Visual Comparison of Biological Taxonomies: A Task Characterization

L. Sancho-Chavarria[1], F. Beck[2], E. Mata-Montero[1] and D. Weiskopf[2]

[1]Instituto Tecnológico de Costa Rica, Costa Rica
[2]University of Stuttgart, Germany

Abstract

Although biological taxonomies are a prevalent use case in hierarchy visualization, there has been little research on the characterization of users' tasks for taxonomy comparison. Task identification is very relevant as a start point to design an effective information visualization solution for taxonomic work. We performed a systematic domain characterization that involved interviews with experts, literature review, the identification of data sources, and a survey of existing software tools. We performed iterative analysis until we reached a satisfactory list of users' tasks. We present the tasks in a two-level model, describe them, and discuss future work.

Categories and Subject Descriptors (according to ACM CCS): Information Visualization [Visualization]: Visualization Design and Evaluation Methods—

1. Introduction

Biological taxonomies have become a prevalent use case in hierarchy visualization [MGT*03, GK10]. When we find a figure that illustrates the species classified by domain, kingdom, phylum, class, order, family, and genus, we might get the impression that everything is stable, or even static. Certainly, the nomenclature system in biology is a standard from the 18th century, but the truth is that the field shows great dynamics at the classification level. Multiple taxonomies lead to misunderstandings. Here, the visual comparison of taxonomies can be a useful mechanism for the identification of differences and the quest for conciliation.

In this paper, we focus on the characterization of users' tasks. As has been pointed elsewhere [Mun09, BM13, KS14, AES05], domain characterization is important as the start point to derive a visualization solution. There has been little research on the characterization of users' tasks for taxonomic work, specially when resolving differences between versions of a taxonomy. We performed a systematic analysis and contribute a proposal for the characterization of tasks in the domain of biological taxonomy comparison. We expect our contribution to be a start point for the design of visualization tools that effectively convey information for taxonomic work.

2. Method

Figure 1 summarizes our method for task characterization. We performed a two-stage analysis. The first stage consisted of interviews with experts and literature review on both taxonomy work and information visualization. We performed a task analysis and obtained
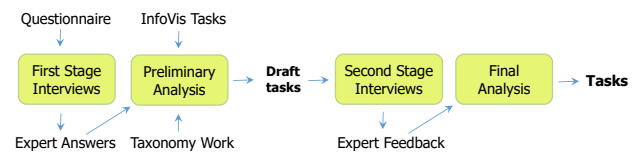

Figure 1: *Method for characterizing users' tasks.*

a list of draft tasks. In the second stage, we went back to experts for feedback and performed a final analysis. We believe that the two-stage analysis as well as the variety of sources of information provides a wide view that reduces bias.

First Stage Interviews: Works such as Kandel et al. [KPHH12] illustrate the importance of interviews for the discovery of needs on visualization and analysis tools. We interviewed six experts gathered from our professional network. They were located in the United States, Spain, and Costa Rica, and held positions either at natural history museums, herbaria or recognized world biodiversity organizations. Experts worked as taxonomists, curators, biodiversity informatics managers, or at the executive level. Before we began, we set up a clear and consistent structure for the interviews. We prepared a questionnaire with questions about the tasks that taxonomists accomplish when doing taxonomic work and also included questions regarding the tasks they would perform with a software tool for taxonomy comparison. In most cases, we were able to send them the questionnaire before conducting a semi-structured interview by video-chat conference. During the meeting,

we also asked open-ended questions to encourage interviewees feature the tasks they could do with a visualization software. Duration of interviews ranged from 65 to 95 minutes. Answers were first recorded in text documents and then organized in a matrix format.

Taxonomy Work: Besides interviews, we investigated literature, databases, and software tools in order to attain a broader view of the tasks that taxonomists accomplish when doing data cleaning. Local or regional organizations have their specific records while international initiatives are working to integrate and standardize biodiversity information in a single repository. Annual versions of checklists and databases are available on-line. Since the taxonomy community has been very successful at standardizing names and recording dates, we found that there is available data for the comparison endeavor; for instance, data provided by Species 2000.

Biodiversity research requires scientists to map their observations to previously known taxonomic concepts. The software Euler [Eul16] implements concept resolution through operations based on set theory relationships [FPW08, FP09], and relies on the user's knowledge to make the alignments; it does not attempt automatic resolution. In contrast, the Taxonomic Tree Tool (TTT) [Bio16] automatically compares two taxonomies and shows the resulting hierarchies in an indented layout, side by side. Interpretation of the visualization requires considerable effort by the user.

Tasks from Information Visualization: Previous information visualization work describes classifications of general and specific visualizations tasks. Graham and Kennedy analyzed three main high-level tasks [GK10]. Amar et al. [AES05] introduced ten specific tasks that denote user's activities while employing information visualization systems. Brehmer and Munzer [BM13] presented a multi-level typology of visualization tasks that includes the description of complex tasks as sequences of simpler lower level tasks. Recent work has been done on the identification of users tasks in specific domains [Rot13, PBK*15, BSIM14]. Our work develops a proposal for the characterization of users' tasks in the domain of biological taxonomy comparison.

Preliminary Analysis: We organized the information in two matrices: one to group the responses from experts and the other one to organize the tasks that come from information visualization literature. We identified the tasks that were more frequently mentioned by experts as well as the most common tasks from literature, and marked them as possible tasks. We reviewed existing taxonomy alignment/comparison software tools—Euler and TTT—, and derived a list of possible specific functions to develop taxonomic work. The analysis was performed iteratively, refining the tasks' names and grouping them in functional categories until we reached at a satisfactory list of draft tasks.

