TEC | Tecnológico de Costa Rica

Costa Rica Institute of Technology

School of Computing

A Multi-view environment to visualize differences between alternative versions of biological taxonomies

Thesis submitted by Manuel Figueroa Montero
to opt for the degree of Magister Scientiae in Computing with
emphasis in Computer Science

Thesis Adviser: Lilliana Sancho Chavarría, Dr.Eng.

2022

Master´s Program in Computing

Dedication

To my wife and parents, who supported me in this journey.

Manuel Figueroa

Acknowledgements

I would like to thank my thesis adviser Dr. Lilliana Sancho Chavarría, for her guidance and valuable advice during the development of this thesis. This work would not have been possible without her knowledge and dedication.

I also want to thank our expert advisors Armando Estrada, MSc., botanist of the National Herbarium of Costa Rica and Luis Acosta MSc. professor of the School of Forestry Engineering of the ITCR for their valuable contributions as domain experts throughout this research, and to professor Alejandro Meza, MSc., Director of the School of Forestry Engineering of the ITCR, who kindly helped us to assemble the test groups.

Abstract

The classification of living organisms into biological taxonomies is fundamental to understand our biodiversity and to support environmental conservation efforts. The biological taxonomy system for the classification of living organisms was proposed by Carl Von Linné in 1735. Organisms are organized into groups that follow a hierarchical structure that is composed at least of the following categories: kingdom, phylum, class, order, family, genus and species. Taxonomists work encompasses the identification and classification of species, as well as the registration and maintenance of the taxonomic records that are kept in museums, herbaria, research centers and institutions dedicated to natural preservation. Those records are constantly changing due to the discovery of new species or the reclassification of the existing ones; therefore, it is common that taxonomists face the challenge of handling large amounts of data that require to be reviewed, compared and corrected. This challenge can be approached through the automatic comparison of the taxonomies.

In this work we investigate whether a cluttering reduction approach that visualizes only the differences (VOD) when comparing two alternative versions of a biological taxonomy is more effective than the visualization of the complete taxonomies. We also investigate whether the use of the Coordinated Multiple Views (CMV) technique contributes to a better identification of differences. For this purpose we take as a baseline Diaforá, a novel interactive software that infers and visualizes the differences between two alternative versions of a biological taxonomy by modelling one view only and using the edge-drawing method [52]. We extended Diaforá by integrating CMV and VOD.

This research has three stages: 1) the design of two different VOD visualization alternatives, 2) a user study with the goal of evaluating the alternatives and selecting

one of them, 3) a final user study that evaluates the original version of Diaforá against one that incorporates the CMV technique with the selected VOD visualization.

We found that the proposed model increases significantly the cognitive value of the visualization, making it easier for test subjects to identify the differences between two versions of the classification.

Resumen

La clasificación de los organismos vivos en taxonomías biológicas es fundamental para comprender nuestra biodiversidad y para apoyar los esfuerzos de conservación del medio ambiente. El sistema de taxonomías biológicas para la clasificación de los organismos vivos fue propuesto por Carl Von Linné en 1735. Los organismos son organizados en grupos que siguen una estructura jerárquica que se compone al menos de los siguientes categorías: reino, phylum, clase, orden, familia, género y especie. El trabajo de los taxónomos consiste en la identificación y clasificación de las especies, así como el registro y mantenimiento de los datos taxonómicos que se conservan en museos, herbarios, centros de investigación e instituciones dedicadas a la preservación natural. Esos registros cambian constantemente debido al descubrimiento de nuevas especies o la reclasificación de las existentes; por lo tanto, es común que los taxónomos enfrenten el reto de manejar grandes cantidades de datos que requieren ser revisados, comparados y corregidos. Este reto se puede abordar mediante la comparación automática de las taxonomías.

En este trabajo investigamos si un enfoque de reducción de desorden que visualiza solo las diferencias (VOD) al comparar dos versiones alternativas de una taxonomía biológica es más efectivo que la visualización de las taxonomías completas. También investigamos si el uso de la técnica Coordinated Multiple Views (CMV) contribuye a una mejor identificación de las diferencias. Para este propósito tomamos como base Diaforá, un novedoso software interactivo que infiere y visualiza las diferencias entre dos versiones alternativas de una taxonomía biológica, modelando una sola vista y usando el método de dibujo de bordes [52]. Ampliamos Diaforá integrando CMV y VOD.

La investigación constó de tres etapas: 1) el diseño de dos alternativas diferentes

de visualización VOD, 2) un estudio de usuario con el objetivo de evaluar las alternativas y seleccionar una de ellas, 3) un estudio de usuario final que evalúa la versión original de Diaforá frente a uno que incorpora la técnica CMV con la visualización VOD seleccionada.

Descubrimos que el modelo propuesto aumenta significativamente el valor cognitivo de la visualización, lo que facilita que los sujetos de prueba hayan podido identificar las diferencias entre dos versiones de la clasificación.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The classification of living organisms via biological taxonomy is indispensable for environmental conservation [43]. In particular, good quality species classification records are necessary for conservation decision making. From a computing perspective, taxonomies are hierarchical data structures (i.e., rooted trees). Biological taxonomies are dynamic, they are constantly being updated as corrections are required due to reclassification or the discovery of new species. Taxonomists often face the challenge of handling large taxonomies that need to be compared to correct previous taxonomy records. Since taxonomies are hierarchical structures, the challenge is related to the visual comparison of hierarchies, which is a research subject in Information Visualization [52].

Several visual approaches can be used to perform visual comparison of hierarchies [31, 33]. Previous study found that when comparing biological taxonomies, the *edge drawing* method was preferred by taxonomists among other hierarchy comparison methods [53]. The edge drawing method was applied and tested on large datasets with Diaforá[52], an interactive tool that automatically infers and visualizes differences when comparing alternative versions of a biological taxonomy. Although taxonomists showed a high level of satisfaction with this method, scalability limitations were present since the method requires the use of scrolling, which forces the user to recall a part of the hierarchy when the hierarchies being compared are very large.

Expert users considered that, complementary to the *edge drawing* method, a multi-view environment might improve Diaforá for the representation of changes and visual identification of differences.

In this work, we investigated whether the use of coordinated multiple views (CMV)[15] can improve users performance when comparing alternative versions of a biological taxonomy. This work aims to enhance the achievements in the Diaforá system by improving its visualization method in two ways: through the design of a multiple view approach that would contribute to the quick identification of differences as well as applying a visual data reduction approach that displays only the nodes affected by changes, which in turn reduces the need of scrolling. The proposed environment will be evaluated through a user study that involves three stages: design of visualization alternatives, a user study to select an adequate design, and a user study to evaluate effectiveness of the CMV technique and the VOD design (this last stage involved the integration of the selected design to the Diaforá tool).

This thesis document is composed of the following chapters:

- Chapter 1 provides an introduction about the problem and the main objective of this work.
- Chapter 2 contains the theoretical framework with concepts and background information that is relevant for this research problem, as well as the evaluation methods used to validate our results.
- Chapter 3 presents the research problem, the originality, depth and impact of the research and the justification statement.
- Chapter 4 describes the thesis hypothesis and its objectives.
- Chapter 5 describes the methodological aspects that were followed to meet the objectives of this thesis.
- Chapter 6 presents and discusses results.
- Chapter 7 enumerates conclusions and gives ideas for future work.

Chapter 2

Background and Related Work

2.1 Taxonomy classification

In the eighteen century, the scientist and naturalist Carl Von Linné published Sistema Naturæ[39], a book were he proposed a hierarchical categorization system for living organisms. Biological taxonomy consists of a classification method for living organisms into the categories of domain, kingdom, phylum or division, class, order, family, genus and species. In spite of newer species identification techniques that use molecular biology, the Linné system is still prevalent [32]. Indistinctly of the identification method, living organisms are classified through large hierarchies, which are denominated taxonomic trees.

The proper identification and classification of living organisms is of vital importance for their conservation [43]. Taxonomists around the world are actively engaged with proper identification, classification, and documentation of species. Costa Rica, a country with a wide biodiversity, also dedicates efforts for the taxonomic classification of its species; for example, the study about Costa Rican snakes by Solorzano [55], the zooplankton study by Vicencio et al. [60], and the work of taxonomists at the National Museum of Costa Rica, including the recently discovered endemic species *Gentiana Bicentenaria* [21]

The task of maintaining these classifications is complex and involves the effort of analyzing and comparing different versions of the taxonomic trees. Versions arise because of the different classification criteria that experts use to organize species and because information is scattered throughout different sources, which requires data integration efforts [8].

2.2 Hierarchy Comparison

Comparison of complex entities is prevalent in the field of visual analytics [30]. Visual comparison difficulty level increases in relation to the size and complexity of the data set to be analyzed, because it is limited by the user's cognitive and perceptual capacity.

The comparison of complex objects is a general problem that is independent of what objects are being compared [31]. Comparison techniques can be considered into three basic layout categories, which can also be combined; these are: juxtaposition, superposition, and explicit encoding. These three categories are based on how the relationships between the related parts of different objects are encoded:

- **Juxtaposition**: objects are placed separately side by side.
- **Superposition:** objects are placed overlapping each other in the same space.
- **Explicit encoding:** the relationships among objects are explicitly shown by providing a visual encoding of them.

The combination of those visualization layout techniques can create hybrid categories that are used to compensate eventual ambiguity issues in any particular one of the basic categories.

Additionally, literature also refers to possible strategies that users apply in order to examine and compare a set of entities [30]:

- **Scan Sequentially:** the user goes through the data collection sequentially.
- **Select group:** the user chooses a subset of data to compare and proceeds sequentially.
- **Summarize the data:** the user applies a strategy of data abstraction that allows comparison.

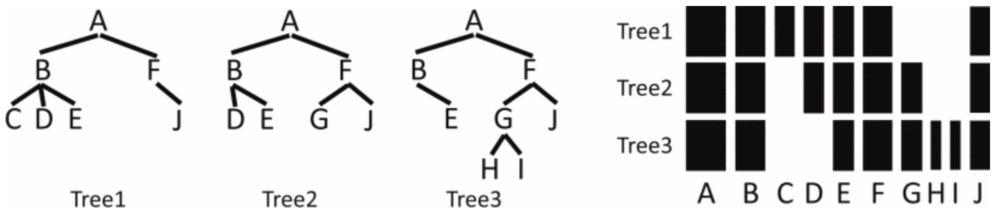


Figure 2.1: BarCode Tree Visualization [38].

The techniques mentioned above have been used for the visualization and comparison of hierarchical data sets. Holten et al. [34] use juxtaposition in a visualization proposal that compares hierarchies. They show the elements that are unique in each hierarchy and those that were rearranged, divided or joined between both hierarchies. Li et al. [38] propose a visualization technique that allows comparison of the topological structure of multiple encrypted trees in a simplified graphical representation similar to a barcode. Figure 2.1 shows an example of this technique where three different hierarchies are compared using the BarCode Tree visualization. BarCode trees use rectangles with different widths that represent each node in the hierarchy. The width encodes the number of children that are under an specific node.

Dong et al. [19] describe PansyTree, a technique that merges multiple hierarchies in a single representation, and uses an icon named *Pansy* that encodes the data items for different datasets in the same hierarchical position . In Figure 2.2 there is a PansyTree that shows the different average entrance examination scores of three different Chinese colleges (Zhongshan *(red)*, Nankai *(blue)*, Wuhan *(yellow)*). Since the colleges have a hierarchical classification with similar disciplines, it is possible to represent the different values for each college using the same hierarchical structure. For example, if we follow the yellow sections in the Pansy icons from section D in the figure 2.2 we see all the disciplines which are also taught in the Zhongshan and Nankai colleges. If we see the section A, we find only the disciplines taught at the Wuhan college.

Bongshin et al. [35] present *CandidTree*, a visualization system that shows the differences between two hierarchical structures. In this system, the differences between the two hierarchies are called uncertainties. These uncertainties make reference to nodes only in the first tree, nodes that moved from first tree, nodes that exist in both trees,nodes moved to the second tree, and nodes that only exist in the second

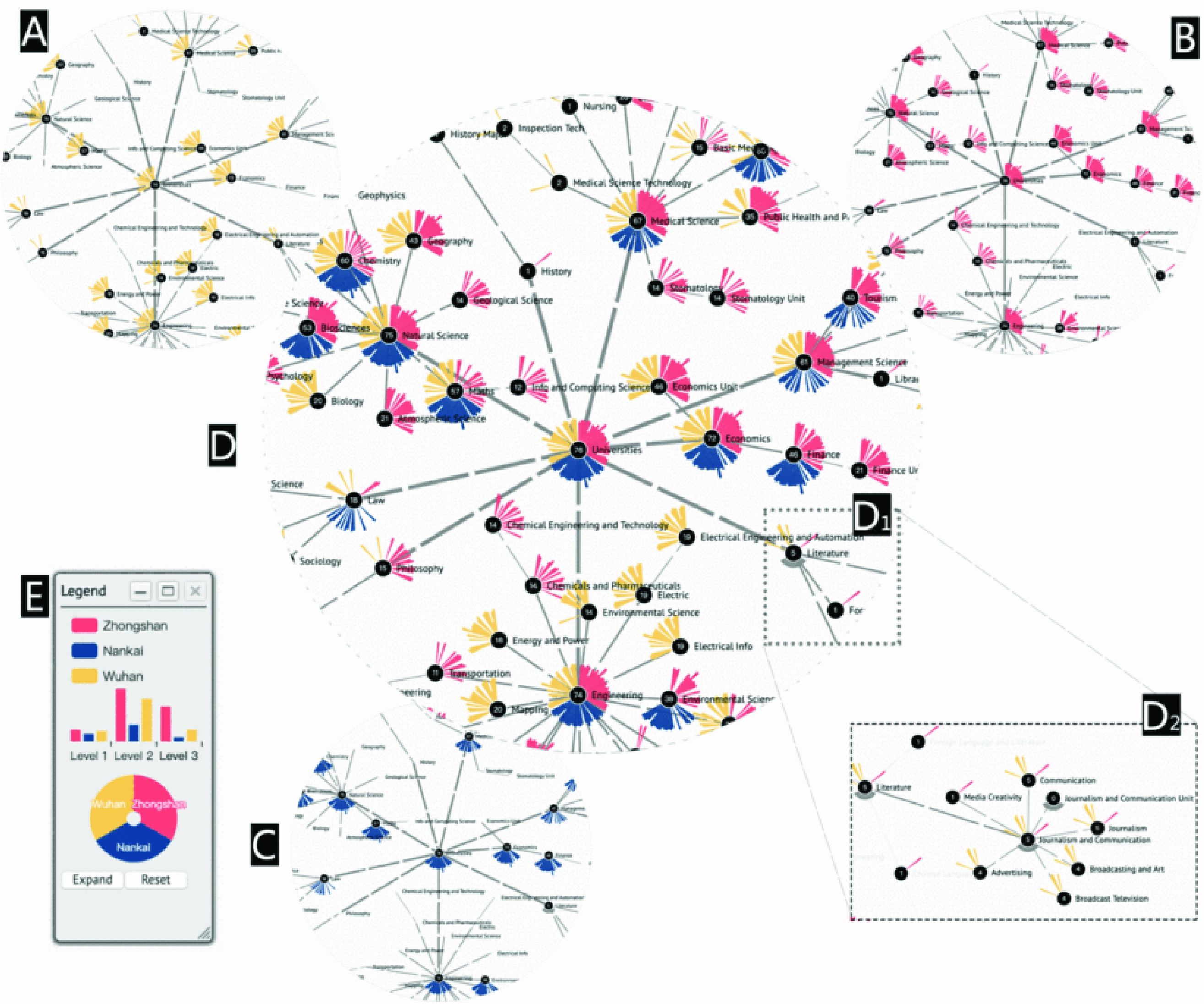


Figure 2.2: PansyTree Visualization [19]

tree. The method consists of merging the two different trees into one and visualizing the structural uncertainty. Through the use of interaction tools it highlights the differences in order to make the analysis easier for users. Figure 2.3 shows the CandidTree tool applied to a biological taxonomy.Specifically a change in the taxa *Megapodiidae* that is moived from the group *Craciformes* to the group *Galliformes*.

Graham et al. [33] survey the visualization of multiple trees and present an overview of various techniques that allow visual comparison of trees. These techniques are illustrated in Figure 2.4 and described as follows:

a. Edge drawing: visualizes corresponding relations between two spatially separated trees.

b. **C**olors: differences and similarities are highlighted by a color encoding between two spatially separated representations.

