

Representación gráfica de la taxonomía para especies del Herbario Nacional

José Gómez Alfaro¹, César Retana Jiménez², and Antonio González Torres³

Escuela de Ingeniería,
Universidad Latinoamericana de Ciencia y Tecnología
ULACIT, Urbanización Tournón, 10235-1000
San José, Costa Rica
José Gómez Alfaro, jgomeza826@ulacit.ed.cr
César Retana Jiménez, cretanaj474@ulacit.ed.cr
Antonio González Torres, agonzalez@ulacit.ac.cr
<http://www.ulacit.ac.cr>

Abstract. El Herbario Nacional de Costa Rica se ha encargado de crear y conservar registros de las diferentes colecciones de especies de la flora del país, con el objetivo de compartir los resultados y las colecciones de especies al público. Por medio del sistema web Ecobiosis se logra publicar dicha información, sin embargo realizar búsquedas y extraer información se vuelve una tarea difícil para aquellos usuarios que no son expertos en la taxonomía de las especies. Por medio del estudio de herramientas de software actual se realizarán comparaciones de sus características, estos resultados se tomarán como base para realizar el diseño y creación de una herramienta de búsqueda dinámica e interactiva de analítica visual, que mejore la experiencia de los usuarios y les permita obtener un conocimiento general sobre biodiversidad de Costa Rica. Este artículo tiene como objetivo investigar y seleccionar el software más adecuado para extraer los registros de colecciones de especies de la base de datos central del Herbario Nacional y el desarrollo de una herramienta web de analítica visual que mejore las experiencias de los usuarios al navegar en un espacio interactivo.

Keywords: Taxonomía, Analítica Visual, Visualización de la Información, Herbario, Evolución de Software

1 Introducción

El Herbario Nacional (HN) es una sección del Departamento de Historia Natural del Museo Nacional de Costa Rica (MNCR) que se dedica a investigar la flora del país. Entre los objetivos del HN se encuentran desarrollar y mantener una colección representativa y actualizada de las especies de la flora del país, generar las bases de datos apropiadas, divulgar los resultados de las investigaciones y proporcionar acceso a las colecciones y su información a la comunidad académica nacional e internacional.

Representación gráfica de la taxonomía para especies del Herbario Nacional

En la actualidad las colecciones del HN cuentan con un gran número de ejemplares que son consultados de forma frecuente por expertos nacionales e internacionales. La información de estas colecciones se encuentra disponible para el público en general a través de un sistema web denominado Ecobiosis (<http://ecobiosis.museocostarica.go.cr/>). Este sistema es el producto de más de 15 años de trabajo del personal del HN en la digitalización de cerca de 350.000 registros de las colecciones, así como de su diseño e implementación.

Ecobiosis es un sistema web de gran atractivo visual, fácil de utilizar y que ofrece posibilidades búsqueda e interacción. Pero en esencia es un sistema bien diseñado, de acuerdo con los principios del diseño web contemporáneo, que busca ofrecer información a los investigadores. Por lo que su uso por parte de los usuarios que no son expertos en la material, se ve limitado por la falta de pericia para realizar búsquedas adecuadas, mediante las opciones de búsqueda regular o avanzada. Ello implica que la riqueza de la información a la cual proporciona acceso el sistema permanece oculta en gran medida.

En consecuencia con lo anterior, este trabajo de investigación tiene como objetivo diseñar e implementar la arquitectura de una herramienta analítica visual para obtener conocimiento general y específico sobre la biodiversidad de Costa Rica, con base en las colecciones del HN y tomando en cuenta la información taxonómica y geográfica (e.g. latitud, longitud y altura) de las diferentes especies, así como las necesidades de información de estudiantes, profesores e investigadores. Lo anterior con el fin de permitir la obtención de conocimiento, mediante el uso de representaciones visuales y técnicas de interacción, tomando en cuenta sus capacidades para hacer visible lo invisible (J. J. Thomas & Cook, 2005).

Los pasos para cumplir el propósito de esta investigación inician con el análisis de las colecciones y los datos que posee la base de datos del HN, luego un estudio de tecnologías que cumplen la misma función que Ecobiosis, posteriormente se investigaran las tecnologías más adecuadas para llevar a cabo el desarrollo de la arquitectura, el diseño de las visualizaciones, implementación de la arquitectura y prueba de la herramienta.

