

Escuela de Computación

Visualización y Análisis de Big Data mediante Ambientes Tridimensionales

Proyecto de Investigación

Código: 5402-1375-4401

Informe Final (Técnico)

Período: 01- 2015 a 12-2016

- Marzo, 2017 -

Dr. Erick Mata M.,
Coordinador
emata@itcr.ac.cr

M.Sc. Lilliana Sancho Ch.
Investigadora
lsancho@itcr.ac.cr

Contenido

Resumen	2
1. Introducción	3
2. Marco teórico	5
3. Metodología	8
4. Resultados	12
5. Discusión y conclusiones	19
6. Recomendaciones	20
7. Agradecimientos	21
8. Referencias	21

APENDICES:

- A. Informe de Avance I semestre, 2015.
- B. Informe de Avance Anual 2015.
- C. Informe de Avance I semestre, 2016.
- D. Reporte Técnico: Estado del Arte de la Visualización de Información Voluminosa con Estructura Jerárquica.
- E. Documento de Diseño.
- F. Documento de Instalación y Operación.
- G. Artículo: Visual Comparison of Biological Taxonomies: A task characterization, poster (EuroVIS).
- H. Artículo: Den3D: A Ludic 3D Visualization and Learning Tool for Biological Taxonomies, demo (WITFOR)
- I. Artículo: Comparación visual de grandes jerarquías: El caso de las taxonomías biológicas (Bienal Centroamericana)
- J. Simposio: Visual Comparison of Biological Taxonomies (TDWG)
- K. Artículo: Visualization approaches for the comparison of biological taxonomies (RIO)
- L. Estudio de Usabilidad de *Den3D*.
- M. Anuncio Charla ProIncomp
- N. Afiche para divulgar y promover *Den3D*

Resumen

El proyecto “*Visualización y Análisis de Big Data mediante Ambientes Tridimensionales*” plantea como problema a resolver la carencia de herramientas informáticas para la creación, visualización, edición y análisis de información cuya organización es jerárquica y voluminosa. El reto principal es mostrar grandes jerarquías en el limitado espacio de la pantalla de una computadora. Utilizamos los árboles de conos para aprovechar las ventajas que, en general, tienen las representaciones en 3D sobre las que usan 2D. El estudio se concentró en el dominio de la biodiversidad, particularmente en las taxonomías de organismos. Para el desarrollo de la investigación utilizamos como base trabajo previo desarrollado por [1, 10] y la metodología iterativa e incremental de ingeniería de software para definir una arquitectura moderna y eficiente basada en el web. Realizamos pruebas con varios conjuntos de datos reales provenientes de organizaciones globales para la conservación de la biodiversidad tales como Catalogue of Life y EOL. Las pruebas indican que con aproximadamente 4,500 nodos o menos, generados aleatoriamente, el ambiente desarrollado (*Den3D*) tiene un buen rendimiento. Se hizo un estudio con docentes de secundaria, quienes indicaron que *Den3D* es útil para visualizar información con estructura jerárquica y para las taxonomías de especies que se estudian en secundaria. Sin embargo se determinó que, en su estado actual, la herramienta es débil para propiciar el análisis. Se recomienda expandir la investigación hacia ambientes colaborativos y poner *Den3D* a disposición de docentes para que los educadores y estudiantes lo utilicen y provean retroalimentación para futuras investigaciones.

Palabras clave: visualización de información, árboles de conos, informática para la biodiversidad, pensamiento visual.

Keywords: information visualization, cone trees, biodiversity Informatics, visual thinking.

1. Introducción

La explosión de información digital disponible en Internet (*big data*) ha hecho indispensable la visualización de esta información de manera sintética y eficiente. Muchas veces, la información tiene relaciones de carácter jerárquico. Por ejemplo, las taxonomías biológicas organizan las especies de organismos por medio de géneros, éstos a su vez por medio de familias y así sucesivamente hasta taxones como *Animalia* y *Plantae*. La visualización de grandes cantidades de información con estructura jerárquica es un gran reto desde el punto de vista computacional, pues conlleva la representación de mucha información en el pequeño espacio de una pantalla de computadora. Los taxónomos biológicos han identificado y descrito aproximadamente 2 millones de especies pero se estima que en nuestro planeta viven alrededor de 10 millones. La visualización de tal cantidad de información es un enorme desafío pero es también una contribución muy valiosa para apoyar los esfuerzos de generación de conocimiento por parte de los científicos y de educación del público en general con el fin último de conservar la biodiversidad.

Nuestro objetivo ha sido explorar el diseño de un ambiente digital tridimensional que posibilite la experimentación con taxonomías; tanto el despliegue adecuado de información con estructura jerárquica como su edición (creación y modificación), en tiempo real, considerando variables tales como la cantidad de información, los mecanismos para la interacción y las capacidades para llevar a cabo análisis de la información. Para tal efecto desarrollamos la herramienta denominada *Den3D*, donde “Den” es diminutivo de la palabra “dendro”, que significa “árbol” en griego.

De acuerdo con estudios de revisiones bibliográficas ([2], [3]) la investigación en Computación se puede organizar en las siguientes categorías:

- Teoría formal: proposiciones, lemas, teoremas, pruebas (12%).
- Diseño y modelamiento: técnicas, sistemas, modelos cuyas propiedades no pueden ser probadas formalmente (70%).
- Trabajo empírico: artículos que recogen, analizan e interpretan observaciones acerca de diseños, sistemas, modelos o teorías abstractas (2%).
- Prueba de hipótesis: artículos que definen hipótesis y describen experimentos para probarlas (2%).
- Otros: artículos que no calzan en ninguna de las categorías anteriores (14%).

Nuestra investigación se enmarca dentro de la categoría de “diseño y modelamiento” con un componente de “trabajo empírico” para la valoración de resultados. Para el diseño del ambiente experimental *Den3D* aplicamos una metodología basada en el enfoque iterativo e incremental de la ingeniería de software [4]. Esta metodología nos permite construir el software – que es eminentemente interactivo – y valorarlo regularmente en cada iteración para probar alternativas e ir precisando las ideas, las decisiones de diseño y el encuadre de la

investigación. El trabajo empírico corresponde al estudio con usuarios que se llevó a cabo con el propósito de evaluar la aplicabilidad del software como herramienta educativa en los cursos de biología en secundaria.