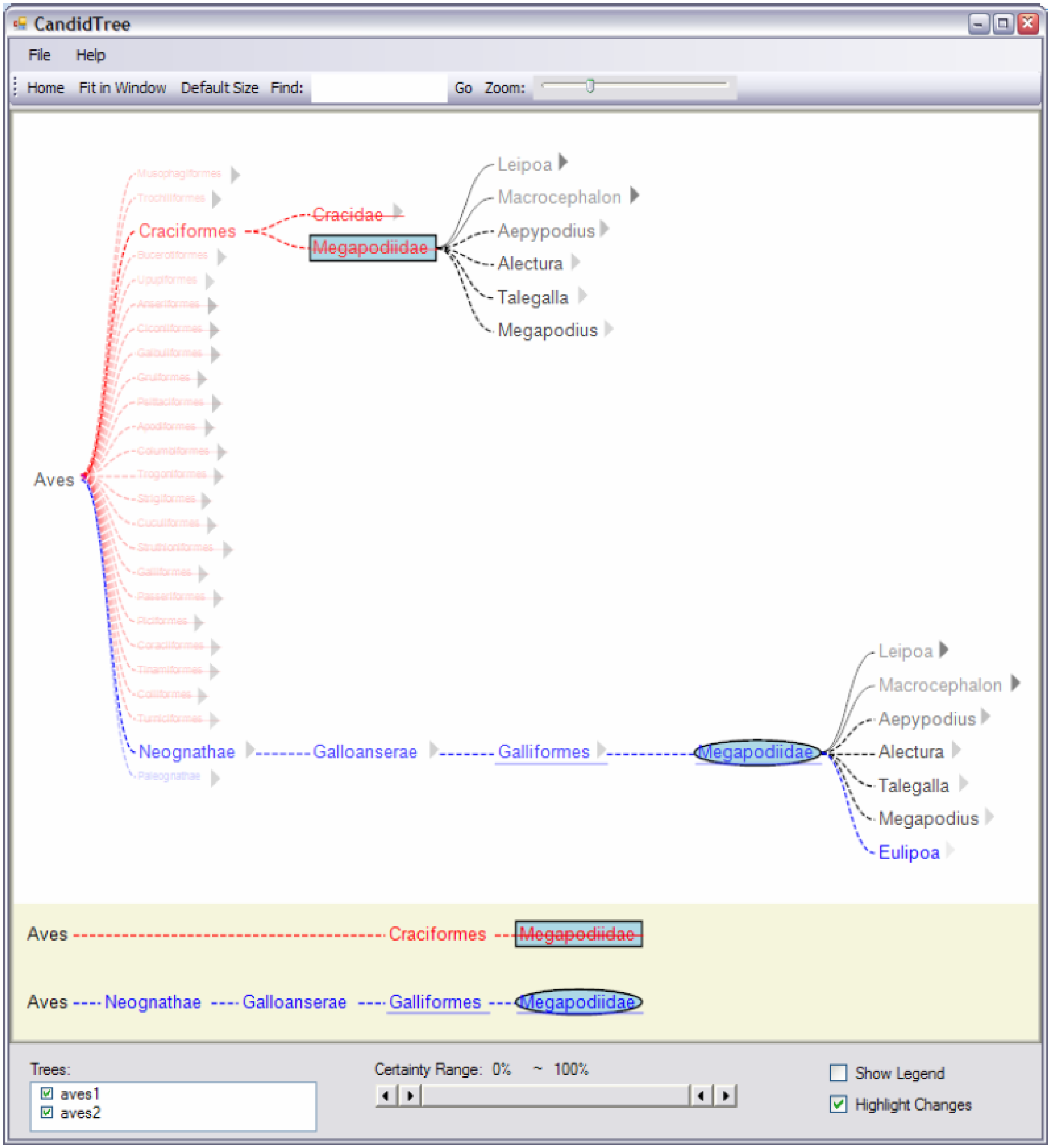
c. Animation: changes are visualized via animation by showing the transformation of nodes in one tree towards the other one (trees are temporarily separated).

d. Matrix comparison: one tree is collocated at one side of the matrix and the other one on top. Either differences or similarities are shown on the cells.

e. Spatial agglomeration: hierarchies are merged into one.

The edge bundling technique is a method used to simplify graph drawings [34, 37]. In a graph, edges are used to encode information such as relations or associations between nodes. Many graphs and trees may have many more edges than nodes and it is very likely that edge crossings will end up producing visual clutter, so by grouping edges together, clearer representations can be obtained. Holten et al. [34] introduced edge bundling to compare two hierarchies and visualize explicitly the relations between matching subhierarchies. Lhuillier et al. [37] present a survey that summarizes techniques used for path bundling as a method to simplify graph drawings.

Even-though, designers have many ways to visually represent data, not all representations are equally perceived by the users [45]. Visual perception has an important



Trees:
☑ aves1
☑ aves2

Certainty Range: 0% ~ 100%

☐ Show Legend
☑ Highlight Changes

Figure 2.3: Candid Tree Visualization Tool used with biological taxonomies [35].

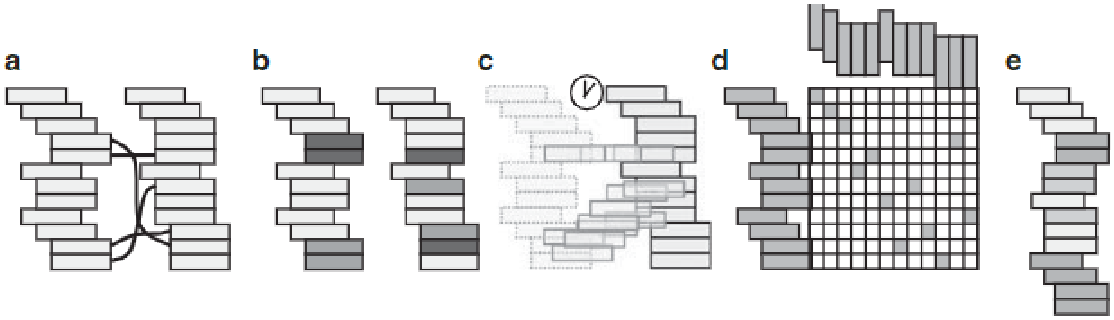


Figure 2.4: Nodes comparison methods: (a)*edge drawing*, (b)colors, (c)animations, (d)matrix representation, (e)Spatial agglomeration. Taken from [33]

role in how the information presented is analyzed and the relative performance of getting the right answer when someone is doing visual comparison tasks; for example, the identification of a maximum delta between data series. The following concepts from perceptual psychology are used to evaluate the performance of a visual comparison representation [45]:

- **Co-location:** The involvement of visual features like length, orientation, and motion, within the same space is very useful to represent distinct features among data sets. Comparison between two regions involves the effort to “remember" parts of the data to compare it like in the “Spot the difference" games. This has a limit given by the cognitive capacity of the users.
- **Symmetry**: This feature can be useful when representing data sets that have a minor difference but when you put a visualization next to the other it is easy to find the differences in images that are otherwise mirror images. Humans are specially sensitive to symmetry.
- **Movement:** Wired as a primitive element of human vision, this can be used to extract information like statistics and structure. It also has a limited span of attention from the user that rapidly decays when several elements are moving simultaneously.

Peña-Araya et al. [48] studied visualization for variables that evolve over space and time. Their work compares techniques that use juxtaposition in either all locations for a given time step or all-time steps for a given location, and a technique that involves the use of symbols to encode the important thematic attributes over a map. They compared techniques such as juxtaposition and small multiples and found that juxtaposition performs better and provide an easier to understand representation than the small multiples representation.

2.3 Focused and Contextual Views

Strategies such as overview+detail, zooming and focus+context are used in order to avoid overcrowded and cluttered visualizations. The use of interactive lenses is widely extended to provide alternative visual representations for selected regions of interest [57]. The relation between the user capacity to attend more information per unit time is stated as *the principle of reducing the cost structure of information* [17]. The idea is to minimize the cost structure of information by placing more relevant information on to a single display.

Lenses are lightweight tools that are included to solve a specific localized problem by altering a selected part of the visual representation of the data[57]. Within the lens boundaries the current information that is visible can be modified and enriched with new data. To avoid visualizations that are overcrowded with data, part of the information can be included in a lens function to be displayed on demand.

According to Björk et al. [11] the basic idea with focus and context visualizations is to enable users to have the object of primary interest presented in detail on the screen while giving and overview or context of the other data at the same time. Focus and context techniques are useful when users require both a detailed visualization and an overview of the data simultaneously.

Borland et al. [13] state that part of important information that can be omitted from the visualization could threat the capacity of users to solve their analytical problems. Those threats are introduced by all the unseen information because that can come from a variety of sources that includes summarizing to reduce the fine-grained information and omitted dimensions of data that is not visually represented within the visualization. Making the missing information visible by a contextual visualization can help to reduce the threats to validity in the users observations.

Phylum	Registered Species
Echinodermata	18 458
Annelida	24 644
Arthropoda	1 325 303

Table 2.1: Sample of registered species in GBIF [29]

2.4 Multi-View Visualizations

A multiple view system uses two or more distinct visualizations to support the investigation of a single conceptual entity [61]. With this approach it is possible to provide a complementary visualization to enhance the efficiency and efficacy of a single view system.

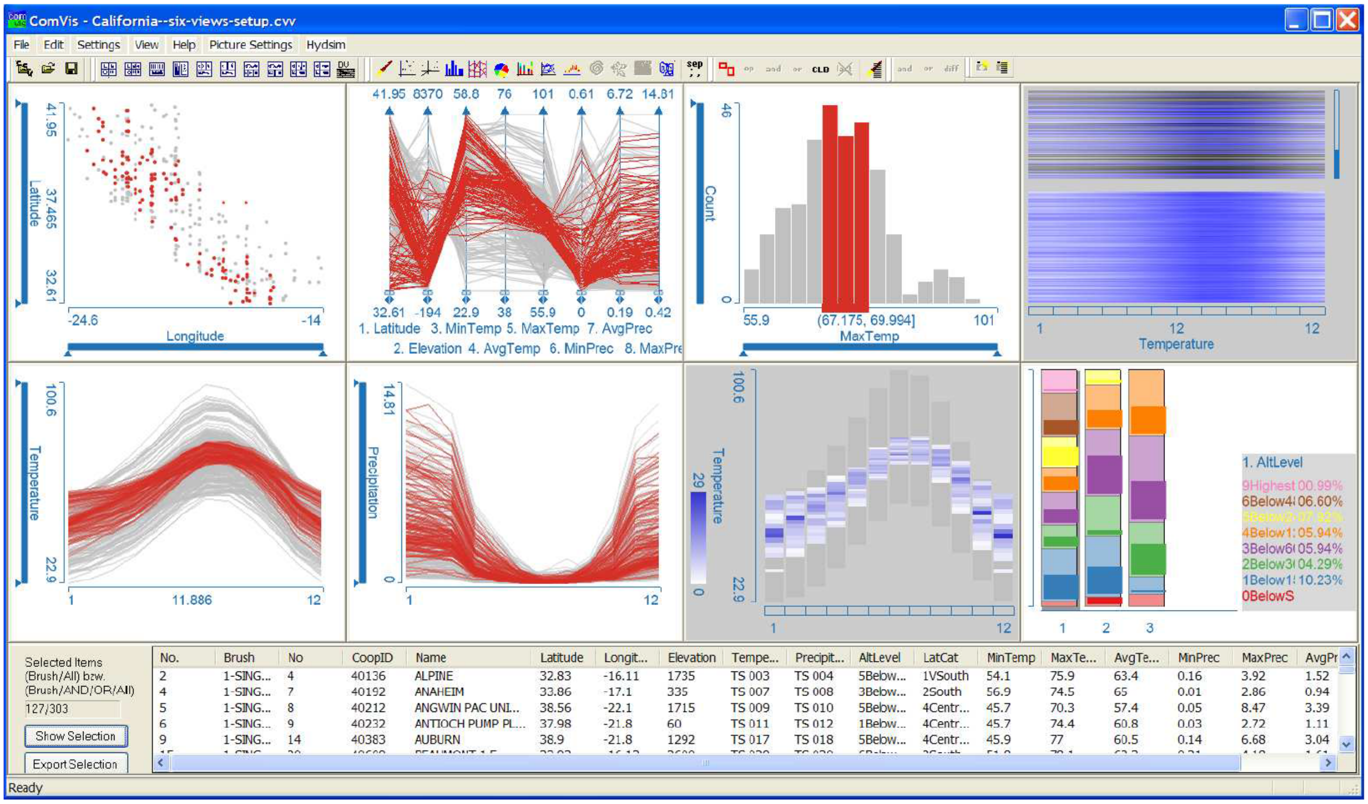
Coordinated Multiple Views (CMV) are used in several user interfaces where contrasting different aspects of the data is needed [15]. CMVs are also used to explore and analyze large datasets by providing different consistent viewpoints and user interactions. CMVs can become a valuable tool for users to achieve a better understanding of their data [50].

Figure 2.5 shows an example of ComVis [42], a tool that presents different views of weather data. The important feature is that all views are coordinated; for example, when selecting one location, all other views are automatically updated showing the corresponding data.

2.5 Taxonomy Comparison

Biological taxonomies [39] are structures in which known living organisms are organized and classified. Groups in the taxonomy can have a considerable number of species; consequently, the corresponding hierarchical structures can consist of a large number of nodes. For example, Table 2.1 presents examples of the amount of species of four different phylums of the kingdom Animalia according to GBIF (Global Biodiversity Information Facility) database.

Due to the large amount of records that may exist in a particular taxonomy, the process of comparison and curation of taxonomies is complex. It involves work with voluminous data sets which might be difficult to visualize and analyze in the space



Selected Items
(Brush/All) bzw.
(Brush/AND/OR/All)
127/303

No.	Brush	No	CoopID	Name	Latitude	Longit...	Elevation	Tempe...	Precipit...	AltLevel	LatCat	MinTemp	MaxTe...	AvgTe...	MinPrec	MaxPrec	AvgPr
2	1-SING...	4	40136	ALPINE	32.83	-16.11	1735	TS 003	TS 004	5Bebw...	1VSouth	54.1	75.9	63.4	0.16	3.92	1.52
4	1-SING...	7	40192	ANAHEIM	33.86	-17.1	335	TS 007	TS 008	3Bebw...	2South	56.9	74.5	65	0.01	2.86	0.94
5	1-SING...	8	40212	ANGWIN PAC UNI...	38.56	-22.1	1715	TS 009	TS 010	5Bebw...	4Centr...	45.7	70.3	57.4	0.05	8.47	3.39
6	1-SING...	9	40232	ANTIOCH PUMP PL...	37.98	-21.8	60	TS 011	TS 012	1Bebw...	4Centr...	45.7	74.4	60.8	0.03	2.72	1.11
9	1-SING...	14	40383	AUBURN	38.9	-21.8	1292	TS 017	TS 018	5Bebw...	4Centr...	45.9	77	60.5	0.14	6.68	3.04

Figure 2.5: An example of the ComVis system using a CMV environment [42].

available on the screen of a computer or electronic device.

Entities such as museums, herbaria, and other institutions dedicated to the conservation of biodiversity keep records of the taxonomic classification of species in conjunction with other physical records with specimen examples, which are constantly updated due to the discovery of new species, or the need to correct or complete previously stored information.

According to Ball-Damerow et al. [9], the number of digitized biologic data sets is increasing and becoming openly accessible to be used and consulted by any researcher. Efforts like *GBIF*, *BOLDSystems* and *SpeciesLink* are part of the most used biodiversity databases. Those databases are constantly cured and corrected by experts. The quality of those databases is very important for biodiversity conservation because taxonomies are primary information to conduct inventories to monitor changes in populations and measure the impact of invasive species, pollution, and climate change, among other threats.

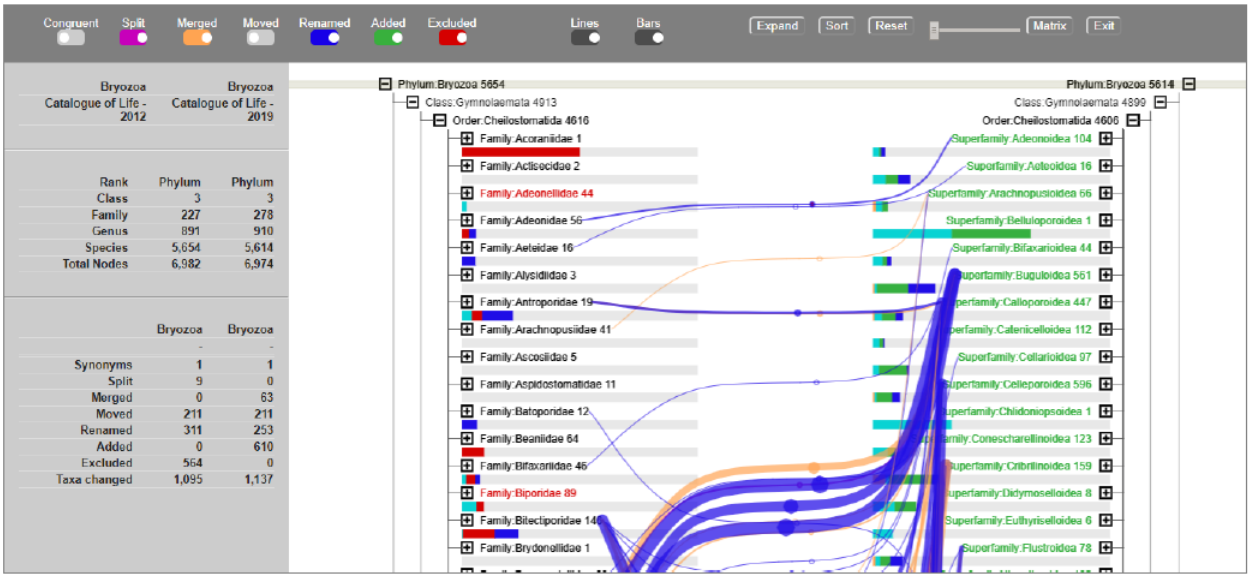
Vaidya et al. [59] show the correction process of a biological taxonomy, and describe how when taxonomists encounter what they believe is an exemplar of a living organism that had not been previously registered, they have to collect evidence to support the hypothesis that it is distinct enough from any other known species. Once they have enough evidence, they have to publish the the discovery in a peer-reviewed journal. Another correction situation occurs when taxonomists have enough evidence to believe that a species had been wrongly categorized within the taxonomy. This process of taxonomy refinement involves lumping and splitting over the existent taxonomy. Taxonomies are being continually developed it is a continuously improving process.