2 Estado del arte

2.1 Taxonomía

Según (Sharma, 1993), la taxonomía es la ciencia que estudia la clasificación de los organismos de acuerdo con sus semejanzas y diferencias, además establece la clasificación, identificación, descripción y nomenclatura como los componentes básicos de la taxonomía que permiten determinar similitudes y realizar comparaciones entre dos especímenes. Los científicos del HN han realizado un gran esfuerzo para registrar las colecciones que poseen actualmente, respetando los niveles jerárquicos de cada taxón. La 1 muestra las categorías de taxones.

Representación gráfica de la taxonomía para especies del Herbario Nacional

Rangos de taxones	Finales de rangos por encima de género	Ejemplos
División	phyta	Pterophyta, Magnoliophyta
Subdivisión	phytina	Pterophytina
Clase	opsida	Pteropsida, Magnoliopsida
Subclase	opsidae	Pteropsidae
Familia	ales	Rosales, Asterales
Subfamilia	ineae	Rosineae
Tribe	aceae	Rosaceae, Asteraceae
Subtribu	oideae	Rosoideae
Género	eae	Roseae
Subgénero	inae	Rosinae
Sección	us, a, um, on, es,	Rosa, Pinus
Subsección	etc.	
Serie		
Subseries		
Especies		
Subespecies		
Variedad		
Subvariedad		
Forma		
Subforma		

Fig. 1. Clasificación taxonómica de las especies

2.2 Analítica visual

Hoy en día la cantidad de información se incrementa rápidamente; en comparación con la última década, los datos se almacenaban sin procesar o filtrar para ser usados. La mayoría de las industrias generaba enormes cantidades de datos, que no podían procesar para la toma de decisiones. Esto ocasiona una sobrecarga de información debido al riesgo de pérdida de esta por procesamiento inadecuado. Debido a la sobrecarga de información, se desperdician tiempo y esfuerzos valiosos por la falta de habilidad de no poder procesar los datos (Andrienko, Andrienko, & Wrobel, 2007).

En muchas industrias obtener la información, en el momento preciso, garantiza el éxito, pero el problema es cómo procesar los datos para que sean confiables y comprobables. Cada vez se implementan más herramientas automatizadas para el análisis de datos, las cuales solo sirven para problemas que están bien definidos y comprendidos. El proceso de tomar los datos para la toma de decisiones es en realidad bastante complejo, los métodos creados para estos análisis representan el conocimiento de sus creadores, si estos métodos están incorrectos, podrían afectar la exactitud de la información. Una de las principales metas de la analítica visual, es convertir el proceso de datos e información en algo transparente y que provea una mejora en el conocimiento y en las decisiones.

Representación gráfica de la taxonomía para especies del Herbario Nacional

Asimismo, la visualización de estos procesos permite que se tomen en cuenta, en lugar de quedarse con los resultados y con ello fomentar la evaluación, corrección y mejora de estos.

En este sentido, la analítica visual es una tecnología que combina las fortalezas humanas y el procesamiento electrónico de datos (Keim et al., 2008). La visualización se convierte en un medio semiautomático de analítica de procesos, en donde los seres humanos y las máquinas cooperan entre sí para lograr los mejores resultados. La investigación analítica visual combina diversas áreas de investigación relacionadas, tales como la visualización, minería, gestión y fusión de datos, además estadísticas, entre otros.

Al respecto, en (J. J. Thomas & Cook, 2005), Thomas y Cook se define la analítica visual como la ciencia del razonamiento analítico facilitado por interfaces visuales interactivas. Es una combinación de técnicas de análisis con visualizaciones interactivas para poder entender, razonar y decidir sobre grandes volúmenes de datos complejos (Keim et al., 2008).

Además, visualizar información es diferente de analítica visual; la visualización tradicional no necesariamente analiza datos o utiliza algoritmos avanzados. A cambio la analítica visual es un enfoque integral a la toma de decisiones, combinando visualización, factores humanos y análisis de datos. La clave es identificar el mejor algoritmo para el análisis y sus límites de lo que no puede automatizar.

Cabe destacar que unos de los problemas más comunes es la escalabilidad con el tamaño de los datos, cuanto más grandes sean son más difíciles de manejar, analizar y visualizar efectivamente. Dividir los datos en volúmenes más pequeños con la información más relevante, beneficia significativamente. La calidad de los datos, se puede ver afectada por una malinterpretación del analista, ocasionando resultados erróneos por elección de un algoritmo no adecuado o los datos propiamente suministrados. Para enfrentar este problema el usuario debe estar consciente de los datos y el análisis en cualquier etapa de este proceso.

