

# Acceso inteligente a sistemas de búsqueda de información por usuarios no especializados

David Bueno Vallejo    Mónica Trella López    Ricardo Conejo Muñoz    José L. Pérez de la Cruz  
[bueno@lcc.uma.es](mailto:bueno@lcc.uma.es)    [trella@apollo.lcc.uma.es](mailto:trella@apollo.lcc.uma.es)    [conejo@lcc.uma.es](mailto:conejo@lcc.uma.es)    [cruz@lcc.uma.es](mailto:cruz@lcc.uma.es)

Dpto. de Lenguajes y Ciencias de la Computacion. Universidad de Malaga, Complejo Tecnológico, Campus de Teatinos, 29071 Malaga (Spain)

## Resumen

El acceso a los sistemas de bases de datos puede ser complejo si los usuarios no tienen un conocimiento amplio sobre sus contenidos. Por eso, es necesario ayudar de alguna forma a ese tipo de usuario a encontrar la información que buscan sin preguntarles directamente por los datos que el sistema contiene. Para esto se pueden utilizar sistemas expertos que emulen a consultores humanos. En nuestra aplicación el sistema intenta emular a un asesor forestal que recomienda las mejores especies arbóreas para reforestar en una región dada formulando preguntas simples y comprensibles por el usuario. Entre otras características el sistema es multilingüe y multialfabeto.

**Palabras clave:** recuperación de la información, sistemas expertos, bases de datos en la www, sistemas multilingües, trabajo cooperativo

## 1. Introducción

Los sistemas de información actuales se caracterizan por su gran volumen de datos y la complejidad en la organización de estos. Esta estructuración permite hacer consultas a la base de datos que no son fácilmente realizables en otro tipo de sistemas. Aunque se disponga de una gran cantidad de información, en la mayoría de los casos nos encontramos con sistemas que necesitan cierta cualificación en dos aspectos. (1) Es necesario conocer el funcionamiento del sistema para poder sacar el mayor beneficio de las opciones y consultas que ofrece, (2) también es necesario que el usuario domine el área de conocimiento de los datos que almacena ese sistema, pues la información puede ser demasiado específica como para que sea comprensible por cualquier tipo de usuario.

En el proyecto europeo Ri.T.A (La RInaturalizzazione del Territorio Agricolo) y concretamente en el subproyecto ATRIS nos encontramos con ese problema y encontramos una solución para la utilización de ese sistema por usuarios no especializados. El objetivo de este proyecto es el de sustituir en cierta medida a un asesor forestal. Cuando un agricultor quiere realizar una reforestación en sus terrenos, sólo podrá plantar unos tipos de especies que se adecuen a las características de la zona. Además le interesará hacer la reforestación con especies subvencionadas por la Unión Europea. Saber qué especies son las más recomendables para una zona no es tarea sencilla, ya que depende de muchos factores que quedan muy lejos del conocimiento del agricultor.

El objetivo prioritario de este proyecto es facilitar el acceso de estas personas a la información que necesitan. Para alcanzar ese objetivo se ha trabajado en dos vertientes. Por un lado se integró el sistema en la WWW haciendo accesible la información del sistema a todos los usuarios de Internet. La segunda opción de accesibilidad es más compleja debido al contenido tan específico de la base de conocimiento. Para hacer esto último se han utilizado técnicas de Ingeniería del Conocimiento en las que mediante diversas entrevistas con un ingeniero forestal se pudo obtener el conocimiento que tiene un agricultor y relacionarlo con el conocimiento de nuestro sistema de información de forma que este pueda generar la respuesta óptima para este tipo de usuarios. Además se ofrecen otras interfaces para usuarios con niveles de conocimiento más avanzados.

## 2. Características del sistema

La integración del sistema dentro de la WWW tuvo varios objetivos, por un lado se trataba de facilitar la inserción de los datos por los distintos expertos. Debido al ámbito en el que se sitúa el proyecto existe una dificultad inherente a la hora de compartir información, pues los componentes de España, Francia, Grecia, Italia y Portugal deben aportar sus datos para formar una única base de conocimiento sobre reforestación en

estos países. Con las herramientas que hemos desarrollado se permite de forma remota introducir y hacer modificaciones de estos datos desde cualquier parte de Internet teniendo los permisos adecuados.

