

# Identificación de Especies Biológicas Utilizando Visualización de Datos y Claves Electrónicas de Multiacceso

Leonardo Rojas León y Antonio González Torres

Escuela de Ingeniería,  
Universidad Latinoamericana de Ciencia y Tecnología,  
ULACIT, Urbanización Tournón, 10235-1000  
San José, Costa Rica  
[leonardo, antonio]@ulacit.ed.cr  
<http://www.ulacit.ac.cr>

**Resumen** Actualmente se vive en la era de la información, donde se manejan una gran cantidad de datos que, con ayuda de herramientas tecnológicas, podemos visualizar de distintas maneras, y convertir estos datos en conocimiento. En la mayoría de los casos en empresas a nivel mundial, es necesario analizar datos y ordenarlos de manera que sean comprensibles visualmente, para así poder transmitir dicho conocimiento a quien se necesite y entender estos datos para ejercer decisiones concisas para tener distintas formas de visualización de los datos, facilitar procesos, estructurar y ordenar los datos, manejar recursos, y en algunos casos hasta mejorar la situación financiera de instituciones.

**Palabras Clave:** Visualización de datos, datos informáticos, especies biológicas, Claves de identificación

## 1. Introducción

La preservación de la biodiversidad es un tema que se discute con mucha frecuencia, tanto en conferencias científicas como internacionales (ej. el cambio climático, y diversas reuniones organizadas o patrocinadas por las Naciones Unidas), debido a la preocupación de científicos y líderes globales, acerca de las consecuencias de la vida humana y la economía que podría provocar la pérdida de especies biológicas, la transformación de los hábitats, los cambios en la composición de los ecosistemas y el cambio climático.

Los ecosistemas son considerados sistemas biológicos vitales, por esto son comunidades tanto de especies vivientes biológicas, y componentes no vivientes que tienen fuertes interacciones entre ellos para ofrecer servicios, como agua fresca, comida, materiales descompuestos, purificación del aire, fertilidad de la tierra, manejo de carbón para reducir emisiones de CO<sub>2</sub>, regulación del clima y

protección en las costas. Consecuentemente, la diversidad y la riqueza de las especies biológicas (ej. heterogeneidad y número biológico de especies) son asociados al tipo y calidad de los servicios que son proveídos por los sistemas biológicos (Thuiller, 2007). Resumiendo, la humanidad depende de la riqueza de los sistemas biológicos para poder contar con recursos sostenibles y servicios para la sobre vivencia.

El término biodiversidad es usado frecuentemente para hacer referencia a la diversidad de los sistemas biológicos (ej. ecosistema o hábitat), pero también involucra una magnitud natural de dicha diversidad, que frecuentemente, es determinada por la tan diversa cantidad de especies (que muchas veces puede contarse de hasta millones). Es por esto, que la identificación biológica de especies que habitan los sistemas biológicos, es un importante elemento para habilitar una contabilización informada de distintas especies. El conocimiento obtenido de determinar la magnitud de la diversidad biológica, puede ser utilizado para ayudar a la preservación de la riqueza de sistemas biológicos, por ende, de los servicios que provee. A su vez, la identificación de especies nos permite la detección potencial de abastecimiento de comida, elementos químicos (Johnson, 1992), agentes colorantes, venenos, elementos para controlar las pestes (Johnson, 1992), fertilizadores, aceites (Heinrich y Gibbons, 2001), medicinas y calidad del aire (Borrell, 2014).

La identificación biológica de especies es llevada a cabo por científicos usando claves de identificación, que involucran su clasificación hacia categorías de la taxonomía (ej. taxa2), en congruencia con sus similitudes y diferencias. Estas claves son construidas desde las características visibles morfológicas de las ya conocidas especies biológicas. Dichas características son organizadas en un orden lógico, de acuerdo a los principios de la taxonomía, que permiten guiar a los científicos por una serie de pasos para decidir si la especie pertenece a un espécimen que está siendo analizado.

La identificación de especies son creadas en formatos electrónicos como impresos, y son clasificados como de acceso-único (claves dicotómicas), multi-accesos y claves tabulares. También, es común encontrar libros de taxonomía que sean exclusivamente dedicados a documentar la identificación de claves para familias y género, mientras que la identificación de claves electrónicas es usada con frecuencia con el soporte de sistemas interactivos especializados.