Por lo demás, mostrar vastas cantidades de datos, no es una tarea fácil, se necesita seleccionar las visualizaciones adecuadas, que permitan enseñar la información relevante y el nivel de detalle necesario. Las visualizaciones deben ser intuitivas para que el usuario pueda entenderlas fácilmente con la menor ayuda posible. Al mismo tiempo, es importante la compatibilidad con diversos dispositivos como computadoras, dispositivos móviles entre otros, y adaptarse a las características de estos para visualizar los datos.

2.3 Herramientas de software

Para seleccionar una adecuada arquitectura y representar las especies de plantas del HN, es necesario estudiar el diferente software que existe actualmente, por lo cual se va recurrir a investigar herramientas tecnológicas que utilizan instituciones, universidades y organizaciones para realizar búsquedas y representar las colecciones biológicas que estas ofrecen al público. Considere el siguiente cuadro de software de las instituciones consultadas.

Estas instituciones ofrecen a los visitantes de sus sitios web, herramientas de búsqueda para las distintas especies que ellas poseen; en su mayoría ofrecen a

Representación gráfica de la taxonomía para especies del Herbario Nacional

Nombre	Localización	Especies	Lenguaje
Herbarium Hamburgense (Hamburgense, 2014)	Alemania	34923	PHP
Royal Botanic Garden Edinburgh (Edinburgh, 2014)	Escocia	15000	PHP
Consortium of North American Lichen Herbaria (of North American Lichen Herbaria, 2014)	North America	3.2 millones	PHP
Illinois State Museum (Museum, 2014)	Estado Unidos	110000	JavaScript
Sienet (Seinet, 2014)	Estado Unidos	-	PHP
Museo de Historia Natural de Florida (de Historia Natural de Florida, 2014)	Estado Unidos	585000	ASP.NET
Missouri Botanical Garden (Garden, 2014)	Estados Unidos	4 millones	ASP.NET
The Linnean Herbarium (T. L. Herbarium, 2014)	Francia	14000	-
University and Jepson Herbaria of the University of California at Berkeley (University & of the University of California at Berkeley, 2014)	Estado Unidos	2200000	-
Instituto de Botanica Darwinion (de Botanica Darwinion, 2014)	Argentina	150000	ASP.NET
Universidad Nacional de Colombia (de Colombia, 2014)	Colombia	940000	-
Jardín Botánico de Río de Janeiro (de Río de Janeiro, 2014)	Brasil	590000	PHP
Instituto de Botanica del Nordeste (de Botanica del Noreste, 2014)	Argentina	400000	-
Museo Ecuatoriano de Ciencias Naturales (de Ciencias Naturales, 2014)	Ecuador	671	-
Instituto Amazónico de Investigaciones Científicas Sinchi (de Investigaciones Científicas Sinchi, 2014)	Colombia	80000	-
Herbario Nacional de Bolivia (de Bolivia, 2014)	Bolivia	300000	-
Museo Nacional de Historia Natural Smithsonian (de Historia Natural Smithsonian, 2014)	Estados Unidos	57000	-
Harvard University (H. University, 2014)	Estado Unidos	430000	PHP
Red Tela Botánica (Botanica, 2014)	Europa	-	-
Museo Canadiense de la Naturaleza (de la Naturaleza, 2014)	Canada	1350000	-
E.C. Smith Herbarium (E. S. Herbarium, 2014)	Canada	550000	-

Table 1. Tabla de Software.

los usuarios páginas de tipo formulario donde se digita ya sea el nombre de una especie, género, familia, orden, clase, división o nombre científico; otros ofrecen búsquedas por localización, país o un espacio general donde se puede digitar cualquiera de los aspectos mencionados.

Al evaluar las diferentes tecnologías, se deben considerar algunos aspectos como la característica de poder adaptarse a cualquier medio o dispositivo; debe permitirse representar cualquier tipo de dato, tomando en cuenta que estos pueden variar durante el tiempo, cosa que no ocurre con estos software, debido a que las páginas fueron diseñadas para ser estáticas y no se ajustan a los distintos dispositivos; el software estudiado permite representar la información, aunque no de una forma amigable para el usuario, listas extensas de resultados de búsqueda dificultan la tarea de recuperar, analizar e interpretar las información, como en el caso de aquellos usuarios que no poseen conocimientos sobre las clasificación de las especies.

