

Herramienta para el desarrollo conjunto de una base de conocimientos a través de Internet

Mónica Trella López
trella@apolo.lcc.uma.es

David Bueno Vallejo
bueno@lcc.uma.es

Ricardo Conejo Muñoz
conejo@lcc.uma.es

Dpto. de Lenguajes y Ciencias de la Computación. Universidad de Málaga, Complejo Tecnológico, Campus de Teatinos, 29071 Málaga (España)

Resumen

El desarrollo de Sistemas Tutoriales Inteligentes (STI) requiere, durante la fase de modelado de la base de conocimientos del dominio, el trabajo de expertos en la materia que se está modelando. A veces los expertos encargados de aportar el conocimiento y los ingenieros del conocimiento dedicados a la adquisición de éste no coinciden en el tiempo ni en el espacio, por lo que el proceso se ralentiza bastante. Para solventar estas dificultades se ha desarrollado un sistema para la adquisición de conocimientos a través de Internet. Esta herramienta ha sido construida para el dominio de conocimiento de morfología botánica.

Palabras clave: Base de conocimientos, Sistemas Tutores Inteligentes, Internet, Botánica, Especies Forestales. Knowledge Base, Intelligent Tutoring Systems, Internet, Botany, Forestry Species.

1. Introducción

El desarrollo de Sistemas Tutoriales Inteligentes (STI) presenta en la actualidad dificultades importantes entre las que destaca la fuerte dependencia del dominio para el que se pretende realizar la instrucción [1]. Esta estrecha relación implica que durante el desarrollo de un STI, y más concretamente durante la fase de modelado de la base de conocimientos del dominio, se requiera con frecuencia la ayuda de expertos en la materia o dominio que se está modelando. Sucede a menudo que los expertos encargados de aportar el conocimiento y los ingenieros del conocimiento dedicados a la adquisición de éste no coinciden en el tiempo y en el espacio, por lo que el proceso se ralentiza bastante. Estos inconvenientes se plantearon durante la planificación del desarrollo de un STI para enseñar a clasificar e identificar especies forestales europeas. Este sistema forma parte del proyecto "*TREE: Training of European environmental trainers and technicians in order to disseminate multinational skills between European countries*" [2], encuadrado en el programa Leonardo Da Vinci de la UE. Para solventar estas dificultades se ha desarrollado un sistema para la adquisición de conocimientos a través de Internet.

En este artículo se explicará en primer lugar como se ha modelado el dominio de conocimientos de morfología botánica (necesario para el desarrollo del módulo del dominio del STI), para pasar a continuación a describir detalladamente el sistema de adquisición de conocimientos (partes y funcionamiento). Dentro de esta descripción se incluye la de una herramienta auxiliar de traducción utilizada para generar las páginas HTML en diversos idiomas.

2. Modelado del dominio.

El dominio sobre el que trabaja el STI que se está desarrollando es el de especies forestales europeas. Se trabaja sobre un conjunto de trescientas treinta y tres especies que viven en Europa. El objetivo del STI es que el alumno aprenda a identificar el máximo número posible de estas especies. Para construir la base de conocimientos es necesario estructurar el dominio y a continuación obtener de los expertos toda la información necesaria para definir y trabajar con las unidades fundamentales de conocimiento, que en este caso son las especies forestales. Para ello se estudiarán todas las características con las que se pueden definir e identificar unívocamente una especie.

2.1 Estructura jerárquica del dominio de especies forestales.

El dominio de especies forestales está organizado siguiendo una estructura jerárquica que está formada por varios niveles: divisiones, clases, órdenes, familias, géneros, especies y subespecies.

La especie es la categoría taxonómica fundamental. De manera simple podemos definir la especie como un conjunto de organismos que:

- poseen un importante número de caracteres en común (comparten un patrimonio genético)
- son interfértiles (forman poblaciones)
- y que en condiciones naturales no intercambian dichos caracteres con el resto de los organismos (aislamiento reproductivo).

Las especies poseen también caracteres en común que sirven para agruparlas en géneros. Los géneros se pueden agrupar en familias y así sucesivamente. Esta ordenación de grupos dentro de grupos de forma creciente constituye pues un sistema jerárquico o jerarquía de clasificación. Para modelar la base de conocimientos se ha seguido la misma estructura, pero omitiendo los niveles de *clases* y *órdenes* (Fig. 1) porque se ha considerado que no aportaban ninguna información relevante a la hora de realizar el reconocimiento e identificación de una especie forestal (objetivo final del STI).