Another study from Sancho-Chavarraía et al. with the participation of experts in biological taxonomies of three different countries evaluated four methods used to compare two hierarchical structures [53]. The evaluated methods are *edge drawing*, *matrix representation*, and *animation and agglomeration*. The results indicated that the participant users preferred the *edge drawing* method while performing taxonomy comparisons.

Diaforá ([52]) is an interactive tool that supports the visual comparison of two versions of a taxonomy. It automatically identifies and shows with visual keys the differences and similarities between the two compared taxonomies. It identifies merges, splits, and rename of taxa, among others. Figure 2.6 presents an example where the Diaforá system is being used to compare two different versions of the *Bryozoa* phylum. The different edges that connect nodes from the two hierarchies represent the changes, using a color scheme to show the different types of changes.

2.6 Taxonomic Databases

Damerow et al. [9] indicated that the goal of creating an online taxonomic database that is constantly updated by taxonomic experts is not reached yet, but there are some important efforts in some specific taxonomic groups that have a very active community and data sources with updated and reliable data. Table 2.2 presents some public access databases that are maintained by distinct communities of biolog-



Bryozoa	Bryozoa
Catalogue of Life - 2012	Catalogue of Life - 2019

Rank	Phylum	Phylum
Class	3	3
Family	227	278
Genus	891	910
Species	5,654	5,614
Total Nodes	6,982	6,974

	Bryozoa	Bryozoa
	-	-
Synonyms	1	1
Split	9	0
Merged	0	63
Moved	211	211
Renamed	311	253
Added	0	610
Excluded	564	0
Taxa changed	1,095	1,137

Figure 2.6: Diaforá Taxonomoy Comparison Tool

ical experts. Those databases contain information about some of the most studied organisms and can be used to study their taxonomic information.

Catalogue of Life [51] is a global initiative that collects data from over 200 expert taxonomic databases. The data exposed through a COL annual checklist contains peer-reviewed taxonomic classification of a vast number of organisms. Containing such amount of valuable information about most of the recognized taxa around the world, COL is a key partner in the major initiatives that inform about global biodiversity. According to the Global Biodiversity Information Facility (*GBIF*), the data sets listed in Table 2.2 are part of the information included in the COL integrated taxonomy data source. The annual checklists of COL are snapshots of the integrated taxonomic data each year, so it can be considered as a reliable data source to study the evolution of taxonomic classification over time. Considering this data, COL is one of the most reliable data sources and it is widely used to share and distribute information related to biodiversity classification among different scientific institutions, becoming a way to achieve compatibility between others that already use COL as the primary data repository. Estimates indicate that by 2020 COL would be tracking 95% of species known to science [51].

Taxonomic Data Set	Access and Data
WoRMS: World Register of Marine Species. Provides an authoritative and comprehensive list of marine organisms[12]	Data is open to access via public API at WoRMS REST webservice.
FishBase: A global biodiversity information system about fish species [28].	Data can be accessed at fishbase.ropensci.org
MilliBase: Offers a global taxonomic database on millipede like organisms, members of the Diplopoda class one of the most species-rich class in the Myriapoda subphylum [54].	Data can be accessed through Catalogue of Life checklist services[51].
MolluscaBase: A global biodiversity information system about molluscan species [20].	Data can be accessed through Catalogue of Life checklist services[51].

Table 2.2: Examples of public access taxonomic databases

2.7 Information Visualization Evaluation

According to Väätäjä et al. [58], heuristics are particularly useful when evaluating information visualization interfaces. Those heuristics are focused on how information visualization techniques enable people to get a better understanding about a dataset[63]. Heuristic evaluations are a useful method to discover possible usability problems, Forsell and Johansson [27] propose a set of ten heuristics used to evaluate information visualization techniques. In table 2.3 there is a summary of those ten heuristics that need to be considered to develop a useful visualization tool.

To evaluate the impact of the mentioned heuristics, a user study is required. User studies must resolve the problem of experiment bias, creating experiments that support the hypothesis from data that is statistically significant. Colin Ware [62] stated that the required number of subjects to perform statistically reliable experiments is between 12 and 20 subjects. According to Nielsen [44], the necessary amount of people in a user study research to test usability is 5, but when we are aiming for statistics at least 20 users are required to gain enough statistically significant numbers. In general, the number of subjects and the number of observations depend on the variability of responses from a single subject and the variability between distinct subjects. Another important factor in the number of test subjects is their specialization and availability.

The tasks that will be part of the experiment design should be both theoretically interesting and common in the field of HCI (*Human-Computer interaction*). In addition, the controls provided to the user for performing a specific task should be simple enough in order to provide a useful measure of performance; for example, an user can use the right mouse button for no and the left one to indicate yes[62].

Heuristic Name	Description
1. Spatial Organization	Describes the overall layout of a visual representation. This heuristic is related on how easy is to locate an information element on the display.
2. Information coding	Related to the mapping of data elements to visual objects like colors and shapes.
3. Orientation and Help	Describes function like support for the user to control levels of detail, redo/undo of actions and representing additional information.
4. Data set reduction	Mechanism to reduce the data set, filtering to focus on data of interest.
5. Recognition rather than recall	Do not depend on user memory to carry out comparison tasks.
6. Remove the extraneous	Remove any extra information than can be a distraction and take the eye away from seeing the data or making comparisons.
7. Prompting	Refers to all the means available to guide the user towards specific actions.
8. Minimal actions	Refers to the amount of steps that are required to perform an action or accomplish a goal.
9. Consistency	Refers to the way design choices are maintained in similar contexts, and are different when applied in different contexts.
10. Flexibility	Refers to the means available to perform a particular tasks in different ways considering the possible working strategies or habits of each user.

Table 2.3: Forsell and Johansson Heuristic set for evaluation in information visualization[27].

Chapter 3

Problem Definition

3.1 Problem

According to the research performed by Sancho et al. the *edge drawing* method has been preferred by taxonomists when comparing biological taxonomies for the identification of similarities and differences [53]. Edge drawing has, however, scalability issues [33] due to the space occupied by the lines that connect the nodes between the different trees, as well as by the chance that related nodes are placed out of the visual focus of the screen. Furthermore, the *edge drawing* method carries the difficulty of determining nodes that belong to only one of the hierarchies; that is, nodes that were removed in one of the structures or nodes that were inserted into one of the hierarchies, because in such cases it is not possible to draw a line between one tree and the other because there is no alternative node that allow to create a connection between both hierarchies.

Taxonomy refinement is a frequent process of taxonomic work [49], which contributes to better identify and document living species, and therefore their conservation. The process of verifying and correcting a given taxonomy requires expert knowledge and meticulous observation of specimens in nature, that is why providing tools that assist the identification of differences between versions of a taxonomic classification can contribute to share the correct data and create a unified global database of taxonomic information [51].

We hypothesize that the user's performance on identifying changes would improve by using multiple views when comparing alternative versions of a biological taxonomy.

This work focuses on extending the Diaforá system by providing an alternative view to the *edge drawing* representation. The alternative view is oriented to visualize only the differences between the two compared taxonomic trees in order to reduce the complexity of the generated visualization, and to give the opportunity to taxonomists to analyze those differences in a simpler way.

3.2 Originality

To the best of our knowledge, this is the first attempt to design and create a coordinated multiple view environment tool for comparison of alternative versions of biological taxonomies considering a set of relevant tasks that taxonomists perform for the identification of specific changes between versions. Moreover, the combination of a previously studied method, *edge drawing,* and a novel visualization that focuses on showing only the differences between two hierarchies has not been approached.

We believe that the proposed method could assist taxonomists to identify changes more easily than the single *edge drawing* view.

3.3 Depth

This work contemplates not only the design and development of a visualization tool that can be used to compare two different versions of a biological taxonomy –a common task in the taxonomy refinement process– but also a two-stage user study. The first stage will evaluate two visualization prototypes focused on displaying only the differences between two alternative versions of a biological taxonomy. The second stage will compare the original Diaforá system with an extended version of the it, which implements the selected visualization from stage one in a CMV approach.

Our hypothesis is that CMV will positively impact the user's performance while identifying differences in biological taxonomies.

3.4 Impact

As a multidisciplinary topic (biology and information visualization), this work pretends to contribute both areas by providing a novel design and a user study that evaluates the effects of coordinated multi-views environments in the biological taxonomy comparison process. We propose a visualization design that focuses on showing only the taxa affected by changes between two alternative versions of a biological taxonomy. With this approach we expect to reduce the actual cognitive complexity of a large *edge drawing* visualization. This visualization can be used to display the changes between hierarchies in other domains as well. By using the coordinated multiple views approach (CMV), we pretend to provide a clearer representation of differences, highlighting taxa that were affected by changes. Furthermore, we expect to make the CMV visualization tool available to be used by the scientific community to perform taxonomic comparisons and to contribute to biodiversity conservation efforts. By increasing the ease of use of a biological taxonomy correction tool, we hope that the effort of maintaining up-to-date and reliable records of biological information will be greatly reduced.

3.5 Justification Statement

The importance of taxonomic information in biological conservation efforts has been extensively documented[40, 46, 56]. The knowledge about the types of “species" in the planet is required and widely used for planning conservationist actions; for example, it is used to recognize threatened species and species that are covered by legislation[40]. Because of that, it is important to provide taxonomists the right tools to carry out their labor of correctly classify all the living organisms; therefore, the comparison of taxonomic classifications is indispensable for the reconciliation of alternative versions of biological taxonomies[53].

Addressing the problem of visualizing differences between two versions of a biological taxonomy is an important topic to help taxonomists to reconcile the classifications. The *Diaforá system* [52] is an effort that presents an interactive tool that infers and visualizes those differences by using the *edge drawing* technique.

This work aims to enhance the Diaforá system, by providing a *coordinated multi-*

ple view environment that can assist taxonomists to analyze the differences between alternative versions of a taxonomy and to focus on taxa that are important for their current task. Using an alternative visualization technique that reduces the amount of information displayed on the screen while keeping the focus on those elements that were affected by changes between the two taxonomic classifications should help taxonomists to be more efficient and with a lower expected error rate when reconciling taxonomies.

Chapter 4

Hypothesis and Objectives

4.1 Hypothesis

The use of a multi-view environment (CMV) that combines a VOD visualization and the edge drawing method, can enhance the expert user's overall performance in terms of efficiency (task-completion time) and effectiveness, compared to the unique usage of edge drawing technique, when identifying changes between two versions of a biological taxonomy.

4.2 Main Objective

To design and develop a visualization environment that improves the efficiency and effectiveness for the comparison of biological taxonomies using a coordinated multiple views (CMV) approach, combining both a visualization that focuses on the differences that occur between two versions of a taxonomy and the edge drawing technique.

4.3 Specific Objectives

a. To design a set of visualization prototypes focused on improving the use of screen space by displaying only the differences between two taxonomy versions.

b. To improve the efficiency and effectiveness of users while identifying differences between alternative versions of biological taxonomies by extending the Diaforá system into a CMV environment that incorporates a VOD design combined with the edge drawing visualization method.

c. To evaluate both effectiveness and efficiency of the proposed model by consulting expert users in the analysis of biological taxonomies through a user study.

4.4 Scope and Limitations

The proposed model should enable the user to visualize the differences between biological taxonomies, but it is necessary to consider the following limitations:

- At most, the comparison would occur at the family level. This means that even if it is possible to perform comparison between upper levels of the taxonomy (that is, kingdom, domain, and phylum levels) because there is very little chance that change occurs at those levels. Also we exclude from comparison changes any intermediate level under the species category like sub-species since these categories are not widely used and introduce extra complexity in the system.

- The comparison must be performed against the same taxonomic group. To display the proposed visualization correctly, both taxonomies must be of the same group, but different revisions or versions.

- It is expected a high level of coincidence between taxonomic versions,since taxonomies are constantly being refined and corrected since its proposition [39].

- Since the tool is proposed to be capable of comparing taxonomies from different data sources, but the data could be in different formats, a layer of normalization is expected to transform the database to a format that can be used by Diaforá System. The transformation layer should be extensible, so more databases can be added, but we will limit the development to the COL Annual Checklist service.

Chapter 5

Research Methodology

This section presents the methodology followed in order to accomplish the proposed objectives of this research and to test our hypothesis.

5.1 Design of the visualization environment

5.1.1 Review of tasks and types of changes

A first step was to understand and analyze the visualizations provided by the Diaforá system[52] and its behavior with real data. Diaforá uses data from Catalogue of Life, specifically, it considers the following data fields:

- **Author**: The name of the scientist(s) who first described the taxa name.
- **Date**: The date the name of the specimen was published.
- **Name**: The scientific name given to the taxa.

Those fields are used by Diaforá to infer the changes between two alternative versions of a biological taxonomy. Using the difference between those data features the system can identify changes like splits, removed, added, or moved.

We also noticed that the databases downloaded from Catalogue of Life might contain inconsistencies, especially in the data that refers to the author. For instance, due to misspelling of the author, a species might be identified as *excluded* or *added*

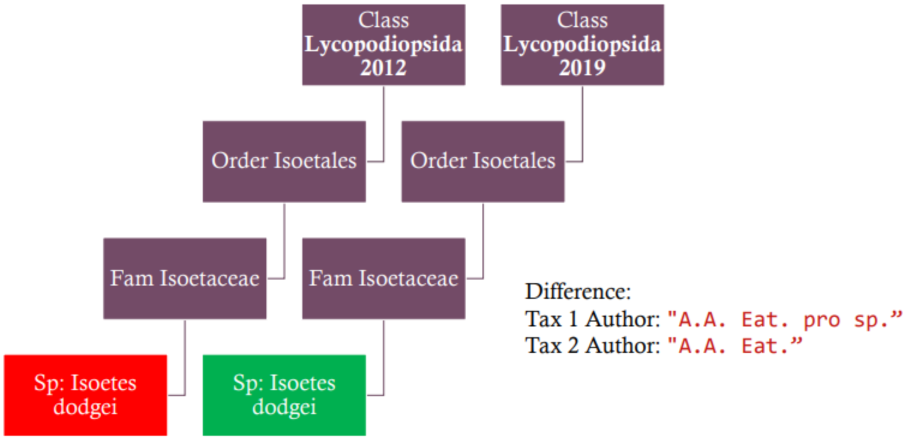


Figure 5.1: Author's change example at *Lycopodiopsida* taxonomy comparison between 2012 and 2019 Catalogue of Life checklists.

in one of the alternative taxonomies. This is a limitation given by the available data sources, it is not due to the changes in the taxonomies. Therefore, recognizing this situation, we considered necessary to add a new change type with the purpose of identifying in the visualization those taxa that remain in the same position of the taxonomic tree but have a difference in their author's data. We named the corresponding task **Author Changed**. This type of change can be explained by the diagram on the image 5.1 where the species *Isoetes dodgei* has a change in his author's field that changed from *"A.A. Eat. pro sp."* to *"A.A. Eat."*. Those changes could be caused by a previous fix in the database format or due to an actual change in the taxa definition.

Therefore, the types of changes of changes to be considered are: *split, merged, renamed, added, removed* and *author changed* [53].

5.1.2 Design of Alternative Views

We propose taxonomy visualization design that focuses on displaying *only* those nodes that were affected by changes (visualization of only differences, VOD). Our rationale is that nodes in the hierarchy that where not affected by changes occupy valuable space in the screen and do not provide as much information as the nodes that changed. By displaying only the affected nodes, we believe to contribute to reduce cluttering, provide the user with a clearer visualization of the changes, and allow them to focus on the identification of the differences between the two biological taxonomies.

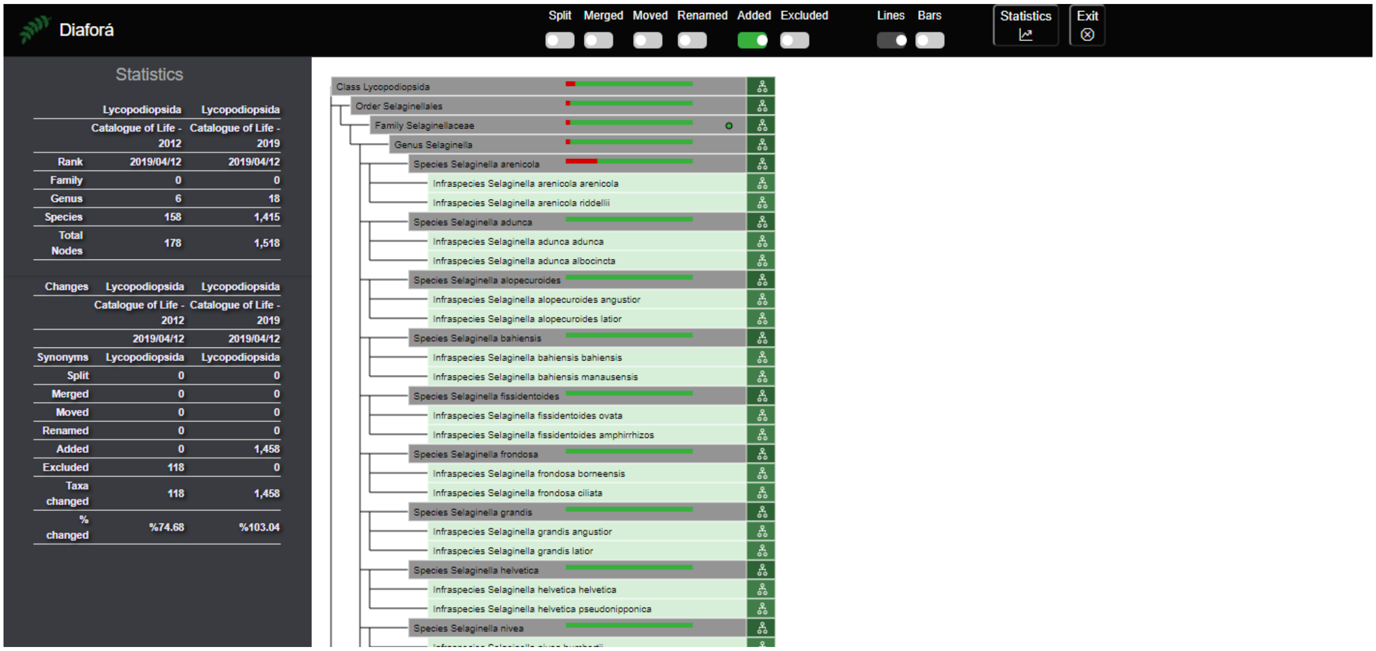
We designed two VOD visualizations. For both, the screen layout follows the proposed design of the Diaforá tool: on top there are selection buttons, on the left a panel that presents statistics, and on the center of the screen is the area for comparison. The buttons on top of the screen allow users to carry out the tasks. Buttons are encoded with colors for the identification of changes: magenta for splits, orange for merges, cyan for moved taxa, pink for author changes, blue for renamed taxa, green for new taxa added and red for excluded taxa. More than one button can be selected.

