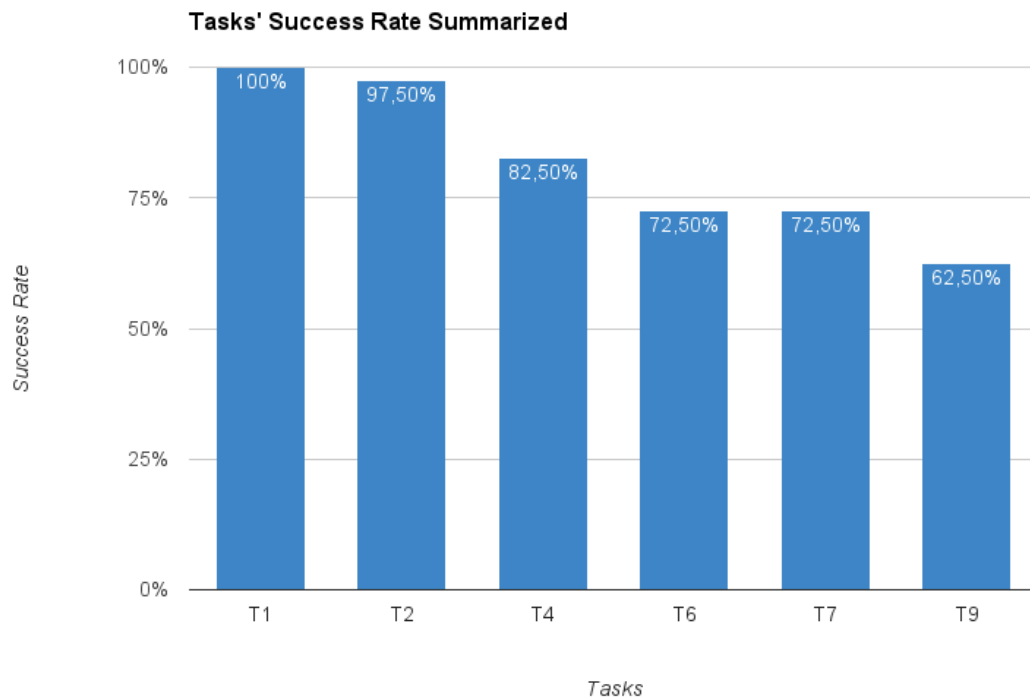
After evaluating the explanation given to each answer, we verified that a considerable amount of wrong answers were probably due to the participant not reading the question thoroughly. For example, in T6 the user is asked to select all species before answering the question. 70% of users that answered the question incorrectly skipped this step and only took into consideration the *Ctenomys Lami* species, selected for the previous question. We made this assumption since the answers correctly described the activities required for the question and the collection ID provided coincides with the lightest and heaviest specimen of the *Lami* species. These results can be considered relevant, mainly because they correspond to a first usage of the system.

Table 6.1 – Summary of success rate results from the tasks with objective answers asked on the questionnaire

Task	Summarized Question	Success Rate
T1	What is the species with most specimens?	100%
T2	How many specimens this species has?	97.5%
T4	How many specimens match the filter?	82.5%
T6	What is the collection ID of the lightest specimen?	72.5%
T7	What is the collection ID of the heaviest specimen?	72.5%
T9	How many specimens have the parameter 'PESO' between 150 and 250?	62.5%

The results from the user satisfaction with the interface design decisions questions (Section 3 of the questionnaire) are summarized in Table 6.2 and Figure 6.2. In general, users liked and understood the proposed visualizations and the layout of the tool. They particularly enjoyed the color scheme, having 97.5% of approval by the participants. The only question slightly controversial was the system's response time, having 72.5% of

Figure 6.1 – Summary of success rate results from the tasks with objective answers asked on the questionnaire



Source: the author

participants satisfied, 15% neutral and 12.5% dissatisfied with the performance. This can be partially related to the variety of hardware possibly used for testing the tool. These results prompted reformulation and tweaking of some functions. More on this issue will be clarified at the end of this section, when we discuss users' suggestions and feedback.