Estas organizaciones ofrecen, en sus sitios web, las grandes colecciones de datos sobre las especies que recopilan; en el caso de CNALH y Missouri Botanical

Representación gráfica de la taxonomía para especies del Herbario Nacional

Garden, estas ofrecen un portal de datos en una red distribuida de universidades, jardines botánicos, museos y agencias proveedoras de información taxonómica y ambiental, además son las que poseen mayor cantidad de especies; ello constituye una desventaja debido a que cada vez que se realiza una consulta se debe recuperar la información de la base de datos, cuyo proceso demora un par de segundos; el tiempo depende de factores como el ancho de banda y la potencia de procesamiento computacional, lo cual debe reducirse para garantizar el acceso a la información consultada (J. Thomas & Kielman, 2009).

3 Desarrollo

La aplicación de los principios de la Analítica Visual (AV) a sistemas de software, se conoce como Analítica de Software Visual (ASV), considerando AV como un proceso integral que incluye el análisis de datos avanzados y el uso de múltiples puntos de vista vinculados. La aplicación de AV a la Evolución de Software (ES) comparte elementos comunes con ASV. Analítica de Software Visual Evolutivo (ASVE) es un proceso que consiste en aplicar la Analítica Visual (AV) a los sistemas de software, para mejorar la comprensión de los cambios de sistemas de software con una activa participación de los usuarios por medio de la interacción del humano y la computadora.

3.1 Arquitectura

La definición de la arquitectura de las herramientas de software es una tarea muy compleja que requiere un análisis cuidadoso. Es un reto determinar qué técnicas utilizar, cómo se interrelacionan y contribuyen a la investigación y los objetivos de diseño de estas. Esta sección busca contribuir a cumplir los objetivos especificados, responder a la investigación y las preguntas formuladas, así como el diseño de herramientas de apoyo en situaciones en donde se aplica la AV a la ES. En consecuencia, utilizando como referencia el proceso ASVE se define una arquitectura para desarrollar una herramienta que se encargue de representar los registros de las especies que posee el HN. Tal arquitectura ha sido implementada utilizando los lenguajes de programación Python y JavaScript, la prueba de los resultados se puede ver más adelante en la sección, Análisis de resultados (González-Torres, Peñalvo, & Theron, 2013).

Por lo tanto, el proceso utiliza un enfoque basado en modular, donde cada módulo es una colección de componentes que son a su vez formados por métodos y técnicas. En consecuencia, los módulos principales de EVSA son: ETC (Extraer Transformar Cargar), este módulo tiene la función de realizar la conexión y recuperación de datos del repositorio de software, registros o cualquier otra fuente de datos disponibles, cuando los datos son extraídos, se limpian, se combinan y son cargados en el Almacén de Datos; acerca del Motor de Análisis para la Evolución de Software Avanzado (MAESA), este módulo se compone de técnicas de análisis cuyo objetivo es extraer hechos de conocimiento; además, el Explorador del Conocimiento Visual para Evolución de Software (ECVES), es módulo compuesto por

Representación gráfica de la taxonomía para especies del Herbario Nacional

tres componentes: Visualización de la Evolución de Software, Vistas Enlazadas y análisis de Hechos, y los componentes de Abstracción de Visualizaciones y Apoyo de Coordinación (González Torres, García Peñalvo, & Theron Sánchez, 2013).