The first design consists of an indented layout, similar to the original Diaforá layout. The second visualization uses an alternative circular layout. The designs were modelled by using prototyping and will be described in the following sections. Then, we performed a user study to validate that these designs, in combination with coordinated multiple views approach, can improve the user's performance while identifying changes between the compared taxonomies.

In a first stage of the user study, the two VOD visualizations were presented to the users in order to determine which one presents the information in an easier-to-understand way. In a second stage we incorporate the best performing VOD design into a coordinated multiple view environment integrated to the original Diaforá's *edge-drawing* visualization.

Design and Development Process

To design the two VOD visualizations, we started by creating several prototypes evaluating different aspects of the way to display the information about taxonomic changes. We vary different features like colors, shapes, font types, interaction points, and additional information displayed by the VOD designs. By refining these early prototypes we select the elements that according to our perspective enhance the experience of identifying the taxonomic changes. To develop these prototypes we started with the previous version of the Diaforá system which was developed using



Statistics

	Lycopodiopsida	Lycopodiopsida
	Catalogue of Life - 2012	Catalogue of Life - 2019
Rank	2019/04/12	2019/04/12
Family	0	0
Genus	6	18
Species	158	1,415
Total Nodes	178	1,518

Changes	Lycopodiopsida	Lycopodiopsida
	Catalogue of Life - 2012	Catalogue of Life - 2019
	2019/04/12	2019/04/12
Synonyms	Lycopodiopsida	Lycopodiopsida
Split	0	0
Merged	0	0
Moved	0	0
Renamed	0	0
Added	0	1,458
Excluded	118	0
Taxa changed	118	1,458
% changed	%74.68	%103.04

Figure 5.2: Indented Tree Taxonomic Differences Visualization

EJS and HTML, then we incorporate the VOD visualization by generating interactive SVG visualizations of the prototypes using the D3.js library. To have a similar experience to the original Diaforá system, we kept the same interaction points in the web portal and include a different look and feel to make it easier to identify the old and the new version of Diaforá.

Taxonomic Difference Indented Tree Design

Figure 5.2 illustrates the first design. Besides visualizing only differences, the design includes colored shadings that provide visual clues about the proportion of types of changes involved as well as detailed on-demand information. This figure illustrates the differences between two alternative versions of the *Lycopodiopsida* taxonomy taken from 2012 *Catalogue of Life* checklist and the 2019 *Catalogue of Life* checklist. Notice that new taxa are displayed in green color and taxa that are not present in the 2019 version of the taxonomy are highlighted in green color.

By taking a closer look we can identify some characteristics of the visualization that provide the user with further information about the changes that are repre-

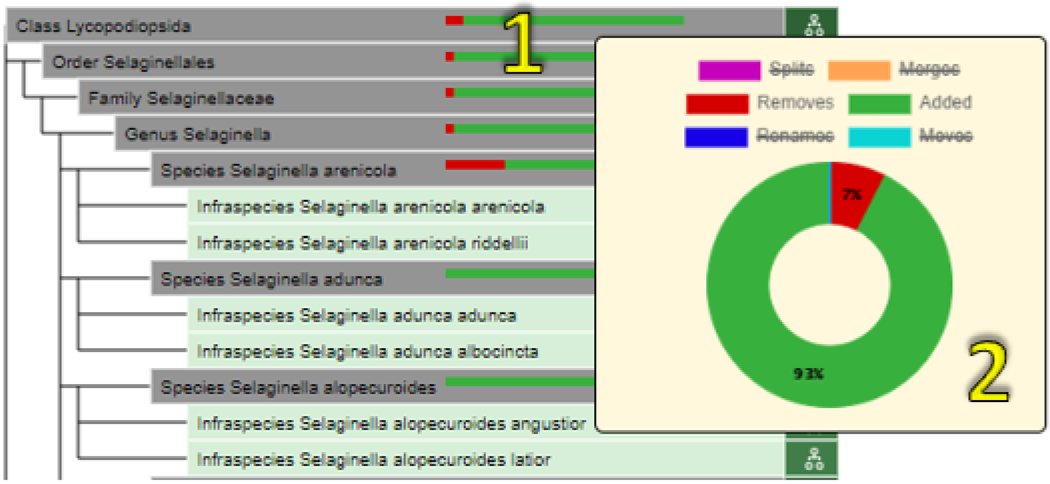


Figure 5.3: Changes distribution on indented tree taxonomic differences visualization

sented in the indented tree view. Figure 5.3 shows an example of the distribution of changes on the taxonomic class *Lycopodiopsida*. The first aspect that we want to explain about the visualization is the distribution bars illustrated in the figure and highlighted with the number 1. The bars tell the user about the nature of changes at any given level of the taxonomy. They reflect the distribution of changes in that taxonomic group; that is, the length of the bar indicates the proportion of changes, and the color of the bar indicates the type of change (new added taxa, missing taxa, splits, etc.). The bar gives the user a quick overview about the changes distribution. By hovering the mouse cursor over the bar, it additionally shows an indicator with the actual detailed distribution on that taxonomic level (see number 2 in the figure).

In order to provide additional valuable information, the user has some interaction points with the visualization. For example by hovering over the taxonomic name, the visualization shows a tooltip with the amount of changes on that specific taxa (see figure 5.4). Moreover if the amount of nodes forces the user to scroll on the visualization, the full taxonomy tree can be consulted with the hierarchy icon on the green area at the left of the node, it provides the full taxonomy from the tree root to the selected node (see figure 5.5).

Finally, a marker that identifies the taxonomic family with the larger amount of

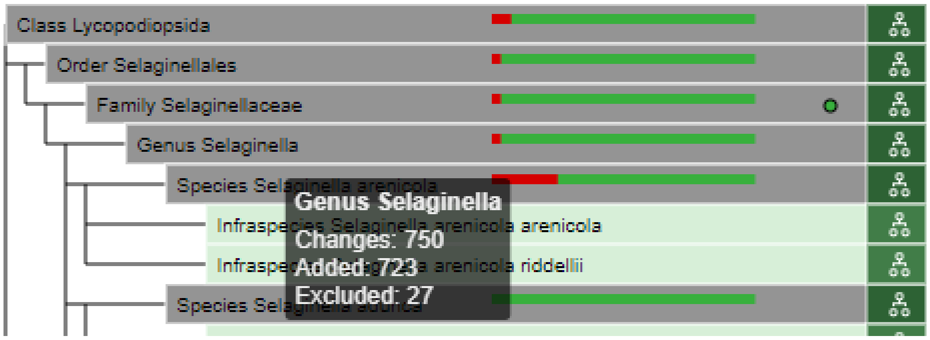


Figure 5.4: Indented tree showing amount of changes in taxa.

changes of any type is given to the node that identifies such a family. For example, in Figure 5.4 on *Family Selaginellaceae*, there is a green dot that informs the user that is the family with the largest amount of additions in the whole taxonomy comparison.

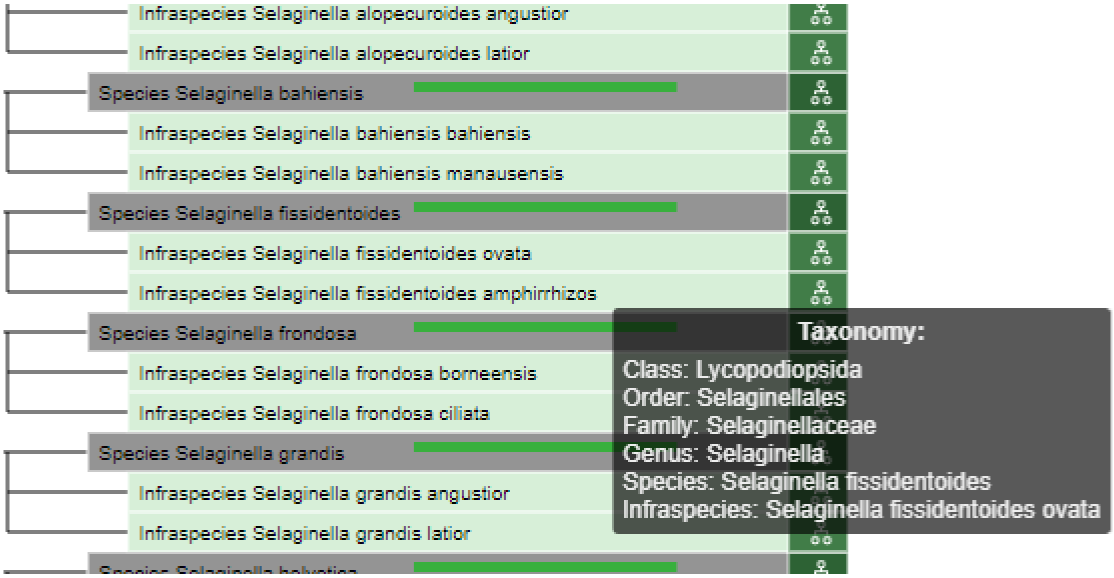
Taxonomic Difference Radial Design

The second visualization design aims to improve the screen space usage not requiring the user to scroll to see the complete difference tree. To offer such functionality, we present a radial representation of the taxonomic difference tree.

The radial representation allows us to show a larger amount of data on the screen and at the same time use aggregation to reduce the visual complexity of our representation by hiding some nodes but allowing the user to drill down on the visualization to review the taxonomic data.

Figure 5.6 presents an example that illustrates the taxonomic difference tree using a radial representation. Similar to the *Indented Tree* explained above, this visualization provides insight about the distribution of changes for every level of the taxonomic comparison. To illustrate that, we introduce the concept of *satellite bars*.

The satellite bars can be a full ring when the user selects the root of the hierarchy or an arc if the user is hovering an internal level of the taxonomy. These satellite bars show the distribution of changes on any particular level of the biological taxonomy. For example, Figure 5.7 shows the distribution of changes for the *Isoetaceae* family, indicating that the 84.7% of changes are additions and the rest 15.3% of changes are exclusions.



Infraspecies Selaginella alopecuroides angustior

Infraspecies Selaginella alopecuroides latior

Species Selaginella bahiensis

Infraspecies Selaginella bahiensis bahiensis

Infraspecies Selaginella bahiensis manausensis

Species Selaginella fissidentoides

Infraspecies Selaginella fissidentoides ovata

Infraspecies Selaginella fissidentoides amphirrhizos

Species Selaginella frondosa

Infraspecies Selaginella frondosa borneensis

Infraspecies Selaginella frondosa ciliata

Species Selaginella grandis

Infraspecies Selaginella grandis angustior

Infraspecies Selaginella grandis latior

Taxonomy:

Class: Lycopodiopsida
Order: Selaginellales
Family: Selaginellaceae
Genus: Selaginella
Species: Selaginella fissidentoides
Infraspecies: Selaginella fissidentoides ovata

Figure 5.5: Indented tree showing full taxonomy data.

Moreover, there is a *tooltip* that indicates the amount of changes for every node, and the family with the larger amount of nodes changed by a particular type of change is painted by the change color indicator, in Figure 5.7 the family *Selaginellaceae* is painted green because it has the larger amount of additions in the whole taxonomy comparison.

Finally, the intensity of the color background for every node in a gray color scale from white to black indicates the total amount of changes for every node at any particular level. That means the darker nodes are the ones with more changes at its taxonomic level.

In general, color coding and data about the species were presented consistently in both designs.

5.2 Design of the user study

5.2.1 Performance Metrics

To investigate the hypothesis, we conducted a user study that assess the effectiveness and efficiency of the proposed display prototypes. We considered the heuristics

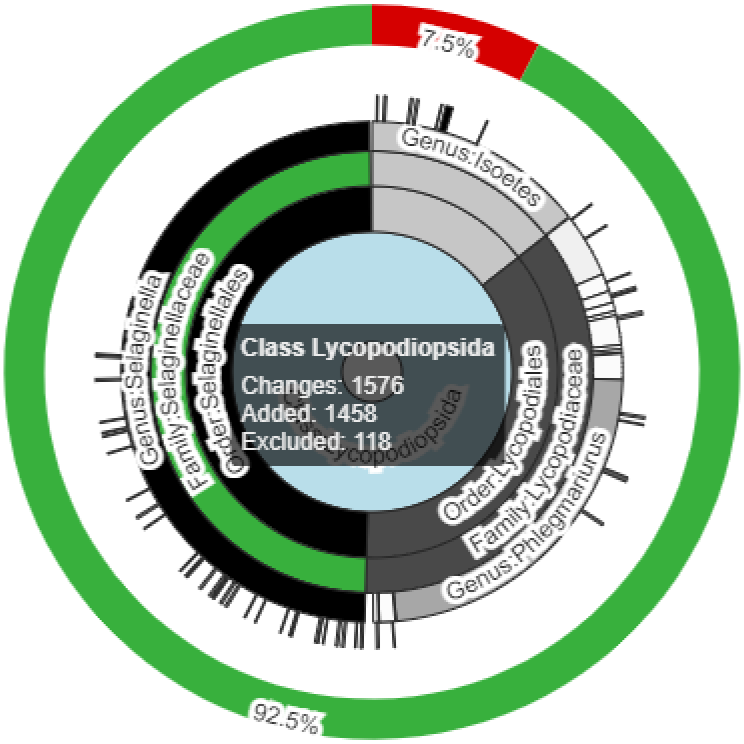


Figure 5.6: Radial taxonomic differences tree visualization

proposed by Forsell et al. [27], which are described in table 2.3. Our vision was to improve the overall users performance, while using and interacting with an integrated version of the proposed design with the Diaforá System, The aspects to be measured in the user study are the following:

Effectiveness

Given a set of different tasks that require the user to identify differences between taxonomic versions, we want to measure the user error rates when using the original Diaforá system and the proposed multi-view environment Diaforá with both *edge drawing* and *difference-focused taxonomic visualization.*

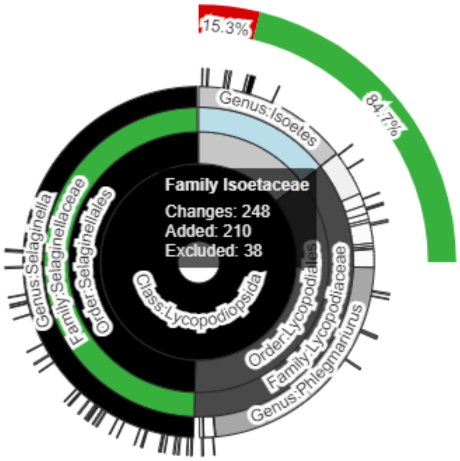


Figure 5.7: Satellite bar indicator for the Family *Isoetaceae*

Efficiency

Efficiency as a metric will be evaluated by the mean completion time, and the mean times per proposed visualization technique. We want to measure the completion time for each task, and since the system will run across different environments and the initial purpose of the Diaforá System is to be available online, the average completion time will suffice as the metric required to measure both versions. Since the time from when the question is displayed to when it is answered was measured and most of the processing time is about the same for any computer, and since the data sets used are relatively small, we can assume that making users use their own computers will not significantly affect the results of this metric.

Confidence Rate

As part of the subjective heuristics that are intended to be evaluated, a self-reported confidence rate will be asked to the users on a 3-point Likert scale (high, medium, low). This confidence rate will be the certainty that the user has that he or she is providing the correct answer to the task, based on the evaluation used in [48]. Our rationale is that the user finds more confidence to answer a given question correctly when the visualization provides a better way to interpret the information.

Overall Engagement Rate

For each set of tasks with a particular method to visualize differences (VOD taxonomic visualization), the user will be asked to give an overall engagement evaluation for the technique. This will measure the cognitive connection between the user and the given visualization and it will be evaluated by using a 5-point Likert scale (Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree). We propose to use the reaction cards proposed in [10] to evaluate this metric and then present the percentages across all tasks per visualization. We followed the same evaluation approached used by Peña-Araya et. al. [48]. For this and the confidence rate metrics according to Lehmann et. al. [36] to get an idea about the average in the population 3-point Likert scales are good enough that's why we are using it for the confidence rate, for individual behavior the 5-point Likert scale has a better performance, therefore we used it for the overall engagement rate.

