

Interlex Project. MLIS-103

Desarrollo de bases de datos generales y multilingües que se explotarán en Internet, a partir de diccionarios de traducción en formato electrónico.

Coordinator:

- UNIVERSIDAD ALFONSO X EL SABIO (SPAIN). (Department of Applied Languages and the Centre for Data Processing). Avda Universidad 1, Villanueva de la Cañada 28691 (Madrid) Spain. Tel.: +34-1-8109737 Fax: +34-1-8109101. Email: diez@uax.es.

Partners:

- LA MAISON DU DICTIONNAIRE(FRANCE). 98 Bld du Montparnasse 75014 Paris France. Tel.: +33-1-43221293. Fax: +33-1-43220177. Email: feutry@aol.com.
- EDITORIAL EVEREST, S.A. (SPAIN). Carretera León-La Coruña km 5, 24080 León, Spain. Tel.: +34-87-802020 Fax: +34-87-801251 Email: publicaciones@everest.es.
- EVEREST EDITORA (PORTUGAL). Parque Industrial Meramar II Armazén nº 1, Portugal. Tel.: +351-1-9152483 Fax: +351-1-9152525 Email: publicaciones@everest.es.
- TECNOLINGUA S.A. (SPAIN). Morillo, 11,7º B Alcalá de Henares 28805 (Madrid) Spain. Tel.: +34-1-8835806 Fax: +34-1-8835806 Email: jsg38746@teleline.es

Persona de contacto: Pedro Luis Díez Orzas. Universidad Alfonso X El Sabio. Avenida de la Universidad, 1, E - 28691 VILLANUEVA DE LA CANADA. Tel.: +34-1-8109737 Fax: +34-1-8109101 Email: diez@uax.es

Dirección de contacto: <http://www.uax.es/imasd/interlex/>

Resumen

Como proyecto MLIS, el objetivo de Interlex es convertir diccionarios bilingües y multilingües (generales y terminológicos), actualmente disponibles en formato papel o en CD-ROM, en recursos electrónicos, a los que se pueda acceder a partir de Internet.

El rápido crecimiento de Internet y la demanda, cada vez mayor, de productos electrónicos, como el CD-ROM, ha hecho que las editoriales dedicadas a la publicación de diccionarios conviertan su recursos léxicos en formato electrónico. Sin embargo, este proceso es muy lento, debido, en primer lugar, al alto coste que requiere la inversión, al miedo a la piratería de productos electrónicos y, finalmente, a la confianza en las formas tradicionales de publicación.

En este contexto, Interlex aparece como un proyecto puntero, no sólo porque es una de los primeros de este tipo en este ámbito, también porque reúne, por primera vez, a un grupo de editoriales, de reconocido prestigio, dedicadas a la publicación de diccionarios. El proyecto hará posible el acceso conjunto a recursos

comunes, la inversión que supone es subvencionada por la Unión Europea.

El socio académico y el industrial aportarán el conocimiento técnico y experto necesario para el desarrollo en las diferentes áreas que incluye el proyecto. La Facultad de Lenguas Aplicadas y el Centro de Proceso de Datos de la Universidad Alfonso X el Sabio de Madrid contribuirán con su experiencia académica, técnica y administrativa. La primera incluye, entre otras aspectos, experiencia en el desarrollo de herramientas léxicas y semánticas, análisis de corpus, diseño de interfaces HTML y explotación de bases de datos. Tecnolingua, el socio industrial, una empresa española puntera en el desarrollo de programas de tratamiento automático de la lengua, desarrollará los programas que requiere el desarrollo del producto y servicio que supone Interlex. El proyecto incluye, finalmente, a una serie de editoriales de reconocido prestigio, como son -la editorial Everest (España); Everest-Editora (Portugal) y la Maison du Dictionnaire (Francia). Estas editoriales pondrán sus diccionarios generales y terminológicos,

bilingües y multilingües a disposición de la explotación conjunta que supone Interlex.

Los objetivos principales del proyecto se podrían resumir en los siguientes puntos:

- Crear una fuente de referencia inestimable para profesionales, estudiantes y, en general, todos aquellos interesados en la traducción y las equivalencias entre lenguas.
- Poner a disposición del usuario recursos léxicos compartidos en inglés, español, alemán, francés, italiano y portugués
- Demostrar que es posible publicar recursos léxico en formato electrónico y, además, ponerlos a disposición del usuario en Internet.
- Introducir a todos los involucrados en la publicación y el uso de recursos léxicos en esta innovadora y útil forma de explotación.

Además de las ventajas que se acaban de señalar, el proyecto supondrá avances en el desarrollo tecnológico y metodológico en el área del tratamiento automático de la lengua que se pueden recoger en los siguientes puntos:

Interlex se propone la conversión de diccionarios bilingües y multilingües en bases de datos léxicos con capacidad de indexación y búsqueda, que serán explotados en la Red. Con este objetivo, se utilizará GENETER; un formato estándar de codificación basado en SGML y con una estructuración en varios niveles, para los recursos terminológicos, a partir de este formato, se desarrollará una extensión del mismo con el objetivo de dar cabida a los diccionarios generales.

- En primer lugar, se desarrollará un formato común de trabajo al que se convertirán las diferentes codificaciones empleadas en editorial y en cada diccionario.
- A continuación, se desarrollará una base de datos dinámica, la herramientas y la

metodología que será necesario aplicar a cada conjunto de datos.

- En este momento, se llevará a cabo un proceso de comprobación y coherencia entre direcciones de lenguas, tanto en diccionarios bilingües como multilingües.
- En el caso de los recursos multilingües y, siguiendo los estándares propuestos por el proyecto Eurowordnet, se generará un índice oculto que se utilizará para los propósitos señalados en el punto anterior.
- Todos estos productos se integrarán en una interfaz que pondrá a disposición del usuario de Internet los recursos léxicos. Se crearán, asimismo, los motores de búsqueda e indexación necesarios y se hará operativa una interfaz de demostración.

Los valores añadidos de este proyecto se pueden concretar en los siguientes puntos:

- En primer lugar, contribución a un mejor entendimiento sobre el funcionamiento de los mecanismos que permiten etiquetar y mejorar los diccionarios bilingües y multilingües, en una base de datos con rápido motor de acceso.
- En segundo lugar, establecimiento de una metodología, que adapta recursos compartidos a un formato común, para explotar diccionarios bilingües y multilingües.
- Finalmente, búsqueda y estudio de fórmulas comerciales que permitan explotar recursos léxicos en Internet; especialmente en lo que se refiere a la relación entre usuario/cliente y desarrollador/proveedor.

El proyecto supondrá una nueva forma de acceder a diccionarios bilingües y multilingües en Internet, en la que el usuario podrá beneficiarse de la continua actualización de estos recursos.