El segundo objetivo de la realización del sistema en la WWW es facilitar el acceso a cualquier usuario que desee realizar una reforestación. Lo único que necesita es una conexión a la WWW y podrá realizar la consulta.

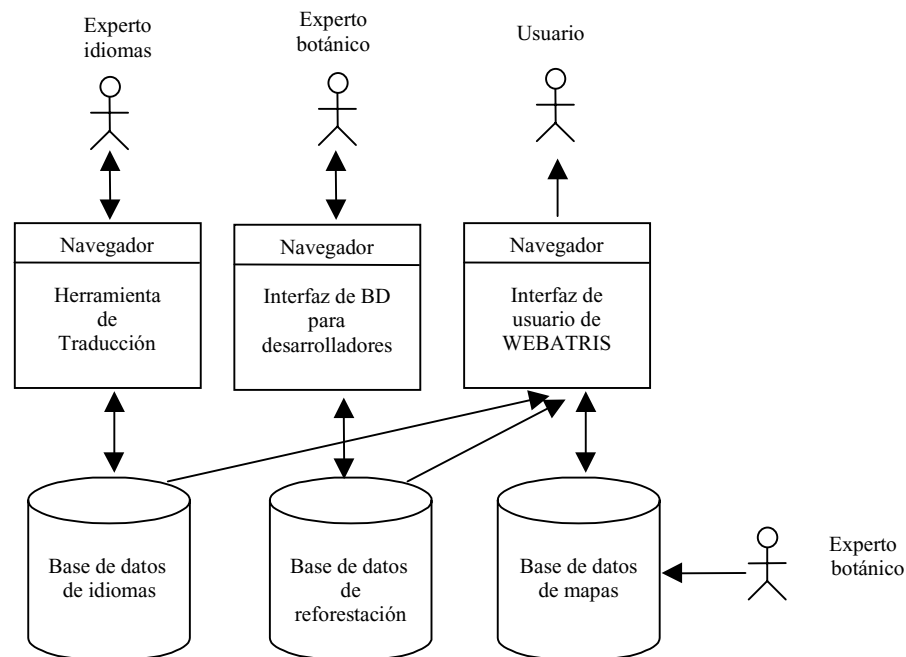


Fig. 1. Estructura general del sistema

En la Fig. 1 se muestra la estructura general del sistema que podría dividirse en dos grupos de componentes, por un lado las bases de datos y por otros las interfaces a través de la WWW.

## 2.1 Bases de datos

Las bases de datos que forman el sistema ATRIS son:

- *BD. de idiomas.* - Una de las características de nuestro sistema es la posibilidad de trabajar con el en multiples idiomas. Esta base de datos contiene una lista de equivalencia de palabras y expresiones. El usuario al comienzo del programa seleccionará un idioma que será utilizado en el resto de la sesión.
- *BD. de reforestación.* - Esta base de datos contiene toda la información técnica sobre los distintos árboles que pueden ser reforestados. Incluye características de estos que pueden ser diferentes dependiendo de las regiones o países. Algunas de éstas son: fitoclima, tipo general de suelo, de suelo, objetivos de reforestación, tipo climático estructural, etc.
- *BD. de mapas.* - En la BD de reforestación hay algunos campos que normalmente no son conocidos por usuarios no expertos. Un ejemplo de esto es el *fitoclima*. Por lo tanto para evitar preguntar directamente por este tipo de campos, como pueden ser calculados automáticamente conociendo la región a reforestar se necesita tener una colección de mapas fitoclimáticos, que irán acompañados de mapas políticos donde un usuario podrá seleccionar con un simple clic donde se localiza el terreno a reforestar.

## 2.2 Interfaces

El sistema ofrece 3 tipos de interfaces, 2 para diseñadores y uno para los usuarios finales del sistema:

- *Herramienta de traducción.* - Con esta herramienta se permite a los expertos de cada país traducir los términos que se utilizan en el sistema a su propio idioma. El sistema está diseñado de forma que no es necesario tener traducido todos los términos a un idioma para que se puedan hacer consultas. En el caso de que un usuario intente interactuar con el sistema en un idioma dado se le ofrecerán todas las frases o palabras que hayan sido traducidas y las otras se darán en el idioma por defecto que es el ingles. Las características de este sistema se detallan en [1].

