Proyecto *RILE*: servidor de recursos para el desarrollo de la Ingeniería Lingüística en español

I. Bermejo, X. Carreras, N. Castell, I. Castellón, E. Coello
J. Gonzalo, N. Kalfon, M. A. Martí, S. Rodríguez,
L. Padró, A. Peñas, T. Read y M. F. Verdejo

Título del proyecto

RILE: Servidor de recursos para el Desarrollo de la Ingeniería Lingüística en Español.

Grupos participantes en el proyecto

- SEMA Group SAE
- Grupo UNED en Procesamiento de Lenguaje Natural
- Grupo de Tratamiento del Lenguaje Natural, UPC
- LaReLC-UB, Laboratorio de Investigación en Lingüística Computacional
- CLIC, Centro de Lingüística Computacional
- Oficina del Español en la Sociedad de la Información, OESI - Instituto Cervantes

Persona de contacto

Felisa Verdejo Maillo

D. Lenguajes y Sistemas Informáticos
E.T.S. de Ingenieros Industriales
UNED
Ciudad Universitaria, s/n
28040 Madrid

Teléfono: 91 398 64 84
`felisa@ieec.uned.es`

Dirección web

`http://rile.sema.es`

Resumen

El objetivo del proyecto RILE, servidor de Recursos para el desarrollo de la Ingeniería Lingüística en Español, es la realización de una prueba piloto para la creación de un servidor con recursos, herramientas e información orientada al desarrollo de aplicaciones en el campo de las tecnologías lingüísticas en español. Esta prueba piloto se lleva a cabo con un número limitado de participantes y de recursos disponibles, a fin de evaluar las posibilidades de un proyecto a mayor escala. La prueba tiene como finalidad estudiar la viabilidad de un servidor público en los que se refiere, principalmente, a tres ámbitos:

- Problemas técnicos de adaptación de recursos y herramientas.
- Respuesta a las necesidades de los usuarios.
- Aspectos, tanto comerciales como legales, derivados de la distribución de recursos y herramientas.

El proyecto piloto se ha realizado en un plazo limitado de tiempo. Por ello, se ha procedido a una selección de recursos y herramientas desarrolladas por los grupos de proveedores participantes en el proyecto, de modo que sea posible desarrollar de una forma rápida las pruebas de integración que se proponen para estudiar la viabilidad del prototipo.

En esta primera fase, sen han incorporado al servidor los siguientes elementos:

Recursos

- Una selección de textos literarios de libre disposición.
- Una selección de corpora anotado morfosintácticamente.
- Un corpus paralelo alineado español/euskara.
- Un corpus anotado semánticamente.
- Una gramática independiente del contexto para análisis parcial del español.

Herramientas y aplicaciones

- Nombre de la aplicación: *Motor de búsqueda ITEM*. Motor de búsqueda conceptual multilingüe (español, inglés y catalán) sobre una colección de noticias de ámbito internacional.
- Nombre de la herramienta: *Analizador morfosintáctico*. Un demostrador web de análisis morfosintáctico en español.
- Nombre de la herramienta: *JCord*. Una herramienta básica de concordancias sobre textos procesados morfosintácticamente.