Es importante recalcar que los especialistas en la identificación de especies son escasos, y que cerca de 3 millones de especies aproximadamente no han sido identificadas. Como consecuencia, se han tomado varios esfuerzos de investigación para desarrollar herramientas con el fin de asistir a científicos con esta tarea, hasta para identificar especies de forma automática. Sin embargo, el uso de visualización de datos no ha sido tomado en cuenta de gran manera, para contribuir a los científicos en la identificación de especies. Consecuentemente, esta investigación propone una visualización de datos para construir una clave electrónica para luego ser usada en la identificación de especies, basado en un simple proceso de toma de decisión.

## 2. Antecedentes

La identificación de especímenes biológicos involucra el reconocimiento de sus características más distintivas (ej. tiene hojas), y también la apreciación de sus atributos que son asociados con cada uno (ej. hojas amarillas, hojas de largas extensiones). Las características de las especies son llamadas caracteres, y los atributos asociados a cada carácter son llamados estados de carácter.

Durante el proceso de identificación, los caracteres y los estados de carácter de una especie son comparados a otras especies usando información por benchmark (ej. incluyendo descripciones e ilustraciones) y si ninguna combinación de carácter y de estados de carácter producen un emparejamiento, la especie puede ser considerada nueva, o sin ninguna identificación (KeyToNature, 2006).

Este proceso requiere un estudio y un análisis de un gran número de posibilidades, y en algunos casos, la participación de varios científicos, inclusive en instituciones diferentes. En consecuencia, claves electrónicas de acceso-único fueron creadas en formato impreso desde hace 200 años para facilitar la identificación de especies.

Las claves de acceso-único son clasificadas como claves dicótomas y claves politómicas y consisten en una secuencia de puntos de decisión que son determinados por los caracteres y estados de carácter. Las estructuras de las claves dicótomas consisten en dos opciones en cualquier punto de decisión, mientras que las claves politómicas tienen más de dos opciones en cualquier punto de decisión de su estructura (KeyToNature, 2006).

Consecuentemente, la Tabla 1 (KeyToNature, 2006) muestra una clave de acceso-único para la identificación de género de la familia Papilionácea. Esta clave consiste de dos caracteres (alas-traseras y alas-delanteras) y siete estados de carácter (Circulares, manchas azules, puntiaguda de extensión larga, puntas de escalope, manchas negras y rayas negras).

La estructura de la clave de identificación en la Tabla 1 (KeyToNature, 2006) se compone de cuatro puntos de decisión, en los cuales el flujo puede tomar distintos caminos. Consecuentemente, el punto de decisión 1(0) evalúa si la especie de la familia Papilionidae tiene ala-trasera o si no tiene ala-trasera redonda. Si el resultado de cada evaluación es positivo, el próximo paso es evaluar el punto de decisión 2(1), por el contrario si el resultado es negativo, entonces el punto de decisión 3(2) debería de ser evaluado.

El punto de decisión 2(1) evalúa si las manchas azules están presentes o no en las alas traseras de la especie a la cual se le hace el análisis, y como resultado, la especie puede pertenecer al Archon (manchas azules presentes) o al género Parnassius (manchas azules no presentes). Mientras tanto, el punto de decisión 3(2) evalúa si la ala-trasera de la especie tiene puntas con extensiones largas (continúe al punto de decisión 4(3)) o puntas de escalope (la especie pertenece a género Zerynthia). Entonces, el punto de decisión 4(3) realiza una evaluación similar para ayudar a determinar si la especie pertenece al género Papilio o Iphiclides.

