

Contexto

Actualmente los científicos se enfrentan al reto de acceder y manejar grandes volúmenes de información. Este proyecto de tesis se desarrolla dentro de una Biblioteca Digital Florística y tiene la misión de proporcionar a sus usuarios una herramienta que les permita enfrentar dicho reto en el auxilio de técnicas de visualización. En este capítulo se presentan algunos conceptos de Bibliotecas Digitales Florísticas, se describen brevemente los antecedentes del proyecto, incluyendo soluciones planteadas anteriormente a la problemática del manejo de información, así como un panorama de la solución desarrollada en este proyecto.

1.1 Bibliotecas Digitales Florísticas

Una biblioteca digital es la versión electrónica de una biblioteca física convencional que además ofrece todo tipo de herramientas y servicios que faciliten la utilización y manejo de la información. Una biblioteca digital botánica posee las características de una biblioteca digital pero está dirigida a la utilización y extensión del conocimiento existente sobre las plantas.

Actualmente se está desarrollando la Biblioteca Digital Florística en el Centro de Informática Botánica (CBI) del Jardín Botánico de Missouri, incluyendo los proyectos de Flora de China (FOC) y Flora de Norte América (FNA). Estos consistirán en una gran colección de documentos en gran variedad de medios y formatos, incluyendo datos en texto, mapas e ilustraciones y ofrecerán toda clase de servicios disponibles por medio de la biblioteca a través de la red mundial. De esta manera, científicos botánicos podrán aprovechar las ventajas que ofrece la computación para el estudio de la flora de estas áreas geográficas.

Para el mejor entendimiento de algunos términos que serán utilizados más adelante en este documento, y para la construcción de una biblioteca digital florística, es necesario saber el tipo de información que la compone. Dentro de esta información se encuentran: las claves taxonómicas, mapas de distribución, ilustraciones y los tratamientos taxonómicos. Las claves taxonómicas son índices jerárquicos utilizados por los botánicos para identificar especímenes basados en las características que puedan ser inspeccionadas. Pueden existir claves para cada nivel taxonómico (por ejemplo: familia, tribu, género, especie, etc.). Los mapas de distribución muestran gráficamente el

área geográfica exacta donde se encuentran grupos taxonómicos de plantas en particular. Los tratamientos son la base de una Flora, ya que proveen las descripciones morfológicas detalladas de las plantas a nivel especie o niveles menores (infra-específico). Cada descripción morfológica consiste en una serie de descriptores a las que se refieren como estructuras, las cuales se utilizan para agrupar un número de plantas con características que pueden ser asociadas [Sánchez et al. 1997, 1998].

1.2 Proyecto Flora de China

Flora de China (FOC) es un proyecto colaborativo que tiene como fin hacer pública la información alrededor de 30,000 especies de plantas vasculares de China (1/8 de la flora del mundo). Para la publicación de la información de las especies, los científicos realizan un proceso complicado de edición y revisión de tratamientos. Este proceso de investigación requiere que los científicos consulten trabajos previos relacionados a los especímenes para los que van a realizar un tratamiento, para después enviarlo a un comité editorial. En este proceso, es de gran importancia contar con fuentes autorizadas de información, como las que se describen a continuación.

1.2.1 Tarjetas de Hu

Flora de China cuenta con una colección de aproximadamente 350,000 fichas de 5X8 pulgadas conocidas como "tarjetas de Hu", las cuales son la única información existente a nivel infra-específico de las plantas de China. Estas tarjetas representan años de trabajo de investigación de la Dra. Hu Shiu-ying. La Dra. Hu trabajó con un equipo de cuatro o cinco personas, buscando en toda la literatura botánica entre los años 1773 y 1955 para localizar todos los nombres que han sido utilizados para las plantas de China.

198

Rinorea bengalensis (Wall.) O. Ktze. Rev. Gen. Fl. 42. 1891.--Merr.
fig. 12
& Chun in Sunyats. 5: 127. 1940.

Alsodeia bengalensis Wall. in Trans. Med. Phys. Soc. Calcutta 7:
224. 1846.--Hook. f. & Thoms. in Hook. f. Fl. Brit. Ind. 1:
186. 1872.

Pentaloba bengalensis Wall. List. No. 4896. 1832, nomen nudum.--
A. Arnott in Mag. Zool. Bot. 2: 544. 1838.

Rinorea glandulosa sensu Merr. in Lingnan Sci. Jour. 6: 328. 1928,
non Merr. 1917.

Hainan: F. C. How 72303, 73903
S. K. Lau 26490, 27116
C. Wang 34112, 34442, 34635
How 70553

Figura 1. Tarjetas escritas a máquina

Existen tarjetas de tres tipos:

- tarjetas escritas a máquina con el nombre de la planta y referencia de la publicación de donde fue utilizada (un ejemplo se ilustra en la Figura 1).

