GERARDO SIERRA MARTÍNEZ
GABRIEL GARDUÑO TORRES
MARLENE MORENO DEL PILAR
Instituto de Ingeniería, UNAM
México, D.F.
{GSierraM, GGardunoT, MMorenoP}@iingen.unam.mx

Banco terminológico para el Grupo de Ingeniería Lingüística

1. Introducción al banco terminológico

Actualmente vivimos en una sociedad donde la manipulación de información es indispensable para cubrir una de las necesidades primordiales en todo ser humano, la comunicación. Estamos en una era donde es posible creer que una computadora es capaz de pensar como un ser humano, y responder de forma adecuada a cualquier pregunta que se le solicite. Contamos con sistemas de información distribuidos en Internet que facilitan la búsqueda de información de cualquier índole y esto ha llevado a los desarrolladores a crear los llamados bancos de datos terminológicos.

El Instituto de Ingeniería de la UNAM cuenta con el Grupo de Ingeniería Lingüística (GIL), el cual tiene entre sus proyectos principales realizar un sistema de búsqueda onomasiológica, esto es, un diccionario que permita la búsqueda de términos a partir de la descripción del concepto mediante el uso de lenguaje natural. Para ello, el GIL busca y extrae la terminología a partir de textos especializados y de diccionarios existentes. Con ello, se forma la base de conocimiento del diccionario integral, la cual es necesario capturar en bases de datos específicas para cada uno de los dominios en que se trabaje, tales como: lingüística, física, fenómenos destructivos, metrología, y desastres. El conjunto de estas bases de datos constituyen el banco terminológico.

Los integrantes del GIL tienen diferentes necesidades de manipular la información con la que se cuenta. Sin embargo, la manera de manejar tal información no es nada fácil y en algunos casos no es posible tener acceso directo a la misma, por lo que el banco terminológico se implantará de manera que sea útil tanto para los integrantes del GIL así como para los usuarios en general.

El banco terminológico lo tenemos clasificado[1] de la siguiente manera: **A)** Por sus objetivos: es un banco informativo, ya que difunde la terminología de uno o varios ámbitos. **B)** Por la organización de sus datos: es un banco organizado por el término, ya que cubre las consultas a partir del término, para que nos guíe hasta el concepto. **C)** Por su temática: es un banco de temática especializada general porque contiene información sobre distintas áreas. **D)** Por el interés prioritario de los datos que contienen: es un banco de términos que va conjuntamente con su definición. **E)** Por el número de idiomas de las informaciones terminológicas: es un banco monolingüe con equivalencias en otros idiomas, esto es, porque la información completa del término se presenta sólo en un idioma y se asignan las equivalencias denominativas en los demás idiomas. **F)** Por el hardware del banco: es un banco desarrollado en una microcomputadora o PC. **G)** Por el tipo de banco: es un banco de segunda generación. **H)** Por el tipo de acceso: es un banco público que podrá ser consultado vía Internet.

Al ser un banco de segunda generación proporciona al usuario las siguientes ventajas: **A)** El tiempo que se requiere para una consulta y para la recuperación de información es mínimo. **B)** Las formas de consulta son variadas, eficaces y amigables para el usuario. **C)** Los usuarios realizan sus consultas de manera autónoma. **D)** Es posible el acceso al banco terminológico vía Internet. **E)** El banco terminológico se va a encontrar disponible en CD-ROM.