El punto de partida fue el Biovisualizador, trabajo desarrollado por [1, 10], el cual tuvo como objetivo visualizar taxonomías mediante árboles de conos. El Biovisualizador fue desarrollado como una aplicación de escritorio (“desktop”) – aunque se tenía la posibilidad de ejecutarlo en web – y fue probado con una cantidad relativamente pequeña de datos (aproximadamente 150 nodos) y tres niveles de información. Para el desarrollo de *Den3D* se hizo un proceso de ingeniería reversa a partir del ejecutable del Biovisualizador y se planteó los siguientes objetivos:

1. OE-1: Identificar algoritmos y nuevas tecnologías relevantes para la visualización y el análisis de datos.
2. OE-2: Demostrar que un “workbench” tridimensional es efectivo para visualización de grandes volúmenes de información.
3. OE-3: Demostrar que un “workbench” tridimensional es efectivo para hacer análisis de información.
4. OE-4: Validar el ambiente para la temática del área de la biodiversidad.

Esto llevó al desarrollo de funcionalidad adicional que resumimos a continuación:

- se diseñó e implementó una arquitectura basada en tecnologías actuales de desarrollo para el web;
- se diseñó e implementó un mecanismo para cargar (importar) grandes volúmenes de datos provenientes de bases de datos globales;
- se investigó y refinó los algoritmos para posicionamiento de los nodos con el objetivo de evitar traslapes al desplegar más de 3 niveles y para lograr un mejor aprovechamiento del espacio y;
- lo más notable, se agregó funciones de edición para que los usuarios puedan crear y editar sus propias taxonomías.

La capacidad de editar y construir taxonomías es un aporte novedoso, especialmente para desarrollar trabajo colaborativo que pueda contribuir con actividades de ciencia ciudadana, *crowd sourcing*, y educación para la conservación de la biodiversidad (bioalfabetización).

Este informe está organizado de la siguiente manera: la Sección 2 presenta el marco teórico; en la Sección 3 se detalla la metodología aplicada en la investigación, tanto para diseñar y modelar la solución, como para validar el modelo con usuarios; la Sección 4 describe los

resultados obtenidos; la Sección 5 discute los resultados y se presentan conclusiones. Finalmente, la Sección 6 incluye recomendaciones y la Sección 7 los agradecimientos.

2. Marco teórico

La clasificación de conceptos de forma jerárquica es común en muchos campos de conocimiento y de la vida cotidiana; por ejemplo, en la organización de los archivos en el directorio de una computadora, el organigrama de una institución, la organización del código fuente de programas computacionales y la taxonomía de especies biológicas. Las jerarquías se representan mediante la estructura de datos conocida como *árbol*. Los árboles son un caso especial de grafo. En términos generales, un grafo $G=(V,E)$ consiste de un conjunto finito de vértices V los cuales se relacionan mediante un conjunto también finito de aristas E . Un árbol es un grafo que no tiene ciclos y es conexo ([5], [6]). En la Figura 1 se muestra un ejemplo de un grafo y de un árbol con el mismo conjunto de nodos.

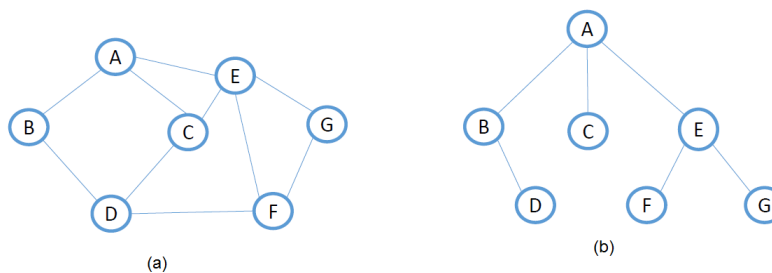


Figura 1: Ejemplo de grafo (a) y de árbol o jerarquía (b).

Varios estudios presentan las diferentes técnicas para la visualización de árboles ([1], [7], [8]). Graham y Kennedy [7] las resumen en cinco categorías que se muestran en la Figura 2. La Figura 2a) presenta la disposición clásica de árboles mediante nodos y arcos, la cual es muy fácil de entender pero presenta la limitación de que consume mucho espacio en la pantalla. El mismo árbol se representa en 2b) mediante una disposición anidada. Los mapas de árboles (“tree maps”) pertenecen a esta categoría y han sido exitosos para, adicionalmente, destacar proporciones. En 2c) se ilustra el árbol con una disposición de nodos adyacentes, también conocida como “icicle plot”, donde los nodos hijos colindan con sus padres. En esta disposición la relación jerárquica no es tan clara. En 2d) se presenta el árbol indentado, una de las disposiciones más populares debido a su uso en los sistemas computacionales que muestran la organización de los archivos en carpetas. La disposición mediante una matriz como se muestra en la figura 2e) es común para grafos pero menos común para árboles ya que por un lado es difícil seguir la conexión entre nodos y es notable desperdicio de espacio (celdas en blanco).

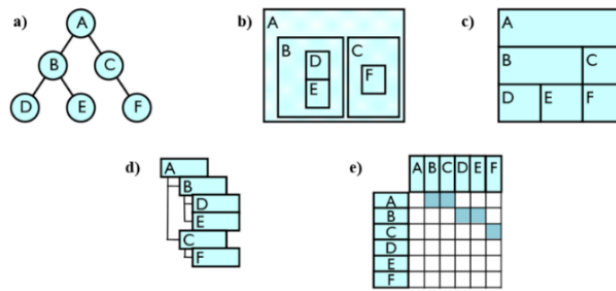


Figura 2: a) nodos y arcos, b) anidado, c) adyacente, d) indentado, e) matriz. Tomado de [7].

En el documento “Estado del Arte de la Visualización de Información Voluminosa con Estructura Jerárquica” (ver apéndice D) presentamos un resumen del estado del arte de la visualización de información voluminosa con estructura jerárquica que amplía los conceptos que se introducen en este apartado. En ese estudio, además de las técnicas de la Figura 2, se presentan y discuten técnicas radiales, árboles hiperbólicos, representaciones inspiradas en la botánica, basadas en anillos y en agrupamiento de bordes. También se discuten las representaciones en tres dimensiones, tal es el caso de los árboles de conos. El sitio web treevis.net cubre de manera muy completa los distintos enfoques y sistemas que han sido descritos en la literatura o implementados hasta el día de hoy.

Los árboles de conos [9] se caracterizan porque son tridimensionales, interactivos y en muchos casos hacen un mejor uso de la pantalla en comparación con la representación clásica de nodos y arcos en dos dimensiones. Hernández et al [10] llevaron a cabo una investigación que produjo el software denominado Biovisualizador, herramienta que utiliza árboles de conos para visualizar taxonomías biológicas. Este fue aplicado como caso de estudio con fines educativos [10] y es un punto de partida para la investigación que se describe en el presente informe. La Figura 3 presenta una imagen con un árbol de conos producido por el Biovisualizador.