The results obtained from the SUS section of the questionnaire are summarized in Table 6.3 and Figures 6.3 and 6.4. The average SUS score was 78.3, which is decently above the average of 68 and close to an A grade score (above 80.3). Feedback was mostly positive, specially regarding how well integrated the system's features were and its overall consistence. Also 87.5% thought that they did not need to learn many things before they could get going 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 amount of participants were not biologists, and therefore would have no practical application for using the tool. When only analysing answers from participants

Table 6.2 – Summary of results from questions on the user satisfaction about interface decisions

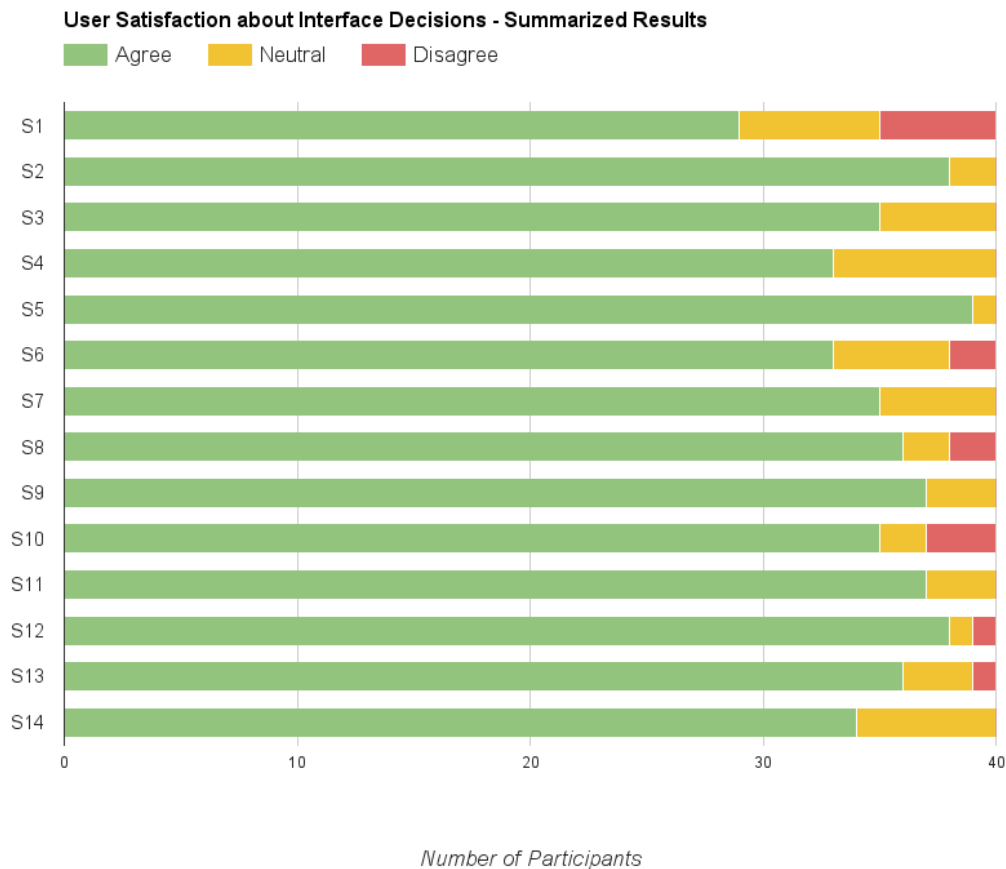
ID	Sentence	Agree	Neutral	Disagreee
S1	I think the tool has good response time.	72.5%	15%	12.5%
S2	I think the chosen visualizations are adequate.	95%	5%	0%
S3	I think the proposed visualizations allow a good understanding of the data base	87.5%	12.5%	0%
S4	I think the menus for each visualization are 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%

from the biology field, 75% agreed they would like to use the system frequently, 12.5% were neutral and 12.5% disagreed. This represents a considerable improvement which represents an increase of 36%.