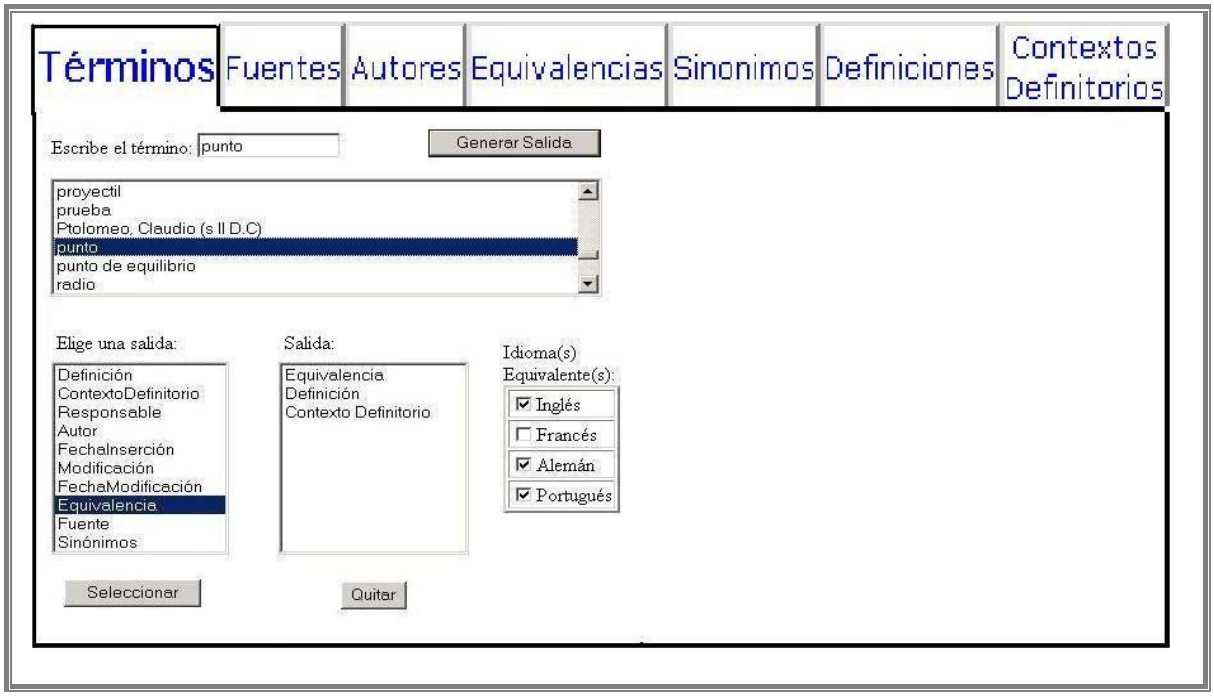
2. Tipos de consulta

Después de dar las características generales del banco, nos enfocaremos en los diferentes tipos de búsquedas que permitirá, ya que es la parte fundamental de su desarrollo. El acceso al banco tendrá dos opciones: **A) Público**, es para todo tipo de usuarios, no tienen restricciones de entrada y pueden hacer todas las consultas disponibles en el banco. **B) Administrador**, permitirá las consultas que tiene el usuario público y además consultas de tipo administrativo; sin embargo, el acceso va a estar restringido por medio de un login y un password para que sea únicamente para usuarios autorizados. Un ejemplo de una consulta de tipo administrativo que se podrá hacer es: ¿Cuántas fichas ha ingresado el responsable X en la fecha Y?