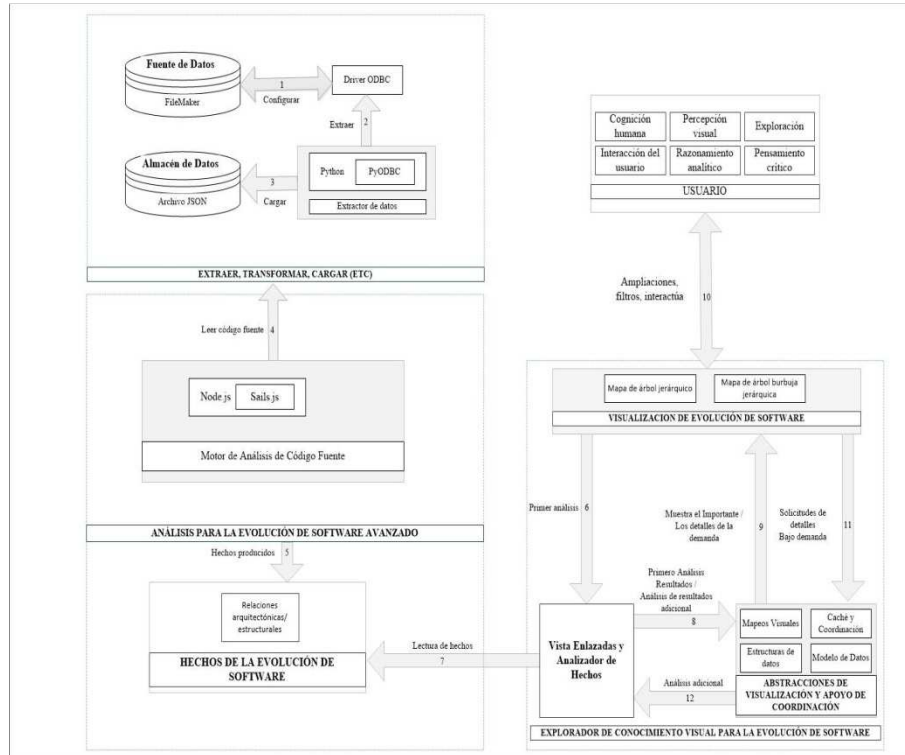


Fig. 2. Arquitectura

En la arquitectura anterior 2, el funcionamiento del módulo ETC y AESA es síncrona mientras que la operación de VKESE es asíncrona, en relación con los otros dos módulos. La arquitectura se basa en el modelo cliente / servidor, en el que el módulos de ETL y AESA son ejecutados por el servidor y VKESE. Esta es una página web consultada por el cliente. Los diferentes módulos y componentes de la arquitectura se describen en el siguiente orden: la recuperación de datos, análisis de datos y la representación visual. El módulo ETL comprende un submódulo (SM) y dos componentes (C), como se muestra en la lista siguiente:

Fuente de datos (C): La fuente de datos utilizada por la arquitectura implementada constar del repositorio de FileMaker. La información que se extrae de este repositorio incluyen las relaciones de taxones de las diferentes especies.

Extractor de Datos (SM): La función de este sub-módulo es extraer los datos

Representación gráfica de la taxonomía para especies del Herbario Nacional

necesarios con el fin de llevar a cabo un estudio, cuyos resultados se utilizan para alimentar las visualizaciones de la herramienta de AV. Además, se compone del código encargado de recuperar los registros de las especies de la fuente de datos.

El módulo de ASEA se compone de un sub-módulo: Motor de Análisis de Código Fuente, como se muestra en la siguiente lista:

Motor de Análisis de Código Fuente (SM): Este sub-módulo es responsable de llevar a análisis del Almacén de Datos generado por el Extractor de Datos.

Hechos de la Evolución de Software (C): Este es el resultado de un análisis que se ha llevado por los otros sub-módulos y componentes del módulo de AESA. Con el fin de hacer esto, se utiliza una jerarquía que muestra la estructura de la taxonomía: Reino -¿ División -¿ Clase -¿ Orden -¿ Familia -¿ Género -¿ Especie, además posee la información general de cada especie.

Los pasos para el proceso de recuperación y análisis de datos (compuesto por los módulos ETC y AESA) se muestran a continuación:

1. Se debe instalar y configurar el driver ODBC a FileMaker, ODBC es un estándar de acceso a las bases de datos que permite extraer los datos de un sistema de gestión de bases de datos como los es FileMaker. (Configurar, Flecha 1)

2. Por medio de la librería PyODBC del lenguaje de programación Python se crea una conexión con ODBC para poder extraer los datos desde la fuente de datos. (Extraer, Flecha 2)

3. Una vez que los datos se han recuperado por medio de la consulta SQL realizada por la librería PyODBC y se guardan los datos en un archivo de texto en formato JSON. (Cargar, Flecha 3)

4. Para realizar su tarea, el Motor de Análisis de Código Fuente (MACF) requiere analizar y crear las relaciones para representar la clasificación de las especies, tomando los datos del código fuente (archivo JSON) que creó en el módulo ETC. (Leer código fuente, Flecha 4)

5. El MACF utiliza la plataforma Node.js para el servidor, desarrollada en JavaScript, esta emplea un modelo de eventos sin bloqueo de I/O, esto permite que sea ligero y eficiente para la ejecución de aplicaciones en tiempo real que manejen gran cantidad de datos. Además, se complementa con el Framework Sails.js, diseñado para emular el patrón MVC y desarrollar aplicaciones web.