5.2.2 Materials

We published a publicly accessible website so that users can register and run the tests using their own equipment. The site used for the asynchronous tests (loop11.com)[2] collects information from the equipment used and we set up some limitations to avoid biases due to slow equipment, outdated software, or slow internet connections. We set as minimum requirements to participate in the tests:

- Internet connection 10Mbps or higher.
- RAM 4gb or higher.
- Chrome 90.0 or higher or Firefox 88.0 or.
- Screen size 13 inches or larger.

We didn't set a restriction for the OS since this is a web application.

Timed tasks will be used to determine the overall user's performance while identifying differences between versions of biological taxonomies.

The website was developed using EJS (Embedded JavaScript templates) [1], which is a platform to develop web sites using reusable template and the visualization prototypes were created with the D3.js development framework [14].

Also, the website contains a layer of telemetry functions that collects the information about the equipment capabilities such as the operating system and the browser version used to execute the tasks.

We included a module where users can perform the different user study tasks and complete the user study in an asynchronous way.

Usability testing platform Loop 11

Loop 11 [2] is an Australian company dedicated to creating solutions for usability testing. They offer a simple way to integrate a web application and easily add questions and tasks that the user need to perform to evaluate and asses the efficiency and effectiveness of a web application. They gave us an extended academic license of their usability testing platform to execute the validations required for this thesis. The usability testing platform provided by Loop11 allowed us to generate the questions required to be answered by each participant and also record their screen while performing the system evaluation this also allow us to measure the time required to answer every question. In addition, the videos of the executions carried out by the users can be downloaded to analyze their behavior when using the platform that is being evaluated.

5.2.3 Methods

We performed a two-stage user study. The first stage evaluates the indented tree vs radial representation. The second stage consists of an evaluation of the users performance of an extended version of the Diaforá system in a CMV environment. In this stage, we used the visualization that performed better in the first user study and integrate it with the Diaforá's *edge-drawing* method in order to test our hypothesis.

The user study was performed using an *asynchronous remote usability testing* platform, in which users and researchers can be separated in terms of location and time. This approach is explained later in this chapter.

Using the *asynchronous remote usability testing* helped us to make it easier for test subjects to participate in our studies, considering that some members of the expert group may be overseas on different time zones and with different availability spans.

Figure 5.8 provides a quick insight about the process that we followed in this research. We can divide the process into three stages that are represented by each level in the figure.

First step: Prototype Design

We created two different visualization designs based on showing only the taxa nodes affected by changes occurred between two alternative versions of a biological taxonomy. We validated those prototypes through a pilot test with two taxonomy experts.

Stage 1: Selecting the best visualization design

We validated our two visualization designs with a user study, analyze the results obtained and selected the design with the highest performance. Then, we integrate it into the Diaforá system.

Stage 2: CMV vs Single View

After integrating our visualization design into Diaforá, we performed a user study to verify if the coordinated multiple view approach increases the user performance while comparing biological taxonomies. We collected the obtained results and present our conclusions further in this document.

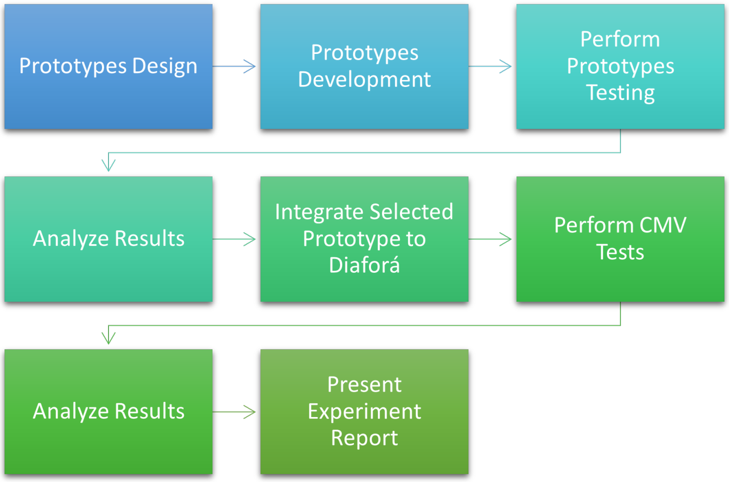


Figure 5.8: Research tasks flow.

Asynchronous Remote Usability Testing

We conducted the study through the asynchronous remote usability testing approach that has been examined for a number of studies [6, 16]. This kind of testing is low cost and can be performed faster than traditional in-laboratory testing, and it is possible to use testing frameworks like the *Hybrid Usability Evaluation Method* that allows to perform unmoderated usability evaluation on web-based systems. This method injects an application into the website pages, and gets information and data from the users while they are performing a set of given tasks in the evaluated web system. [18].

Conducting in-lab studies faces issues like the difficulty of reproducing the user's everyday environment and the unavailability of representative end-users.

The asynchronous remote testing separates users from the researcher in terms of location and time. All the participants use their own equipment to perform the tests, providing a cost-efficient alternative to the standard in-lab and remote synchronous testing [5].

Alghamdi et al. [5] described a comparison between synchronous and asynchronous remote testing and did not find any statistical significant differences between the number of usability issues identified in both cases.

Pilot Test with Experts

Before evaluating the visualization designs we asked for advice to two professionals with experience in biological taxonomies, one is a botanical taxonomist from the Natural History Department of the National Museum of Costa Rica [3], and the other one is a professor of the School of Forestry Engineering of the Costa Rica Institute of Technology [4]. They tested both the visualization designs and the asynchronous remote test platform to validate the effectiveness of the presented solution before asking our testing group to perform the system evaluations.

The goal of the test set that evaluates the proposed prototypes is to decide on the visualization method that has better performance and use that visualization for the CMV environment in the extended Diaforá system.

The test task measures the mean completion times and error rate while identifying specific changes between different versions of a taxonomy. Ranking the results will lead us to the chosen prototype that will be integrated into the CMV environment.

Asynchronous Usability Study

We designed a set of tasks and used the Cochran's Q-test method to evaluate the results obtained from the visualization designs evaluation. This method can be used when you have a group of people performing tasks where the outcome is dichotomic (eg. success or failure) [53].

The test consists of nine timed questions related to comparison of the taxonomies. After each question, participants rated the level of certainty and the level of satisfaction with the answer provided, using a three-level Likert scale. Questions are described in table 5.1.

These questions were validated by our expert taxonomists, to avoid any bias in the design of the evaluating instrument and to adjust the question to the actual work done by taxonomist while conducting taxonomy correction process.

After answering the 9 question instrument, participants should answer a 5-level Likert scale evaluation about the user's satisfaction level about the visualization, that means an engagement rate with the visualization according to Attfield et al. [7].

Id	Question
1	What is the most common change type?
2	List the three families with the most splits?
3	What is the most common change in the family "f"?
4	What species did merged for species "s"?
5	Identify the current species name "s"
6	Identify the family with the most new species.
7	What is the percentage of added species in family "f"?
8	How many changes are in the genus "g"?
9	Identify the complete taxonomy of the species "s".
	Level of certainty (3-level Likert scale) After each question.
	Level of satisfaction (5 level Likert scale) After each visualization test set.

Table 5.1: Prototypes evaluation questions

Qualitative User Study with the testing group

We conducted a user study with a group of expert users that are familiar with taxonomic classifications. In order to avoid bias due to the specific expertise of the participants, we chose taxonomies that are outside the usual participant's field of expertise.

The set of experimental tasks were to be performed in both versions of the Diaforá System [52], the original one with only the *edge drawing* visualization and the new one that combines that visualization with our new visualization design, to measure the effectiveness and error rate with and without the new proposed views.

The users should give feedback about the difficulty of each task and at the same time, the asynchronous usability testing platform (Loop11)[2] described in the *Materials* section should allow to measure the user's mean execution time per task.

After that, we want to compare the global results obtained in both cases and determine if the new visualizations contributed expert users to perform better while using the Diaforá system.

5.3 VOD Prototypes Evaluation

We used an online asynchronous remote usability testing platform [2] for the evaluation of the prototypes designed to visualize only differences(VOD), that means the user and the researcher are separated both by location and time. Users can perform the tests by accessing a *url* that guides them over the test set of questions defined in table 5.1.

5.3.1 Indented design evaluation

The evaluation instrument consisted of ten questions, they are presented in Table 5.2. The questions must be filled by the participants in the *Loop11.com* [2] user study online platform.

The questions are based on the comparison of alternative versions of *phylum Annelida* taxonomy. The *phylum Annelida* also known as the segmented worms is a large phylum with over $22,000$ species[47]. We selected this taxonomic group because

Index	**Question**
1	What is the most common change in this taxonomy? (The one with most occurrences).
2	Please identify the three families that have the largest amount of split changes.
3	What is the most common change in the *Tubificidae* family.
4	Select Merges and look up for species *Megasyllis inflata*. Which species did merged for species *Megasyllis inflata*?
5	Look at renames and please enter the new name for the renamed species *Quistadrilus multisetosus*.
6	Please enter the family with most new species added?
7	What is the percentage of added species in family *Naididae*?
8	How many changes are in the genus *Paranais*?
9	Please enter the complete taxonomy tree (Phylum to Species) of new species *Syllis boggemanni*.
10	Please rate this visualization prototype. Did you like the data visualization? (5-level likert scale).
*	After every question, there is a 3-level likert scale rating question about a confidence level that the given answer was correct.

Table 5.2: Taxonomic differences indented tree prototype evaluation questions.

it is one of the most actively changed and presents a good distribution of the changes highlighted by Diaforá. The questions were reviewed and improved by two expert taxonomists that we consulted.

The data collected for each participant is discussed in the *Results and Discussion* section of this document.

5.3.2 Radial design evaluation

To perform the same process for the second prototype (*Radial taxonomic differences tree design visualization*), users also required to complete a set of ten questions about a taxonomic comparison. The purpose was to have a comparable scenario to the one used in the indented tree visualization evaluation.

Table 5.3 presents the questions that were prompted to the user to evaluate the radial tree prototype. Those questions are based on changes occurring between alternative versions of the *phylum Cnidaria*. The *phylum Cnidaria* that includes over 11.000 species of corals, sea anemones, and jellyfishes[47]. The questions were also reviewed and improved by two expert taxonomists that we consulted. We selected this group because it contains all the types of changes described by Diaforá and has a similar distribution of changes than the *Annelida* group described above.

5.3.3 User Study Methodology

Expert Assessment Validation

As part of the calibration of the user study instrument, to evaluate and adjust the user study questions and the online tool, we consulted two expert taxonomists. One is a botanical taxonomist from the Natural History Department of the National Museum of Costa Rica [3]. The other is a professor of the School of Forestry Engineering of the Costa Rica Institute of Technology [4]. The results obtained in the expert assessment can be found in chapter 7 *Results and Discussions*. The purpose of this evaluation was to verify that both the evaluation questions and the usability testing platform made sense to evaluate the Diaforá system.

Index	**Question**
1	What is the most common change in this taxonomy? (The one with most occurrences).
2	What is the percentage of splits in the *Alcyonidae* family?.
3	Which is the most common change in the *Nephtheidae* family?
4	Select Merges and look up for species *Lytreia plana*. Which species did merged for species *Lytreia plana*?
5	Look at renames and please enter the new name for the renamed species *Orbicella franksi*.
6	Please enter the family with most new species added?
7	What is the percentage of added species in family *Actiniidae*?
8	How many changes are in the genus *Actinia*?
9	Select Splits and look for species *Aegina citrea*. Please enter the complete taxonomy tree (Phylum to Species) of species *Aegina citrea*.
10	Please rate this visualization prototype. Did you like the data visualization? (5-level likert scale).
*	After every question, there is a 3-level likert scale rating question about a confidence level that the given answer was correct.

Table 5.3: Taxonomic differences indented tree prototype evaluation questions.

Participant ID	Major	Years of study
TS1	Forestry engineering student	3
TS2	Forestry engineering student	4
TS3	Forestry engineering student	4
TS4	Forestry engineering student	5
TS5	Forestry engineering student	3
TS6	Forestry engineering student	5

Table 5.4: Participants' profiles

Test Subjects

The test subjects for the user study are students from Forestry Engineering career of the Costa Rica Institute of Technology. The group of students had already taken the Taxonomy class. The list of participants was suggested by professors of the Forestry School. All participants were familiar with classification of biological taxonomies. Table 5.4 summarizes the participants' profiles. We assigned to each participant a unique id that ranges from TS1 to TS6.

5.4 User Study of Coordinated Multiple Views Integrated in Diaforá

After selecting the VOD design, we integrated it to a new extended version of Diaforá in a CMV environment. We then evaluated the extended version in contrast to the original Diaforá system that uses the *edge drawing* visualization exclusively. We used the same testing approach that was used to evaluate the prototypes.

According to the results obtained in the prototypes evaluation, the selected prototype will be the *Radial taxonomic difference tree*. The detailed results are shown in chapter 7 *Results*. To facilitate this study, we created a user guide [24] for the system and a video [25] that allowed users to familiarize themselves with the use of

Diaforá in order to perform the tests.

5.4.1 Single view Diaforá evaluation

In order to evaluate the original Diaforá version (i.e., the one that has only the *edge drawing* visualization to represent the changes between the alternative versions of a biological taxonomy), participants had to answer a 10 question instrument. The questions that were asked to the participants are based on the comparison of alternative versions of *phyulm Cnidaria* taxonomy. We used the same metrics that were used in the prototype phase (error rate, efficiency, confidence rate and engagement rate). Table 5.5 presents the questions used to evaluate the original version of Diaforá system.

5.4.2 CMV Diaforá version evaluation

In this extended version we included the usage of a visualization that is focused on only displaying the taxa affected by changes, along with the edge drawing and a details section that explains in text format about the changes occurred for every selected taxa. According to the results obtained in the prototypes evaluation, the selected visualization that is coordinated with the *edge drawing* is the *Radial taxonomic difference tree*. Table 5.6 shows the questions used to evaluated the extended version of Diaforá system.

5.4.3 Test Subjects

This second stage evaluates the two versions of the Diaforá system, version 1.0 that is the original version [52] that uses *edge drawing* as the main visualization to show the changes between the alternative versions of a biological taxonomy and the version 2.0 that combines the *edge drawing* visualization with the *radial taxonomic difference tree*. For this stage the test subjects are a group of advanced students of Forestry Engineering of the Technological Institute of Costa Rica, a plant taxonomist, a biologist and a professor and forestry engineer. The table 5.7 summarizes the test subjects profiles, also we assigned an unique identifier for each participant that ranges

Index	**Question**
1	What is the number of excluded taxa in this taxonomic comparison?
2	What is the percentage of renames in the *Nephtheidae* family?.
3	Which is the most common change in the *Anthozoa* class?
4	Select Moves and look up for genus *Cubaia*. What is the original family and the new family to which this genus belongs?
5	Look at Added and look up for the Order *Leptothecata*, How many added taxa are included in this order?
6	Please enter the family with most new species added?
7	What is the percentage of excluded species in family *Acroporidae*?
8	How many changes are in the genus *Acropora* ?
9	Select Renames and look for species *Dipsastraea rotumana*. Please enter the complete taxonomy tree (Phylum to Species) of species *Dipsastraea rotumana*.
10	Please rate this visualization prototype. Did you like the data visualization? (5-level likert scale).
*	After every question, there is a 3-level likert scale rating question about a confidence level that the given answer was correct.

Table 5.5: Original Diaforá evaluation questions.

Index	**Question**
1	What is the number of merged taxa in this taxonomic comparison?
2	What is the percentage of merges in the *Clausophyidae* family?.
3	Which is the most common change in the *Halcampidae* family?
4	Select Moves and look up for genus *Euphellia*. What is the original family and the new family to which this genus belongs?
5	Look at Added and look up for the Order *Zoantharia*, How many added taxa are included in this order?
6	Please enter the family with most species excluded?
7	What is the percentage of excluded species in family *Ellisellidae*?
8	How many changes are in the genus *Viminella* ?
9	Select Renames and look for species *Filigorgia schoutedeni*. Please enter the complete taxonomy tree (Phylum to Species) of species *Filigorgia schoutedeni*.
10	Please rate this visualization prototype. Did you like the data visualization? (5-level likert scale).
*	After every question, there is a 3-level likert scale rating question about a confidence level that the given answer was correct.

Table 5.6: Extended Diaforá evaluation questions.

from TS1 to TS10. To motivate the participation of the forest engineering students, we gave them a 20$ participation bonus.

Participant ID	**Major**	**Years in professional career/ work experience**
TS1	Forest engineering student	5
TS2	Forest engineering student	4
TS3	Forest engineering student	6
TS4	Forest engineering student	6
TS5	Forest engineering student	6
TS6	Forest engineering student	5
TS7	Plant taxonomist	24
TS8	Forest engineering student	5
TS9	Biologist	2
TS10	Forestry engineer	21

Table 5.7: Participants' profiles

Chapter 6

Results and Discussion

As presented in the *Research Methodology* section, this thesis involves three stages: 1) the design of two proposals for the visualization of differences only (VOD), 2) the selection of the best performing VOD design (evaluated through a user study), and 3) the evaluation of an extended version of Diaforá that integrates the selected VOD design within a CMV environment. This chapter presents and discusses the results obtained in the user studies of stages 2 and 3.