- *Interfaz para desarrolladores.* - Los expertos botánicos de los diferentes países rellenan la BD utilizando la WWW. Para ello cada uno debe tener unas limitaciones de acceso y unos permisos de escritura para modificar solamente las partes de la BD que le correspondan.
- *Interfaz de usuario (WEBATRIS).* - Éste podría considerarse la interfaz del programa final, pues los dos anteriores están orientados a los desarrolladores del sistema, ya sean expertos botánicos o traductores. A través de esta interfaz el usuario podrá realizar las consultas al sistema que darán como resultado los árboles más indicados para reforestar en un determinado lugar en función de las características de este y de los objetivos de la repoblación del usuario. En el apartado siguiente se analiza este sistema en detalle.

### 3. WEBATRIS

El usuario final que accede a nuestro sistema a través de la WWW debe realizar en un primer momento la selección del país en el que desea hacer la reforestación y del idioma en el que quiere interactuar con el sistema. Para permitir al usuario trabajar con el sistema en diferentes idiomas y no tener que diseñar la misma página en diferentes idiomas, los documentos HTML del sistema son plantillas que tienen un código asociado para cada expresión o palabra. Cuando el usuario carga una página, el sistema convierte ese código en la expresión correspondiente en el idioma seleccionado por el usuario.

Después de la selección del país y del idioma el sistema dispone de 3 posibles tipos de consulta: Consulta directa, consulta ampliada, consulta asistida.

**Consulta directa.** - Este tipo de consulta está orientado a usuarios finales cuyo conocimiento sobre botánica o reforestación es alto. Lo que se realiza es una consulta directa a la base de conocimiento para localizar las características de un determinado tipo de árbol indicando o no un fitoclima determinado. En esta consulta se puede utilizar el nombre común, nombre científico y/o fitoclima.

**Consulta ampliada.** - Esta opción ha sido desarrollada para usuarios con un nivel medio/alto sobre reforestación y que comprenden los campos con los que esta hecha la base del conocimiento. Se ofrece al usuario la mayoría de los componentes de la base de datos de reforestación y se permiten hacer consultas tan complejas como se quiera. Para la realización de las consultas se utiliza la técnica de *coincidencia de patrones (pattern matching)*. Aquí el usuario selecciona sólo los campos que desea que cumplan unas determinadas condiciones y con eso se realizará la consulta al sistema.

Para poder realizar esta consulta, se crean los siguientes grupos: nombre científico, nombre común, fitoclima, fisionomía, tipo climático estructural, tipo general de suelo, nivel evolutivo de la fitocenosis recomendada para su instalación, carácter autóctono, especies acompañantes, exposición y objetivos de la repoblación.

El criterio para realizar la consulta consiste en realizar una O entre los elementos que forman parte del mismo grupo y una Y entre elementos de diferentes grupos. Por ejemplo, si un usuario selecciona dentro del grupo *exposición (vaguada solana y red de drenaje)* y del grupo *tipo general de suelo (calizo y silíceo)*. La consulta que se realizaría al sistema sería: "Dame todos los árboles (cuya exposición pueda ser *vaguada solana* O *red de drenaje*) Y (cuyo tipo general de suelo pueda ser *calizo* O *silíceo*).

Todas esas interacciones con el sistema se realizan utilizando una sencilla e intuitiva interfaz. Tanto este tipo de consulta como el anterior han sido probadas y validadas por diferentes expertos encontrándolas como una herramienta interesante de fácil manejo y con datos que no se pueden encontrar fácilmente.