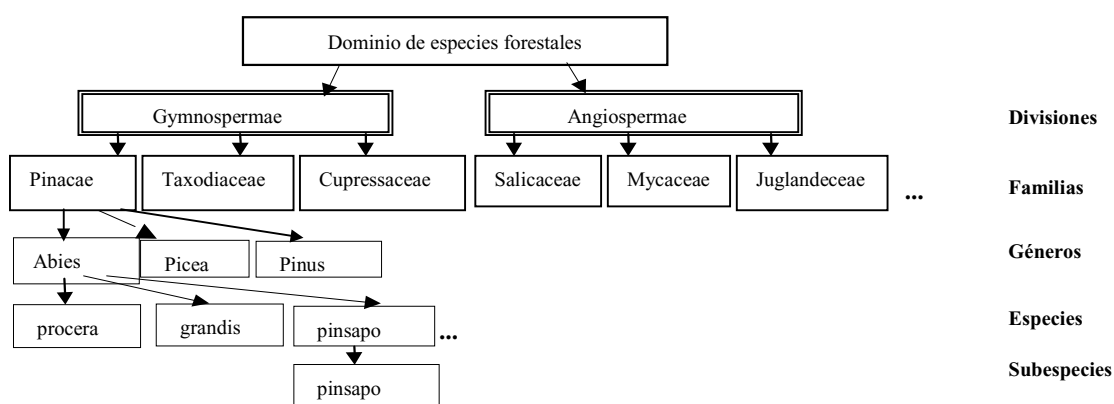


Fig. 1 Estructura jerárquica del dominio

2.2 Características morfológicas de las especies forestales.

Para identificar una especie forestal y diferenciarla del resto de especies basándose en la observación directa, ya sea al natural o mediante fotografías, es necesario conocer sus características morfológicas.

Los expertos en botánica han considerado un conjunto de ciento cuarenta y tres características necesarias para clasificar el total de especies que abarca el proyecto TREE. Estas características han sido agrupadas en doce categorías (Porte, Copa, Ramas, Persistencia, Corteza, Ramillas, Yemas, Hojas, Brácteas, Flores, Frutos, Semilla) correspondientes a las principales partes de una especie forestal que son observables a simple vista y que pueden servir para su identificación.

Cada una de estas características lleva asociado un conjunto de valores que es suficiente para cubrir todo el conjunto de especies que se está utilizando. Por ejemplo, la característica *FISONOMÍA DEL PORTE* puede tomar cualquier valor de los dados por el siguiente conjunto:

- Sin fisonomía característica
- Con posibilidad de brotes de cepa
- Con posibilidad de brotes de raíz
- Estolonífero
- Sin caracterizar

Existen dos tipos de características, desde el punto de vista del modelo del dominio:

- Univaluadas: son aquellas que pueden tomar sólo un valor de su conjunto asociado para una especie determinada.
- Multivaluadas: las características de este tipo pueden tomar varios valores para la misma especie. Estos valores pueden darse simultáneos en un mismo individuo de la especie (p. ej. un árbol presenta a la vez hojas de disposición helicoidal y alterna) o a la vez dentro de la misma especie pero no en el mismo individuo.

Existen características que no son aplicables a algunas especies, por lo que tomaran en algunos casos valores nulos.

3. Sistema de adquisición de conocimientos.

Una vez determinada la información que se necesita obtener de los expertos es necesario diseñar una metodología de trabajo y un soporte de almacenamiento que minimice el esfuerzo de los expertos y que permita que el tratamiento de la información se realice de forma automática evitando así los errores debido a la manipulación de los datos por personas no expertas en la materia.

Se ha desarrollado un sistema de adquisición de conocimientos a través de Internet. De este modo es posible coordinar el trabajo de distintos expertos proporcionando una plataforma común y de fácil acceso sobre la que realizar el trabajo y estar al tanto del progreso del proyecto, pudiendo así colaborar y realizar aportaciones al trabajo de los demás.

Desde el punto de vista del manejo de los datos y del desarrollo de la herramienta informática, se trabaja con información validada, actualizada y no duplicada que puede ser usada desde el mismo momento en el que el experto la almacena a través de Internet.

3.1 Estructura de la base de conocimientos.

La base de conocimientos ha sido desarrollada siguiendo el modelo jerárquico descrito en el apartado 2.1. Cada nivel o categoría taxonómica de las consideradas en el modelo representa una clase (según el paradigma de programación orientada a objetos) y cada elemento de una clase es un objeto. Según esto consideraremos que existen cuatro clases: división, familia, género y especie (los niveles de especie y subespecie se unen formando uno sólo) y serían objetos de la clase división *Gymnospermae* y *Angiospermae*; de la clase familia *Pinaceae*, *Taxodiaceae*...; etc. Las características morfológicas son los atributos de las clases. Se han implementado mecanismos de herencia, de modo que un objeto hereda los atributos de su padre o predecesor en la jerarquía. Por ejemplo, el experto asigna valores a todas las características para la familia *Pinaceae* y automáticamente todas las características de los géneros pertenecientes a esa familia (*Abies*, *Picea*,...) pasan a tener los mismos valores. De este modo para definir por ejemplo el género *Abies*, sólo habría que asignar valores a aquellas características que no hayan sido definidas para su familia. Debido a que existen excepciones y que no todos los objetos de una clase tendrán el mismo valor para la misma característica, los expertos pueden redefinir valores heredados.