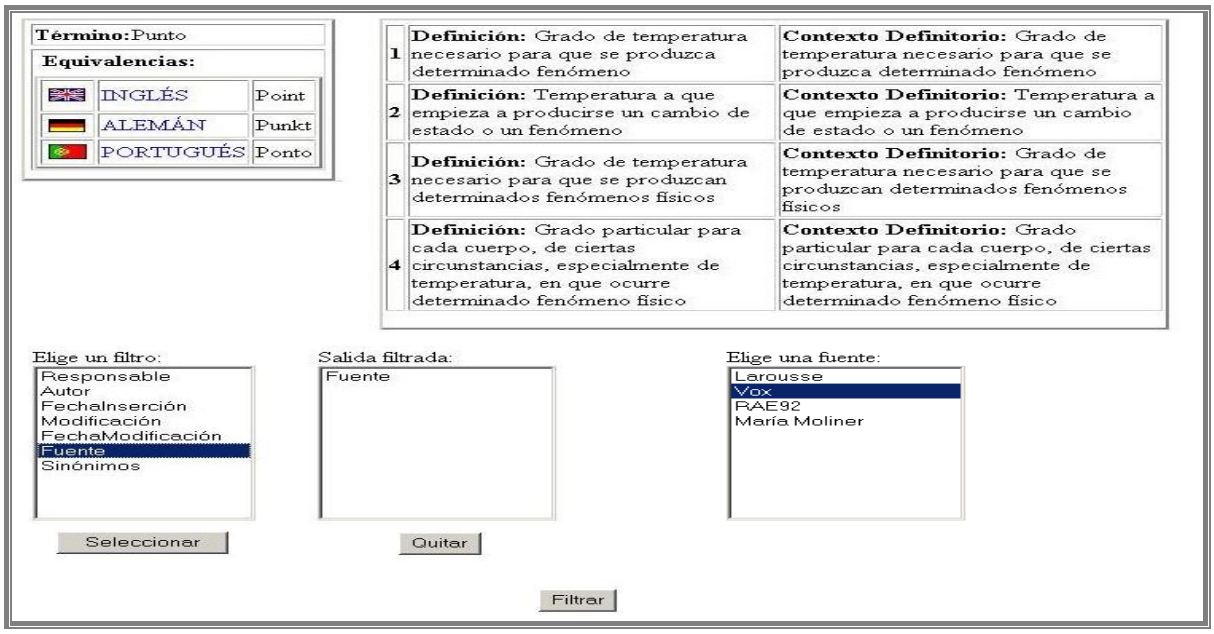
2.1 Consulta por términos

[1] Cabré, M. Teresa, “La terminología. Teoría, metodología, aplicaciones”, Editorial Antártica / Empúries: Barcelona, 1993.

Al haber ingresado al sistema, le aparecerá al usuario una interfaz en la cual tendrá la opción de elegir de un catálogo la opción que considere más importante para generar sus búsquedas, que pueden ser múltiples. En primera instancia aparece la opción del dominio y en éste puede hacer búsquedas por término, ya que es la más importante de todas. En los ejemplos se consulta en el banco de física.



En el ejemplo que se ilustra en la imagen anterior, le pedimos al banco que nos muestre las equivalencias en inglés, alemán y portugués, las definiciones y los contextos definitorios del término **punto**. En el caso de elegir como salida las equivalencias, se habilita una pantalla para seleccionar entre las existentes (Ingles, Francés, Alemán o Portugués), por lo que la salida de nuestra búsqueda en el banco es la siguiente:



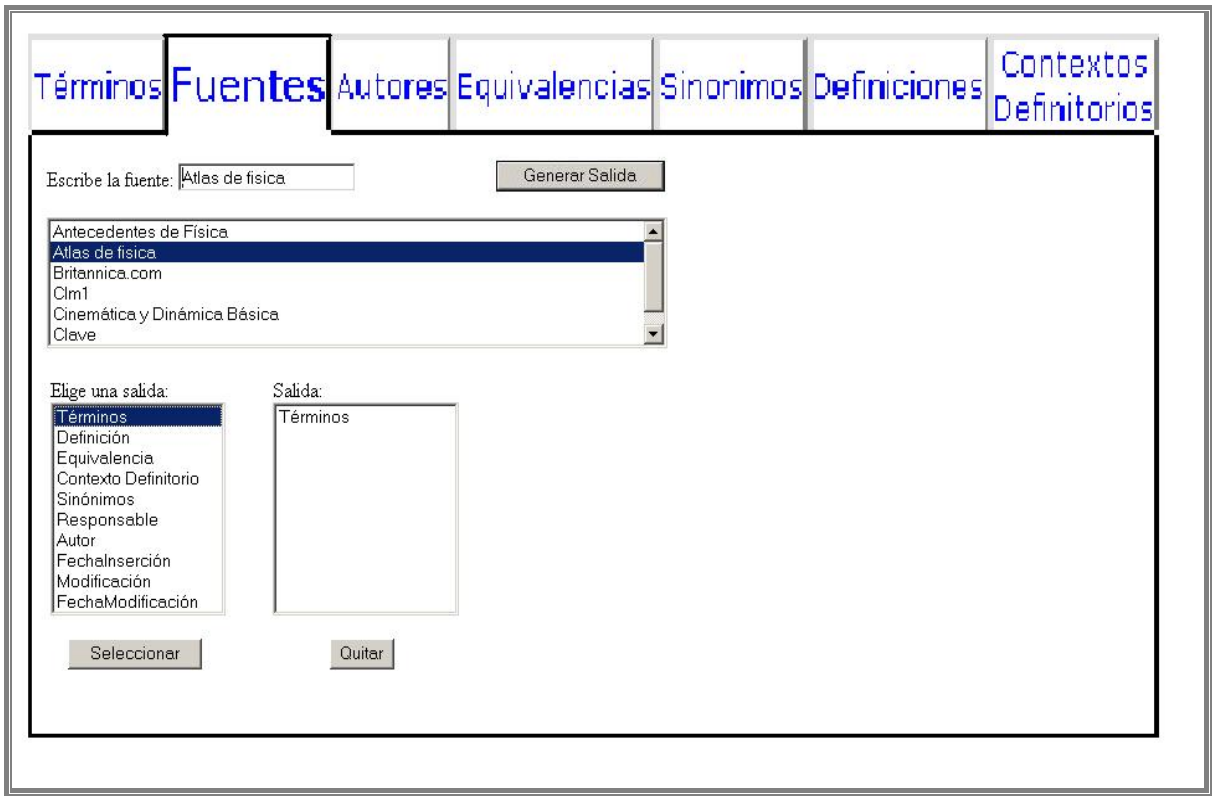
El sistema nos presenta la respuesta de la búsqueda y nos brinda la opción de hacer un filtrado sobre ésta misma. Cabe denotar que el filtrado no es múltiple, es decir, si deseamos filtrar la salida tenemos que hacerlo filtro por filtro. En el ejemplo se desea filtrar por fuente y nos aparecen las opciones de fuente que se tienen disponibles.