**Tabla 1.** Identificación de claves para el género de la familia Papilionidae (The Teacher's Handbook)

1(0). Alas-traseras son redondeadas	2
Alas-traseras no son redondeadas	3
2(1). Manchas azules presentes en las alas traseras	Archon
Manchas azules no presentes en las alas traseras	Parnassius
3(2). Alas-traseras tienen punta con largas extensiones	4
Alas-traseras con punta de escalope	Zerynthia
4(3). Alas-delanteras con manchas negras	Papilio
Alas delanteras con rayas negras	Iphiclides

La representación visual, usando un árbol dicotómico, para esta clave de identificación está prevista en la **Figura 1**. La ventaja de la representación gráfica es que el flujo de decisión puede ser seguido con facilidad para la identificación de especies. Sin embargo, esta representación gráfica puede ser valiosa para asistir en el reconocimiento de especies, vale la pena destacar que su uso puede ser abrumador cuando se está lidiando con un complejo y amplio número de claves, ya sea en formato electrónico o impreso.

**Figura 1.** Representación dicotómica para la identificación de claves para el género de la familia Papilionidae.

En general, la desventaja de las claves dicotómicas es que el proceso de identificación puede requerir que se vayan por varios puntos de decisión, mientras que las claves de multi-acceso, proveen un acceso directo a los caracteres, y los científicos rara vez necesitan seguir más de 3 puntos de decisión (desde la raíz de las hojas de la estructura) para obtener resultados. Sin embargo, la representación de claves multi-acceso, en los formatos impresos es muy complicada, su fortaleza no podría ser explotada con la ayuda de sistemas interactivos, que son intuitivos y fáciles de utilizar.

Consecuentemente, el primer paso para crear claves de identificación multi-acceso, es representar la clave de identificación como una matriz, que después sería utilizada como una entrada a sistemas especializados que están listos para la construcción de claves de identificación interactivas de tipo multi-acceso, como se ve en la **Tabla 2**.

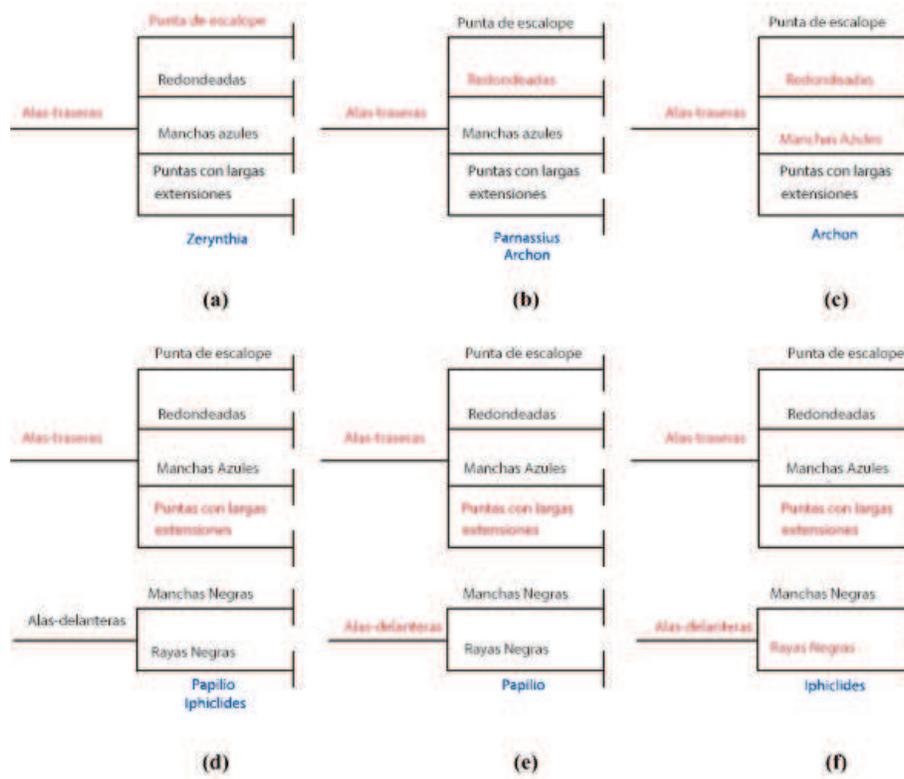
**Tabla 2.**

Genero	Alas-traseras				Alas-delanteras	
	Punta de escalope	Circulares	Manchas azules	Puntas - Largas extensiones	Manchas negras	Rayas negras
<b>Parnassius</b>	No	Si	No	No	No	No
<b>Archon</b>	Si	No	No	No	No	No
<b>Papilio</b>	No	No	No	Si	Si	No
<b>Iphiclides</b>	No	No	No	Si	No	Si

