

Modelo de lentes interactivos para la visualización y comparación de taxonomías biológicas

Manuel Figueroa, *ITCR*, Nathalia Gonzalez, *ITCR*, Esteban Leandro, *ITCR*

MC-7201 Introduction to Research

Instituto Tecnológico de Costa Rica

{mfigueroacr, natgondou, elc790}@gmail.com

Resumen—Se presenta un modelo de visualización alternativo para la comparación de taxonomías biológicas, que busca fortalecer el avance logrado en el sistema Diaforá [1]. Permitiendo a los taxónomos enfocarse en aspectos importantes de los árboles de clasificación y manteniendo al mismo tiempo un mapa de la totalidad de los árboles de taxonomía que están analizando. Dicha propuesta pretende ser evaluada por un panel de expertos en taxonomía, para verificar la eficacia de esta extensión al sistema Diaforá, de manera similar al análisis presentado en [2].

Index Terms— \LaTeX Visualización, Taxonomías biológicas, Diaforá, Enfoque y Contexto.

I. INTRODUCCIÓN

El problema descrito se deriva de la investigación realizada por los profesores Lilliana Sancho Chavarría y Erick Mata Montero del ITCR, como parte de su investigación en la comparación y visualización de taxonomías biológicas [1]. Las taxonomías biológicas son estructuras donde las especies son clasificadas de acuerdo a un sistema jerárquico propuesto por Linnaeus en el siglo 18 [3], y que incluye las categorías de dominio, reino, filo o división, clase, orden, familia, género y especie. La información de todos los seres vivos conocidos se agrupa en árboles taxonómicos, que han sido creados y mantenidos por taxónomos a lo largo del mundo durante siglos. La reciente revolución digital ha permitido que gran parte de esa información pueda ser compartida y revisada por expertos. Debido a la naturaleza dinámica de estos datos es común que los taxónomos se enfrenten a distintas versiones de los datos que pueden ser corregidas y unificadas mediante la comparación de árboles taxonómicos. Las herramientas que ayuden a este grupo a analizar e identificar estas diferencias y facilitar el proceso de curación de las taxonomías permitiría un avance significativo en la calidad y fiabilidad de las clasificaciones biológicas de los seres vivos.

II. TRABAJOS RELACIONADOS

III. ANÁLISIS DEL USO DE LENTES INTERACTIVOS EN TAXONOMÍAS BIOLÓGICAS

Uno de los sentidos más importantes de los seres humanos es la visión. Ésta es empleada para obtener la información visual del entorno, y en este caso específico la visualización se ha convertido en un medio para ayudar a las personas de diversos campos a obtener información relevante sobre los datos.

Sin embargo, dado a que el tamaño de los datos aumenta constantemente, los enfoques de visualización tienen que resolver el problema de representaciones visuales exponenciales que dificultan la visualización de contenido relevante en una sola imagen de visualización [4]. Algunos investigadores como C. Tominski han tratado de abordar el desafío de la visualización con enfoque a través de exploraciones con grandes volúmenes de datos. Una de las técnicas para resolver los problemas de visualización son los lentes interactivos, una clase de métodos que permiten la exploración de datos con múltiples facetas. Se busca con el uso de lentes interactivos una vista alternativa de los datos presentes en una área específica de la pantalla, con el fin de enfatizar parte de esta información de una manera más clara para los usuarios [4]. Los datos estructurados en árboles son comunes en muchas disciplinas; este trabajo se enfocará específicamente en las clasificaciones biológicas para la detección de diferencias y detalles relevantes en una única pantalla, por ejemplo, los árboles filogenéticos que a diferencia de las categorizaciones taxonómicas estudian las relaciones de parentesco entre las especies. Se han estudiado diferentes técnicas de visualización que permiten enfatizar las similitudes y resaltar las diferencias existentes entre los árboles, como árboles de consenso y debido a que estos árboles cuentan en promedio con más de 50 nodos es necesario la utilización de estrategias para ordenar los árboles de manera automática entre estas se destacan la diferencia mínima de tripletas (MDT), y la semejanza máxima de ramas (MBS). Estos algoritmos buscan maximizar el alineamiento de las hojas de los árboles en una comparación cara a cara [5].

III-A. Lentes Interactivos

Según la definición encontrada en [4], un lente interactivo es una herramienta ligera, que intenta resolver un problema localizado de visualización, alterando temporalmente una parte seleccionada de la representación de los datos.