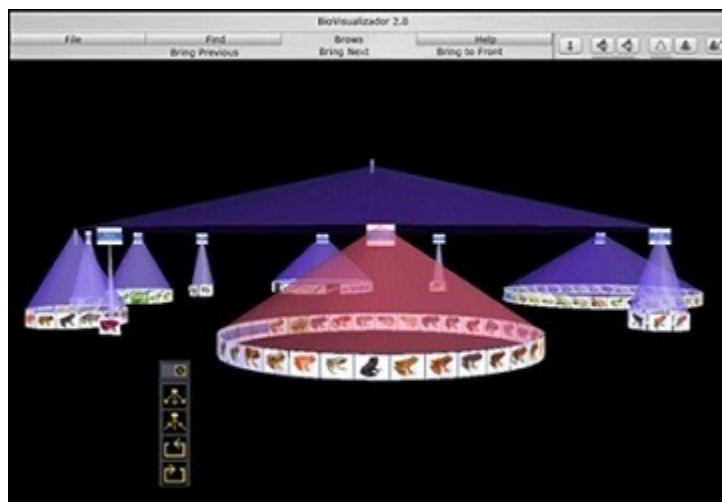


Figura 3: Arbol de conos generado con el Biovisualizador. Tomado de [10].

Como se puede observar en la Figura 3, en los árboles de conos la jerarquía se presenta en tres dimensiones, esto con la intención de maximizar el uso del espacio disponible en pantalla, lo cual permite muchas veces la visualización de toda la estructura. Con esta técnica, aunque los árboles estén muy poblados, el desplazamiento mediante “scroll” no es indispensable ya que los nodos están colocados en tres dimensiones. Mediante comandos interactivos (pulsando teclas), la estructura y cualquier subárbol se pueden rotar, de tal manera que el usuario puede visualmente buscar información y puede enfocarse en subárboles o nodos de interés, sin perder contexto. Los conos son transparentes para no obstaculizar totalmente la vista de otros conos e información que están detrás (alejados del usuario). Los árboles de conos se pueden administrar y reestructurar mediante operaciones de “jardinería”, tales como: *podar* (“prune”) que sirve para eliminar subárboles que no interesa tener dentro del foco de atención, o mediante operaciones para *extraer* (“snip”) definitivamente un subárbol del árbol completo, el cual no se va a requerir más. También se tienen operaciones para contraer y expandir nodos, lo cual permite dejar temporalmente en el espacio visual solamente los nodos de interés. En general, las operaciones de jardinería sirven para administrar y lograr una mejor comprensión y administración de un árbol grande (i.e., que no cabe completamente en la pantalla).

Entre las ventajas de los árboles de conos están que para árboles medianos la estructura se puede mostrar de manera completa en la pantalla, mostrando tanto información de contexto como el detalle en cada momento; que tiene un buen aprovechamiento del espacio; y que utiliza la perspectiva, transparencia y otras pistas visuales que son cognitivamente naturales para que el ser humano pueda hacer “visual thinking”. Tiene como desventaja la oclusión parcial de los nodos y que al ser una representación tridimensional el despliegue puede ser lento si el equipo usado no tiene un GPU (Graphics Processing Unit).

Den3D tomó el Biovisualizador como punto de partida, por lo cual aprovecha los resultados de esa investigación en cuanto a mecanismos de visualización, navegación y consultas. Al igual que en el Biovisualizador, *Den3D* contempla tres niveles de información (ver Figura 4). El primero es a nivel del árbol, donde se puede visualizar una imagen representativa del dato y al pasar el “mouse” encima del nodo, el software despliega una descripción (*tooltip*) que lo identifica así como metadatos tales como el número de descendientes de ese nodo. Al hacer “doble click” en el nodo se despliega la información de segundo nivel (ficha) que contiene más datos tales como el nombre científico, una imagen, la descripción de la especie y un enlace hacia la información más detallada. El tercer nivel consiste en un acceso a un sitio web de alguna organización sobre biodiversidad – externa a *Den3D* – en el que se accede a mayor detalle de información.



Figura 4: Los tres niveles de información.

Como se indicó de manera resumida en la Sección Introducción, *Den3D* se diferencia de otros sistemas porque incorpora funcionalidad adicional (más allá de la visualización) y un enfoque más robusto desde el punto de vista ingenieril. *Den3D* se desarrolló con una arquitectura cliente/servidor basada en tecnologías web, la cual distribuye la carga de procesamiento entre el cliente y el servidor. La aplicación se instala en un servidor, se puede ejecutar desde cualquier navegador de Internet, y pueden acceder múltiples usuarios simultáneamente. Además, en *Den3D* los datos se manejan mediante un sistema de bases de datos. Se seleccionó MongoDB, base de datos NoSQL, flexible, gratuita, escalable y que utiliza documentos en notación JSON, la cual es frecuentemente utilizada por organizaciones globales para el intercambio de datos de biodiversidad, tales como Catalogue of Life, EOL y GBIF. Asimismo, se diseñó e implementó un módulo para cargar (importar) datos provenientes de bases de datos globales, el cual aporta gran flexibilidad para la alimentación del software. También se refinaron los algoritmos para posicionamiento y despliegue de los nodos con lo cual se logró un mejor aprovechamiento del espacio y se evitan traslapes cuando el árbol tiene más de tres niveles. Además, una de las contribuciones más importantes de *Den3D* es la capacidad de edición de árboles y el manejo de bosques (conjuntos de árboles) para que los usuarios puedan crear y editar sus propias taxonomías.

3. Metodología

Esta investigación en Computación se enmarca en la categoría “diseño y modelamiento”, la cual según [3] representa aproximadamente el 70% de las investigaciones publicadas en Ciencias de la Computación. En este apartado presentamos la metodología usada para diseñar y modelar una solución para el problema de visualización de información voluminosa con estructura jerárquica en el dominio de las taxonomías biológicas y también la metodología aplicada para validar el modelo con usuarios.

a) Metodología para diseñar y modelar una solución.