La principal ventaja de implementar claves de multi-acceso usando sistemas interactivos, especialmente cuando se tratan de claves extensas, es que la interacción entre el usuario y el sistema se convierte en un elemento central que hace que la exploración de las características y procesos de toma de decisión sean más efectivos. Este tipo de claves puede facilitar a los científicos la identificación de especies, basados en su conocimiento y experiencia. Con su uso ellos podrían seleccionar una o más características sin estar limitados a seguir un proceso secuencial de evaluación, el cual puede requerir la evaluación de características desconocidas a los científicos, en lugar de tomar ventaja de su experiencia.

La **Figura 2** muestra una secuencia de figuras con la opción que pueden ser seleccionadas desde una clave multi-acceso sencilla, el cual está basado en la tabla 2. El color rojo en las figuras es usado para destacar los caracteres y los estados de carácter que son seleccionados, y así, los géneros que compartan estas características aparecen en la parte inferior de la figura correspondiente en esta serie. Además, los caracteres de la clave son Alas-traseras (controlador) y Alas-delanteras (dependiente). Entonces, cuando una clave es cargada, aparece sólo el carácter Alas-traseras, donde Alas-delanteras se mantiene oculto hasta que la selección de caracteres y estados de carácter sea apropiada.

De acuerdo con esto, la identificación del género *Zerynthia* se realiza al seleccionar el carácter Alas-traseras y el estado de carácter Punta de escalope (ver figura 2 (a)). Mientras que para activar el carácter Alas-delanteras es necesario seleccionar el carácter Alas-traseras y el estado de carácter Punta con largas extensiones, como se muestra en la figura 2 (d). En dicha figura además se puede observar el carácter Alas-delanteras resaltado en color verde y los géneros *Papilio* y *Iphiclides*, los cuales reúnen las características de Alas-traseras – Punta con largas extensiones – Alas delanteras. En tanto que las figuras 2 (e) y 2 (f) muestran como resultado uno u otro género, dependiendo si de forma adicional



**Figura 2.** Representación de una clave multi-acceso para el género de la familia Papi-lionidae: interacción e identificación de opciones.

se ha seleccionado el estado de carácter Manchas negras o Rayas negras. Esto es utilizando un formato de matriz (ver Tabla 2). Esta representación permite ver la clave desde una perspectiva tabular, la cual ofrece otro punto de vista sobre el proceso de toma de decisiones durante la identificación, pero también puede servir como base para construir una representación de árbol dicotómico, o politómico, según sea el caso.

### 3. ¿Qué son las herramientas de visualización de datos?

A lo largo de los años, muchos de los problemas de organizaciones que manejan grandes datos, es el cómo almacenar esta información, y cómo hacer que la información sea íntegra, confidencial, y segura, mediante distintas herramientas que ayudan a esta cuestión. Con el tiempo, no se ha tenido gran problema a la hora de almacenar los datos, sino más bien a la hora de recuperarlos, ya que cada segundo se genera una gran cantidad de datos a nivel mundial. Al crecer los datos, se requiere de distintos métodos tanto físicos como lógicos y digitales,

para hacer que la recuperación de estos sea más rápida, y poder así mostrar una solución a altos jerarcas de compañías o a público en general, acerca de los datos que se están analizando. Un ejemplo de este tipo de optimizaciones para visualizar la información, son las bases de datos no relacionales, estas poseen una lógica distintas de cómo almacenar los datos, permitiendo el uso de agregados (listas, registros, entre otros.), permitiendo agregar o quitar elementos a la base de datos de manera sencilla.

A su vez, las bases de datos no relacionales, se logran acoplar a esquemas que son poco convencionales, que se salen de lo que conocíamos antes como bases de datos relacionales, donde un conjunto de datos pertenece a otro por medio de identificadores, construyendo tablas que se comparten entre distintas entidades de información. Cabe resaltar, que estas últimas formas de almacenamiento de datos NoSQL para tipos de almacenamiento que no cumplen con un esquema relacional, surgió a finales de los años 90, y no fue hasta hace 8 años que una de las más populares (MongoDB) fue acogida de gran manera por empresas de alto calibre a nivel mundial (G. Hansel, 2012).