También siguiendo el trabajo de Tominski, se definen como propiedades importantes de los lentes interactivos:

- **Forma:** La forma del lente virtualmente no tiene restricción, sin embargo, es común que muchos sistemas intenten emular el modelo de un lente del mundo real, en su mayoría circulares, no obstante esta forma puede adaptarse según la naturaleza de los datos que se están

explorando. La importancia radica en que el usuario pueda identificar el lente fácilmente y sobre cuales datos quiere que el lente realice su función.

- **Posición y tamaño:** Se consideran atributos parametrizables, y que el usuario pueda ubicar el lente y ajustar su tamaño sobre cualquier parte de los datos en el área de exploración.
- **Orientación:** Cuando se emplea el recurso de visualización en tres dimensiones, la orientación toma relevancia en la forma en la que se observan los datos, ya que dependiendo del ángulo de visión del punto de observación el modelo de datos presentado en pantalla puede variar.

III-B. Lentes Interactivos para Visualización

Las técnicas de lentes son herramientas que nos permiten enfocarnos temporalmente en un punto de interés, un lente es una selección de una visualización base donde se buscan localizar un punto específico y una vez que se llega al punto de interés la visualización vuelve a su estado original. La selección captura lo que debe ser resaltado por un lente. Normalmente el usuario controla la selección a través de movimientos sobre la representación visual de los datos.

III-C. Lentes Interactivos para Visualización de Taxonomías

La taxonomía tiene su origen en un vocablo griego que significa “ordenación”. Se trata de la ciencia de la clasificación que se aplica en la biología para la ordenación sistemática y jerarquizada de los grupos de animales o plantas. Es importante establecer además que la taxonomía está muy en relación con lo que se conoce por el nombre de sistemática. Esta puede definirse como la ciencia que se encarga de llevar a cabo el estudio de las relaciones de parentesco, también llamadas afinidades, que se producen entre las distintas especies. En este paper nos enfocaremos en la taxonomía biológica, la cual forma parte de la biología sistemática, dedicada al análisis de las relaciones de parentesco entre los organismos. Una vez que se resuelve el árbol filogenético del organismo en cuestión y se conocen sus ramas evolutivas, la taxonomía se encarga de estudiar las relaciones de parentesco. La visión más extendida entiende a los taxones como clados (ramas del árbol filogenético, con especies emparentadas por un antepasado común) que ya fueron asignados a una categoría taxonómica.

El proceso de la taxonomía continúa con la asignación de nombres (de acuerdo a los principios de la nomenclatura), la elaboración de las claves dicotómicas de identificación y la creación de los sistemas de clasificación.

Los taxones permiten clasificar a los seres vivos a partir de una jerarquía de inclusión (cada grupo abarca a otros menores mientras está subordinado a uno mayor). Las categorías fundamentales, desde la más abarcativa hasta la menor, son el dominio, el reino, el filo o división, la clase, el orden, la familia, el género y la especie.

IV. DISEÑO DE VISUALIZACIÓN INTERRING PARA EL SISTEMA TAXONÓMICO DIÁFORA

Las técnicas conocidas como *RSF* o *Radial, space-filling* por sus siglas en inglés, tienen ciertas ventajas para la visualización de jerarquías, utilizan el espacio en pantalla de

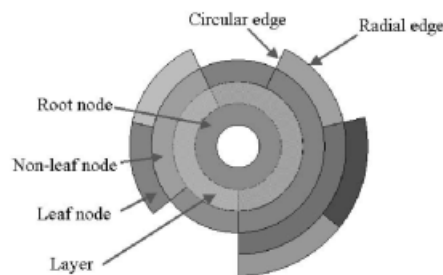


Figura 1. Ejemplo de visualización RSF. Tomado de [6]

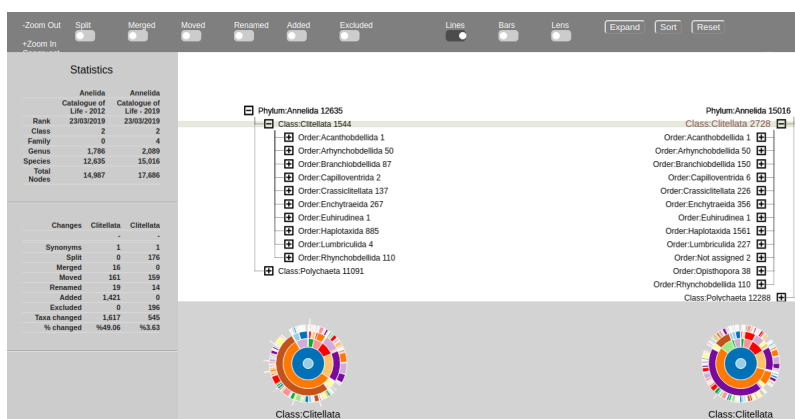


Figura 2. Sistema Diaforá en conjunto con visualización InterRing.

manera eficiente mientras que proveen una vista intuitiva de la estructuras jerárquicas.