La base de conocimientos contiene toda la información morfológica sobre el conjunto de especies forestales europeas que el STI enseñará a identificar, además de fotos de las principales características de cada especie y mapas con la localización geográfica aproximada.

3.2 Generación dinámica de páginas HTML.

El sistema expuesto en esta comunicación ha sido desarrollado para construir una base de conocimientos a través de Internet. Una parte de la base de conocimientos, como se explica en el apartado anterior, es fija y la constituye la estructura de objetos creados que se corresponden a todos los géneros, especies, etc. Esa estructura es un esqueleto que los expertos irán completando a través de Internet.

Las páginas HTML del sistema de desarrollo de la base de conocimientos, consisten fundamentalmente en un conjunto de formularios que los expertos deberán ir rellenando para cada especie. Si se tiene en cuenta el elevado número de especies y de características que hay que "rellenar" para cada una de ellas, se llega a la conclusión de que lo mejor es utilizar la información de la estructura de la base de conocimientos para que las páginas se generen de forma dinámica según las opciones que elija el usuario. En vez de tener formularios preparados en el servidor para cada especie y grupo de características, cada página contiene instrucciones para buscar en la base de conocimientos los datos que el usuario determine en cada momento de su *navegación* (qué características quiere consultar, qué conjunto de valores tiene asociado, qué valores tiene para la especie indicada por el usuario, etc.).

De este modo se facilita el mantenimiento de las páginas ya que si la estructura de la base de datos se modifica (se añade alguna característica, se modifica algún conjunto de valores,...) no es necesario revisar las páginas HTML y modificarlas ya que las páginas se cargan en el *navegador* utilizando la información que encuentren en la base de conocimientos en el momento mismo en el que el usuario las ha solicitado.

3.3 Sistema de traducción .

En el proyecto TREE (apdo. 1) participan diferentes países de la Unión Europea por lo que el sistema debe estar traducido a al menos todos los idiomas de los países participantes.

Las páginas de este sistema contienen un vocabulario muy técnico específico del dominio de aplicación, pero sin embargo el volumen de texto requerido no es muy grande y suelen emplearse frases aisladas que no forman parte de un discurso completo. En estas circunstancias, y en especial debido al fuerte componente técnico del lenguaje empleado, los traductores automáticos de lenguaje natural de propósito general no son una solución factible. Para realizar esta tarea se requiere la cooperación de los distintos grupos participantes en el proyecto, ya que precisamente son estos grupos los mas indicados para realizar traducciones técnicas.

Una solución a este problema es un sistema de traducción que consiste en crear unas *tablas de traducción* donde se almacenan los términos y frases que se han introducido en la base de conocimientos, así como sus correspondientes traducciones

El proceso de traducción consiste en la traducción frase a frase y término a término de los registros de la tabla de traducción. Para dar soporte a esta tarea se ha desarrollado una interfaz Web a través de la cual los distintos grupos acceden a esta tabla. Este enfoque tiene la ventaja de que la traducción puede llevarse a cabo de forma simultánea a la realización del proyecto sin que tenga que esperarse a la conclusión del mismo para generar las versiones en otros idiomas distintos del idioma usado durante la fase de desarrollo. Además se pretende con este módulo que los datos puedan ser utilizados de forma automática y sencilla por los esquemas de páginas Web del sistema que los requieran.

Su funcionamiento es el siguiente: el experto se conecta a través de Internet a la interfaz de traducción. En la primera página el usuario debe elegir los idiomas origen y destino de la traducción, es decir la lengua a la que va a traducir y el/los idiomas base en los que leerá los términos que debe traducir. Además se ofrece la posibilidad de filtrar la lista de términos que se quieren traducir introduciendo un patrón de búsqueda y seleccionando todos los términos, o bien los que quedan por traducir, o bien los ya traducidos (si se quiere realizar una modificación sobre ellos). Una vez que el usuario haya determinado qué términos desea traducir debe identificarse introduciendo su nombre de usuario y contraseña. Este sencillo sistema de control de usuarios evita que personas no autorizadas modifiquen datos que son compartidos en el desarrollo global de la aplicación. Cada experto sólo tendrá permiso para modificar los datos en su idioma pero puede consultar los términos en los demás idiomas.