La metodología para diseñar y modelar una solución se basó en el desarrollo iterativo e incremental y por prototipos [4]. Con este enfoque se identifican los componentes medulares, se hacen prototipos, los cuales se discuten y validan para precisar las

características del software. Los prototipos se refinan en una siguiente iteración y el proceso se repite hasta obtener una versión satisfactoria. En este proceso, la evaluación constante por parte del equipo de trabajo es vital para definir el rumbo que toma la investigación. En nuestro caso, la validación se llevó a cabo mediante reuniones semanales de revisión y avance. Los pasos seguidos son los siguientes:

1. Investigación del estado del arte en visualización de información con estructura jerárquica.

Se hizo una revisión bibliográfica en el tema de visualización de información con estructura jerárquica y se elaboró un reporte técnico.

2. Selección de tecnologías a utilizar que mejor se adapten a los objetivos del proyecto para desarrollar el ambiente experimental.

Se hizo un análisis de tecnologías actuales para determinar cuáles eran útiles para cumplir con los objetivos del proyecto. Para lograr que *Den3D* se pueda implementar de manera económica y en múltiples ambientes, se dio prioridad a herramientas gratuitas, open-source y multiplataforma.

3. Análisis del Biovisualizador, tomando en cuenta la documentación disponible y el software ejecutable para iniciar el proceso de ingeniería reversa.

El Biovisualizador, único software ejecutable y disponible para visualizar árboles de conos, representa un insumo fundamental para el proyecto, por lo cual era importante analizar los resultados de esa investigación y tomarlo como base para establecer los requerimientos de visualización.

4. Diseño de la arquitectura base del software.

Una vez entendidos los requerimientos de visualización, se procedió con el diseño de la arquitectura base para iniciar con el proceso iterativo e incremental para el modelamiento de una solución de visualización.

5. Proceso iterativo e incremental de modelamiento de la solución de visualización:

Esta etapa de la metodología se repite cuantas veces sea necesario, hasta lograr un resultado satisfactorio. Consiste de los siguientes pasos:

- a. Ajustes al diseño (de ser necesario)
- b. Programación de prototipo
- c. Preparación de datos de prueba
- d. Validación de prototipo, identificación de necesidades y generación de propuestas para modelar la solución

6. Proceso iterativo e incremental de modelamiento de la solución de edición:

- a. Diseño (ajustes al diseño) de la solución de edición
- b. Programación de prototipo
- c. Preparación de datos de prueba
- d. Validación de prototipo, identificación de necesidades y generación de propuestas para modelar la solución

7. Modelamiento de la solución de análisis.

Para modelar la solución para hacer análisis de información con *Den3D* se consideró lo siguiente: el tipo de usuarios potenciales (público en general, docentes, estudiantes, taxónomos), las posibilidades de análisis que ofrecen los datos que maneja *Den3D*, el análisis que podría hacer un usuario con las herramientas de navegación para distinguir visualmente, por ejemplo, diferencias de tamaño entre subárboles, la generación de estadísticas (por ejemplo, cantidad de familias, géneros, especies), las operaciones de búsqueda de información como herramientas de análisis y la comparación de taxonomías como una de las formas de análisis más relevantes.

Es importante mencionar que la preparación de datos de prueba involucró la generación de algoritmos para la generación de datos al azar con el objetivo de probar los resultados con diferentes volúmenes de datos y determinar límites que deberían ser superados en una siguiente iteración. Asimismo, involucró la generación de un componente de software para importar datos provenientes de bases de datos globales. Este módulo se diseñó con el objetivo de usar los web services de Catalogue of Life y EOL (Encyclopedia of Life). Estas dos iniciativas mundiales tienen cantidades masivas de información taxonómica. Por ejemplo, Catalogue of Life da acceso gratuito a la taxonomía biológica digital más completa del mundo y EOL brinda acceso – gratuito también – a casi 4 millones de fotografías. Las taxonomías globales que se importaron para hacer pruebas tanto del módulo de importación como de *Den3D* fueron: Mamíferos del mundo (5,825 especies, Catalogue of Life), anfibios del mundo (6,439 especies, Catalogue of Life), y Anfibios de CR (201 especies, INBio). Además, utilizando las funciones de edición de *Den3D* y el generador de taxonomías al azar, se crearon jerarquías al azar (de 150 a 6,500 nodos, árboles de 5 niveles con una cantidad variable de hijos entre 5 y 9), una taxonomía vernácula de anfibios de CR (201 especies), y una taxonomía de Pokémon para niños (108 especies).

Para la validación de los prototipos se siguió la metodología de reuniones semanales del equipo de trabajo, en las cuales se revisaron aspectos de corrección, calidad de la visualización estática (render), calidad de la visualización dinámica (animaciones suaves) y de rendimiento.

b) Metodología para validar el modelo con usuarios.

La validación de *Den3D* en la temática de biodiversidad se llevó a cabo de la siguiente manera:

1. Selección de población para el estudio.

Dado el potencial en educación de *Den3D*, se revisaron los programas de estudio en Biología de la Educación Diversificada del Ministerio de Educación Pública para constatar que la Unidad III trata los temas de “Evolución y Biodiversidad” y el cual incluye “La Biodiversidad y los Inventarios de las Formas de Vida” donde se trata la clasificación de los seres vivos. Por lo tanto, determinamos como la población meta para el estudio a los docentes de Ciencias y Biología de secundaria. Se hizo una invitación a profesores a través de contactos personales y de una red social. Participaron de forma voluntaria 12 profesores provenientes de 6 centros educativos, principalmente de la provincia de Cartago.

2. *Diseño de un instrumento de evaluación.*

Se llevó a cabo el diseño de un instrumento de evaluación con el propósito de obtener retroalimentación por parte de docentes de Biología y Ciencias con respecto a la utilidad del ambiente *Den3D* para visualizar y analizar información con estructura jerárquica, así como para desarrollar actividades de aprendizaje en temas de Biodiversidad. El instrumento estaba conformado por 10 preguntas, 8 de las cuales con una escala de valoración de 1 a 5, 1 pregunta de respuesta Sí/No y 1 pregunta abierta donde se solicitaron sugerencias.

3. *Taller de capacitación de usuarios en el uso de Den3D.*

Se llevó a cabo un taller de capacitación en el uso de *Den3D* para los docentes participantes. El taller tuvo una duración de 4 horas. Se diseñó una guía de trabajo. El taller fue práctico, con un enfoque de “aprender haciendo”.

El taller se llevó a cabo en un laboratorio de computadoras, de tal manera que cada docente tenía asignada una computadora para trabajo individual.