For the analysis of effectiveness, we used Cochran's Q-test method, which is commonly used when you have a group of people performing a series of tasks where the outcome is dichotomic, which means it can be a failure or success.

The levels of satisfaction and certainty of the participants are analyzed using the Friedman test method that is suitable for ordinal data like the Likert scale questions [41].

6.1 Selection of a method for displaying only differences

6.1.1 Pilot Test Results

Table 6.1 summarizes the results obtained by the experts for the indented tree prototype evaluation and Table 6.2 summarizes the results obtained for the radial prototype evaluation. Experts are labeled *E1* and *E2*. The engagement rate value is represented by ★ symbols. The number of filled stars is the user's selected value to rate the visualization prototype as it is represented below:

- ★☆☆☆☆ = Strongly disagree.
- ★★☆☆☆ = Disagree.
- ★★★☆☆ = Neither agree nor disagree.
- ★★★★☆ = Agree.
- ★★★★★ = Strongly agree.

To represent correct and incorrect answers we use the following symbols:

- ✓= represents a correct answer.
- ✘ = represents an incorrect answer.

Question	**E1**		**E2**	
	Confidence Level	Result	Confidence Level	Result
1	High	✓	High	✓
2	Medium	✗	High	✓
3	High	✓	High	✓
4	High	✓	High	✓
5	High	✓	High	✓
6	High	✗	Low	✗
7	High	✓	High	✓
8	High	✓	Medium	✗
9	High	✓	High	✗
Engagement rate	★★★★★		★★★★☆	

Table 6.1: Expert assessment results using the Indented Tree Visualization.

Question	**E1**		**E2**	
	Confidence Level	Result	Confidence Level	Result
1	High	✓	High	✓
2	Medium	✓	Low	✗
3	High	✗	High	✓
4	High	✓	High	✓
5	High	✓	High	✓
6	High	✓	High	✗
7	High	✓	High	✓
8	High	✓	High	✓
9	High	✓	High	✓
Engagement rate	★★★★☆		★★★★☆	

Table 6.2: Expert assessment results using the Radial Tree Visualization.

According to these results, our expert advisors were confident enough to let us continue with the usability study with our test participants. They found both prototypes usable and understandable and they also felt familiar with the proposed tasks and how the information was displayed by our visualization designs. Although expert advisors performed better using the radial visualization design, they found valuable the way that the information was presented in both designs. After we evaluated the visualization designs we found that both the experts and the testing group performed slightly better with the radial visualization design as explained below.

6.1.2 User study for selecting VOD design

Six test subjects participated in this stage of the study. Table 6.3 summarizes the results obtained from the test subjects using the taxonomic differences indented tree visualization, and Table 6.4 shows the results obtained from the test subjects using the radial visualization prototype. The tables contain the results obtained from every test subject, who are identified as *TS1*, *TS2*, ..., *TS6*. Each row shows the selected confidence level and the time required by the participant to answer each question. The goal of this stage was to select the best-performing visualization design. According to the obtained results, it was the radial visualization design.

Question		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
TS1	CL	M	H	H	H	L	H	H	H	H
	R	✘	✘	✔	✔	✘	✘	✔	✔	✔
	T	00:38	01:49	00:33	02:47	18:24	00:57	01:09	00:37	04:00
	ER	★★★★☆								
TS2	CL	M	L	H	L	H	H	H	M	H
	R	✔	✘	✔	✘	✔	✘	✔	✘	✔
	T	02:21	04:06	01:51	05:25	01:01	03:17	00:40	00:51	04:56
	ER	★★★☆☆								
TS3	CL	M	L	H	H	H	H	H	H	H
	R	✘	✔	✘	✔	✔	✘	✔	✔	✔
	T	01:49	01:31	00:04	05:08	01:23	03:50	00:37	00:44	02:18
	ER	★★★★☆								
TS4	CL	M	H	H	H	H	M	H	H	H
	R	✔	✘	✔	✔	✔	✘	✔	✔	✔
	T	01:41	04:38	01:21	07:12	02:10	07:28	01:23	01:06	01:26
	ER	★★★★★								
TS5	CL	M	H	H	H	H	M	H	H	H
	R	✔	✔	✔	✔	✔	✘	✔	✘	✔
	T	02:36	04:39	00:40	03:11	01:28	13:24	01:27	01:41	05:10
	ER	★★★★☆								
TS6	CL	M	H	H	L	M	L	H	M	H
	R	✔	✘	✔	✘	✔	✘	✔	✔	✔
	T	02:01	08:26	00:35	02:24	01:32	03:51	00:26	02:36	01:15
	ER	★★★★☆								

Table 6.3: Results for the indented tree visualization prototype evaluation.(CL: Confidence level [L: Low, M: Medium, H: High], R: Result, T: Time (Minutes: Seconds), ER: Engagement rate)

Question		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
TS1	CL	H	L	H	H	H	H	H	H	H
	R	✓	✗	✓	✓	✓	✓	✓	✓	✓
	T	00:34	03:39	00:35	01:16	00:37	00:57	00:17	00:37	02:30
	ER	★★★★☆								
TS2	CL	L	H	H	L	L	H	H	H	H
	R	✗	✓	✓	✗	✓	✗	✓	✓	✓
	T	00:54	00:33	00:43	01:55	00:54	01:21	00:29	00:19	02:31
	ER	★★★★★								
TS3	CL	L	M	H	H	H	M	H	H	H
	R	✗	✗	✓	✓	✓	✓	✓	✓	✓
	T	00:29	04:41	01:53	01:18	01:28	04:10	00:40	00:23	00:48
	ER	★★★★☆								
TS4	CL	M	L	L	H	H	H	H	H	H
	R	✓	✗	✓	✓	✓	✓	✓	✓	✓
	T	01:24	04:55	03:36	01:26	01:04	03:16	00:29	00:30	00:51
	ER	★★★★★								
TS5	CL	H	H	H	H	H	M	H	H	H
	R	✓	✗	✓	✓	✓	✗	✓	✗	✓
	T	00:45	01:45	05:47	01:29	00:59	02:07	00:59	00:31	01:01
	ER	★★★★☆								
TS6	CL	H	L	H	M	H	M	H	M	H
	R	✓	✗	✓	✓	✓	✓	✓	✗	✗
	T	00:38	02:37	02:16	01:13	00:34	01:06	00:25	00:24	03:16
	ER	★★★☆☆								

Table 6.4: Results for the radial tree visualization prototype evaluation.(CL: Confidence level [L: Low, M: Medium, H: High], R: Result, T: Time (Minutes: Seconds), ER: Engagement rate)

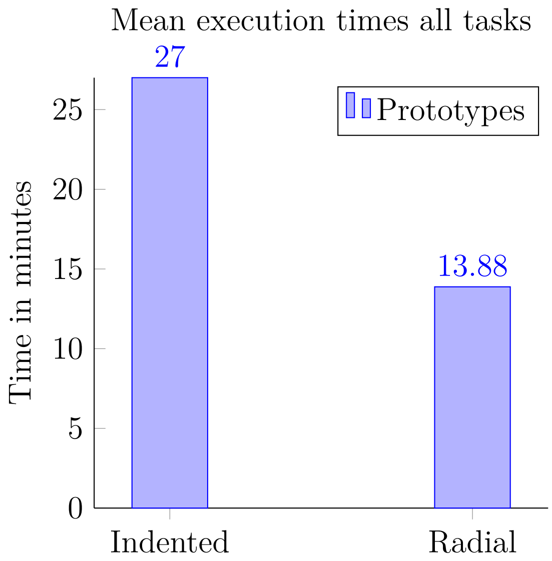


Figure 6.1: Indented tree tasks were completed in a mean time of 27 min. and the Radial tree tasks were completed in a mean time of 13.88 min.

Overall results across tasks

Completion time: Figure 6.1 shows the meantime of prototype task execution. The tasks were completed with the radial prototype in a mean time of 13.88 minutes, and the tasks with the indented tree were completed in a mean time of 27 minutes, which means that the participants' performance was 52% faster with the radial tree than with the indented tree representation. The mean difference between the two designs is -12.774 minutes.

Error rate: Figure 6.2 shows the effectiveness rate for the indented tree and the radial tree designs. The error rate was 8% lower when using the radial tree. By conducting a Cochran Q test on the obtained results we obtained the values **(P= 0.196706, Q = 1.66667 with α = 0.05)**, which means there is no statistically significant difference between both prototypes and can be considered equivalent on the error's rate metric.

Confidence: Figure 6.3 shows the self-reported confidence for each visualization. Confidence is relatively high for both designs but slightly 3% higher on the Radial tree visualization. For the Indented tree, the self-reported confidence values are 69% of High confidence and 72% of self-reported high confidence for the Radial tree. By

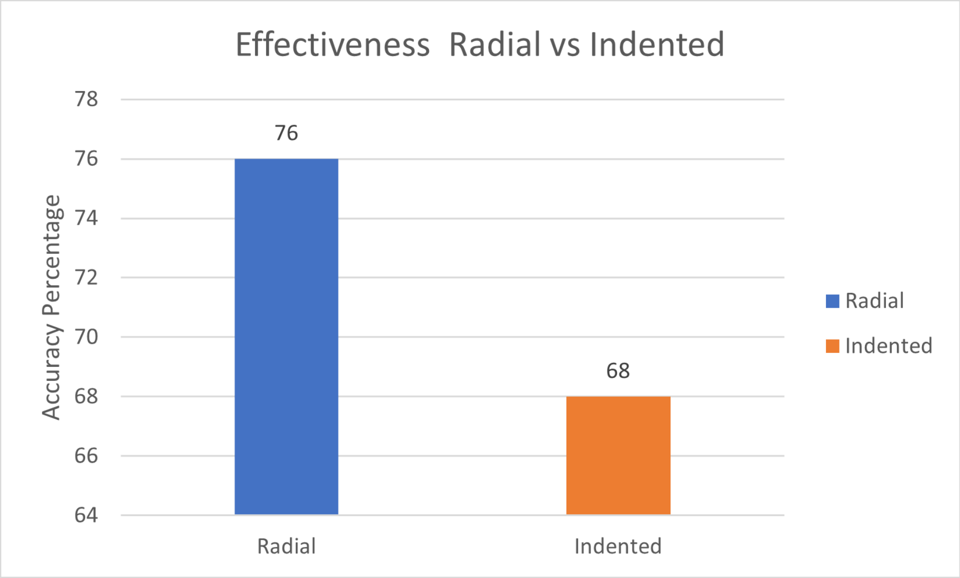


Figure 6.2: Effectiveness rate between Indented tree and Radial tree designs.

giving a numeric value to each confidence evaluation (1: low, 2:medium, 3: high) and perform a Friedman test the result are the same for both prototypes **(Q= 0.00000, p= 1.00000)**. That indicates that both prototypes provide similar self-perceived confidence.

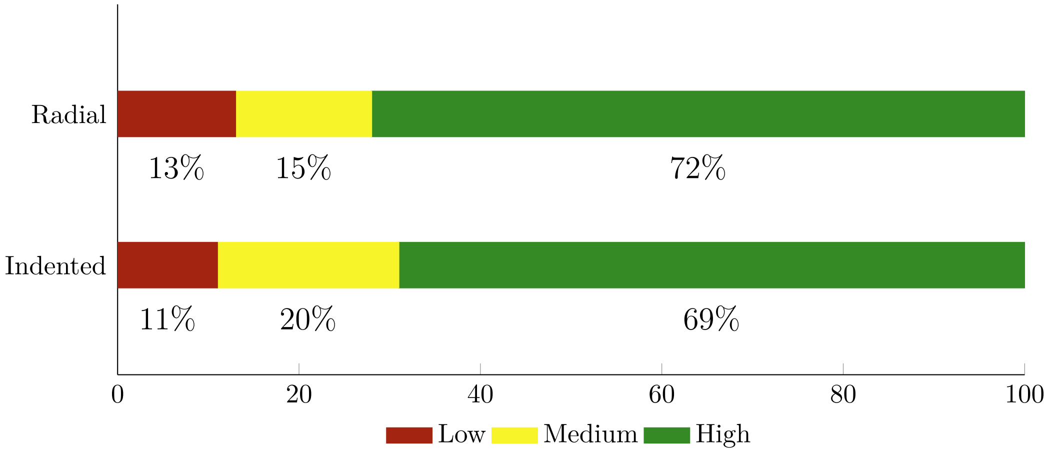


Figure 6.3: Self reported confidence for each prototype.

Engagement rate: Figure 6.4 shows the engagement rate results for Indented and Radial tree visualization designs. Both prototypes obtained a very positive review with 87% positive evaluation (*Agree and Strongly agree*), the radial prototype has a slightly better evaluation with 33% of strongly agree reviews. By performing a Friedman test on the obtained results, we calculate values of **(Q= 0.00000, p= 1.00000).** That means there is no a statistically significant difference between the engagement rate values of both prototypes.

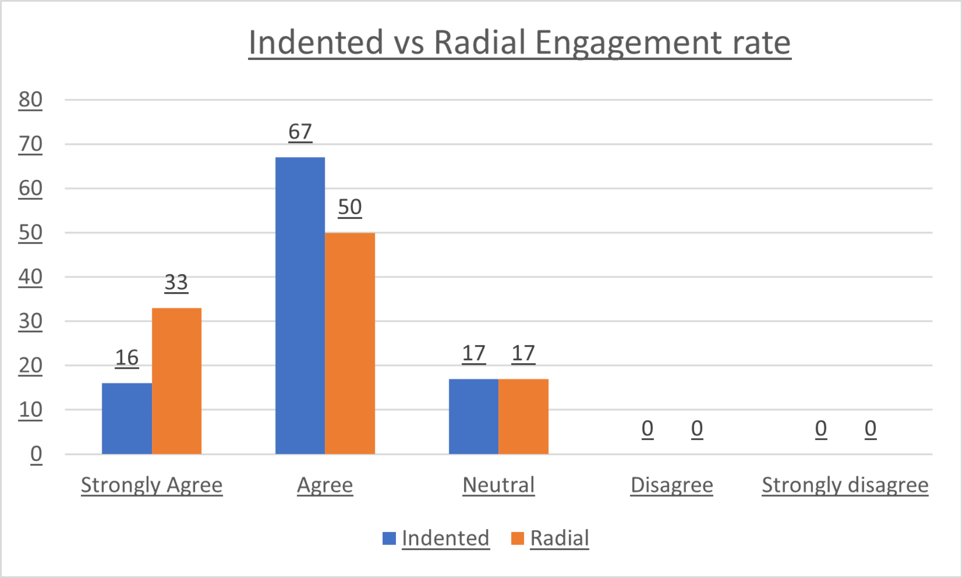


Figure 6.4: Indented tree vs Radial visualization engagement rate results.

As stated by the obtained results, even if there is no statistically significant difference between both visualization designs, since participants performed better with the radial visualization, we selected this design to integrate it with the original Diaforá's *edge drawing* visualization in order to create our coordinated multiple views environment for taxonomic comparisons.

6.2 Multiple Coordinated Views Evaluation Results

As stated in the methodology section we evaluated the *edge drawing* visualization method that uses a (*Single View*) against a CMV environment that combines *edge drawing* with a selected visualization method for displaying only differences (*CMV Environment*).

From the previous stage of the study, we obtained that the selected method for displaying only differences was the radial visualization described above. We choose that visualization because it performed better than the indented visualization method.

6.2.1 Avoiding Familiarity Bias

At this stage, we managed to gather 10 test subjects to participate in the usability study. Five of them participated in the first stage of the study, so they were familiar with the VOD prototypes, but had no previous experience with the CMV environment or with the Single View version.

To avoid any familiarity bias since both the single view and the CMV versions of Diaforá were new to all the participants we also decided to use a different set of taxonomies that do not include species known by the participants(i.e, not using plants). The tasks for this stage, were described in chapter 6 in tables 5.5 and 5.6.

We found that consistent with the obtained results described later in this chapter and apparent general improvements in performance affected all participants, including those who had no prior experience with Diaforá.

Table 6.5 summarizes the results obtained by our test subjects using the Single View version of Diaforá system, and table 6.6 summarizes the results obtained while using the CMV environment of Diaforá with the coordinated multiple views environment.