Cuando el usuario ha introducido todos los datos requeridos para realizar la traducción aparece una página Web que contiene una tabla con tantas columnas como idiomas haya seleccionado en el formulario anterior. Cada tabla tendrá como máximo veinte filas con términos que se van a traducir. La última columna estará formada por cuadros de texto para que el experto escriba la traducción de los términos correspondientes a esa fila. Una vez que el usuario ha realizado su trabajo pulsará el botón que aparece al final de la página y todos los nuevos términos serán introducidos en la tabla de traducción. A continuación la interfaz presentará una nueva página con los veinte siguientes términos y se repetirá el proceso. Cada término lleva asociado un identificador que será el mismo para cada idioma al que haya sido traducido y que se usará en la generación dinámica de páginas. Cuando el experto se conecte al sistema de desarrollo de la base de conocimientos, verá todas las páginas en el idioma seleccionado. Si algunos términos aun no han sido traducidos aparecerán en inglés y resaltados en rojo.

3.4 Funcionamiento.

El sistema de desarrollo de la base de conocimiento es de fácil manejo con el fin de que personas no familiarizadas con el uso de los ordenadores no encuentren dificultades añadidas a la tarea que deben realizar.

Cuando el usuario se conecta a la página principal¹ debe seleccionar el idioma en el que va a trabajar pulsando una de las banderas que se encuentran en ella. A continuación aparecerá una página a través de la cual se puede acceder al sistema de desarrollo de la base de conocimientos, o bien, a otras herramientas de consulta de acceso no restringido.

El usuario experto que desee entrar en el sistema de desarrollo deberá identificarse mediante un nombre de usuario y una clave que le permitirán modificar las especies de la base de conocimientos asignadas a su grupo de trabajo. No todos los usuarios están autorizados a

¹ <http://www.lcc.uma.es/TREE> (Herramientas de desarrollo de la base de conocimientos, Login: *demo*, Password: *demo*)

3.5 Diferencias y parecido s

Para establecer estrategias tutoriales eficientes en el STI, es necesario conocer las principales diferencias entre las especies. Esta información puede obtenerse automáticamente de la base de conocimientos, pero esto no sería del todo útil para los propósitos didácticos del STI. Se ha construido una herramienta Web, como parte del sistema de desarrollo de la base de conocimientos, para señalar las diferencias más importantes entre especies (divisiones...).

Realizar la comparación de todas las especies dos a dos es una tarea muy tediosa ya que los expertos tendrían que revisar una tablas de 333x333 especies, sin contar los otros niveles. Para reducir este número se ha decidido realizar sólo las comparaciones entre objetos de la misma clase inmediatamente superior en la jerarquía, es decir, entre especies del mismo género, géneros de la misma familia y familias de la misma división.

El entorno de este sistema es muy parecido al explicado hasta ahora. El usuario en lugar de una lista de especies tendrá dos. Cuando seleccione una especie de la lista de la izquierda, aparecerán a la derecha todas las especies de su mismo género. Eligiendo una de ellas, el experto accederá a una lista de todas las diferencias obtenidas de comparar la información guardada en la base de conocimientos de las dos especies y él deberá señalar las más significativas, entendiendo por éstas las que permitan diferenciar mejor a simple vista una especie de otra.

Del mismo modo el experto almacenará los parecidos más importantes entre dos especies, es decir, aquellas características que pueden hacer que dos especies se confundan entre sí a la hora de identificar una de ellas.



Fig. 4 Envío y anotación de fotografías

4. Conclusiones

En este trabajo se ha realizado el modelado del dominio de conocimientos de morfología botánica sobre el que se está creando un STI [3] para ayudar a la identificación de especies forestales Europeas. Con el fin de ayudar a la construcción de la base de conocimientos se ha desarrollado un sistema de adquisición de conocimientos a través de Internet, así como una herramienta auxiliar de traducción utilizada para generar las páginas HTML en diversos idiomas. En un futuro se pretende generalizar este sistema de adquisición de conocimientos para cualquier dominio y en especial para dominios estructurados jerárquicamente.

5. Referencias

[1] Murray, T. (1998): Authoring Knowledge Based Tutors: tools for content, instructional strategy, student model and interface design In Journal of the Learning Sciences, Vol. 7, N°1, 1998, pp-5-64

[2] Trella, M., Conejo, R., Pérez de la Cruz, J.L. & Triguero, F. (1997): Un sistema concha para WWW. In II Jornadas de Transferencia Tecnológica de Inteligencia Artificial, 1997.

CESAR:

[3] Lelouche, R. (1998): *The successive contributions of computers to education: a survey.*
In European Journal of Engineering Education, 1998.