La primera parte del taller consistió de una introducción a la herramienta. Mediante una sesión guiada los docentes experimentaron cargar un árbol, rotar el árbol en diferentes direcciones (arriba, abajo, derecha, izquierda), alejar y acercar el árbol, contraer y expandir subárboles, mostrar la ficha de una especie, hacer búsquedas por palabras clave, marcar organismos y generar reportes. En la segunda parte del taller los docentes realizaron ejercicios para experimentar individualmente el uso de las funciones de navegación y consulta. Para ello se utilizó la base de datos de anfibios de Costa Rica. La tercera parte del taller consistió en una sesión guiada para crear y editar taxonomías. Los docentes buscaron imágenes de Internet e hicieron sus propias clasificaciones. En la cuarta parte, los docentes trabajaron en grupos de 2 personas para diseñar una actividad de aprendizaje en temas de biodiversidad utilizando la herramienta *Den3D*. En esta última actividad cada grupo hizo una discusión de ideas, un diseño de la actividad de aprendizaje, la creación de la actividad utilizando *Den3D* y utilizando datos de la base de datos de anfibios y datos propios. El taller concluyó con la presentación de la actividad de aprendizaje al resto del grupo.

4. *Aplicación y procesamiento del instrumento de evaluación.*

Al finalizar el taller se aplicó el instrumento de evaluación. El procesamiento se hizo utilizando Microsoft Excel para obtener cantidad y porcentaje de respuestas para cada pregunta según la escala de valoración.

4. Resultados

Se desarrolló *Den3D* v1.0, un ambiente experimental en tres dimensiones con árboles de conos para la visualización, la edición y análisis de grandes taxonomías de especies. *Den3D* es un ambiente novedoso ya que, de acuerdo con la revisión del estado del arte, podemos afirmar que actualmente las herramientas de administración de información jerárquica tienen una o más de las siguientes limitaciones: se limitan a la visualización (no incluyen edición), no manejan grandes volúmenes de datos, no están disponibles vía web, usan software propietario y no intercambian web services con otros sitios. Se obtuvo *Den3D*, un ambiente basado en herramientas gratuitas, open-source y multiplataforma que no tiene ninguna de esas limitaciones. Más específicamente, se obtuvo lo siguiente:

- a) El diseño e implementación de una arquitectura de software basada en tecnologías web que permite la distribución de la carga de procesamiento entre el cliente y el servidor. Después de estudiar el estado del arte y las herramientas disponibles para el diseño e implementación de *Den3D*, se optó por: Three.js para la programación de la visualización de las jerarquías en tres dimensiones, MongoDB para el manejo de las bases de datos, y Node.js para ejecutar los cálculos de los ángulos y las posiciones de los conos en el servidor web con el objetivo de reducir la carga de trabajo del lado del cliente. Node.js tiene las condiciones para implementar micro servicios con el propósito de hacer más eficiente la ejecución asíncrona del código. Aunque para la versión actual de *Den3D* no se implementaron micro servicios, es importante contar con esa capacidad en caso de que en un futuro se tenga un número alto de usuarios concurrentes. La herramienta three.js demostró su flexibilidad y eficiencia para la implementación de ambientes gráficos en 3D basados en el web. Se pudo desplegar y manejar adecuadamente árboles de hasta 4,500 especies. MongoDB resultó una muy buena opción como administrador de base de datos por su flexibilidad para relacionar diferentes tipos de datos. Esta base de datos es popular por su capacidad de crecimiento por lo que será útil a futuro para implementar bases de datos más grandes.
- b) El desarrollo de un algoritmo para el manejo de datos a partir de la carga de datos provenientes de bases de datos globales, lo cual flexibiliza la alimentación de datos al software. Los datos se almacenan en bases de datos que trabajan con archivos en formato JSON, aprovechando los estándares para el intercambio de datos a partir de bases de datos de biodiversidad.
- c) El diseño y desarrollo de la visualización de árboles en tres dimensiones. En relación con otros trabajos, los algoritmos implementados evitan el traslape de nodos cuando

se despliegan árboles de más de tres niveles, manteniendo la razonabilidad en el aprovechamiento del espacio en pantalla.

d) El desarrollo de herramientas para la navegación, inspección y consulta del árbol, tales como:

- rotar el árbol en diferentes direcciones
- acercar y alejar el árbol
- traer al frente un nodo
- comprimir (esconder) un subárbol para dejar visibles las partes del árbol que sean de interés
- expandir (mostrar) un subárbol que estuviera comprimido
- mostrar la ficha de un organismo
- buscar una especie por nombre común o por nombre científico
- marcar nodos y dejarlos en una lista de nodos de interés
- generar un reporte con información a partir de los nodos marcados
- variar el nivel de transparencia de los conos, lo cual ayuda a definir visualmente diferentes partes del árbol

e) Algoritmos para visualizar árboles grandes, de aproximadamente 5 mil nodos con un rendimiento en la carga y despliegue razonable.

Se implementó algoritmos y pruebas para la generación aleatoria de árboles grandes y así probar la capacidad de visualización y de rendimiento de la herramienta. Las pruebas iniciaron con aproximadamente 150 nodos hasta llegar a aproximadamente 7 mil nodos. Se fue haciendo ajustes a los algoritmos para mejorar el rendimiento.

En el caso de las bases de datos reales, las pruebas con la mayor cantidad de datos se hicieron con mamíferos obtenidos de Catalogue of Life. Esta base de datos está conformada por 5,825 especies de mamíferos del mundo y genera un árbol de más de 6 mil nodos.

Consideramos que el caso de generación del árbol con 4,375 nodos tuvo un rendimiento razonable, principalmente considerando el tiempo inicial de carga de la estructura, el cual fue de aproximadamente 30 segundos.

f) El desarrollo de herramientas para la creación y edición de árboles:

Para la creación y edición de árboles destaca el modelo para el manejo de bosques, ya que la edición produce que naturalmente se tengan que manejar varios árboles a la vez.

Para la creación de un árbol o de un subárbol se diseñó un mecanismo ágil mediante el cual los usuarios pueden indicar la cantidad de hijos del árbol, de tal manera que, si se quisiera, se puede crear rápidamente la topología.