Haciendo uso de una visualización de tipo *InterRing* [6], que emplea el concepto de distorsión circular extendemos la capacidad actual del sistema *Diaforá* para mantener el contexto de la estructura jerárquica que esta siendo desplegada en el árbol taxonómico.

V. DESARROLLO DEL MODELO DE VISUALIZACIÓN EN EL SISTEMA DIAFORÁ

Según lo investigado en [1], el método *edge drawing* puede comunicar de manera clara las diferencias entre dos versiones de una taxonomía. El uso de colores y líneas permite de manera clara detectar los cambios, además de que la interacción del usuario con la visualización permite enfocarse en aquellos cambios que puedan llamar su atención.

La principal desventaja detectada sobre el sistema de visualización *edge drawing* consiste en la pérdida de contexto de la estructura jerárquica del árbol taxonómico debido a la gran cantidad de nodos presentes una taxonomía. El propósito de utilizar una visualización *InterRing* [6] secundaria pretende apoyar al usuario a conocer el ámbito local del árbol taxonómico presentando de manera gráfica parte de la jerarquía circundante que puede no ser visible debido a la cantidad de datos y que para poder visualizarlos requiere que el usuario haga *scroll* sobre la gráfica *edge drawing*.

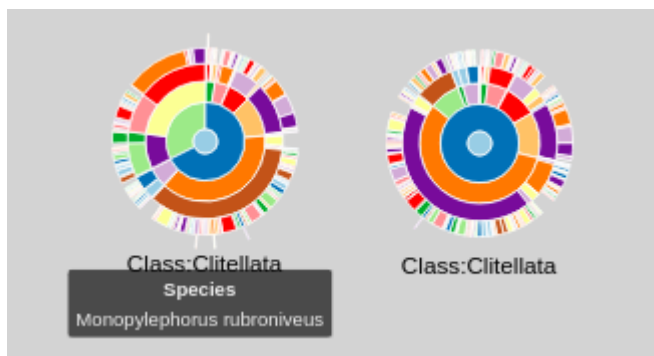


Figura 3. Comparación de dos clases usando la visualización InterRing

El uso del método InterRing permite de manera compacta y simple renderizar una gran cantidad de nodos y permite al usuario navegar sobre las ramas del árbol taxonómico y explorar su composición, también al estar sincronizadas ambas gráficas se pueden realizar comparaciones visuales sobre las imágenes resultantes de las visualizaciones InterRing que destacan las mayores diferencias entre distintas versiones de una taxonomía biológica.

Como se puede observar en la figura 5, es posible detectar diferencias entre las figuras que representan una misma clase (*Clitellata*) mediante las variaciones existentes en ambas figuras, de esta manera se espera poder contribuir con el trabajo de refinamiento de las taxonomías al resaltar y hacer más evidentes las diferencias entre versiones de una taxonomía.

VI. VALIDACIÓN DEL MODELO PROPUESTO

VII. TRABAJO FUTURO

VIII. CONCLUSIONES

REFERENCIAS

- [1] L. Sancho-Chavarría, C. Gómez-Soza, F. Beck, and E. Mata-Montero, "Diaforá: A visualization tool for the comparison of biological taxonomies," *Communications in Computer and Information Science High Performance Computing*, p. 423–437, 2019.
- [2] L. Sancho-Chavarria, F. Beck2, and E. Mata-Montero1, "An expert study on hierarchy comparison methods applied to biological taxonomies curation," Aug 2019. [Online]. Available: <https://peerj.com/preprints/27903/>
- [3] L. C. Von, "Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis /," 1767.
- [4] C. Tominski, S. Gladisch, U. Kister, R. Dachsel, and H. Schumann, "Interactive lenses for visualization: An extended survey," *Computer Graphics Forum*, vol. 36, no. 6, p. 173–200, 2016.
- [5] W. Zainon and P. Calder, "Visualising phylogenetic trees," *Conferences in Research and Practice in Information Technology Series*, vol. 50, pp. 145–152, 01 2006.
- [6] J. Yang, M. Ward, and E. Rundensteiner, "Interring: an interactive tool for visually navigating and manipulating hierarchical structures," *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002*.