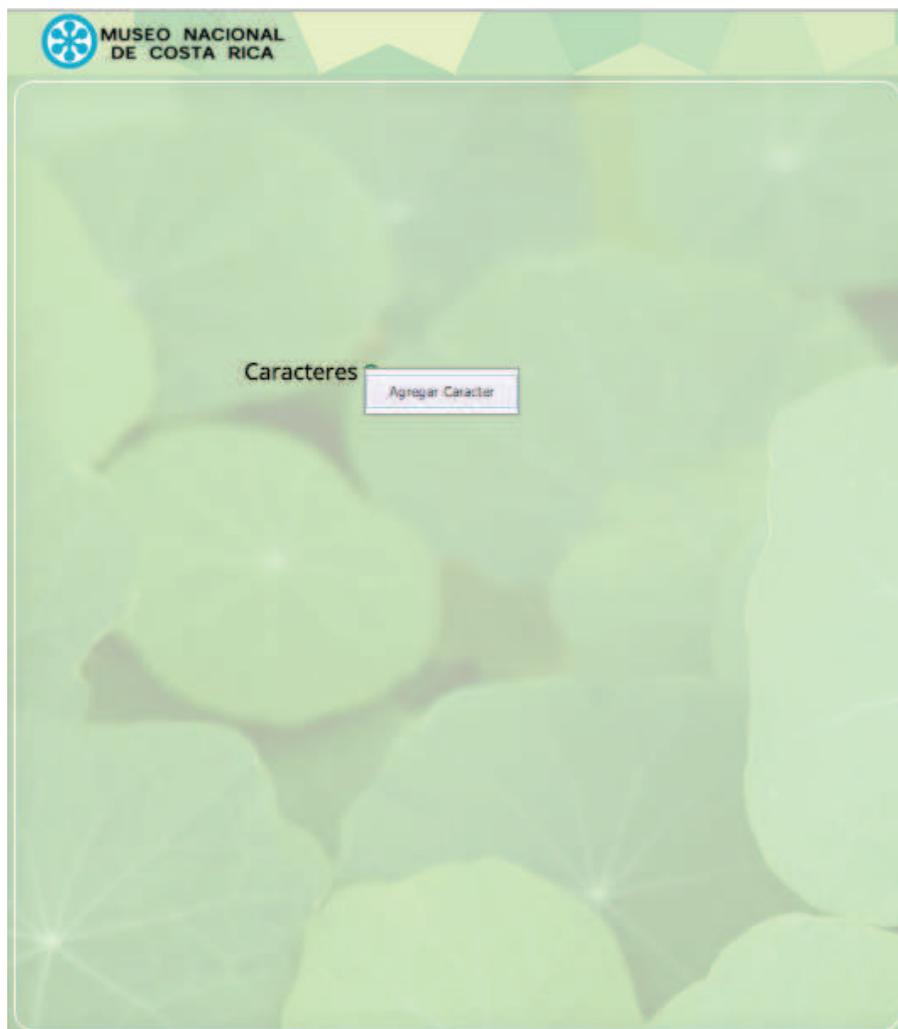
Es importante saber que los datos están correctos y pensados para que puedan crecer ampliamente para poder visualizarlos y pasarlos por una herramienta para renderizar dichos datos en resultados. Es por esto que para la creación del “Sistema de Identificación de Especies por Medio de Claves Electrónicas” se utilizó un modelo no relacional, junto con la ayuda del lenguaje más omnipresente de la historia, llamado JavaScript.

JavaScript es un lenguaje que se utiliza todos los días por millones de usuarios, ya que se encuentra en cualquier teléfono celular, o dispositivo que pueda cargar una página web, llámese computadora desktop, portátil, hasta relojes y computadores de automóviles. Sabiendo esto, existen herramientas de visualización de grandes datos que nos facilitan la interacción entre tener los datos limpios y de fácil acceso, y poder manipularlos para lograr un resultado superior al que se podría obtener únicamente teniendo los datos almacenados (D. Giovanni, 2015).