El diseño incluyó el manejo de las siguientes operaciones:

- hacer un árbol nuevo (desde cero)
 - actualizar los datos de un nodo
 - insertar un nodo nuevo (ya sea la inserción de un nodo hijo a un subárbol que estaba vacío o la inserción cuando el subárbol tenía hijos, en este último caso se asume que la inserción se va a hacer hacia la derecha de uno de uno de los nodos hijos)
 - enlazar un árbol existente (que está en el bosque) como hijo de un nodo del árbol
 - podar un árbol (el subárbol que se corta pasará automáticamente al área de bosque)
 - recortar un árbol (el subárbol seleccionado permanece y el resto del árbol se elimina)
 - borrar un nodo
 - deshacer la última operación
 - intercambiar el árbol que está en el área principal de trabajo por un árbol del bosque
- g) La exploración de condiciones y alternativas para llevar a cabo análisis de información. Se planteó estrategias para hacer análisis de información mediante trabajo colaborativo, haciendo algunos ajustes a futuro en el software desarrollado. Se llegó a la conclusión de que el análisis más relevante para taxonomías biológicas es la comparación visual de diferentes versiones de una taxonomía, tema que se sale del alcance del presente proyecto pero que dio pie para plantear un nuevo proyecto de investigación con vigencia y relevancia mundial.
- h) Documentación técnica de diseño, instalación y operación del software. Ver apéndices “Documento de Diseño” y “Documento de Instalación y Operación”.

- i) El desarrollo de un estudio con usuarios para validar el ambiente para la temática de biodiversidad:

El estudio se llevó a cabo con 12 docentes de secundaria del área de Ciencias pertenecientes a 6 instituciones educativas, en su mayoría de la zona de Cartago, quienes participaron de forma voluntaria. En el apéndice “Estudio de Usabilidad de *Den3D*” se describe el estudio y se presentan los resultados, y a continuación se presenta un resumen:

- El 100% de los usuarios indicó que el ambiente *Den3D* es bueno para visualizar información con estructura jerárquica.
- El 91,7% manifestó estar “muy de acuerdo” y el 8.3% indicó estar “de acuerdo” con que para el caso de las taxonomías de especies que se requieren para la enseñanza en secundaria, la visualización en tres dimensiones con *Den3D* es una mejor opción que la visualización en dos dimensiones.
- El 75% indicó estar “muy de acuerdo” y el 25% indicó estar “de acuerdo” con que el ambiente *Den3D* es efectivo como herramienta de aprendizaje en temas de biodiversidad.
- El 58.3% indicó estar “muy de acuerdo” y el 41.7% indicó estar “de acuerdo” con que *Den3D* se puede utilizar como herramienta para propiciar el análisis de la información de parte de los usuarios (estudiantes).
- El 66.7% manifestó estar “muy de acuerdo”, el 25% manifestó estar “de acuerdo” y el 8.3% indicó estar conforme (ni de acuerdo ni en desacuerdo) con respecto a que La cantidad de información que se puede visualizar en la pantalla es suficiente con relación al tamaño de las taxonomías que se estudian en secundaria.
- El 50% indicó estar “muy de acuerdo”, el 41.7% manifestó estar “de acuerdo” y el 8.3% indicó estar conforme (ni de acuerdo ni en desacuerdo) con que *Den3D* es fácil de usar.
- El 83.3% indicaron estar muy satisfechos con *Den3D* y el 16.7% indicaron estar satisfechos.
- El 100% de los usuarios indicaron que sí recomendarían *Den3D* a sus colegas.

Den3D modela e implementa 23 casos de uso para cargar datos, visualizar, navegar, consultar y editar taxonomías de especies. En el “Documento de Diseño” se presentan con detalle y ejemplos el funcionamiento. El Documento de Diseño también se describe la arquitectura y el modelo para el manejo de datos.

En la Figura 5 se presenta una imagen de la pantalla principal de *Den3D*. Nótese las tres áreas A, B y C. El área A corresponde al espacio “de trabajo” donde se puede visualizar, navegar y editar el árbol principal con que se está trabajando. El área B corresponde al menú principal. En el área C – que denominamos “el bosque” – se mantienen árboles de trabajo, los cuales en la mayoría de los casos se generan durante el proceso de edición. La edición de árboles y el manejo de bosques son herramientas novedosas para el manejo de jerarquías.



Figura 5: Ventana principal de *Den3D*

Como se indicó anteriormente, la herramienta presenta la información en tres niveles de detalle. La Figura 6 presenta un ejemplo donde se muestra el primer nivel de información (el árbol completo, en este caso es un árbol pequeño con solamente un cono) que indica la cantidad de descendientes y la ficha técnica (segundo nivel de información).

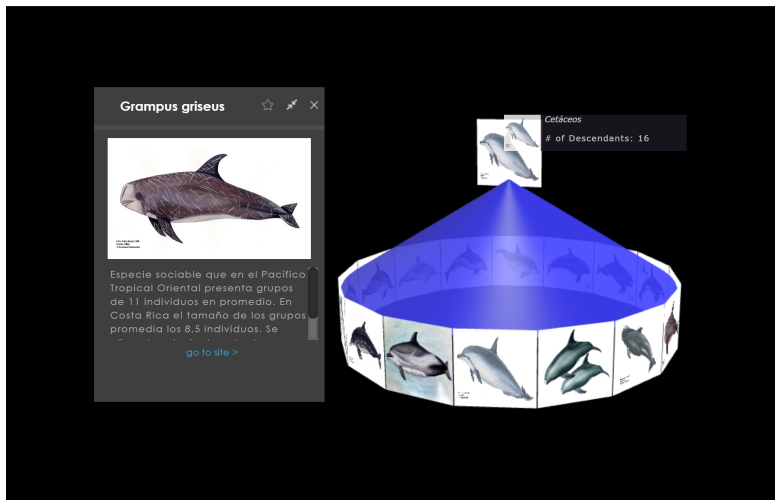


Figura 6: Ejemplo con primer y segundo nivel de información.

Las figuras 7 y 8 presentan un ejemplo que ilustra la creación de un árbol nuevo. Como se puede observar en la Figura 7, el usuario indica el nombre del árbol, una descripción y la dirección url donde se podrá consultar la información de tercer nivel (por ejemplo, EOL). Asimismo se indica la cantidad de hijos con que se creará el árbol. La Figura 8 muestra el árbol creado con 5 hijos sin información. Posteriormente, mediante la operación de actualización, el usuario podrá agregar la información específica para cada nodo.