The analysis of additional comments left by 18 participants allowed us to understand how they feel about the tool, and what modifications they perceived as important. One participant remarked that the way selection is demonstrated on the Sunburst was not clear to the user. To help in this matter a stroke was added to each partition selected, as can be seen in Figure 6.5.

Many users complained about performance issues when viewed on their personal computer, particularly when selecting many specimens. This issue has been noticed and investigated, and we concluded that the number of DOM elements in the website, created for each circle printed on the screen, was greatly slowing performance. This issue was addressed by changing the Specimen's View visualization to use canvas instead of SVG, when there is a considerable amount of specimens selected. The canvas approach,

Figure 6.2 – Summary of results from questions on the user satisfaction about interface decisions



Source: the author

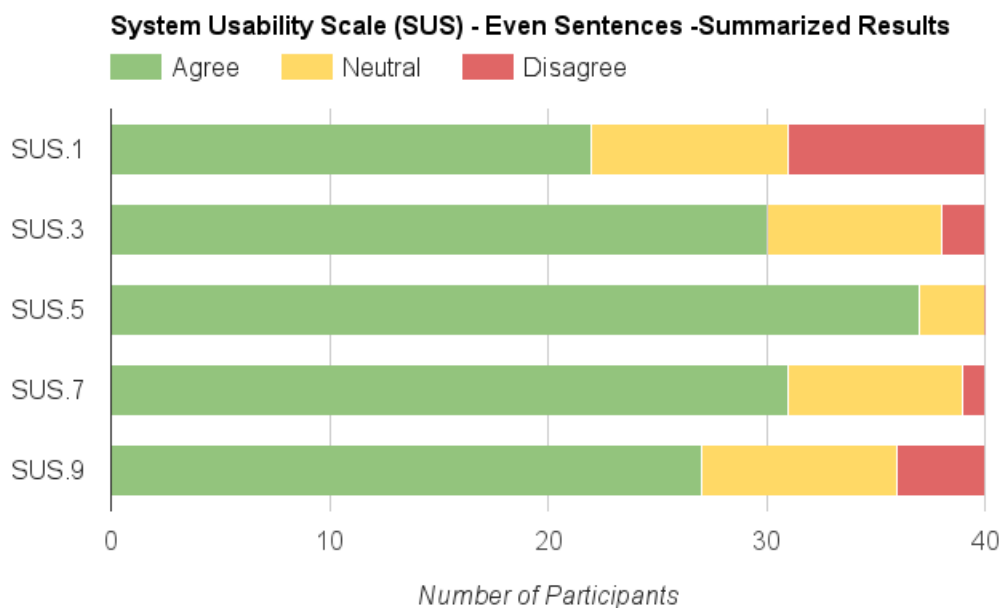
even though has a lower quality and cannot handle event interactions on each circle, is considered as a single DOM element, substantially improving the overall performance. Since when there is a large amount of specimens selected interactions such as clicking for checking specific information and dragging were considered not substantially relevant to the user; the responsiveness obtained by this alteration were deemed worthwhile.

One participant suggested adding the number of parent nodes in the central circle of the Sunburst after zooming, to give an idea of how many of them are not being displayed. To solve this issue, a breadcrumb approach was used above the Sunburst visualization, adding all parent nodes after any zoom is applied, as can be seen in Figure 6.6. When clicking on each one of these labels, the user can go back to the visualization of the selected taxon, reversing the zoom more efficiently than clicking on the center of the Sunburst for each level. Another interesting suggestion was to implement double-click on the Sunburst visualization in order to deselect all taxa and select only the one double-clicked, but it will still be considered since this could compromise the selection's response time.

