



LIDER – Les dades enllaçades com a element clau per a facilitar l’anàlisi de continguts multimèdia i multilingüe a empreses d’Europa

Rerafons

L’objectiu de LIDER és ocupar-se dels reptes que suposa el gran augment de contingut multilingüe i multimèdia a la Web: en volum, velocitat i varietat, aquest augment requereix noves formes i avenços en el camp del seu anàlisi de contingut. Per a aquest objectiu, la comprensió i interpretació del llenguatge d’aquest contingut, tant com a contingut textual sense estructura, com a contingut lingüístic present en diferents conjunts de dades de transmissions multimèdia són de vital importància.

LIDER construirà un camí per l’establiment d’un nou ecosistema de Linked Open Data (LOD), dades enllaçades i obertes, basat en els recursos lliures, intervenculats i semànticament interoperables en el camp del llenguatge (dades lingüístiques enllaçades que representen corpus, diccionaris, metadades de lèxica i sintàctica, etc...) i dels media (metadades en imatges, vídeos, etc...). LIDER estudiarà com aquests recursos poden servir com a tecnologia que faciliti l’anàlisi de contingut dins de la Web multilingüe, inclòs el contingut multilingüe present en dades multimèdia.

LIDER: Èxits durant el primer any

➤ **Apropament a diverses comunitats, i definició de les millors pràctiques de negoci.** LIDER ha organitzat quatre tallers dedicats a l’elaboració d’un full de ruta, en els que s’ha treballat conjuntament per tal de recollir idees sobre les millors pràctiques a l’empresa, i requisits que les dades lingüístiques enllaçades deuen acomplir. Els tallers anaven dirigits a quatre grups: els professionals en matèria de dades, els professionals de la Web, els experts en localització i els experts en l’anàlisi de dades.

Els comentaris i opinions han estat recollits a través d’un grup de l’unitat de LD4LT W3C (“Linguistic Linked Data for Language Technology”), han estat documentats a diferents informes dels workshops i han servit de guia per als tallers. A més, LIDER ha estat present en un gran nombre d’esdeveniments de recerca científica i del sector empresarial, en diferents tutorials, i a presentacions sobre dades lingüístiques enllaçades i d’anàlisi de contingut

➤ **Establiment de directrius i bones pràctiques.** LIDER ha comptat amb el grup de treball BPMOLD (millors Pràctiques per a

dades lingüístiques enllaçades) del consorci W3C, dedicat a recollir directius i bones pràctiques a l'hora de crear fonts de dades lingüístiques enllaçades. Aquestes directrius s'han fet servir en el disseny de mètodes de migració de recursos lingüístics existents a les dades enllaçades. En primer lloc, s'han transformat les metadades dels recursos lingüístics. LIDER ha treballat estretament amb la comunitat de META-NET i CLARIN per a elaborar un model de descripció de metadades lingüístiques i així assegurar que els esquemes de metadades ja existents siguin reutilitzats. En segon lloc, LIDER ha generat un conjunt de bones pràctiques per a generar dades lingüístiques enllaçades, inclús diccionaris biligües o multilingües, i ha preparat directrius per a la migració de còrpora i de recursos terminològics (TBX).

➤ **Disseny d'una arquitectura de referència i proposta de full de ruta.**

LIDER ha creat un primer borrador d'una arquitectura de referència per a dades lingüístiques enllaçades per a l'anàlisi de contingut, juntament amb el que proporciona un resum de les eines i iniciatives rellevants. La mateixa arquitectura en defineix tant un model general per a la construcció de serveis sensible a les dades lingüístiques enllaçades com a diversos patrons per al disseny d'aplicacions d'anàlisi.

Propers Objectius

LIDER presentarà els avenços aconseguits durant el primer any de projecte als agents interessats. En concret, es tractaran els aspectes relacionats amb la migració de

recursos lingüístics i les seves metadades, l'arquitectura de referència per a dades lingüístiques enllaçades i el seu ús en l'anàlisi de continguts així com el full de ruta per a l'ús i creació dels mateixos. A la vegada, LIDER s'aproparà a nous grups d'usuaris, especialment als interessats en l'apertura de dades en el sector públic.

Coordinadora

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

Tel:+34 913367417

Email: asun@fi.upm.es

Organització: Universidad Politécnica de Madrid (UPM)

Participants

- 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 projecte

Durada: 24 mesos

Inici: 1 de noviembre de 2013

Final: 31 de octubre de 2015

Projecte 610782, CSA

Cap de projecte: Susan Fraser

Per a més informació o per com participar:

- Pàgina web del projecte: <http://www.lider-project.eu/>
- Portal de la comunitat: <http://www.multilingualweb.eu/>
- Grup de treball W3C sobre “Linked Data for Language Technology”:
<http://www.w3.org/community/ld4lt/>
- Llista de correu lider-community@listas.fi.upm.es per a butlletins informatius y suggerències.
- Subscripcions <http://www.lider-project.eu/get-involved>.
- Activitat en les xarxes socials mitjançant el hashtag #LiderEU. Twitter:
<https://twitter.com/multilingweb>