Cuando el sistema encuentra que el resultado de la salida del filtrado fue solamente uno ya no nos da la opción de seguir filtrando; en caso contrario podremos generar tantas búsquedas como sea posible de forma sucesiva hasta que el usuario encuentre los datos que le interesan.

2.2 Consulta por fuentes

El usuario tiene la posibilidad de realizar búsquedas a partir de Fuentes, Autores, Equivalencias y Sinónimos. En esta etapa el sistema aporta grandes beneficios a los usuarios ya que es posible obtener como resultado un listado de términos partiendo de la opción que hayamos elegido, además de hacer filtrados sobre las respuestas.

La siguiente imagen muestra la interfaz encargada del proceso de búsqueda por Fuente, así como la salida seleccionada para generar la respuesta.



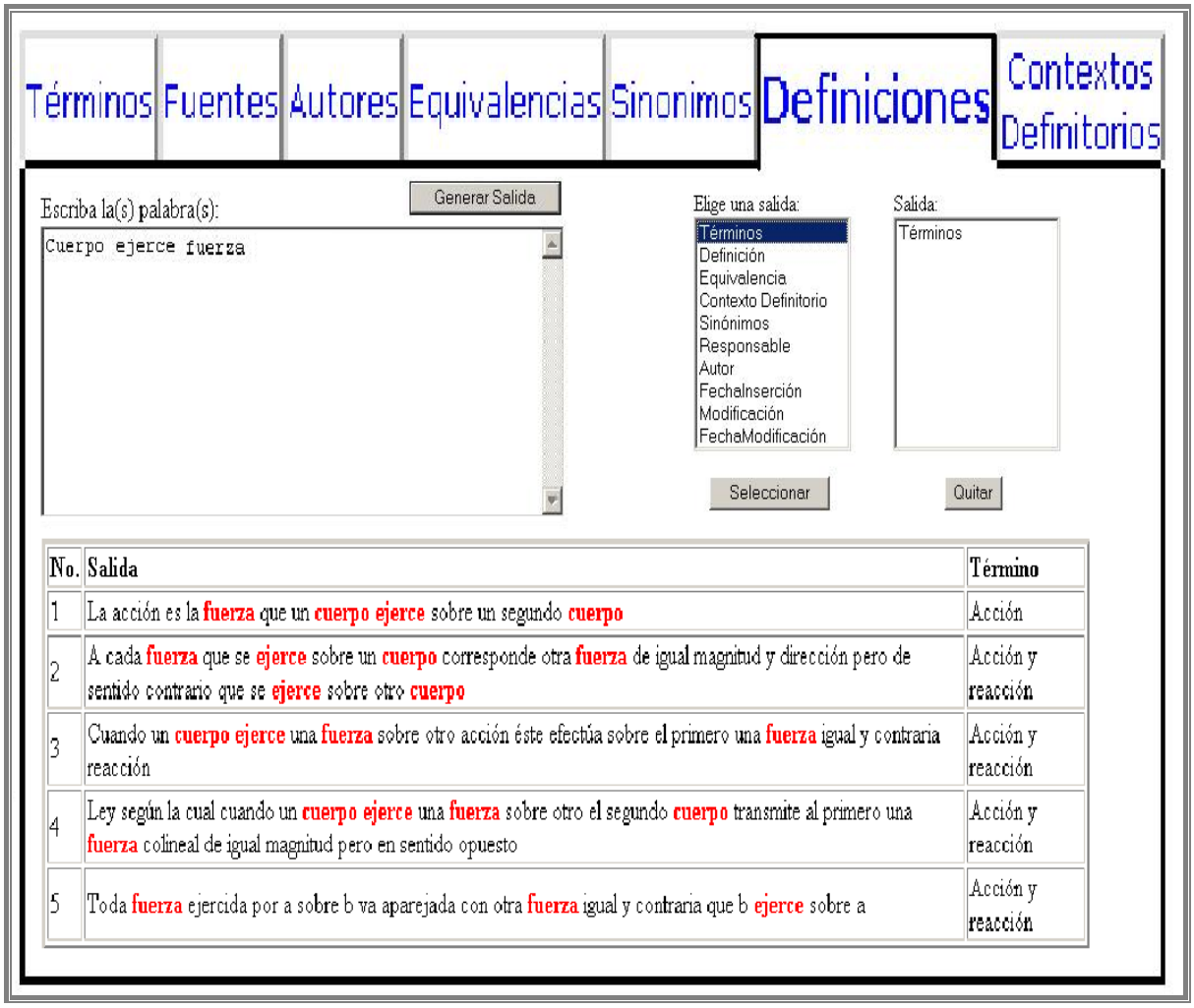
En este ejemplo estamos buscando la fuente “Atlas de física” y queremos que nos muestre los términos, así que el banco arroja la respuesta que cumpla con esta característica así como la posibilidad de generar un filtrado sobre los resultados.