Table 6.3 – Summary of results from the SUS section

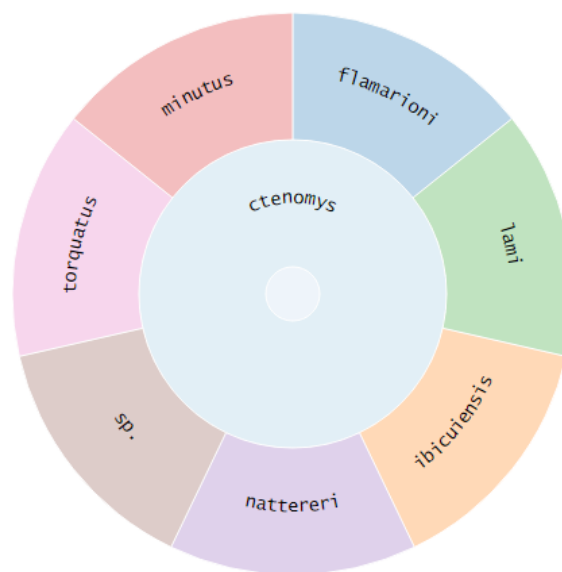
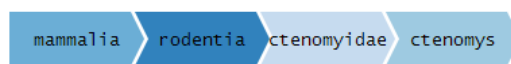
ID	Sentence	Agree	Neutral	Disagree
SUS.1	I think that I would like to use this system frequently.	55%	22.5%	22.5%
SUS.2	I found the system unnecessarily complex.	7.5%	15%	77.5%
SUS.3	I thought the system was easy to use.	75%	20%	5%
SUS.4	I think that I would need the support of a technical person to 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%

Figure 6.3 – Summary of results from the SUS section



Source: the author

Figure 6.6 – Breadcrumbs and the Sunburst visualization, showing all parent nodes of the zoomed-in taxa.



Source: the author

7 CONCLUSIONS AND FUTURE WORK

This work presents a new interface for the *TaxonomyBrowser* relational database based on Shneiderman's Visual Information Seeking Mantra: overview first, zoom and filter, then details on demand. A Sunburst visualization is implemented for displaying an overview of the taxonomy tree and for managing the database information, such as its taxa 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. Displayed items are only available after checking the access privilege of the users, depending on information such as their account's category and which research groups they participate. This encourages users to record their collections, since access to their data will be properly managed.

The interactions provided by the new version were compared with the previous version of the interface using Sheiderman's seven tasks for information visualization, namely overview, zoom, filter, details-on-demand, relate, history and extract. The new version was superior or equal in features and usability in each one of these tasks. Also the number of interactions needed to perform certain tasks was greatly minimized.

The interface was validated via a remote survey based on a questionnaire, which 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 tool. Considering the wide range of hardware used by the participants, it was possible to receive feedback on performance issues and tweak the system accordingly.

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

In order to improve this work, more visualizations, specially the ones extensively used in the field of biology, can be embedded on the website. 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.

REFERENCES

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

BARVE, V.; OTEGUI, J. bdivis: visualizing biodiversity data in R. **Bioinformatics**, p. btw333, 2016. ISSN 1367-4803. Disponível em: <<http://bioinformatics.oxfordjournals.org/lookup/doi/10.1093/bioinformatics/btw333>>.

BOSTOCK, M.; HEER, J. Protovis: A graphical toolkit for visualization. **IEEE Transactions on Visualization and Computer Graphics**, v. 15, n. 6, p. 1121–1128, 2009. ISSN 10772626.

BOSTOCK, M.; OGIEVETSKY, V.; HEER, J. D³ data-driven documents. **IEEE transactions on Visualization and Computer Graphics**, v. 17, n. 12, p. 2301–2309, 2011. ISSN 1077-2626.

BROOKE, J. SUS - A quick and dirty usability scale. **Usability evaluation in industry**, v. 189, n. 194, p. 4–7, 1996. ISSN 1097-0193. Disponível em: <<http://hell.meiert.org/core/pdf/sus.pdf>>.

CAÑETE, S. d. C. Interface de gerenciamento e consultas visuais em banco de dados de biodiversidade. 2011. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/29014>>.