**Consulta asistida.** - Esta consulta está orientada a usuarios cuyos conocimientos sobre reforestación son bastante bajos y necesitan de la ayuda de un experto para saber que especie es la más recomendable para reforestar en su terreno. Hay preguntas información de la base de datos que es difícil de obtener realizando preguntas al usuario. Pero si es fácil de saber su valor si se conoce la situación exacta de la zona a reforestar. Por lo tanto, si un usuario es capaz de localizar su terreno en un mapa, el sistema puede deducir esos parámetros de forma automática. Esta es la primera parte de la consulta asistida. El usuario deberá localizar un punto de su terreno accediendo a través de mapas que van limitando las regiones. Primero se selecciona el país, después la región dentro del país, y por último en un mapa con una escala mucho mayor se selecciona el punto donde se encuentra la zona a reforestar. Para diseñar esta facilidad se han realizado mapas políticos y específicos de toda Europa. Los mapas políticos son los que el usuario utiliza en su

tarea de localización y los mapas específicos (como el fitoclimático) son los que utiliza el sistema para extraer información. La obtención de información se realiza mediante applets de Java. Un ejemplo puede verse en la Fig. 2

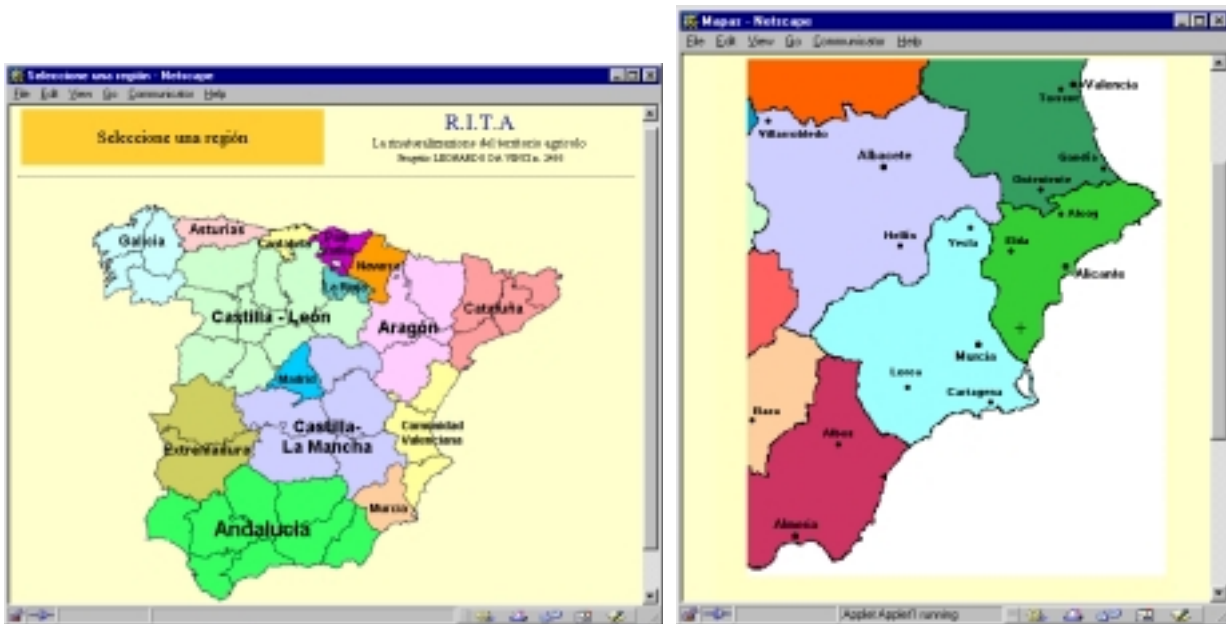


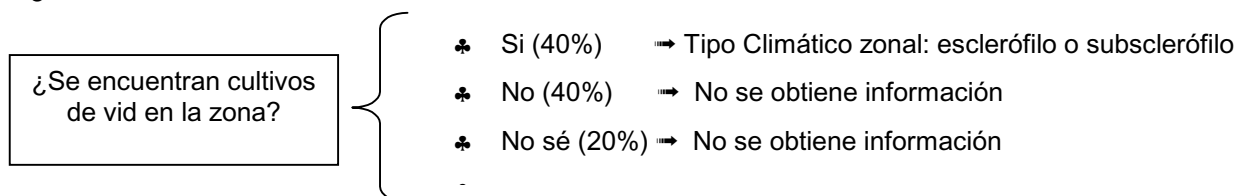
Fig. 2 Ejemplo de acceso a los datos mediante mapas

Después de seleccionar la zona a reforestar el sistema a podido reducir en gran medida el número de especies candidatas para la reforestación ya que dependiendo de la región habrá un conjunto de especies que son incompatibles. Cuando el sistema no puede obtener más información de los mapas es el momento de interactuar, mediante preguntas, con el usuario. De eso se encarga el sistema experto cuyo funcionamiento se detalla en el apartado siguiente.