#### 4. Acerca de la herramienta desarrollada

La herramienta desarrollada está pensada para que sea utilizada por una persona del ámbito de la ciencia, propiamente a investigar especies biológicas, siendo esta, un apoyo para poder crear nuevas claves electrónicas de especies biológicas y poder visualizarlas fácilmente. Esta herramienta propone una nueva forma de almacenar y observar las especies que ya existen, también se encuentran otras herramientas que también generan el mismo resultado, pero la parte de visualización y de cómo insertar y seleccionar especies es distinta y de poco valor agregado. Para agregar una nueva especie con sus caracteres y estados de carácter, basta con agregar un primer carácter con sus nodos hijos en la siguiente pantalla:

Añadimos más estados de carácter a ese carácter que se va a ingresar, para que se muestren como aparecen en la **Figura 4**.



**Figura 3.** Para agregar un caracter se le da click derecho al círculo a la par de “Caracteres”

Seguidamente procedemos a ingresar la información de la especie con los campos para llenar su ficha. Estos campos se pueden observar ya llenos en la **Figura 5.**

Se selecciona una imagen, y se le indican la cantidad de posiciones en la parte superior izquierda para agregar la cantidad de posiciones en el mapa donde se encontró dicha especie.

Al hacer click en “Crear ficha” se almacenará su información, tanto con los caracteres como estados de carácter elegidos en la base de datos. Para acceder



**Figura 4.** Una vez ingresados los demás caracteres se seleccionan con click izquierdo sobre el nombre de cada carácter, su nombre cambiará a color naranja, indicando que ya están seleccionados para ingresar a la base de datos.

a esta información previamente ingresada, se hace click en el botón superior izquierdo “Identificar especies” donde se mostrarán todas las especies ya existentes en la base de datos, junto con la nueva especie creada con su información y posiciones en el mapa, a como se muestra en la **Figura 7**.