6. Cuando MACF termina, lleva el análisis de datos al módulo de Hechos de la Evolución de Software; es esta parte donde la información es cargada por la librería en Java Script D3.js lo cual permite manipular los documentos basados en datos, por medio de tecnologías web HTML y CSS emplea componentes de visualización de gran alcance. (Hechos Producidos, Flecha 5)

7. El proceso que llevan a cabo los módulos ETC y AESA funciona sin importar las actualizaciones que se hagan en los registros de FileMaker.

Carga de estructura y mapeo de la visualización: En esta parte se incluye la carga de los Hechos de la Evolución de Software, crear las estructuras de datos, asignaciones visuales y cargar las visualizaciones.

Representación gráfica de la taxonomía para especies del Herbario Nacional

8. El usuario carga el componente Visualización de la Evolución de Software (VES), donde utiliza los dos tipos de visualizaciones, estas visualizaciones se explican a continuación.

La presencia de datos jerárquicos, en entornos empresariales y académicos, es tan abundante que ha dado lugar a una gran cantidad de investigación orientada hacia su tratamiento y una representación adecuada. En consecuencia, las próximas visualizaciones son seleccionadas para las representaciones visuales que se utilizan para representar la jerarquía de datos de la taxonomía de las especies del HN:

Mapa de árbol jerárquico: Esta visualización permite la representación de información jerárquica en un espacio rectangular en 2D, utilizando 100 por ciento de la superficie disponible. También proporciona controles interactivos, facilita una rápida recuperación de la información con un bajo nivel de percepción y de carga cognitiva, mientras que proporciona una presentación estéticamente agradable (Johnson & Shneiderman, 1991).

Mapa de árbol burbuja jerárquico: Otra variante de esta es la propuesta por Karl Wetzel para la visualización de archivos en un disco duro. En lugar de rectángulos, cuadrados y polígonos, propone el uso de componentes de color circulares codificados. Sin embargo, aunque esta visualización volvió a ser atractiva, no utiliza el espacio para la máxima ventaja ni es posible determinar a primera vista la relación jerárquica entre los componentes (Hlawatsch, Burch, & Weiskopf, 2014).

9. Cuando el componente VES es cargado se solicita la estructura de datos requeridos por la visualización por las Vistas Enlazadas y el Analizador de Hechos (VEAH). (Primer análisis, Flecha 6)

10. El componente VEAH lee los Hechos de la Evolución de Software y los pasa a las Abstracciones de Visualización y Apoyo de Coordinación (AVAC). (Primer análisis de resultados, Flecha 8)

11. El componente AVAC crea y pasa el modelo apropiado, la estructura de datos y el mapeo de visualizaciones al componente VES. (Muestra importante los detalles en demanda, Flecha 9)

La interacción del usuario y detalles bajo demanda es la etapa final del proceso de transformación de datos en conocimiento. Después la recuperación, sigue el análisis y mapeo visual de la información; esta fase hace posible un bucle de retroalimentación entre el usuario y el sistema; el usuario solicita datos adicionales al sistema por medio de la interacción, las posibilidades disponibles, y el sistema proporciona los datos solicitados. De acuerdo con las interacciones del usuario, el proceso de descubrimiento de conocimiento se refina y progresa hacia el hallazgo de ideas y respuestas útiles.

12. Durante el proceso de descubrimiento de conocimiento, el usuario navega, utiliza filtros y explora diferentes perspectivas sobre los datos, selecciona elementos de una o más de las visualizaciones (Ampliaciones, filtros, interactúa, Flecha 10)

13. Según las necesidades y las interacciones del usuario, la visualización solicita nuevas estructuras de datos y asignaciones visuales para pro-

Representación gráfica de la taxonomía para especies del Herbario Nacional

porcionar información adicional para el usuario de acuerdo con las opciones seleccionadas (Solicitud de detalles bajo demanda, Flecha 11).

14. Si los detalles adicionales que han sido solicitados están disponibles en la forma de estructuras de datos, los datos de asignaciones visuales se pasan al componente VES (Muestra importante los detalles en demanda, Flecha 9). Sin embargo, si estos datos no están disponibles, la solicitud se pasa a el componente VEAH que lee los hechos adicionales, (Lectura de hechos, Flecha 7), transforma los datos y luego (Mas análisis de resultados, Flecha 8) los pasa al componente AVAC por lo que puede proceder para crear estructuras de datos de datos y asignaciones visuales.