Question		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
TS1	CL	L	H	H	H	H	H	H	H	H
	R	✘	✘	✓	✘	✓	✘	✘	✘	✓
	T	00:49	06:16	01:31	08:46	02:34	03:39	01:39	01:57	07:27
	ER	★★★☆☆								
TS2	CL	M	M	M	L	H	M	M	L	L
	R	✓	✘	✓	✘	✘	✘	✘	✘	✘
	T	00:59	07:11	03:39	18:24	01:51	00:50	02:44	01:28	07:28
	ER	★☆☆☆☆								
TS3	CL	H	L	H	H	H	L	H	H	H
	R	✘	✘	✓	✓	✘	✘	✘	✘	✓
	T	00:32	03:03	00:36	02:52	01:38	04:09	01:49	00:39	05:29
	ER	★★☆☆☆								
TS4	CL	M	H	H	H	H	L	M	M	H
	R	✓	✘	✘	✓	✓	✘	✓	✘	✓
	T	00:16	02:13	04:50	04:19	01:59	11:07	02:00	00:14	10:50
	ER	★★☆☆☆								
TS5	CL	H	M	M	H	H	H	M	H	H
	R	✘	✘	✘	✘	✘	✘	✘	✘	✓
	T	00:18	04:07	00:43	07:20	01:42	00:56	02:04	01:37	05:56
	ER	★★☆☆☆								
TS6	CL	M	M	M	H	H	M	M	L	H
	R	✓	✘	✓	✓	✓	✘	✘	✘	✓
	T	03:19	07:48	02:12	08:49	02:30	06:50	03:26	05:02	11:13
	ER	★★☆☆☆								
TS7	CL	H	H	H	H	H	H	H	H	H
	R	✓	✘	✓	✘	✓	✘	✘	✘	✘
	T	00:09	00:40	01:00	00:17	00:44	00:46	00:15	00:23	00:11
	ER	★☆☆☆☆								

TS8	CL	L	L	L	L	M	L	L	L	L
	R	✓	✘	✓	✘	✓	✘	✘	✘	✘
	T	00:24	01:55	01:16	02:52	00:39	01:39	00:26	00:27	00:50
	ER	★★☆☆☆								
TS9	CL	M	M	H	L	H	M	M	L	M
	R	✘	✘	✘	✘	✘	✘	✘	✘	✓
	T	00:44	02:03	00:37	01:47	00:59	00:24	02:10	02:36	04:54
	ER	★★★☆☆								
TS10	CL	H	H	H	H	H	H	H	M	H
	R	✓	✘	✘	✓	✓	✘	✘	✘	✓
	T	00:44	04:21	00:39	18:36	01:16	02:07	03:18	01:37	10:00
	ER	★★★★☆								

Table 6.5: Results of tests using the *Single View* Diaforá version (CL: Confidence level, R: Result, T: Time (Minutes: Seconds), ER: Engagement rate).

Question		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
TS1	CL	H	H	H	H	H	H	H	H	H
	R	✔	✔	✘	✔	✔	✔	✔	✘	✔
	T	01:21	02:14	02:34	01:30	00:43	00:55	00:34	01:44	01:25
	ER	★★★★☆								
TS2	CL	M	M	M	H	M	H	H	H	H
	R	✔	✔	✔	✔	✔	✔	✔	✔	✔
	T	00:21	01:35	00:59	01:00	01:43	00:49	00:27	06:05	06:07
	ER	★★★★★								
TS3	CL	H	H	H	H	H	H	H	L	H
	R	✔	✘	✘	✔	✔	✔	✔	✘	✔
	T	01:09	01:01	00:29	01:00	01:07	00:42	00:36	02:58	01:07
	ER	★★★★★								
TS4	CL	H	L	M	H	H	M	M	H	H
	R	✔	✔	✔	✔	✔	✘	✘	✔	✔
	T	00:31	02:12	04:15	02:36	01:39	04:06	00:41	01:37	01:20
	ER	★★★★★								
TS5	CL	M	H	H	H	M	M	H	H	H
	R	✘	✘	✘	✔	✘	✔	✔	✔	✘
	T	00:31	04:30	01:58	03:38	00:58	04:27	02:25	06:47	07:16
	ER	★★★☆☆								
TS6	CL	H	H	L	H	M	H	H	M	H
	R	✔	✔	✘	✔	✘	✔	✔	✔	✔
	T	00:30	02:53	02:35	01:51	02:48	01:21	01:12	03:47	01:11
	ER	★★★★★								
TS7	CL	H	H	H	H	H	H	H	H	H
	R	✔	✔	✔	✔	✔	✔	✘	✘	✔
	T	00:15	00:37	00:58	00:43	00:41	00:58	00:23	01:22	01:02
	ER	★★★★★								

TS8	CL	L	L	L	L	L	L	L	L	M
	R	✓	✘	✓	✘	✘	✓	✘	✘	✓
	T	01:10	00:52	01:07	03:37	00:32	03:22	01:50	01:08	04:04
	ER	★★★★☆								
TS9	CL	L	M	M	M	H	M	M	L	H
	R	✘	✘	✘	✘	✘	✓	✘	✘	✓
	T	00:54	01:43	01:18	02:12	01:03	00:43	03:03	00:31	01:02
	ER	★★★★☆								
TS10	CL	H	H	H	H	H	H	H	H	H
	R	✓	✓	✓	✓	✓	✓	✓	✓	✓
	T	00:45	01:12	03:00	02:19	01:44	01:15	00:54	02:17	01:31
	ER	★★★★★								

Table 6.6: Results of tests using the CMV environment.(CL: Confidence level, R: Result, T: Time (Minutes: Seconds), ER: Engagement rate)

Overall results across tasks

Completion time: Figure 6.5 shows the mean time with each task execution version. Task executions with the CMV environment were completed in a mean time of 16.395 minutes and the task executions with the Single View version were completed in a mean time of 28.74 minutes. That means that the use of the CMV environment to solve the proposed tasks was 54.7% faster than using the *edge drawing* visualization only. The obtained values for the Single View version show a standard deviation of 15.7 and the results for the CMV environment have a standard deviation of 6.94. During the execution of this experiment, we found that in some tasks, some users got stuck finding the results, like TS2 in question number 4, which took him over 18 minutes to finally find an answer. We believe that having two different perspectives, using the coordinated multiple views, improves the user experience and makes it easier to find specific answers to common taxonomic comparison tasks.

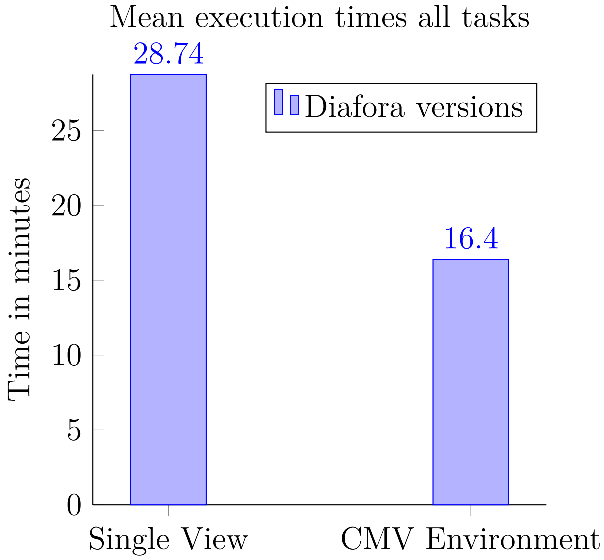


Figure 6.5: Participants executed the tasks with the CMV environment in less time than with a Single View.

Error rate: Figure 6.6 shows the participants effectiveness rate for Single View, and the results for CMV Environment. The error rate was 35% lower when using the CMV environment version. By performing a Cochran's Q-test on the obtained results we obtained the values of **(P= 0.000002, Q=22.26087, with** α**=0.05)** that means there is a statically significant difference between the two Diaforá system versions. This confirms our hypothesis that the user will have a better performance by using the CMV environment approach to analyze the taxonomic differences.

Confidence: Figure 6.8 shows each participant self-reported confidence for each version. Confidence is higher while using the CMV environment. Participants report a 9% more of high confidence while using the CMV environment. We believe that this is because it was easier for them to find the correct answer to the evaluation questions. By calculating a Friedman test on the obtained results by giving a numeric value for each self-confidence reported by the users we obtained **(Q=2.5, p = 0.11385, with** α**=0.05)** meaning that we did not detect a statistically significant difference on this metric. In practice, we found that users tended to give a high confidence value when selecting the "don't know" option in the exercises. This means that even if there is no significant difference, the higher confidence in the CMV environment

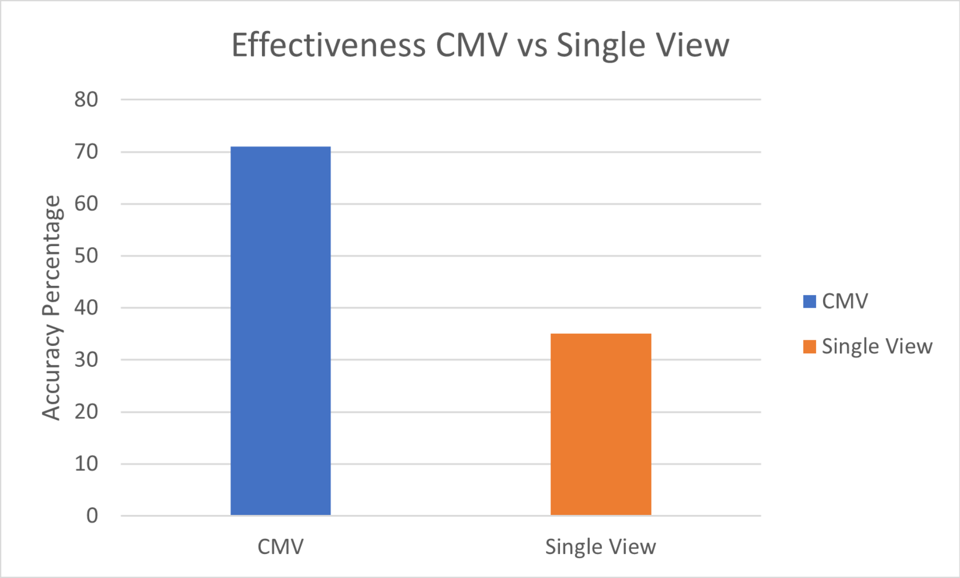


Figure 6.6: Effectiveness rate between CMV vs Single view versions.

coincides with a lower error rate, which makes us think that improves the general confidence of users to answer a question in the correct way.

Engagement rate: Figures 6.7 show the engagement rate results for Single View and CMV environment. The CMV version was notably better rated. It receives a combined 90% of positive reviews, 60% of strongly agree rates and a 30% of agree rates. A Friedman Test was conducted on the obtained results for the engagement rate. Results showed that the system version used led to statistically significant differences in the engagement rate values **($Q = 10$, $p = 0.00157$, with α=0.05)**.

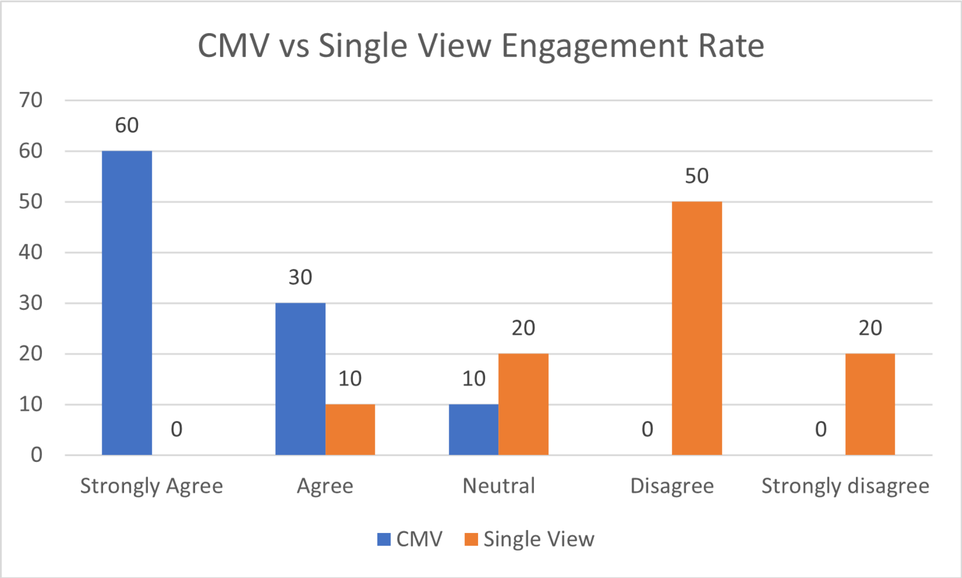


Figure 6.7: Engagement rate comparison.

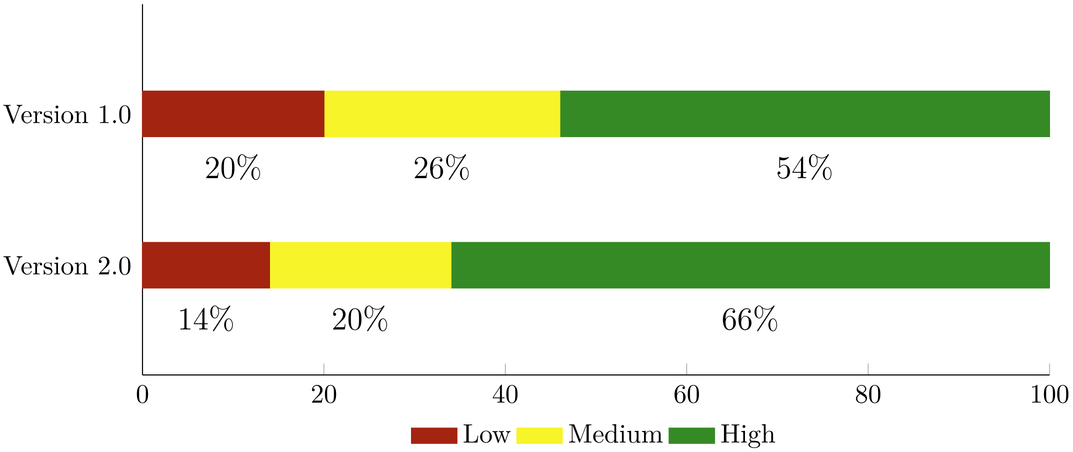


Figure 6.8: Self reported confidence for each version.

6.2.2 Discussion

According to the obtained results we observed that having a multiple view environment can provide a significant improvement in the overall user experience while performing taxonomic comparison tasks against having a Single View environment.

When we observed the videos of the tests carried out [26], that are available to be consulted online, we confirmed that the users did use both visualizations (*edge drawing and radial visualization design*) in the CMV environment, which leads us to think that our hypothesis was correct by assuming that multiple coordinated views can improve user performance by identifying the differences between alternative versions of biological taxonomy.

Results also indicate a statistically significant difference in both error rate and the engagement rate metrics meaning that we can successfully think that according to our tests, these results back up our affirmation that having the CMV environment impacts positively the experience of using the Diaforá system.

Another important consequence of this research is that the use of an online platform to perform the test can support the researchers to assemble a test group more easily, even if they are in different cities or with different availability schedules. The usage of Loop11 [2] in our case results in a very convenient way to deliver our tests.

We believe that the use of an asynchronous remote test approach to evaluate similar research topics can help both the investigators and the test subjects to easily provide a test environment in which users can interact and provide feedback on their experience with the system, solution, or tool that is being evaluated.

We considered recruiting forest engineering students due to their familiarity with biological taxonomies, however, due to their academic load, it was difficult to get a greater number of participants. We encouraged the students to participate in this study and we also received support from the Forestry Engineering School professors at the ITCR to share the study details among different groups. We would have liked to have more participants, however, finding people with experience in managing biological taxonomies is difficult. Considering the scarcity of professionals with knowledge of biological taxonomies, it makes us think about the importance of having tools, such as Diaforá, to support the work they do.

To avoid familiarity bias, we carefully selected the data for the experiments. We avoided using groups of species that might be familiar to participants (i.e, we did not use plants). Also, the taxonomic groups to test the CMV environment and the Single View environment were different. This motivated them to use the visualizations and to look for the answers to the tasks.

Overall, having a coordinated multiple view approach not only made things easier in general, but our visualization designs improved the cognitive value of the tool. Having a visualization with only the nodes affected by the changes made it easier to find specific changes which stand out in the results obtained in the experiments.

In general terms, the results show that the CMV environment outperforms the Single View environment, in all the proposed metrics. This strengthens our conclusion that this approach is a considerable improvement over Single View and that users will benefit from the design of our visualization and the CMV environment.

Furthermore, having our design that only shows the nodes affected by changes can effectively reduce the time spent by taxonomist when they are searching for specific changes in a taxonomy comparison. By reducing the amount of information displayed on the screen we also highlight the changes and by having the CMV environment we also keep the complete taxonomy information that can also be consulted by the users.

We believe that these findings can also be applied in other fields, but it is particularly useful to display hierarchical information like taxonomies. CMV environments are really practical for highlighting particular information displayed in different visualizations at the same time that can also help people to discover patterns and relevant information in their data.

Another important detail about our results is that both VOD designs presented in this work did not have a huge difference in performance so we believe that the indented taxonomic tree design can be useful for displaying taxonomic information or any other hierarchical information that needs to be displayed considering the nodes that represent a difference between two hierarchical structures.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

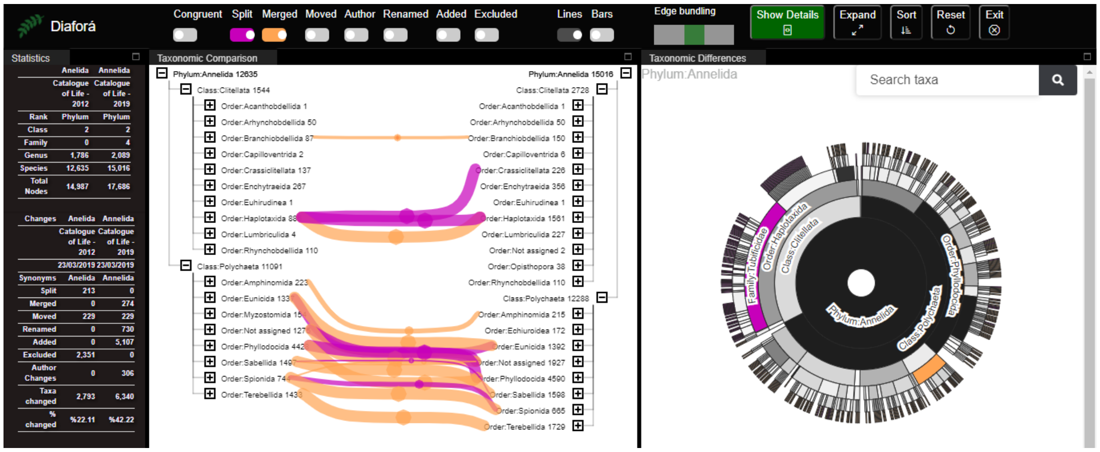
As stated by the results obtained in the first stage of user testing that evaluates the two novel visualization designs, the radial visualization performed better than the indented tree visualization for showing only the differences when comparing two alternative versions of biological taxonomy.