No.	Término
1	acción y reacción
2	aceleración
3	amplitud
4	calor
5	cinemática
6	difracción
7	dinámica
8	elongación
9	energía
10	energía cinética

Los filtrados son los mismos que se ilustraron en la búsqueda por términos.

2.3 Consultas por definiciones

Por último, el banco contará con búsquedas que permiten las consultas por palabras que se encuentren inmersas en definiciones y contextos definitorios; en este caso presentaremos un ejemplo a partir de Definiciones, en donde el usuario solicita la búsqueda de palabras dentro de una definición o de un contexto definitorio. Se proporcionan respuestas de tipo texto donde el usuario visualiza la palabra clave resaltada en el párrafo y puede elegir entre todas las posibilidades para tener información adicional de las definiciones. Enseguida se muestra el ejemplo:



La ilustración anterior nos muestra una consulta, la cual pide al sistema que encuentre las palabras “cuerpo, ejerce y fuerza” en todas las definiciones y que además nos indique a que término pertenece cada definición. Podemos indicarle que nos extraiga más información adicional con las opciones que nos presenta.

Este tipo de búsqueda es el preámbulo al diccionario onomasiológico, ya que las búsquedas podrán introducirse en lenguaje natural y el sistema se va a encargar de encontrar la(s) palabra(s) inmersa(s) entre todas las definiciones y contextos definitorios.

3. Conclusiones

La cantidad de búsquedas que se podrán realizar en nuestro banco supera a la que se permite en los bancos existentes; si bien hay bancos que pueden hacer algunas de estas exploraciones, no hay ninguno que pueda hacer todas.

Con este banco brindaremos una gran ayuda a los especialistas de la comunicación, la información y la documentación, expertos en lexicografía, terminología y traducción, ingenieros lingüistas y lingüistas computacionales, editores, profesores de lengua, investigadores en lingüística aplicada, profesores de áreas especializadas y a todos los usuarios que deseen consultar un diccionario técnico y científico con diversas áreas temáticas que contengan suficiente información almacenada en un mismo lugar.

En este momento el banco terminológico se encuentra en su etapa de desarrollo; sin embargo ya se cuenta con varios diccionarios que nos permiten introducir la información al banco de datos y hacer algunas consultas. Este diccionario se encuentra en la dirección http://imsai.iingen.unam.mx/diccionarios

Una vez terminado el banco y puesto en marcha en la gran telaraña mundial, el siguiente objetivo del GIL será tenerlo en CD-ROM.