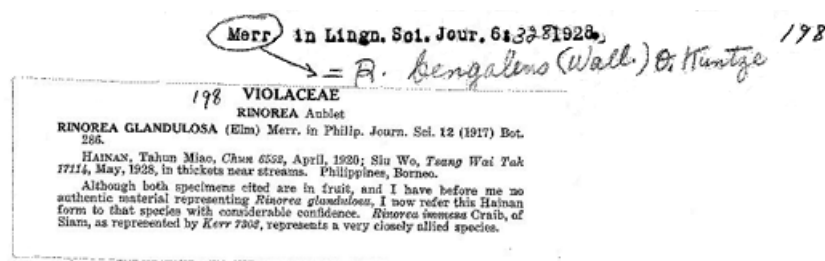


Figura 2. Tarjetas con recortes de revistas

- recortes de revistas que contienen el nombre de la planta y alguna cita bibliográfica (Figura 2).

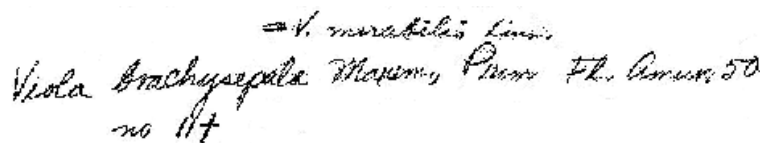


Figura 3. Tarjetas escritas a mano

- tarjetas escritas a mano con el nombre de la planta y alguna cita bibliográfica (Figura 3)

Muchas tarjetas contienen notas adicionales escritas a mano acerca de la planta. Las "tarjetas de Hu" están actualmente archivadas en el herbario de la Universidad de Harvard y son consultadas frecuentemente por investigadores de la Flora de China en Harvard y otros lugares del mundo.

Es necesario entonces, que este extraordinario recurso esté disponible en forma electrónica para que pueda ser utilizado de manera efectiva en el proyecto de Flora de China. Con este fin se ha iniciado un proyecto que incluye la digitalización de cada una de las tarjetas.

1.2.2 Lista para comprobación de nombres de Flora de China (FOCC)

Este es otro de los recursos actualmente disponible a los investigadores de FOC. Esta lista es la base en datos de taxa (plural de taxón) y sinónimos de taxa aceptados para Flora de China. El

desarrollo de esta lista ha sido realizado por varios años en el Jardín Botánico de Missouri. La meta es desarrollar una sinopsis computarizada de las 30,000 especies de plantas de China.

1.2.3 Lista anotada para comprobación de nombres de Flora de China (AFCC)

Uno de los objetivos del proyecto de FOC es unificar los recursos de información existentes sobre la flora de China. En particular, las Tarjetas de Hu y FOCC serán unificados en un solo producto, el cual se ha llamado "Lista Anotada para Comprobación de Nombres de Flora de China" (AFCC). Esto requerirá que los proyectos mencionados anteriormente sean terminados y por lo tanto que se desarrollen herramientas que permitan utilizarlo de forma adecuada.

1.3 Problemática

En el proyecto de Flora de China la información existente que representa la parte más importante de la biblioteca digital florística debe estar disponible por medio de herramientas que permitan accederla y manejarla. Como ya se ha mencionado anteriormente, en el caso de las "tarjetas de Hu", el acceso es lento y poco favorable para la investigación realizada en la edición de tratamientos, ya que las tarjetas se encuentran centralizadas en la Universidad de Harvard y son requeridas por un gran número de científicos distribuidos en diferentes partes del mundo. La búsqueda de información en las tarjetas puede ser tediosa y complicada pues sólo puede realizarse por los niveles taxonómicos de familia y género, dado que así están organizadas físicamente. Para obtener la información completa de cada planta, es necesario leer cada una de las tarjetas. Como parte del proyecto AFCC se ha decidido digitalizar las "tarjetas de Hu" y desarrollar mecanismos para facilitar su uso para los investigadores. Actualmente existe un sistema desarrollado en Java que permite visualizar las "tarjetas de Hu" así como otros sistemas que permiten acceder a la información de la flora de China. Estos serán descritos a continuación.

1.3.1 Navegador de Hu versión 0.1

Realiza el despliegue de cada una de las plantas a través de listas desplegadas por familia, género y especie en HTML (Figura 4). Estas plantas pertenecen a la Lista Anotada de la Flora de China. Permite el enfoque de la información de las tarjetas de Hu que han sido digitalizadas. El desarrollo del navegador esta basado en esquema de servicios en imágenes de la Biblioteca Digital Florística.