#### 4. El sistema experto

La selección de la región en los mapas ayuda al sistema a conocer algunas características intrínsecas del suelo que son propias de una región e independientes del estado del suelo. Además de estas características hay otras que no pueden obtenerse de otra forma que no sea consultando a alguien que tenga información sobre el estado actual del terreno (nivel evolutivo del suelo), algunas características propias no disponibles en los mapas (tipo general de suelo, textura, tipo climático estructural zonal e intrazonal, exposición) y algo muy personal del usuario como los objetivos de la reforestación (obtención de madera, leña, resinas, corcho o para la caza, fauna, apicultura, etc.).

El gran problema que aparece es que la persona que posee el terreno no tiene que conocer todos los términos específicos, por lo que no se le podría preguntar directamente si el tipo climático es Esclerófilo, halohidrófilo o hiperxerófilo o si el tipo de suelo es calizo, dolomítico o silíceo entre otros. Por eso, utilizando técnicas de ingeniería del conocimiento se entrevistó a un experto ingeniero forestal para ver qué preguntas realizaban los asesores forestales para poder reconocer esos parámetros. De ahí se obtienen reglas como la siguiente:



En el desarrollo del sistema y después de realizar algunas pruebas, algunos usuario querían dejar de contestar preguntas, o ver, en función de la información que tenía el sistema cuales eran las posibles especies a reforestar. Para eso el sistema permite en cualquier momento ver la solución actual hasta el momento. Se seguirán haciendo preguntas mientras haya soluciones posibles y tenga más datos que calcular o cuando el usuario desee dejar de consultar pues piense que la información actual es suficiente. En la Fig. 3 se puede ver un ejemplo de ejecución del experto y de una solución parcial.

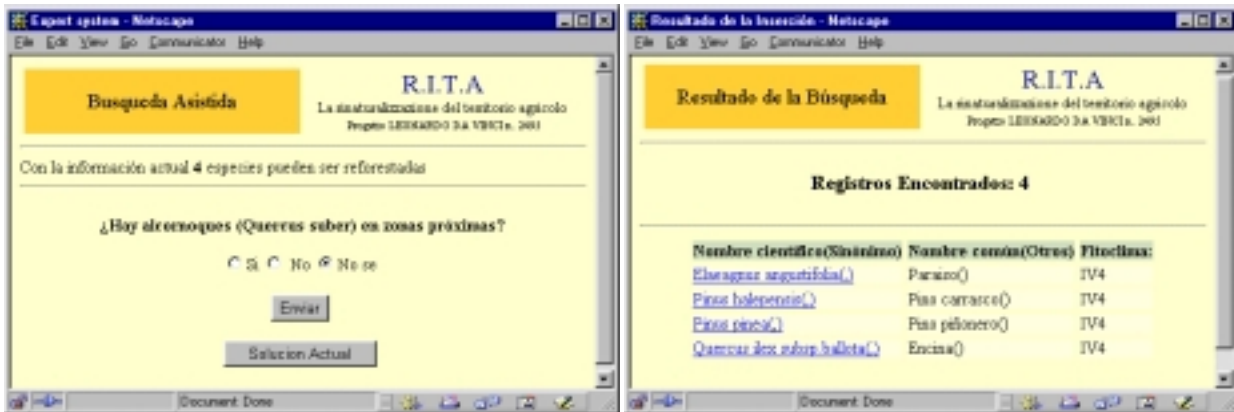


Fig. 3 Ejemplo de pregunta y solución parcial del sistema experto

## 5. Algoritmo de evaluación

El algoritmo utilizado para resolver el problema de la selección de la siguiente pregunta utiliza una aproximación probabilística. El objetivo del algoritmo es dar una puntuación a todas las preguntas y seleccionar aquella con mejor puntuación. Para la valoración de una pregunta es necesario tener en cuenta dos parámetros asociados ambos a cada una de las posibles respuestas para una pregunta dada. Los dos parámetros son:

- Por un lado el número de registros de la BD que quedarían en el caso de que esa opción fuese elegida. ( $n_i$ )
- Por otro lado la probabilidad de que esa opción sea elegida entre las otras opciones de la pregunta ( $p_i$ )

$$Valor = \sum n_i p_i$$

Ecuación 1

La mejor forma de comprender el funcionamiento es mediante un ejemplo. Supóngase que tenemos las preguntas siguientes:

1.- ¿Se encuentran cultivos de vid en zonas próximas?