CANHOS, V. P. Sistema de Informação Distribuído para Coleções Biológicas : A Integração do Species Analyst e SinBiota Integração do Species Analyst e SinBiota. n. 2001/02175-5, p. 2–51, 2005. Disponível em: <<http://smlink.cria.org.br/docs/outubro2005.pdf>>.

ENGHOFF, H. What is taxonomy? - An overview with myriapodological examples. **Soil organisms**, v. 81, n. 3, p. 441–451, 2009. ISSN 1864-6417.

FRIENDLY, M.; DENIS, D. The early origins and development of the scatterplot. **Journal of the History of the Behavioral Sciences**, v. 41, n. 2, p. 103–130, 2005. ISSN 00225061.

HENKIN, R. Interface de consultas analíticas para bases de dados de biodiversidade. 2010. Disponível em: <<http://www.lume.ufrgs.br/bitstream/handle/10183/26354/000757817.pdf?sequence=1>>.

HENKIN, R. A study on visual analysis of georeferenced haplotype networks. n. October, 2013.

HUSON, D. H. et al. Dendroscope: An interactive viewer for large phylogenetic trees. **BMC bioinformatics**, v. 8, n. August 2016, p. 460, 2007. ISSN 14712105.

HUSON, D. H.; SCORNAVACCA, C. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. **Systematic Biology**, v. 61, n. 6, p. 1061–1067, 2012. ISSN 10635157.

INSELBERG, A.; DIMSDALE, B. Parallel coordinates: A tool for visualizing multi-dimensional geometry. In: **Proceedings of the 1st Conference on Visualization '90**. Los Alamitos, CA, USA: IEEE Computer Society Press, 1990. (VIS '90), p. 361–378. ISBN 0-8186-2083-8. Disponível em: <<http://dl.acm.org/citation.cfm?id=949531.949588>>.

INTRO.JS - Step-by-step Guide and Feature Introduction for your Website. <<http://introjs.com/>>. (Accessed on 18/11/2016).

ITOL: Interactive Tree Of Life. <<http://itol.embl.de/>>. (Accessed on 10/06/2016).

JOHNSON, B.; SHNEIDERMAN, B. Tree-maps: a space-filling approach to the visualization of hierarchical information structures. **Proceeding Visualization '91**, p. 284–291, 1991. ISSN 0818622458. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=175815>>.

LEE, B. et al. How users interact with biodiversity information using TaxonTree. **Interface**, p. 320–327, 2004. Disponível em: <<http://portal.acm.org/citation.cfm?doid=989863.989918>>.

LETUNIC, I.; BORK, P. Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. **Bioinformatics**, v. 23, n. 1, p. 127–128, 2007. ISSN 13674803.

LIKERT, R. A technique for the measurement of attitudes. **Archives of Psychology**, v. 22 140, p. 55, 1932. ISSN 0006-8993. Disponível em: <<http://psycnet.apa.org/psycinfo/1933-01885-001>>.

MCDONNELL, K. T.; MUELLER, K. Illustrative parallel coordinates. **Computer Graphics Forum**, v. 27, n. 3, p. 1031–1038, 2008. ISSN 01677055.

ONDOV, B. D.; BERGMAN, N. H.; PHILLIPPY, A. M. Interactive metagenomic visualization in a Web browser. **BMC Bioinformatics**, v. 12, n. 1, p. 385, 2011. ISSN 1471-2105. Disponível em: <<http://www.biomedcentral.com/1471-2105/12/385>>.

SHNEIDERMAN, B. The eyes have it: a task by data type taxonomy for information visualizations. **Proceedings 1996 IEEE Symposium on Visual Languages**, p. 336–343, 1996. ISSN 1049-2615.

SILVA, J.; FREITAS, C. M. D. S. Projeto e Desenvolvimento de Sistema Web para Armazenamento de Coleções de Espécimes. 2007.

SOBERÓN, J.; PETERSON, T. Biodiversity informatics: managing and applying primary biodiversity data. **Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences**, v. 359, n. 1444, p. 689–698, 2004. ISSN 0962-8436. Disponível em: <<http://rstb.royalsocietypublishing.org/content/359/1444/689.abstract%delimitar%026E30F%http://rstb.royalsocietypublishing.org/content/359/1444/689.full.pdf>>.