Family groups :

❖ Families A-F

❖ Families G-K

❖ Families L-O

❖ Families P-Z

Figura 4. Navegador de Hu versión 0.1

1.3.2 Navegador de Hu versión 1.0

El Laboratorio de Tecnologías Interactivas y Cooperativas de la UDLAP, en cooperación con CBI, ha desarrollado un sistema prototipo de navegación que permite al usuario visualizar las tarjetas de Hu, el sistema permite hacer la búsqueda escogiendo todas las tarjetas pertenecientes a una familia o bien todas las de una familia y algunos géneros seleccionados. El usuario puede realizar su búsqueda en imágenes ordenadas en grupos de 3 tarjetas (Figura 5).

La interfaz proporciona: un botón de despliegue de todas las familias existentes (Families) y un botón de despliegue de todas los géneros de la familia escogida (Genera). Como puede observarse, el usuario puede ir avanzando en las tarjetas de tres en tres y de la misma forma retroceder (por medio de los botones Next y Previous Cards). El sistema no proporciona una idea de la cantidad de tarjetas que cumplen con las características de familia y/o género escogidas por el usuario. No existe otra interacción entre el sistema y el usuario más que el hecho de escoger las tarjetas que desee visualizar.

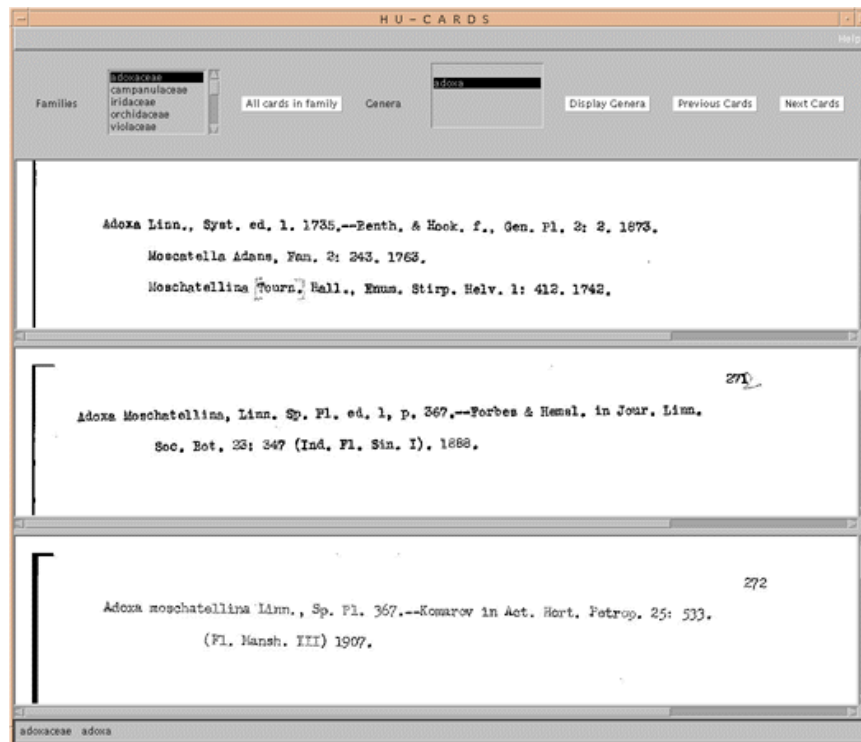


Figura 5. Sistema prototipo de navegación para las tarjetas de Hu.

1.3.3 Reconocimiento de texto en las tarjetas de Hu

Actualmente para la biblioteca digital se está desarrollando un proyecto de tesis paralelo a este, el cual realizará el reconocimiento del texto contenido en las tarjetas de Hu. Este proyecto trabajará con la interfaz de visualización de tarjetas. La interfaz podrá proporcionar al usuario la información completa contenida en las tarjetas que por ahora no está disponible. En ella se podrá realizar una búsqueda más específica y mucho más rápida a la que existe.

1.3.4 Taxonomías Múltiples

La taxonomía vegetal o botánica sistemática es la ciencia que estudia la identificación, clasificación, nomenclatura y evolución de las plantas en todas sus variedades [Jones et al. 1986]. La taxonomía representa un dispositivo humano para estructurar ordenadamente la información del total de las especies [Sivarajan 1991]. Los principales aspectos de la taxonomía son la identificación, la clasificación y la nomenclatura. La identificación es el proceso de reconocer a una identidad ya sea como semejante o diferente a una identidad conocida. La clasificación involucra el establecimiento y definición de grupos taxonómicos, y colocar apropiadamente a los elementos en tales grupos. La nomenclatura consiste en asignar un nombre científico a tales grupos, de tal manera que puedan ser referidos y comunicados por ese nombre y no por largas frases descriptivas.

