# Diccionario de diccionarios electrónico da lingua galega

Organismo financiador: Dirección Xeral de Política Lingüística (Xunta de Galicia).
Grupos participantes en el proyecto: imaxin, Instituto da Lingua Galega (USC).
Personas de contacto:

- José Ramom Pichel Campos. *Coordenador Área Productos lingüísticos. Departamento Enxeñería do Software*. imaxin software. Santiago de Compostela. jramompichel@imaxin.com
- Antonio Fernández Cabezas: *Departamento Enxeñería do Software*. imaxin software. Santiago de Compostela. afernadez@imaxin.com

**Dirección de contacto: http://www.imaxin.com**

**RESUMEN**

Este proyecto consiste en la realización de un diccionario electrónico formado por diccionarios convertidos previamente a formato SGML. Estos diccionarios son en total 11 diccionarios históricos gallegos desde el Rodríguez de 1863 hasta el de D. Eladio Rodríguez de 1963. Sobre este se podrán realizar **Consultas simples** (consultas de cada palabra), **Consultas complejas** (Consultas sobre las entradas con operadores lógicos en la que se pueden utilizar caracteres especiales), **Visualización de entradas** (Podremos visualizar las entradas viendo las definiciones en los diferentes diccionarios en los que aparece), **Búsquedas por refranes** (búsquedas de todas las entradas que contienen refranes), **Búsquedas por poemas** (búsquedas de todas la entradas que contienen poemas), Búsquedas de ejemplos (entradas que contienen ejemplos ilustrativos), **Histórico**, **Cesta** (Almacenamiento de entradas de trabajo), **Hipertexto**, **Impresión** (podremos imprimir cualquier tipo de consulta), etc.

**Textos en SGML**

Los diccionarios son diccionarios antiguos conteniendo cada una de sus entradas desde sinónimos y afines hasta nomenclaturas científicas. Se realizaron aplicaciones ad hoc que a partir de una DTD realizada previamente marcasen los textos y los convirtiesen a SGML. Debido al carácter heterogéneo de los diccionarios se tuvo que implementar una DTD lo suficientemente amplia para poder abarcar a todos.

Luego hubo que hacer una revisión de los textos usando herramientas de validación de textos SGML como el Corel WordPerfect 8.0

**Codificación de los textos SGML**

Para evitar que cualquier persona ajena al proyecto pudiera acceder a la información maracada en SGML se desarrollaron programas de encriptado de la información, por lo que hubo que implementar en la aplicación desencriptadores de los textos.

**Indexación de los textos**

Se desarrolló un indexador que partiendo de la Gramática DTD y poniéndole restricciones de lo que debía ser indexado o no, de los diferentes casos que se muestran como ejemplos, stop list de español y gallego, generó todos los índices sobre una base de datos en formato .mdb. Para localizar cada uno de los lemas dentro de los textos SGML se usó el paradigma de los índices inversos.

**Visualización de cada entrada**

En el Diccionario de diccionarios electrónico vamos a poder ver cada entrada del diccionario (en total 88.000 entradas) visualizando las diferentes definiciones presentes en los 11 diccionarios. De este modo p.e. una entrada como “alén” vamos poder consultarla en los cinco diccionarios

donde aparece. Se tuvo que implementar un visualizador SGML quién a partir de la DTD y de especificaciones de salida genera las definiciones.

**Consultas simples**

Al ser textos antiguos y por motivos de conflicto lingüístico los diccionarios normalmente tienen las entradas en gallego y las definiciones en español. Debido a esto podremos realizar consultas de "en que lemas se encuentra una determinada palabra en español". Lo mismo sucedería para gallego "en que lemas se encuentra una determinada palabra en gallego", ya que existen dentro de las definiciones palabras en esta variedad lingüística galego-portuguesa. Además podremos hacer consultas de sinónimos y de palabras afines.

**Consultas complejas**

Podremos realizar consultas de carácter complejo sobre los lemas utilizando para ello operadores lógicos presentes en cualquier consulta sobre una base de datos. Estas consultas tienen además la posibilidad de incluír caracteres especiales ya que en las entradas y definiciones existen grafías de caracter etimológico de propia creación de los autores.

**Histórico**

Todas las entradas que vayamos visitando van a ser registradas en un histórico que se perderá al acabar la sesión.

**Cesta**

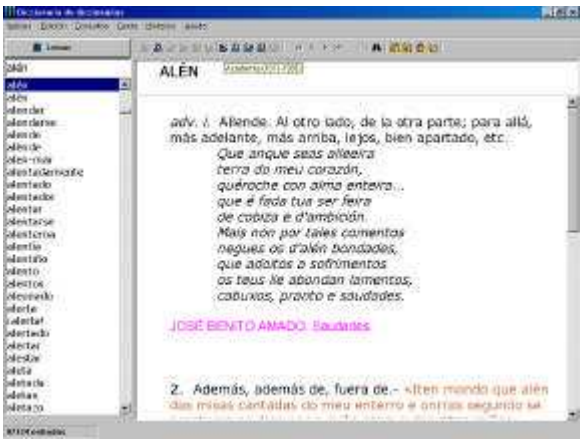
Podremos incluír en esta estructura de datos todas aquellas entradas que juzguemos necesarias para nuestro trabajo.

**Hipertexto**

A partir de cada una de las palabras presentes en la definición podremos navegar por hipertexto a los lemas asociados.

**Impresión**

Podremos hacer una impresión de todas las consultas que realicemos sobre la base de datos. Además existe la posibilidad de imprimir para cada resultado de la consulta, una entrada en un diccionario o en todos e incluso un intervalo de entradas.

**Visualización de Prólogos y portadas originales**

Los prólogos de los textos originales así como las portadas se pueden visualizar en formato PDF.