STASKO, J. et al. An evaluation of space-filling information visualizations for depicting hierarchical structures. **International Journal of Human-Computer Studies**, v. 53, n. 5, p. 663–694, 2000. ISSN 10715819.

STASKO, J.; ZHANG, E. Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In: **Proceedings of the IEEE Symposium on Information Visualization 2000**. Washington, DC, USA: IEEE Computer Society, 2000. (INFOVIS '00), p. 57–. ISBN 0-7695-0804-9. Disponível em: <<http://dl.acm.org/citation.cfm?id=857190.857683>>.

TAHIR, S.; AFZAL, M. T. A novel phylogenetic tree data visualization application for researchers. **Proceedings of 2014 Science and Information Conference, SAI 2014**, n. August, p. 93–99, 2014.



TAVARES, D.; CANETE, S. TaxonomyBrowser: a biodiversity data management system. **Journal of ...**, v. 2, p. 37–46, 2011. Disponível em: <<http://www.inf.ufrgs.br/~dlmtavares/JCIS11-art.30.p>>.

YANG, J. et al. InterRing: A Visual Interface for Navigating and Manipulating Hierarchies. **Information Visualization**, v. 2, n. 1, p. 16–30, 2003. ISSN 14738716. Disponível em: <<http://ivi.sagepub.com/cgi/content/abstract/2/1/16>>.

APPENDIX A — INTERFACE POP-UPS

Figure A.1 – Pop-up with information from a taxon. When clicking on the edit button, information such as the taxon’s name, information and characteristics can be altered.

nattereri

chordata -> mammalia -> rodentia -> ctenomyidae -> ctenomys -> nattereri

Name:	nattereri	unit undefined	text	undefined
Rank:	species	unit undefined	text	undefined
Specimens:	1	unit undefined	text	undefined
Information:		unit undefined	text	undefined

CITOGENETICOS

OUTROS

PROCEDENCIA

LOCAL DE DEPOSITO DE ESPECIMES

BIOMETRIA

MORFOMETRICOS

OBSERVACOES

REPRODUCAO

AMOSTRAS COLETADAS

ETARIOS

RECAPTURAS

<input checked="" type="checkbox"/>	2N	unit undefined	number	numero diploide
<input checked="" type="checkbox"/>	ACROS	unit undefined	number	numero de cromossomos acrocentricos
<input checked="" type="checkbox"/>	METAS	unit undefined	number	numero de cromossomos metacentricos
<input checked="" type="checkbox"/>	NF	unit undefined	number	numero fundamental
<input checked="" type="checkbox"/>	PAR1	unit undefined	text	(AA, MA ou MM) - identificacao do tipo do par 1
<input checked="" type="checkbox"/>	PAR2	unit undefined	text	(AA, MA ou MM) - identificacao do par 2

Close

Source: the author

Figure A.2 – Pop-up with information from a specimen. Each tab represents a group of characteristics that are populated for this particular specimen. Data can be altered or edited with the edit button on the upper right of the pop-up.

ibicuiensis (TR.1507)




specimen's visibility: public

Species: ibicuiensis	unit undefined	text	
Collection ID: TR.1507	unit undefined	text	
Collected by: Freitas, T.R.O.; Roratto, P.; Langone,	unit undefined	text	
Data: 2011-07-12	unit undefined	text	
Latitude: -29.156	unit undefined	number	
Longitude: -55.4204722222	unit undefined	number	
Information:	unit undefined	text	

BIOMETRIA

PROCEDENCIA

CITOGENETICOS

SEXO: m	unit undefined	text	sexo do espécime. m = macho, f = femea
PESO: 250	unit undefined	number	Peso em gramas do espécime

Source: the author

Figure A.3 – Pop-up with information from the selection made, such as the number of specimens selected for each species and an analysis on a certain characteristic selected by the user.

