



## LIDER – Los datos enlazados como elemento clave para facilitar el análisis de contenidos multimedia y multilingüe en empresas en Europa

### Descripción

El objetivo del proyecto LIDER es afrontar los retos que supone el crecimiento exponencial de contenidos multilingüe y multimedia en la Web: el aumento de volumen, velocidad y variedad en los contenidos requiere nuevas formas de abordar su análisis. Para este fin, la comprensión e interpretación del lenguaje, tanto en contenido textual no estructurado como en contenido lingüístico presente en distintos conjuntos de datos, resultan de vital importancia. LIDER pretende marcar las pautas para la creación de un innovador ecosistema de recursos libres, interoperables semánticamente y enlazados entre sí, dentro del marco de los datos abiertos enlazados (*Linked Open Data*). Este ecosistema integrará recursos lingüísticos (diccionarios, corpus, metadatos léxicos y sintácticos, etc.) y recursos multimedia (metadatos de imágenes, video, etc.). LIDER estudiará cómo estos recursos pueden constituir una tecnología que facilite el análisis de contenido dentro de la Web multilingüe, incluyendo el contenido multilingüe presente en datos multimedia.

### LIDER: Principales logros en el primer año

➤ **Acercamiento a diferentes tipos de comunidades y definición de los posibles casos de uso en la empresa.** LIDER ha organizado cuatro talleres dedicados a la elaboración de una hoja de ruta, en los que se recabaron ideas sobre los posibles casos de uso en la empresa, así como sobre los requisitos que los datos lingüísticos enlazados deberían cumplir. Estos talleres se han orientado a cuatro grupos: los profesionales en materia de datos, los que trabajan en temas de la Web, los expertos en localización y los expertos en el análisis de contenidos. El grupo de trabajo LD4LT (*Linguistic Linked Data for Language Technology*) del

consorcio W3C recogió los comentarios expuestos en las sesiones, que se detallan en los informes de los talleres. Asimismo, LIDER ha participado en un gran número de encuentros de la comunidad científica y del sector empresarial, ha organizado sesiones formativas y ha realizado presentaciones sobre datos lingüísticos enlazados y sobre análisis de contenidos.

➤ **Establecimiento de directrices y buenas prácticas.** LIDER ha contado con el grupo de trabajo BPMLOD (*Best Practices for Multilingual Linked Open Data*) del consorcio W3C, dedicado a recopilar directrices y buenas prácticas a la hora de crear fuentes de datos lingüísticos enlazados. Dichas directrices se han seguido en el diseño de métodos de migración de recursos lingüísticos existentes a datos enlazados. En primer lugar, se han transformado los metadatos de los recursos lingüísticos. LIDER ha trabajado estrechamente con la comunidad de META-NET y de CLARIN para elaborar un modelo consensuado de descripción de metadatos lingüísticos y así asegurar que los esquemas de metadatos ya existentes sean reutilizados. En segundo lugar, LIDER ha generado un conjunto de buenas prácticas para generar datos lingüísticos enlazados, incluyendo diccionarios bilingües o multilingües, y ha preparado directrices para la migración de corpora y de recursos terminológicos (TBX).

➤ **Diseño de una arquitectura de referencia y propuesta de una hoja de ruta.** LIDER ha creado un primer borrador de una arquitectura de referencia para datos lingüísticos enlazados para el análisis de contenidos, junto con el que proporciona un resumen de las herramientas e iniciativas relevantes. La arquitectura en sí define tanto un modelo general para la construcción de servicios sensibles a los datos lingüísticos enlazados como diversos patrones



para el diseño de aplicaciones de análisis de contenidos. Asimismo presenta un panorama de las iniciativas y herramientas relevantes. La arquitectura se ve complementada por una primera versión de la hoja de ruta para la creación de datos lingüísticos enlazados. En concreto, la hoja de ruta integra ideas y comentarios de distintas fuentes, tanto de actividades de LIDER (como encuestas o talleres), como de informes y actas, organismos de financiación y diversas comunidades investigadoras. La hoja de ruta está formulada en torno a las principales áreas de aplicación: compromiso de los clientes, sector público y sociedad civil, y ciclo de vida y cadena de publicación de los datos lingüísticos enlazados. Para cada una de estas áreas, la hoja de ruta recoge distintos casos de uso, sus plazos, predicciones y factores significativos.

### **Próximos objetivos**

LIDER presentará los avances conseguidos durante el primer año de proyecto a los agentes interesados. En concreto, se tratarán los aspectos relacionados con la migración de recursos lingüísticos y sus metadatos, la arquitectura de referencia para datos lingüísticos enlazados y su uso en el análisis de contenidos así como la hoja de ruta para el uso y creación de los mismos. De igual forma, LIDER se acercará a nuevos grupos de usuarios, especialmente a los interesados en la apertura de datos en el sector público.

### **Coordinador**

Persona de contacto: Prof. Dr. Asunción Gómez-Pérez

Tel:+34 913367417

Correo electrónico: [asun@fi.upm.es](mailto:asun@fi.upm.es)

Organización: Universidad Politécnica de Madrid (UPM)

### **Participantes**

- Universidad Politécnica de Madrid (UPM)
- Trinity College Dublin (TCD)
- German Research Center for Artificial Intelligence (DFKI)
- INSIGHT National Center for Data Analytics, National University of Ireland, Galway
- Institute for Applied Informatics (InfAI), University of Leipzig
- University of Bielefeld
- Sapienza University of Rome
- W3C / Geie ERCIM

### **Programa del proyecto**

Duración: 24 meses

Fecha de comienzo: 1 de noviembre de 2013

Fecha de finalización: 31 de octubre de 2015

Jefe de proyecto: Susan Fraser

Proyecto 610782, CSA

Para solicitar más información o participar:

- Página web del proyecto: <http://www.lider-project.eu/>
- Portal de la comunidad: <http://www.multilingualweb.eu/>
- Grupo de trabajo W3C sobre “Linked Data for Language Technology”: <http://www.w3.org/community/ld4lt/>
- Lista de correo [lider-community@listas.fi.upm.es](mailto:lider-community@listas.fi.upm.es) para boletines informativos y sugerencias.
- Suscripciones a través de <http://www.lider-project.eu/get-involved>.
- Actividad en las redes sociales mediante el hashtag #LiderEU. Twitter: <https://twitter.com/multilingweb>