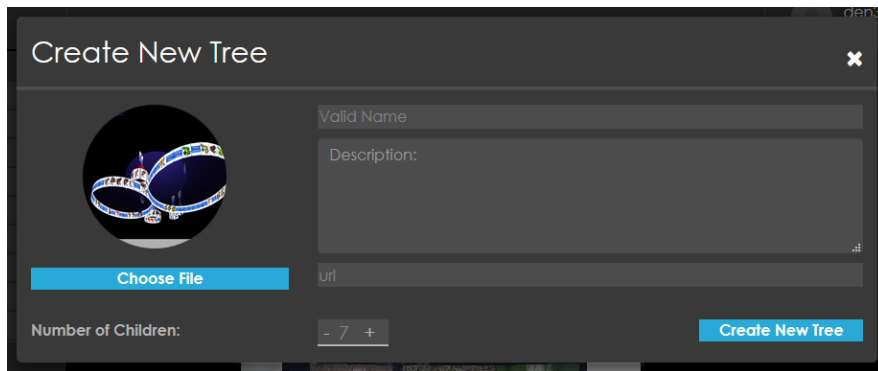


Figura 7: Ventana para indicar la creación de un árbol nuevo.

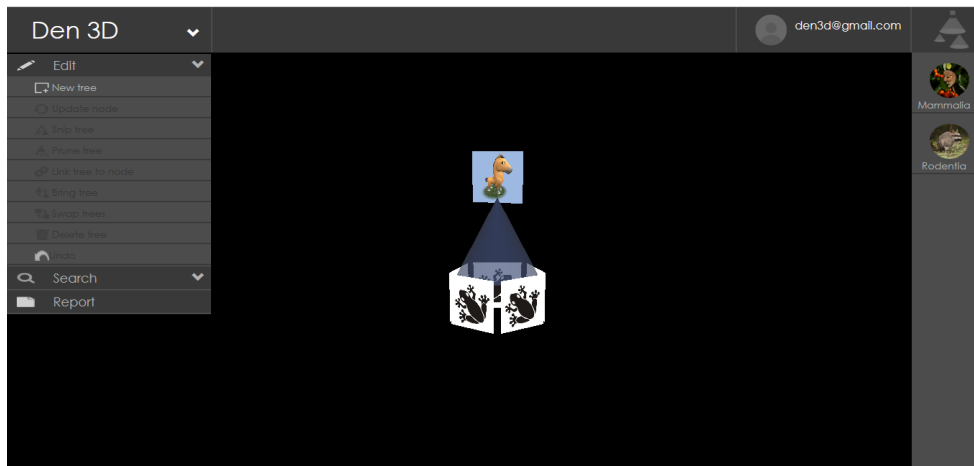


Figura 8: Arbol creado con 5 hijos que todavía no tienen información.

Finalmente, como resultado de esta investigación también obtuvimos los siguientes productos (ver apéndices e informes previos con detalles) en eventos y revistas internacionales:

Nombre de obra	Tipo de obra	Estado (aceptado, por publicar y publicado)	Base de datos de indexación	Nombre de Evento	Contó con Comité científico (Sí ó NO)
Visual Comparison of Biological Taxonomies: A tasks characterization (ver apéndice G)	Poster	Aceptado y presentado	–	Eurographics Conference on Visualization, EuroVIS 2016, The Netherlands.	Sí
Den3D: A Ludic 3D Visualization and Learning Tool for Biological Taxonomies (ver apéndice H)	Demo	Aceptado y presentado	–	World Information Technology Forum, WITFOR 2016, San José, Costa Rica.	Sí
Comparación visual de grandes jerarquías: El caso de las taxonomías biológicas (ver apéndice I)	Ponencia	Aceptado y presentado	–	Tercer Encuentro Bienal Centroamericano de Investigación y Posgrado, Ciudad de Guatemala, Guatemala.	Sí
Visual comparison of biological taxonomies	Ponencia	Aceptado y presentado	–	TDWG 2016 Annual Conference, Santa Clara, San Carlos, Costa Rica.	Sí

Den3D: A 3D environment for the visualization, creation and edition of hierarchically structured biological information	Ponencia	En preparación	IEEE Explorer	XLIII Conferencia Latinoamericana de Informática, CLEI 2017.	Sí
Visualization approaches for the comparison of biological taxonomies (ver apéndice K)	Artículo	Aceptado	DOAJ	RIO, Research Ideas and Outcomes.	Sí

Adicionalmente, gracias a la colaboración del programa eScience, al cual está adscrito el proyecto, la Oficina de Comunicación y Mercadeo elaboró un video que describe el proyecto (ver: <https://www.youtube.com/watch?v=cdbDvU8WF6I>).

Finalmente, con el fin de divulgar promover el uso de Den3D, se elaboró un afiche que se incluye como apéndice N en este informe.

5. Discusión y conclusiones

En términos generales, consideramos que los resultados demuestran el cumplimiento de los objetivos planteados. Se desarrolló una herramienta novedosa de visualización de jerarquías en tres dimensiones, con capacidad para visualizar una cantidad grande de datos (aproximadamente 5 mil nodos) con un rendimiento muy aceptable y los usuarios que valoraron la herramienta manifestaron tener una opinión muy favorable respecto a la capacidad del software y a su potencial para ser utilizado con fines educativos en temas de biodiversidad. Además, se basa en herramientas gratuitas, open-source y multiplataforma, lo cual amplifica su potencial impacto.

Como lo indicamos en la sección anterior, de acuerdo con la revisión del estado del arte, se puede afirmar que las herramientas existentes para gestionar jerarquías tienen una o más de las siguientes limitaciones: solo visualizan información (no incluyen edición), no manejan grandes volúmenes de datos, no están disponibles vía web, usan software propietario y no intercambian web services con otros sitios. *Den3D* es un ambiente basado en herramientas gratuitas, open-source y multiplataforma que no tiene ninguna de esas limitaciones.

Los algoritmos desarrollados para la visualización de los árboles logran desplegar adecuadamente la información en la pantalla. Desde el punto de vista gráfico, los árboles están bien contruidos, no ocurre traslape de nodos, se pueden visualizar árboles de N niveles, las herramientas de navegación permiten que el usuario pueda inspeccionar fácilmente el árbol, cuando ocurren cambios en el árbol se mantiene información de contexto para que el usuario no se pierda y se utilizan animaciones para suavizar la transición entre estados de un árbol (por ejemplo, al contraer o expandir un subárbol).

La creación y edición de árboles grandes se logra mediante mecanismos sencillos para la creación de la topología del árbol. La metáfora del bosque como repositorio de árboles que apoyan los procesos de creación y edición resultó natural en las pruebas de usabilidad.

Es importante reconocer que el modelo desarrollado aún no implementa suficientes facilidades para llevar a cabo análisis de información, y así consta en la valoración que hicieron los usuarios durante el estudio de valoración de la herramienta. Las facilidades para llevar a cabo análisis de información taxonómica podrían enfocarse de tres maneras:

- a) que el usuario utilice las operaciones de navegación y consulta para hacer visualmente su propio análisis con las facilidades que el software ofrece en la actualidad,
- b) agregar a *Den3D* funciones para trabajo colaborativo de manera tal que los usuarios puedan compartir taxonomías que pudieran comparar visualmente, y
- c) generar herramientas para la comparación automática de múltiples taxonomías. Este es el caso que consideramos más relevante para futura investigación.