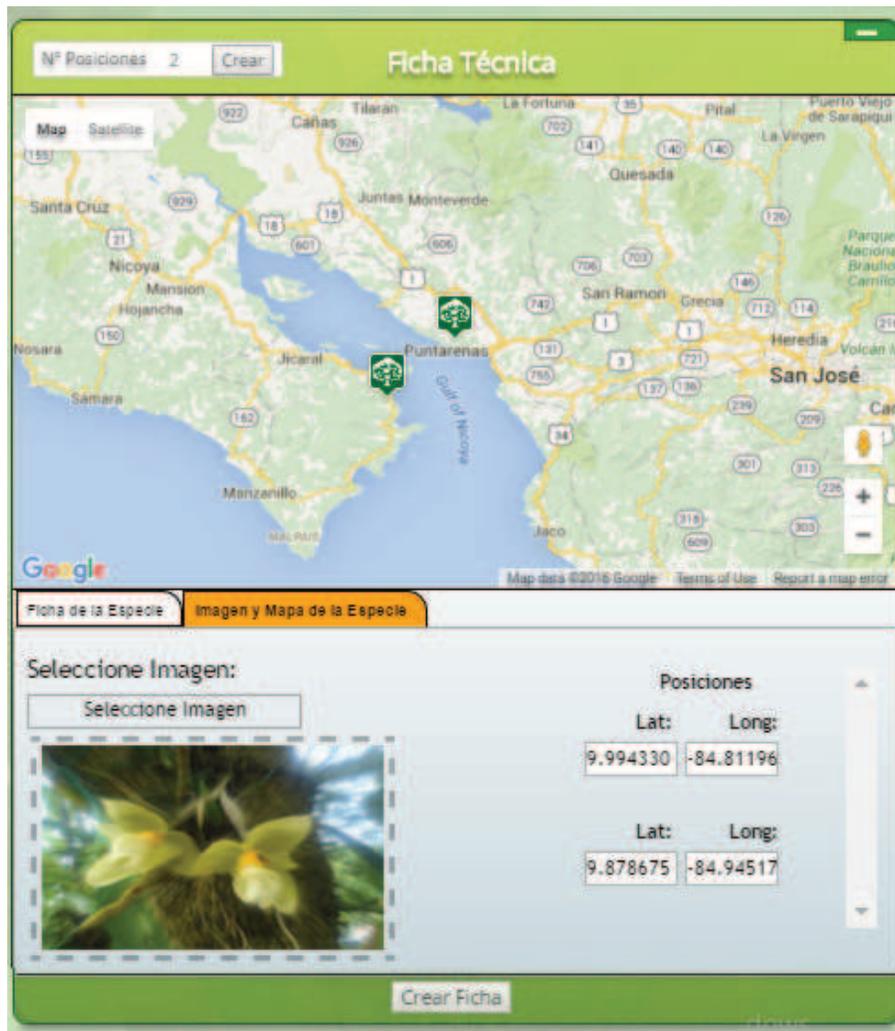


Figura 5. Una vez ingresados procedemos a agregar la imagen y el mapa de la especie.

## 5. Anexos

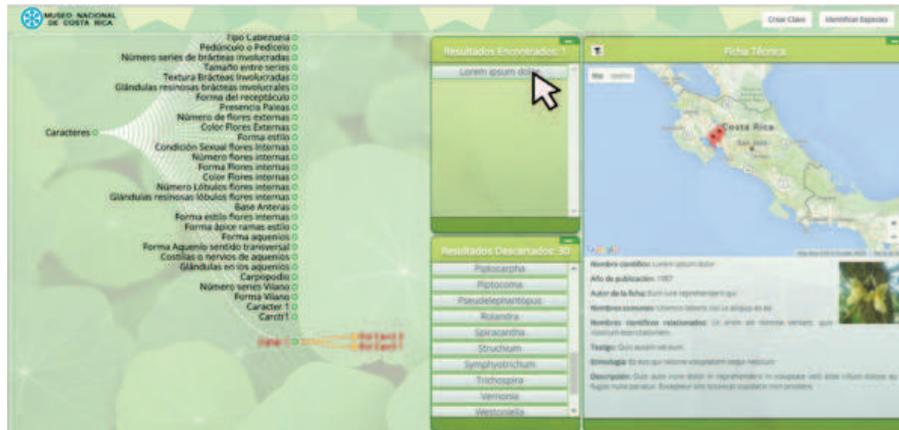
### 5.1. Instalación y pasos de configuración

- Se optó por montar el software en una máquina virtual montada en VMWare Workstation 12 Player, únicamente se debe de instalar la máquina virtual y ya todos los archivos junto con la configuración están listos para continuar el desarrollo.



**Figura 6.** Cuando se van ingresando las posiciones de latitud y longitud, se va marcando un indicador en el mapa con su posición.

- El software fue creado utilizando MongoDB como motor de base de datos, Node.js para su programación, D3 como librería de visualización de datos hecha en JavaScript, y Express Server como servidor.
- En la dirección "C:/MongoDB" se encuentra la base de datos junto con los archivos de datos para poder inicializarla. Las colecciones de datos "ClavesE-lectronicas.ns" y "ClavesElectronicas.0" en sí están en "C:/MongoDB/bin/data/db/".



**Figura 7.** Para ver la información de cada especie en el sistema, se debe de posicionar el cursor sobre la especie tanto en resultados encontrados, como en resultados descartados.

- Antes de inicializar el servicio de MongoDB, es importante irse a la carpeta “C:/MongoDB/bin/data/db/” y asegurarse que no exista el archivo “mongod.lock” el cual impide inicializar el servicio.
- Para inicializarlo, se debe de abrir una ventana de línea de comando y abrirla con click derecho seleccionando la opción de “abrir como administrador”, una vez abierta, se procede a ir a la dirección “C:/MongoDB/bin/” y se ejecuta el comando “mongod -dbpath ./data/db” (sin comillas) y un mensaje de “Waiting for connections on port 27017” debería de aparecer.
- Para visualizar la data de MongoDB, se utilizó RoboMongo como cliente para hacer esto, al abrir esta aplicación una ventana de “MongoDB Connections” aparece y basta con seleccionar “connect” en la nueva conexión que sale establecida con un “Adress” de “localhost:27017” para poder examinar las colecciones (Importante: debe de estar inicializado el servicio del punto anterior).
- Para inicializar el servidor con la aplicación, se debe de abrir una ventana de línea de comando y abrirla con click derecho seleccionando la opción de “abrir como administrador”, una vez abierta, se procede a ir a la dirección “C:/ClavesElectronicas/” y se ejecuta el comando “node app.js”.
- Para ver la aplicación se debería de ir a cualquier navegador e ingresar el URL “localhost:3000”.

