



LIDER: FP7 – 610782

Linked Data as an enabler of cross-media and multilingual content analytics for enterprises across Europe

LIDER - Les Web de données ouvertes et liées pour faciliter l'analyse de contenu multimédia et multilingue en Europe.

Challenge

La croissance continue de données sur le Web, tant en volume qu'en variété, appelle à la conception et mise en œuvre de nouvelles approches pour l'analyse de contenu. Ces approches constituent un préalable indispensable pour l'analyse et l'interprétation de données hétérogènes à grande échelle, émanant de différents médias et exprimées en différentes langues.

Les récentes avancées du Web Sémantique, dont l'adoption des technologies du web des données ouvertes et liées, ont permis de poser les bases de la nécessaire interopérabilité entre les différents outils développés et utilisés pour l'analyse de la *datamasse* (données structurées ou non) générée quotidiennement par les entreprises et les utilisateurs. Néanmoins, un point essentiel fait encore défaut, à savoir la compréhension et l'interprétation des aspects linguistiques de ces données, qu'il s'agisse de contenu textuel non structuré d'une part, ou de données langagières attachées à divers médiums d'autre part. La représentation sémantique de telles données langagières -- incluant leurs aspects linguistiques -- est nécessaire pour la mise en œuvre d'une analyse de contenu sur un Web toujours plus varié et multilingue.

Objectifs et innovation

Le projet LIDER veut établir les fondations d'un nouvel écosystème pour le web des données ouvertes et liées. Cet écosystème sera centré sur des ressources libres, interliées et sémantiquement interopérables avec, d'une part, des ressources linguistiques ("web des données liées linguistiques" avec représentations de corpus, dictionnaires, lexiques etc.) et, d'autre part, des ressources médias (image, vidéo, etc.). Le projet étudiera comment ces ressources peuvent faciliter et stimuler l'analyse de contenu

d'entreprise sur le Web sémantique multilingue, y compris les données délivrées *via* différents médias.

LIDER favorisera l'établissement d'une communauté d'intérêt autour du nuage du « Web des Données Linguistiques Liées », ou "Linguistic Linked Licensed Data" (3LD) cloud, lequel comprend des ressources linguistiques ouvertes ou sous licence en format de données liées.

Groupes visés

Un large panel de communautés bénéficiera du projet LIDER: consommateurs et fournisseurs de contenu multilingue et multimédia, entreprises multinationales, institutions européennes, PME, fournisseurs de services linguistiques, ainsi que la communauté des technologies du langage dans son ensemble. Par le biais d'une coopération active avec ces communautés, le projet LIDER identifiera des cas d'usages et mènera à bien des analyses de rentabilité qui détermineront et guideront les cas d'utilisation du projet.

Résultats attendus

Le projet LIDER entend créer:

1. Une large et solide communauté autour de la question de l'analyse de contenu basée sur l'utilisation du web des données ouvertes et liées dans un contexte multilingue et multimédia..
2. Un jeu de recommandations et de bonnes pratiques pour la construction et l'exploitation de ressources sur le web des données liées pour l'analyse de contenu et pour le développement de services de traitement automatique du langage basés sur des technologies propre au web des données liées.
3. Une architecture de référence pour le web des données liées linguistiques, intégrant des plateformes et ressources existantes et à venir.
4. Une feuille de route pour l'utilisation de technologies propre au web des données liées pour l'analyse de contenu multimédia et multilingue en entreprise.

Impact

LIDER permettra aux communautés mentionnées ci-avant de développer une compréhension commune des enjeux d'une représentation sémantique des informations relatives au langage. La conception de la feuille de route à long terme sera basée sur la prise en compte des contributions de diverses entreprises et organismes de recherche.

Le projet LIDER posera les bases pour la réutilisation de solutions d'analyse de contenu existantes pour d'autres langues ou médias, et la feuille de route exercera une influence positive sur les partenaires impliqués, en contribuant à l'excellence scientifique et à la coordination, tout en assurant la pertinence industrielle de leurs activités de recherche.

Coordinateur

Personne de contact: Prof. Dr. Asunción Gómez-Pérez

- Tél:+34 913367417

- E-mail: asun@fi.upm.es
- Organisation: 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

Organisation du projet

- Durée: 24 mois
- Début: 1 Novembre 2013
- Fin: 31 Octobre 2015
- Responsable de projet: Susan Fraser
- Projet 610782, CSA

Comment contribuer et être informé

- Site internet du projet: <http://www.lider-project.eu/>
- Portail communauté: <http://www.multilingualweb.eu/>
- Group W3C en constitution sur le thème "Linked Data for Language Technology"
- Liste de diffusion lider-community@listas.fi.upm.es pour les retours anonymes et être informé. Souscription *via*: <http://www.lider-project.eu/get-involved>
- Présence sur les réseaux sociaux avec le hastagh #LiderEU, par ex. sur twitter: <https://twitter.com/multilingweb>