Second Stage Interviews: An important step in our method was to ask experts for their feedback. We contacted experts by email so that they could get acquainted with the tasks before performing the interviews. Experts considered the list of draft tasks was complete but also provided important feedback to refine them. At this stage, the information they provided made us gain a deeper understanding of the taxonomic work and their needs for visualization.

Final Analysis: A matrix format was useful again to organize the answers and proceed with the analysis. We observed the need to

modify the names of some tasks in order to make them more consistent with the taxonomy work vocabulary. For instance, we had considered the draft task *Identify exclusions*. We had noticed that node names in the hierarchy could disappear from a version of the taxonomy to a newer one. However, when taxonomists decide that the current name of a specie is incorrect, they change it but this cannot be interpreted as if they excluded the specie and added a new one. It might be that the name has become a synonym. Therefore, we decided to change the task name to a more familiar one for taxonomists: *Identify corrections*. While doing this, we were able to re-group tasks and clarify their descriptions.

3. Results

Through our method, we obtained a better understanding of the challenges of taxonomic cleansing, the specific needs for a visualization/analysis tool, and the role of hierarchy comparison. We identified ten user visualization tasks and organized them in a two-level model (see Figure 2).

Categories	User Visualization Tasks	Description
Pattern Identification	1. Identify congruence	Identify topological and taxonomic naming congruence.
	2. Identify corrections	Identify splits, merges, moves, and naming corrections.
	3. Identify additions	Identify new nodes added.
	4. Overview changes	Obtain an overview of different types of change.
	5. Summarize	Obtain a numerical understanding of change.
Query	6. Find inconsistencies	Recognize violations of rules (e.g., repeated or missing names).
	7. Filter	Find cases that satisfy certain conditions.
	8. Retrieve details	Retrieve attributes of a particular node.
	9. Focus	Navigate to an area of interest.
Edit	10. Edit	Perform taxonomic alignment and data updates.

Figure 2: *Characterization of users' tasks.*

The pattern identification category corresponds to visualization tasks that allow users to recognize general patterns of change, the query category involves tasks that provide detailed information, and the edit category groups tasks for performing taxonomic alignments. These tasks can be mapped to some previously identified tasks from the information visualization research. For instance, we believe that specific tasks 1 to 5 fit within Graham and Kennedy high-level task *exploration of differences* [GK10], task 6 has a correspondence with both *exploration of differences* and *filtering data*, tasks 7 to 9 fit within *filtering data*, and task 10 can be considered part of *mapping relationships*. It is valuable to see correspondences with these previously defined high-level tasks as it is also valuable the specific low-level characterization that we introduced in our two-level task proposal for the domain of taxonomy.

It is important to observe that expert users highlighted the benefit of editing tasks. We believe this is particularly interesting because the output from automatic comparison can still be refined through further experts' assessment.

4. Future Work

In the near future, we intend to analyze how already known visualizations fit our task characterization. We expect that such analysis can provide valuable insights for future development of visualization proposals for the taxonomy domain.

References

[AES05]   AMAR R., EAGAN J., STASKO J.: Low-Level Components of Analytic Activity in Information Visualization. In *Proceedings of the 2005 IEEE Symposium on Information Visualization (INFOVIS'05)* (2005), IEEE, pp. 111–117. 1, 2

[Bio16]   BIODINFO: Taxonomic Tree Tool, 2016. URL: http://ttt.biodinfo.org/TF/. 2

[BM13]   BREHMER M., MUNZNER T.: A Multi-Level Typology of Abstract Visualization Tasks. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2376–2385. 1, 2

[BSIM14]   BREHMER M., SEDLMAIR M., INGRAM S., MUNZNER T.: Visualizing dimensionally-reduced data. In *Proceedings of the Fifth Workshop on Beyond Time and Errors Novel Evaluation Methods for Visualization - BELIV '14* (2014), ACM Press, pp. 1–8. 2

[Eul16] Euler Project, 2016.   URL: https://github.com/EulerProject. 2

[FP09]   FRANZ N., PEET R.: Perspectives: Towards a language for mapping relationships among taxonomic concepts. *Systematics and Biodiversity 7*, 1 (2009), 5–20. 2

[FPW08]   FRANZ N., PEET R., WEAKLEY A.: On The Use Of Taxonomic Concepts In Support Of Biodiversity Research And Taxonomy. In *The New Taxonomy*, Wheeler Q., (Ed.). CRC Press, 2008, pp. 63–86. 2

[GK10]   GRAHAM M., KENNEDY J.: A survey of multiple tree visualisation. *Information Visualization 9*, 4 (2010), 235–252. 1, 2

[KPHH12]   KANDEL S., PAEPCKE A., HELLERSTEIN J. M., HEER J.: Enterprise Data Analysis and Visualization: An Interview Study. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2917–2926. 1

[KS14]   KANG Y.-A., STASKO J.: Characterizing the intelligence analysis process through a longitudinal field study: Implications for visual analytics. *Information Visualization 13*, 2 (2014), 134–158. 1

[MGT*03]   MUNZNER T., GUIMBRETIÈRE F., TASIRAN S., ZHANG L., ZHOU Y.: TreeJuxtaposer: Scalable Tree Comparison Using Focus+Context with Guaranteed Visibility. *ACM Transactions on Graphics 22*, 3 (2003), 453–462. 1

[Mun09]   MUNZNER T.: A Nested Model for Visualization Design and Validation. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 921–928. 1

[PBK*15]   PAPERS S., BERTINI E., KENNEDY J., PUPPO E., NUSRAT S., KOBOUROV S.: Task Taxonomy for Cartograms. *Eurographics Conference on Visualization (EuroVis)* (2015). 2

[Rot13]   ROTH R. E.: An empirically-derived taxonomy of interaction primitives for interactive cartography and geovisualization. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2356–65. 2