Dentro de las jerarquías taxonómicas, la unidad básica de la clasificación biológica es la especie y alrededor de esta se construye toda la estructura taxonómica. Una especie consiste en todos los

individuos similares en la suma de las características esenciales. Las especies se agrupan en géneros, los géneros en familias, y así sucesivamente en grupos superiores de elementos similares. Cada grupo se denomina genéricamente un taxón, y estos se sitúan en diferentes niveles y rangos de la taxonomía llamados categorías [Flores 1997].

Las principales categorías usadas en la taxonomía son, en orden ascendente: especie, sección, género, tribu, familia, orden, clase, división y reino, donde el reino Plantae es la categoría superior que agrupa a todos los elementos del reino vegetal. Cuando en un grupo se incluyen varios elementos diferentes que son tan distintos como para ser separados en grupos inferiores, se pueden agrupar en subgrupos de categorías intermedias, agregando el prefijo sub a la categoría inferior inmediata, ejemplos de subcategorías son: subgénero, subfamilia y subclase. También existen infracategorías, de las cuales los principales son: subespecie, variedad y forma.

La diferencia de opiniones entre los taxonomistas que asignan las posiciones diferentes de un mismo elemento dentro de una jerarquía taxonómica origina naturalmente taxonomías múltiples. El avance de los conocimientos biológicos y la evolución en el concepto de taxón o grupo taxonómico, hace que se realicen cambios por múltiples autores a la clasificación y por lo tanto a la nomenclatura de cada una de las plantas, siendo esta también la razón por la que existen las taxonomías múltiples.

El principio básico de la clasificación es la identificación, dado que es el medio a través del cual los sistemas de clasificación se construyen. La identificación es el proceso de decidir en la opinión del taxonomista si dos cosas son o no semejantes. Para clasificar plantas, se compara y se usa el juicio personal para decidir si un objeto es similar a otro para colocarlo en el mismo grupo. Debe notarse que la identificación involucra tomar una decisión, lo cual es obviamente una fuente de posible desacuerdo entre taxonomistas, puesto que distintas personas piensan de manera diferente. Esto significa que la decisión tomada es un elemento subjetivo que varía de persona a persona. Cada persona tiene su propio concepto del significado de semejanza, y depende del criterio de cada persona decidir si dos cosas son o no semejantes.

1.3.5 Resumen de la problemática

El procedimiento de acceso a información taxonómica utilizado actualmente en la Biblioteca Digital Florística es lento y tedioso para los usuarios. Esto se debe a que deben realizar la búsqueda de información a través de una serie de listas en páginas de HTML como es el caso del navegador de Hu versión 0.1 (Figura 4) o a través de un conjunto de tarjetas desplegadas de tres en tres como el navegador de Hu versión 1.0 (Figura 5). Los problemas de acceso son mayores si se considera la posibilidad de que existan taxonomías múltiples. Dados

el volumen y complejidad de la información mencionadas es deseable contar con mecanismos que faciliten al usuario la contextualización de los datos y su visualización en forma gráfica.

1.4 "3Dtree": Visualización de grandes volúmenes de información en una Biblioteca Digital Florística.

Como parte del presente proyecto, se desarrolló 3Dtree, un prototipo navegador que despliega la información de la Biblioteca Digital Florística de una forma gráfica. Esta información, que está almacenada en la base de datos, esta organizada jerárquicamente, como se mencionó en la sección 1.3.4 (familia, género, especie, etc.). Esta estructura se representa como un árbol en 3 dimensiones en el cual el usuario puede navegar a través de los nodos que él escoja, hasta llegar a aquella planta de la que requiere información específica. Se puede encontrar una explicación más amplia de 3DTree en los capítulos 3 y 4 de este documento.

1.5 Organización del documento

El documento será organizado de la siguiente forma: en el capítulo 2 se mencionarán y explicarán las características más importantes y aplicables a este proyecto de tesis acerca de otros trabajos que se han realizado para la visualización de información. En el capítulo 3 se describirá el diseño del navegador propuesto. El capítulo 4 se encarga de la presentación del prototipo de 3Dtree y por último en el capítulo 5 se concluye con una síntesis del proyecto y se revisa el trabajo a desarrollar en el futuro.

índice resumen 1 2 3 4 5 A B C referencias

Amavizca Ruiz, L. M. 1998. **3DTree: Visualización de grandes volúmenes de información en una Biblioteca Digital Florística.** Tesis Licenciatura. Ingeniería en Sistemas Computacionales. Departamento de Ingeniería en Sistemas Computacionales, Escuela de Ingeniería, Universidad de las Américas-Puebla. Diciembre.
Derechos Reservados © 1998, Universidad de las Américas-Puebla.