According to these results, the radial tree visualization was selected to be displayed in conjunction with the *edge drawing* visualization in a coordinated multiple views environment.

Figure 7.1 shows the Diaforá system displaying both visualization mechanisms, *edge drawing* and *radial tree*. Those visualizations are coordinated, which means that the user actions like navigation across taxa are reflected in both visualizations. That means that users always have the detail and context about the taxon (i.e, node in tree terms) of the taxonomic group of interest.

The results of phase 2, the one that evaluates the single view vs the CMV environment, lean in favor of the use of CMVs, which was implemented as of version 2.0 of Diaforá. The CMV method presented a better performance in terms of the average time required by the user to identify the information from the visualizations when comparting alternative versions of a biological taxonomy as well as to find an answer to the questions in the instrument.

According to our methodology, this means that our design and usage of a multiple



	Anelida	Annelida
	Catalogue of Life - 2012	Catalogue of Life - 2019
Rank	Phylum	Phylum
Class	2	2
Family	0	4
Genus	1,786	2,089
Species	12,635	15,016
Total Nodes	14,987	17,686

Changes	Anelida	Annelida
	Catalogue of Life - 2012	Catalogue of Life - 2019
	23/03/2019	23/03/2019
Synonyms	Anelida	Annelida
Split	213	0
Merged	0	274
Moved	229	229
Renamed	0	730
Added	0	5,107
Excluded	2,351	0
Author Changes	0	306
Taxa changed	2,793	6,340
% changed	%22.11	%42.22

Figure 7.1: Diaforá Coordinated Multiple Views environment displaying Annelida's phylum taxonomy comparison

coordinated views environment significantly improves the user experience on the system.

Our visualization design provides a quick way to interact with large taxonomies and extract statistical information like the distribution of changes and its impact on alternative versions of biological taxonomies.

We managed to significantly improve the user's ability to find and identify changes and to focus on modified taxa that should be fixed or modified in a biological taxonomy curation process.

We expect that this extended version of the system will improve the productivity of taxonomists and that the increase in the cognitive value of the visualization will allow a simpler taxonomic correction process so that the automatic inference of the system can concentrate the efforts of professionals in the points that require special interest to obtain a better classification of living organisms.

7.2 Future Work

As part of the improvements and next steps in the evolution of Diaforá, we identified some user needs that involve the requirement of a taxonomic database cleanup tool, that is, a module that enables the user to identify and modify common differences that occur between taxa records from diverse databases; for example, the author name can be registered using different formats, and wants to consolidate the records using the same format.

Another improvement to the Diaforá tool is to enable a way to correct the taxonomy online and generate a newer version with all fixes that can be uploaded or shared with other taxonomic systems.

It would be useful to have a module to expand the integration of different databases so that the user can add the structure of their databases and use it to load the taxonomic information in the Diaforá system. That module can be shared with the taxonomy community to contribute to support for the different publicly available taxonomic databases.

Participants also mentioned that enabling multilanguage features could be useful since this can be used by people involved in biology conservation around the globe that will require the software to be presented in their native languages to make it easier to use.

A better search engine can be introduced in the system that allows users to find taxa without selecting changes in the comparison portal, in addition, the concept of advanced search can be introduced that includes filters and search by similarity or by metadata attributes, this can improve the user experience by providing faster access to inspect species or groups of interest.

We can also perform some improvements in the GUI overall performance by using some sort of SSR *(Server Side Rendering)* on the system, which will improve the display process of large taxonomies, which means that we can process larger comparisons and provide the user with the visualizations in a shorter time.

Since this research was performed using the asynchronous remote testing approach, we can also think to use the old-fashioned in-lab experiments in the future, just to validate our findings and conclusions in this thesis against this testing approach.

Finally, we would like to publish this tool in a publicly available server that can be used, extended, and modified by anyone interested in the value that this platform can offer for the refinement process of their taxonomic databases [22, 23].

Bibliography

[1] Embedded JavaScript templating language. URL `https://ejs.co/`. Accessed: 2020-08-23.

[2] Loop11. Real people.Real tasks.Unreal websites! URL `https://www.loop11.com/`. Accessed: 2021-11-20.

[3] Museo Nacional de Costa Rica website, 2021. URL `https://www.museocostarica.go.cr/`. Accessed: 2020-08-23.

[4] Insituto Tecnológico de Costa Rica website, 2021. URL `https://www.tec.ac.cr/`. Accessed: 2020-08-23.

[5] Ahmed Alghamdi, Ali Al-Badi, Roobaea Alroobaea, and P.J. Mayhew. A Comparative Study of Synchronous and Asynchronous Remote Usability Testing Methods. *International Review of Basic and Applied Sciences (IRBAS)*, 1:61–97, 01 2013.

[6] Chris Andrzejczak and Dahai Liu. The effect of testing location on usability testing performance, participant stress levels, and subjective testing experience. *Journal of Systems and Software*, 83:1258–1266, 07 2010.

[7] Simon Attfield, Gabriella Kazai, Mounia Lalmas, and Benjamin Piwowarski. Towards a science of user engagement (Position Paper). In *WSDM Workshop on User Modelling for Web Applications*. 01 2011.

[8] John C. Avise and Jin-Xian Liu. On the temporal inconsistencies of Linnean taxonomic ranks. *Biological Journal of the Linnean Society*, 102(4):707–714, 2011.

[9] Joan Ball-Damerow, Laura Brenskelle, Narayani Barve, Pamela Soltis, Petra Sierwald, Rüdiger Bieler, Raphael Lafrance, Arturo Ariño, and Guralnick Robert. Research applications of primary biodiversity databases in the digital age. *PLOS ONE*, 14:e0215794, 09 2019.

[10] Carol M. Barnum and Laura A. Palmer. More than a feeling: Understanding the desirability factor in user experience. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, page 4703–4716, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781605589305. URL `https://doi.org/10.1145/1753846.1754217`.

[11] S. Bjork and J. Redstrom. Redefining the focus and context of focus+context visualization. In *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings*, pages 85–89, 2000. doi: 10.1109/INFVIS.2000.885094.

[12] WoRMS Editorial Board. World Register of Marine Species (WoRMS). =http://www.marinespecies.org, 2020. URL `http://www.marinespecies.org`. Accessed: 2020-08-23.

[13] David Borland, Wenyuan Wang, and David Gotz. Contextual Visualization. *IEEE Computer Graphics and Applications*, 38(6):17–23, 2018.

[14] Mike Bostock. Data-driven documents. URL `https://d3js.org/`. Accessed: 2020-08-23.

[15] N. Boukhelifa, J.C. Roberts, and P.J. Rodgers. A coordination model for exploratory multiview visualization. In *Proceedings International Conference on Coordinated and Multiple Views in Exploratory Visualization - CMV 2003 -*, pages 76–85, 2003.

[16] Anders Bruun, Peter Gull, Lene Hofmeister, and Jan Stage. Let your users do the testing: A comparison of three remote asynchronous usability testing methods. pages 1619–1628, 04 2009.

[17] Stuart Card, Jock Mackinlay, and B. Schneiderman. *Readings in Information Visualization: Using Vision to Think: Morgan Kaufmann.* 05 2014.

[18] Piotr Chynał and Janusz Sobecki. Statistical verification of remote usability testing method. In *Proceedings of the Mulitimedia, Interaction, Design and Innnovation*, MIDI '15, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450336017. URL `https://doi.org/10.1145/2814464.2814476`.

[19] Yu Dong, Alex Fauth, Maolin Huang, Yi Chen, and Jie Liang. Pansytree: Merging multiple hierarchies. In *2020 IEEE Pacific Visualization Symposium (PacificVis)*, pages 131–135, 2020.

[20] MolluscaBase eds. MolluscaBase. =http://www.molluscabase.org, 2020. URL `http://www.molluscabase.org`. Accessed: 2020-08-23.

[21] Armando Estrada CH. and Alexander Fco. Rojas Alvarado. Una nueva especie de Gentiana(Gentianaceae) de Costa Rica. *Phytoneuron*, 60(2153 733X):1–8, 2021.

[22] Manuel Figueroa. Sistema Diaforá Versión 2.0, . URL `https://diafora-taxonomies.herokuapp.com/`.

[23] Manuel Figueroa. Github repository Sistema Diaforá Versión 2.0, . URL `https://github.com/figuemon/DiaforaExt`.

[24] Manuel Figueroa. Guía para las pruebas con usuarios para el sistema Diaforá, . URL `https://bit.ly/3lyLIH6`.

[25] Manuel Figueroa. Demostración de uso sistema Diaforá, . URL `https://bit.ly/3Gj6PFB`.

[26] Manuel Figueroa. Diaforá system test subjects videos, . URL `https://drive.google.com/drive/folders/1BfwNP2Mf6CR3Yd1kAU4IxvDpcNFElEZn?usp=sharing`.

[27] Camilla Forsell and Jimmy Johansson. An heuristic set for evaluation in Information Visualization. *Proceedings of the Workshop on Advanced Visual Interfaces AVI*, pages 199–206, 01 2010.

[28] R. Froese and D. Pauly. FishBase. =http://www.fishbase.org, 2019. URL `http://www.fishbase.org`. Accessed: 2020-08-23.

[29] Gbif.org. GBIF Home Page. =http://www.gbif.org, 2020. URL `https://www.gbif.org`. Accessed: 2020-08-23.

[30] Michael Gleicher. Considerations for Visualizing Comparison. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):413–423, 2018.

[31] Michael Gleicher, Danielle Albers Szafir, Rick Walker, Ilir Jusufi, Charles Hansen, and Jonathan Roberts. Visual Comparison for Information Visualization. *Information Visualization*, 10:289–309, 10 2011.

[32] Michal Grabowski, Anna Wysocka, and Tomasz Mamos. Molecular species delimitation methods provide new insight in taxonomy of the endemic gammarid species flock from the ancient Lake Ohrid. *Zoological Journal of the Linnean Society*, zlw025, 04 2017.

[33] Martin Graham and Jessie Kennedy. A Survey of Multiple Tree Visualisation. *Information Visualization*, 9(4):235–252, 2009.

[34] Danny Holten and Jarke J. Van Wijk. Visual Comparison of Hierarchically Organized Data. *Computer Graphics Forum*, 27(3):759–766, 2008.

[35] Bongshin Lee, George Robertson, Mary Czerwinski, and Cynthia Parr. CandidTree: Visualizing Structural Uncertainty in Similar Hierarchies. *Information Visualization*, 6, 10 2007.

[36] Donald R. Lehmann and James Hulbert. Are Three-Point Scales Always Good Enough? *Journal of Marketing Research*, 9(4):444, 1972.

[37] Antoine Lhuillier, Christophe Hurter, and Alexandre Telea. State of the Art in Edge and Trail Bundling Techniques. *Computer Graphics Forum*, 36:619–645, 06 2017.

[38] Guozheng Li, Yu Zhang, Yu Dong, Jie Liang, Jinson Zhang, Jinsong Wang, Michael J. Mcguffin, and Xiaoru Yuan. BarcodeTree: Scalable Comparison

of Multiple Hierarchies. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1022–1032, 2020.

[39] C. Linnaeus. Systema Naturae per Regna Tria Naturae, Secundum Classes, Ordines, Genera, Species, cum Characteribus, Differentiis, Synonymis, Locis. *Editio Decima 1, Holmiae*, 1:634–635, 01 1758.

[40] Georgina Mace M. The role of taxonomy in species conservation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 359: 711–9, 05 2004.

[41] I. Scott MacKenzie. *Human-computer interaction: An empirical research perspective.* Elsevier, 2013.

[42] Krešimir Matković, Wolfgang Freiler, Denis Gracanin, and Helwig Hauser. ComVis: A Coordinated Multiple Views System for Prototyping New Visualization Technology. *Proceedings of the International Conference on Information Visualisation*, pages 215–220, 08 2008.

[43] Gilberto J. De Moraes. Importance of taxonomy in biological control. *International Journal of Tropical Insect Science*, 8(4-5-6):841–844, 1987.

[44] Jakob Nielsen. How many test users in a usability study?, Jun 2012. URL `https://www.nngroup.com/articles/how-many-test-users/`.

[45] Brian Ondov, Nicole Jardine, Niklas Elmqvist, and Steven Franconeri. Face to Face: Evaluating Visual Comparison. *IEEE Transactions on Visualization and Computer Graphics*, PP:1–1, 08 2018.

[46] Pierre Paquin, Nadine Duperre, James Cokendolpher, Kemble White, and Marshal Hedin. The fundamental importance of taxonomy in conservation biology: The case of the eyeless Cicurina bandida (Araneae:Dictynidae) of central Texas, including new synonyms and the description of the male of the species. *Invertebrate Systematics - INVERTEBR SYST*, 22, 01 2008.

[47] N. Wilson P. Leary K. S. Schulz K. Lans L. Walley J. A. Hammock A. Goddard J. Rice M. Studer J. T. G. Holmes Parr, C. S. and Jr. R. J. Corrigan. The

Encyclopedia of Life v2: Providing Global Access to Knowledge About Life on Earth. *Biodiversity Data Journal 2: e1079*, 2014.

[48] Vanessa Peña-Araya, Emmanuel Pietriga, and Anastasia Bezerianos. A Comparison of Visualizations for Identifying Correlation over Space and Time. *IEEE Transactions on Visualization and Computer Graphics*, PP:1–1, 08 2019.

[49] Helen Regan, Mark Colyvan, and Mark Burgman. A Taxonomy and Treatment of Uncertainty for Ecology and Conservation Biology. *Ecological Applications*, 12:618–628, 04 2002.

[50] Jonathan Roberts. State of the Art: Coordinated 'I&' Multiple Views in Exploratory Visualization. *Coordinated and Multiple Views in Exploratory Visualisation: 2 July 2007; Zurich, Switzerland*, pages 61–71, 08 2007.

[51] Y. Roskov, G. Ower, T. Orrell, D. Nicolson, N. Bailly, P.M. Kirk, T. Bourgoin, R.E. DeWalt, W. Decock, J. Nieukerken, E. van Zarucchi, and L. Penev. Species 2000 'I&' ITIS Catalogue of Life, 2019 Annual Checklist, 2019. ISSN 2405-884X. URL `www.catalogueoflife.org/annual-checklist/2019`. Accessed: 2020-08-23.

[52] Lilliana Sancho-Chavarría, Carlos Gómez-Soza, Fabian Beck, and Erick Mata-Montero. Diaforá: A Visualization Tool for the Comparison of Biological Taxonomies. pages 423–437, 2020.

[53] Lilliana Sancho-Chavarría, Fabian Beck, and Erick Mata-Montero. An expert study on hierarchy comparison methods applied to biological taxonomies curation. *PeerJ Computer Science*, 6:e277, 06 2020.

[54] P Siewald and J Spelda. MilliBase. =http://www.millibase.org, 2020. URL `http://www.millibase.org`. Accessed: 2020-08-23.

[55] Alejandro Solorzano. *Serpientes de Costa Rica: distribución, taxonomía e historia natural = Snakes of Costa Rica: distribution, taxonomy, and natural history.* Instituto Nacional de Biodiversidad, 2004.

[56] Gippoliti Spartaco and Giovanni Amori. Mammal diversity and taxonomy in Italy: Implications for conservation. *Journal for Nature Conservation*, 10:133–143, 06 2002.

[57] Christian Tominski, S. Gladisch, Ulrike Kister, Raimund Dachselt, and H. Schumann. Interactive Lenses for Visualization: An Extended Survey. *Computer Graphics Forum*, 36, 05 2016.

[58] Heli Väätäjä, Jari Varsaluoma, Tomi Heimonen, Katariina Tiitinen, Jaakko Hakulinen, Markku Turunen, Harri Nieminen, and Petri Ihantola. Information visualization heuristics in practical expert evaluation. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, BELIV '16, page 36–43, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450348188. URL `https://doi.org/10.1145/2993901.2993918`.

[59] Gaurav Vaidya, Denis Lepage, and Guralnick Robert. The tempo and mode of the taxonomic correction process: How taxonomists have corrected and recorrected North American bird species over the last 127 years. *PLOS ONE*, 13: e0195736, 04 2018.

[60] Maricela Vicencio and María Ana Fernández. Zooplancton del Domo de Costa Rica: taxonomía y biogeografía. *International Journal of Tropical Biology and Conservation*, 44(2), 1996.

[61] Michelle Q. Wang Baldonado, Allison Woodruff, and Allan Kuchinsky. Guidelines for Using Multiple Views in Information Visualization. page 110–119, 2000. URL `https://doi-org.ezproxy.itcr.ac.cr/10.1145/345513.345271`.

[62] Colin Ware. *Information Visualization: Perception for Design.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3 edition, 2012. ISBN 9780123814647.

[63] Torre Zuk, Lothar Schlesier, Petra Neumann, Mark Hancock, and Sheelagh Carpendale. Heuristics for Information Visualization Evaluation. *Proceedings*

of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization, pages 1–6, 01 2006.