Selected Specimens

Counting

- ibicuiensis: 34 selected specimen(s)
- lami: 131 selected specimen(s)
- flamarioni: 301 selected specimen(s)
- nattereri: 1 selected specimen(s)
- sp.: 1 selected specimen(s)
- torquatus: 405 selected specimen(s)
- minutus: 727 selected specimen(s)
- Total: 1600**

Selected Variable Analysis

PESO ▼

- Number of specimens with this measure populated: 1370
- Number of specimens with this measure's value equal to zero: 177
- Average: 183.5839416058394
- Standard Deviation: 89.40495441257623

Close

Source: the author

Figure A.4 – User login pop-up. Here you can also register a new account.

Login

Username

Password

Sign In

[Forgot Password?](#) [Register](#)

Cancel

Source: the author

APPENDIX B — SURVEY

Table B.1: Personal information questions from the survey

P1	Age	
P2	Gender	<input type="checkbox"/> Male <input type="checkbox"/> Female <input type="checkbox"/> Other
P3	Education	<input type="checkbox"/> High School Student <input type="checkbox"/> Graduated School <input type="checkbox"/> Graduation Student <input type="checkbox"/> MSc Degree <input type="checkbox"/> PhD Degree
P4	Field of work	<input type="checkbox"/> Computer Science <input type="checkbox"/> Biology <input type="checkbox"/> Engineering/Mathematics/Physics/Chemistry <input type="checkbox"/> Social Science <input type="checkbox"/> Health
P5	Do you have any experience with interactive systems? (check all applicable options)	<input type="checkbox"/> I have experience with usual web systems (shopping, social networks, banking) <input type="checkbox"/> I have experience with systems necessary for my professional activities <input type="checkbox"/> I have experience with computer games <input type="checkbox"/> I navigate and make search on the web
P6	Do you have any experience with biodiversity information systems? (check all applicable options)	<input type="checkbox"/> No Experience <input type="checkbox"/> Professional Experience <input type="checkbox"/> Other <input type="checkbox"/> Unprofessional Experience
P7	Do you know what a taxonomy tree is?	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Maybe

Table B.2: Practical Tasks from the survey

T1	What is the species with most specimens?
T2	How many specimens this species has
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?
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)?
T7	What is the collection ID of the heaviest specimen?
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)?
T10	Describe briefly how you reached this answer

Table B.3: User satisfaction questions from the survey

S1	I think the tool has good response time	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
S2	I think the chosen visualizations are adequate	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
S3	I think the proposed visualizations allow a good understanding of the data base	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
S4	I think the menus for each visualization are adequate	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
S5	I think the color scheme is pleasant	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
S6	I liked the taxonomy tree visualization (sunburst)	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
S7	I liked the visualization of the selected specimens (circles on the center of the screen)	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
S8	I liked and understood the Scatterplot visualization	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
S9	I liked and understood the map visualization	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
S10	I found the filtering method intuitive	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree

S11	I found the options of visualization, search and comparison of data adequate?	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
S12	I think the layout of the system is appealing?	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
S13	I found the tool interesting	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
S14	I found the tool effective in its features	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree

Table B.4: SUS questions from the survey

SUS1	think that I would like to use this system frequently	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
SUS2	I found the system unnecessarily complex	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
SUS3	I thought the system was easy to use	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
SUS4	I think that I would need the support of a technical person to be able to use this system	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
SUS5	I found the various functions in this system were well integrated	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
SUS6	I thought there was too much inconsistency in this system	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
SUS7	I would imagine that most people would learn to use this system very quickly	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
SUS8	I found the system very cumbersome to use	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
SUS9	I felt very confident using the system	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree

SUS10	I needed to learn a lot of things before I could get going with this system	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree
SUS10	I found the filtering method intuitive	<input type="checkbox"/> Strongly Disagree <input type="checkbox"/> Disagree <input type="checkbox"/> Neutral <input type="checkbox"/> Agree <input type="checkbox"/> Strongly Agree