15. Finalmente, el usuario sigue interactuando con el sistema hasta que obtiene el conocimiento necesario o se considera que es imposible llegar a una conclusión determinada usando los datos y las representaciones disponibles.

4 Resultados

La herramienta de analítica visual fue desarrollada en JavaScript por medio de la implementación de tecnologías como FileMaker, ODBC, Python, Node.js y D3.js. Se usó de una técnica de visualización que interactúa con los demás componentes, en la 3 se muestra la herramienta que está compuesta por tres partes: la sección de búsqueda, la visualización y la columna de información o datos relacionados con cada especie.

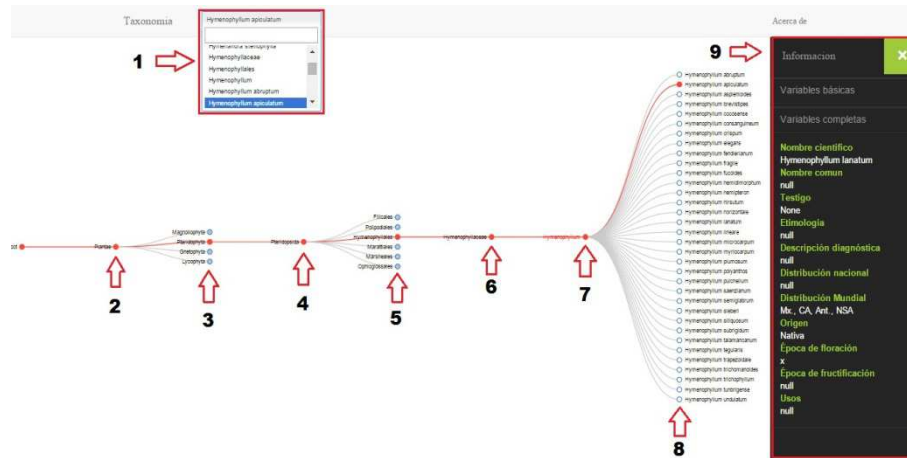


Fig. 3. Herramienta de Analítica Visual

La fecha 1 muestra la sección de búsqueda de especies, donde se puede ingresar el nombre de una especie o se puede buscar alguna otra con base en la lista que despliega el cuadro de texto; cuando se ha escogido y se selecciona la

especie, esta se representa en la visualización por medio de un camino marcado en rojo; asimismo, la visualización permite opción de zoom por medio del scroll del mouse, además permite arrastrar la visualización por medio del clic izquierdo presionado; la técnica empleada por la visualización es un árbol jerárquico que posee siete niveles, la flecha 2 indica el reino animal, la 3 muestra la división, la 4 revela la clase, la 5 evidencia el orden, la 6 la familia, la 7 indica el género y la flecha 8 muestra la especie que es el nombre científico compuesto por el género y el nombre de la especie.

Al seleccionar una especie en la sección de búsqueda la visualización, se despliega hasta llegar al nodo final donde se encuentra la especie buscada; sin embargo, es posible interactuar con la visualización, sin utilizar la búsqueda; al inicio solo se mostrará el nivel de reino, señalado por la flecha 2, al clic sobre ese nodo, se muestra automáticamente las divisiones que pertenecen a ese reino; posteriormente, se sigue dando clic sobre los nodos hasta llegar al final de la visualización, lo cual es un especie indicada por la flecha 8; de esta manera el usuario puede interactuar con la aplicación para descubrir las distintas clasificaciones.

La flecha 9 indica la columna de información de cada especie, esta posee dos menús el de información básica de la especie que muestra el nombre científico y el nombre común de esta y un segundo menú el cual contiene la información completa donde se muestran datos como el testigo, la persona que descubrió la especie, la etimología, la descripción diagnóstica, la distribución nacional y mundial; origen, época de floración, época de fructificación y usos de la especie. Esta columna de información es dinámica y su contenido cambia, dependiendo de la especie que se busca, sin embargo, algunas especies no poseen la información completa, debido a la falta de actualización de la base de datos del Herbario.

5 Conclusiones

La principal contribución de este trabajo corresponde a la creación de un prototipo que se encarga de representar las especies del HN. En relación con los objetivos de la investigación, no se logra cargar toda la información de la base de datos FileMaker debido a que la arquitectura desarrollada hace la consulta total de estos, ello provoca que el proceso de búsqueda dependa de la capacidad de procesamiento del hardware en el cual se ejecuta la aplicación. Para solucionar el problema, se extraen 2000 mil registros de fuente de datos; se recomienda en un trabajo futuro segmentar las consultas de los datos, solicitando aquella información según el momento en que el usuario la requiera o la carga de datos por demanda.