## 6. Recomendaciones

1. Tener las versiones que se van desarrollando en repositorios de datos donde se puedan visualizar los adelantos y cambios que va teniendo la aplicación por las distintas personas que contribuyen en él, para poder hacer de su desarrollo un lugar centralizado donde se encuentren los datos para agilizar el acceso al código y demás archivos necesarios para desarrollar la aplicación.

2. Detallar una lista de requerimientos y ordenarlos por prioridad, para esto se pueden utilizar herramientas como JIRA (versión de prueba) o Trello (versión gratuita) para el control de proyectos por tareas con tiempos, que a su vez sirve para saber quién trabaja sobre cuál tarea. Esto porque al tratarse de un proyecto grande y que tiene continuidad a través de los años, puede resultar difícil de documentar.
3. A la hora de tratar con proyectos de ésta índole, siempre es idóneo comenzar sacando una lista de posibles requerimientos, y basados en estos, trabajar sobre la base de datos abarcando cualquier situación o escenario para saber que a nivel de datos la aplicación está cubierta para realizar lo que se pidió en un principio. Esto haría que a la hora de desarrollar, los datos se puedan editar, eliminar, y agregar, en su totalidad, facilitando el desarrollo hacia la parte visual.
4. Es importante tratar de desarrollar pensando en los dispositivos móviles de una vez, ya que una aplicación como esta requiere de estar en muchos lugares remotos descubriendo especies, y si se hubiera pensado en un principio para ser “móvil primero” ya no haría falta la necesidad de volver a estructurar la parte de la vista para que se adapte a una amplia gama de dispositivos.

## 7. Conclusiones

En conclusión, en las últimas décadas se ha visto que emergen nuevas tecnologías, y con el pasar de los años nos encontramos con mejores máquinas de procesamiento que nos permiten visualizar mejor los datos con rapidez y agilidad, para así transformarlos en conocimiento. Sin embargo, hay una fuerte inmersión de muchas tecnologías para visualizar datos, proporcionando una amplia gama de opciones para escoger. Es por esto, que ya no es tanto el problema de no tener una herramienta de analítica visual, si no más bien cuál de todas las herramientas disponibles en el mercado, satisface mejor la solución al problema que se requiere. Un proceso de analítica visual correctamente empleado, puede brindar múltiples mejoras a procesos de una organización, a ver en tiempo real cuántos datos se trasladan de un lugar a otro, o facilitar tareas engorrosas, y transformarlas en un proceso amigable para el usuario. Es importante recalcar, que la analítica visual puede hacer un cambio de paradigma de un día para otro, de cómo utilizamos un servicio, de cómo identificamos una especie biológica, y hasta cómo nos trasladamos de un lugar a otro (como es el caso de Uber), visualizar el espacio desde nuestro celular (google sky map), ver nuestras cuentas bancarias desde la palma de la mano, y hasta salvar vidas por medio de cucarachas con chips que crean un mapeo y permiten saber donde pueden estar sobrevivientes, y así salvar sus vidas.

## Referencias

- Borrell, B. (2014). Seeds of a cure.  
pages 2

- D. Giovanni, L. M. (2015). Utilización de nosql para resolución de problemas al trabajar con cantidades masivas de datos.  
pages 7
- G. Hansel, Y. O. (2012). Bases de datos nosql.  
pages 7
- Heinrich, M., y Gibbons, S. (2001). Ethnopharmacology in drug discovery: an analysis of its role and potential contribution.  
pages 2
- Johnson, R. (1992). Reflections of a plant pathologist on breeding for disease resistance, with emphasis on yellow rust and eyespot of wheat. plant pathology.  
pages 2
- KeyToNature. (2006). Teacher's handbook.  
pages 3
- Thuiller, W. (2007). Biodiversity: climate change and the ecologist. nature.  
pages 2