Si ( $p_1=50\%$ ,  $n_1=15$ )      No ( $p_2=20\%$ ,  $n_2=30$ )      No se ( $p_3=30\%$ ,  $n_3=32$ )

2.- ¿Hay palmito (*Chamaerops humilis*) en las proximidades?

Si ( $p_1=66\%$ ,  $n_1=5$ )      No ( $p_2=14\%$ ,  $n_2=35$ )      No se ( $p_3=20\%$ ,  $n_3=40$ )

La pregunta que el sistema se plantea es cual de las dos es mejor. Será mejor pregunta aquella que en promedio reduzca al máximo el número de registros posibles de la base de datos, independientemente de la respuesta que elija el usuario. Para saberlo con exactitud se calcula *valor* para ambas:

$$Valor_1 = 0.5 \cdot 15 + 0.2 \cdot 30 + 0.3 \cdot 32 = 23.1$$

$$Valor_2 = 0.66 \cdot 5 + 0.14 \cdot 35 + 0.2 \cdot 40 = \underline{16.2}$$

El resultado nos indica que la segunda pregunta deja como media 16 registros de la base de datos cuando el usuario responde a esa pregunta, independientemente de la opción que elija, mientras que la pregunta primera deja 23 registros. Como el objetivo de las preguntas es reducir el número de especies candidatas, la segunda pregunta es mejor. Esta selección ayuda también a que el sistema realice al usuario el mínimo número de preguntas necesarias para conocer cual es la mejor especie para reforestar en esa región dada.

## 5.1 Justificación de la elección

El algoritmo que se implementó en un principio seleccionaba aquella pregunta cuya respuesta afirmativa diese el mínimo número de registros, es decir, sólo se miraba en el ejemplo anterior el 15, 5 para las preguntas 1 y 2 respectivamente. Según ese criterio si se respondía de forma afirmativa se obtenía el mejor resultado, pero después de las pruebas se pudo ver que al no existir una probabilidad máxima de que el usuario eligiese de forma afirmativa, se realizaban muchas preguntas innecesarias al usuario. Por eso se decidió que el experto humano diese unas probabilidades en función de las respuestas más habituales que dan los usuarios para obtener los mejores resultados.

## 6. Conclusiones

En este sistema se ha conseguido imitar a un asesor forestal y debido a que el sistema esta desarrollado por expertos botánicos y forestales, la información que ofrece es bastante precisa y la base del conocimiento que contiene es bastante amplia. Interesante es el hecho de facilitar el acceso al sistema a través de la WWW y la posibilidad de que funciones en diferentes idiomas. Importante también es destacar las características del sistema que permitieron realizar todo el desarrollo de una forma distribuida en toda Europa trabajando casi exclusivamente con la WWW.

## 7. Referencias

- [1] M. Trella, D. Bueno, R. Conejo. Interfaz Web para el desarrollo cooperativo de sistemas. *IV Jornadas de Informática. Las Palmas de Gran Canaria Julio 98*
- [2] Amos David, "Modelisation de l'utilisateur et recherche cooperative dans les systemes de recherche d'informations", ISKO '97, Lille(Francia)
- [3] M. Trella, D. Bueno, R. Conejo "Herramienta para el desarrollo conjunto de una base de conocimientos a través de Internet". ISKO'99 International Society of Knowledge Organization. Granada
- [4] Mei-Mei Wu, Saracevic, T., Spink A., "Users and Intermediaries in Information Retrieval: What are they talking about", UM97
- [5] Spink,A., Ford, N., Wilson, T. "Modeling Users' Successive Searches in Digital Environments", D-Lib Magazine 1998.