La implementación de esta arquitectura ha permitido poner a prueba el proceso de AV para la representación de las especies del HN, con el objetivo de obtener un conocimiento general y específico sobre la biodiversidad de Costa Rica, además de determinar si esta herramienta contribuye con el descubrimiento de conocimiento, por medio de la técnica de visualización empleada. El uso de este software por los científicos del Departamento de Historia Natural ha demostrado que la AV puede contribuir de una forma rápida y efectiva la exploración de la biodiversidad de Costa Rica, sin embargo para un trabajo futuro se requiere de que la

Representación gráfica de la taxonomía para especies del Herbario Nacional

aplicación, aparte de mostrar las distintas especies, visualice la información de las colecciones, las localizaciones donde se han encontrado los ejemplares de una misma especie.

La aplicación de visualización de la información junto con AV, permite captar fácil y rápido aquellos detalles que permanecen ocultos de los usuarios. Por lo tanto, uno de los puntos fuertes del proceso de ASVE en tareas de descubrimiento de conocimiento es el uso de técnicas de visualización.

References

- Andrienko, G., Andrienko, N., & Wrobel, S. (2007). Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, 9(2), 38–46. pages 3
- Botanica, R. T. (2014). pages 5
- de Bolivia, H. N. (2014). pages 5
- de Botanica Darwinion, I. (2014). pages 5
- de Botanica del Noreste, I. (2014). pages 5
- de Ciencias Naturales, M. E. (2014). pages 5
- de Colombia, U. N. (2014). pages 5
- de Historia Natural de Florida, M. (2014). pages 5
- de Historia Natural Smithsonian, M. N. (2014). pages 5
- de Investigaciones Científicas Sinchi, I. A. (2014). pages 5
- de la Naturaleza, M. C. (2014). pages 5
- de Río de Janeiro, J. B. (2014). pages 5
- Edinburgh, R. B. G. (2014). pages 5
- Garden, M. B. (2014). pages 5
- González Torres, A., García Peñalvo, F. J., & Theron Sánchez, R. (2013). How evolutionary visual software analytics supports knowledge discovery. pages 7
- González-Torres, A., Peñalvo, F. J. G., & Theron, R. (2013, March). Human-computer interaction in evolutionary visual software analytics. *Computers in Human Behavior*, 29(2), 486-495. Retrieved from <http://www>

- Representación gráfica de la taxonomía para especies del Herbario Nacional
[.sciencedirect.com/science/article/pii/S0747563212000155](http://www.sciencedirect.com/science/article/pii/S0747563212000155) doi:
 10.1016/j.chb.2012.01.013 pages 6
- Hamburgense, H. (2014).
 pages 5
- Herbarium, E. S. (2014).
 pages 5
- Herbarium, T. L. (2014).
 pages 5
- Hlawatsch, M., Burch, M., & Weiskopf, D. (2014). Bubble hierarchies. In
 Proceedings of the workshop on computational aesthetics (pp. 77–80). pages
 9
- Johnson, B., & Shneiderman, B. (1991). Tree-maps: A space-filling approach
 to the visualization of hierarchical information structures. In *Visualization*,
 1991. visualization'91, proceedings., iee conference on (pp. 284–291).
 pages 9
- Keim, D., Andrienko, G., Fekete, J.-D., Gorg, C., Kohlhammer, J., & Melançon,
 G. (2008). *Visual analytics: Definition, process, and challenges*. Springer.
 pages 4
- Museum, I. S. (2014).
 pages 5
- of North American Lichen Herbaria, C. (2014).
 pages 5
- Seinet. (2014).
 pages 5
- Sharma, O. (1993). *Plant taxonomy*. Tata McGraw-Hill Publ. Retrieved from
<http://books.google.es/books?id=uWg76rCqA68C> pages 2
- Thomas, J., & Kielman, J. (2009). Challenges for visual analytics. *Information
 Visualization*, 8(4), 309–314. pages 6
- Thomas, J. J., & Cook, K. A. (2005). *Illuminating the path: The research and
 development agenda for visual analytics*. IEEE Computer Society Press.
 pages 2, 4
- University, & of the University of California at Berkeley, J. H. (2014).
 pages 5
- University, H. (2014).
 pages 5