El estudio con 12 usuarios es restringido, pero obedece al interés de los docentes de participar en forma voluntaria y a los recursos con que contaba el proyecto.

6. Recomendaciones

- a) Desarrollar un proyecto de extensión con el cual se pueda evaluar la aplicabilidad de los productos de este proyecto a una mayor escala, particularmente en el estudio de la rica biodiversidad de CR en la educación secundaria. Conversaciones preliminares con la Fundación Omar Dengo (Dra. Natalia Zamora, co-Directora de PRONIE) sugieren que se puede transferir productivamente la tecnología asociada a *Den3D* a docentes y estudiantes en ambiente de trabajo colaborativo.
- b) Desarrollar futuros estudios con usuarios que permitan obtener retroalimentación que enriquezca la investigación en análisis de información con estructura jerárquica.
- c) Agregar las capacidades para implementar un ambiente colaborativo que permita a los docentes de diferentes comunidades compartir taxonomías y hacer análisis colaborativo de la información.

- d) Ofrecer web services en *Den3D* para que sea usado como interfaz en iniciativas mundiales tales como EOL, GBIF, Catalogue of Life y otras que brindan información taxonómica pero mediante interfaces en 2D poco atractivas para el público en general.

7. Agradecimientos

Agradecemos a:

- El Instituto Nacional de Biodiversidad, en particular a su directora de TI, María Auxiliadora Mora, por su valiosa colaboración con las bases de datos de anfibios y mamíferos de Costa Rica.
- Los estudiantes asistentes Ricardo Brenes Camacho, Melvin Sandoval Quirós, Eros Hernández Romero, Silvia Alpízar Rivera y Kevin Ramírez Vega por su valiosa colaboración en las labores de programación, pruebas y documentación en las diferentes etapas del desarrollo del proyecto. A los estudiantes Melvin Sandoval Quirós y Kevin Ortega Núñez así como a la profesora Ing. María Estrada Sánchez, MSc. por su colaboración a través del curso IC-7841- Proyecto de Ingeniería de Software, en el cual se desarrolló el primer prototipo funcional para el módulo Editor de Den3D.
- Los docentes Olger Alpízar Montero, Isabel Arguedas Fernández, Katherine Granados Calderón, Graciela Lobo Salazar, Cristian Marrero Solano, Dayana Mora Fallas, Ingrid Navarro Rivera, Melissa Ortiz Navarro, Gabriela Robles Camacho, Nuria Rojas Cordero, Marcela Venegas Leiva y Jenny Villalobos por su valiosa participación en el estudio de valoración de *Den3D*.
- La Vicerrectoría de Investigación y Extensión del TEC por el apoyo y el financiamiento para llevar a cabo esta investigación.

8. Referencias

- [1] F. Hernández-Castro and J. Monge-Fallas, “Visualización Tridimensional de Estructuras Jerárquicas,” MSc Thesis, 2006. ITCR.
- [2] W. F. Tichy, P. Lukowicz, L. Prechelt, and E. A. Heinz, “Experimental evaluation in computer science: A quantitative study,” *J. Syst. Softw.*, vol. 28, no. 1, pp. 9-18, Jan. 1995.
- [3] J. Wainer, C. G. Novoa Barsottini, D. Lacerda, and L. R. Magalhães de Marco, “Empirical evaluation in Computer Science research published by ACM,” *Inf. Softw. Technol.*, vol. 51, no. 6, pp. 1081-1085, Jun. 2009.
- [4] I. Sommerville, *Software Engineering*, 9th ed. Addison-Wesley, 2010.
- [5] T. H. Cormen, C. Stein, R. L. Rivest, and C. E. Leiserson, *Introduction to Algorithms*, 2nd ed. McGraw-Hill Higher Education, 2001.
- [6] J. Kleinberg and E. Tardos, *Algorithm Design*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005.
- [7] M. Graham and J. Kennedy, “A survey of multiple tree visualisation,” *Inf. Vis.*, vol. 9, no. 4, pp. 235-252, 2010.
- [8] S. Jurgensmann and H. J. Schulz, “A Visual Survey of Tree Visualization,” *Proc. IEEE Inf. Vis. (Salt Lake City, USA, 2010)*, IEEE Press, 2010.
- [9] G. G. Robertson, J. D. Mackinlay, and S. K. Card, “Cone Trees: Animated 3D Visualization of Hierarchical Information,” in *CHI*, 1991, pp. 189-202.
- [10] F. Hernández-Castro, E. Mata- Montero, and J. Monge-Fallas, “Biovisualizador: Visualizando los anfibios de Costa Rica,” *Rev. Tecnol. en Marcha*, vol. 22, no. 1, p. 15, 2009.

APENDICES

- A. Informe de Avance I semestre, 2015.
- B. Informe de Avance Anual 2015.
- C. Informe de Avance I semestre, 2016.
- D. Reporte Técnico: Estado del Arte de la Visualización de Información Voluminosa con Estructura Jerárquica.
- E. Documento de Diseño.
- F. Documento de Instalación y Operación.
- G. Artículo: Visual Comparison of Biological Taxonomies: A task characterization, poster (EuroVIS).
- H. Artículo: Den3D: A Ludic 3D Visualization and Learning Tool for Biological Taxonomies, demo (WITFOR)
- I. Artículo: Comparación visual de grandes jerarquías: El caso de las taxonomías biológicas (Bienal)
- J. Simposio: Visual Comparison of Biological Taxonomies (TDWG)
- K. Artículo: Visualization approaches for the comparison of biological taxonomies (RIO)
- L. Estudio de Usabilidad de Den3D (incluye: Programa Taller de Visualización y Análisis de Taxonomías Biológicas Usando el Software Tridimensional Den3D, Invitaciones al Taller, Guía de Trabajo para el Taller Visualización y Análisis de Taxonomías Biológicas haciendo Uso del Software Tridimensional Den3D Dirigido a Docentes del Area de Ciencias, Instrumento de Evaluación del Ambiente Den3D, Resultados del Estudio de Usabilidad de Den3D, Lista de Participantes y Fotografías del Taller).
- M. Anuncio Charla ProIncomp
- N. Afiche para divulgar y promover Den3D