



LIDER – Linked Data como um facilitador para cross-media e análise de conteúdo multilíngue para empresas europeias

Descrição

O objetivo do projeto LIDER é abordar os desafios colocados pelo crescimento exponencial de conteúdo multilíngue e multimídia na Web: o aumento de volume, velocidade e variedade de conteúdo requer novas abordagens para a análise. Para este fim, a compreensão e interpretação da linguagem, tanto conteúdo não estruturado textual como este conteúdo linguístico em diferentes conjuntos de dados, são vitais. O projeto LIDER visa estabelecer diretrizes para a criação de um ambiente de inovação, semanticamente interoperável e ligados entre si no âmbito de dados abertos ligados (Linked Open Data) recursos livres. Este ambiente irá integrar os recursos linguísticos (dicionários, corpora, metadados léxica e sintática, etc.) e recursos multimídia (metadados de imagens, vídeo, etc.). O LIDER irá estudar como esses recursos podem ser uma tecnologia que facilite a análise da Web multilíngue, incluindo dados multimídia no conteúdo multilíngue.

LIDER: Principais realizações no primeiro ano

➤ Introdução aos diferentes tipos de comunidades e definição de possíveis casos de uso de negócios.

LIDER organizou quatro workshops dedicados ao desenvolvimento de um roteiro, no qual foram coletadas ideias sobre os possíveis casos de uso nas empresas, bem como os requisitos de dados linguísticos que deviam ser respeitados. Estes workshops têm-se centrado em quatro grupos: profissionais da área de dados, trabalhando em questões da Web, especialistas em localização e especialistas na análise de conteúdo. O grupo de trabalho LD4LT (*Linguistic Linked Data for Language Technology*) do consórcio W3C recolheu os

comentários feitos nas reuniões, que são detalhados nos relatórios dos workshops. Da mesma forma, o LIDER participou de numerosas reuniões da comunidade científica e do setor empresarial, organizou sessões de treinamento e fez apresentações sobre dados linguísticos vinculados e análise de conteúdo.

➤ Estabelecendo diretrizes e melhores práticas.

O LIDER tem contado com o grupo de trabalho BPMLOD (*Best Practices for Multilingual Linked Open Data*) do consórcio W3C, dedicado a coletar diretrizes e melhores práticas na criação de fontes de dados linguísticos ligados. Estas orientações foram seguidas na elaboração de métodos de migração de recursos linguísticos de dados ligados existentes. Em primeiro lugar, têm transformado os metadados de recursos linguísticos. LIDER tem trabalhado em estreita colaboração com a comunidade de META-NET e CLARIN para desenvolver um modelo de consenso da descrição linguística de metadados e, assim, assegurar que os sistemas de metadados existentes são reutilizados. Em segundo lugar, o LIDER tem gerado um conjunto de melhores práticas para a geração de dados linguísticos ligados, incluindo dicionários bilíngues ou multilíngues, e preparou as diretrizes para a migração de corpora e recursos terminológicos (TBX).

➤ Projeto de uma arquitetura de referência e proposta de um roteiro. O LIDER criou um primeiro esboço de uma arquitetura de referência para dados linguísticos ligados e para análise de conteúdo linguístico, juntamente com um resumo das ferramentas e iniciativas relevantes. A arquitetura se define tanto um modelo geral para a construção de serviços sensíveis aos dados ligados à linguagem como padrões diferentes para a concepção de serviços de aplicações de análise de conteúdo.



Assim apresentamos uma visão geral das iniciativas e ferramentas relevantes. A arquitetura é complementada por uma versão inicial do roteiro para a criação de dados linguísticos ligados. Especificamente, o roteiro integra ideias e comentários de várias fontes, ambas as atividades LIDER (tais como pesquisas ou workshops), relatórios ou registros, agências de financiamento e de várias comunidades de pesquisa. O roteiro é formulado em torno das principais áreas de aplicação: envolvimento do cliente, do setor público e da sociedade civil e do ciclo de vida e da cadeia de publicação de dados linguísticos ligados. Para cada uma destas áreas, o roteiro inclui casos de uso, prazos, previsões e fatores significativos.

Próximos objetivos

LIDER irá apresentar os progressos realizados durante o primeiro ano do projeto aos agentes interessados. Em particular, ser questões relacionadas com a migração de recursos linguísticos e seus metadados, arquitetura de referência para os dados linguísticos ligados e seu uso na análise de conteúdo e será discutido o roteiro para o uso e criação dos mesmos. Da mesma forma, o LEADER vai abordar novos grupos, especialmente aqueles interessados em dados abertos no setor público.

Coordenador

Pessoa para contato: Prof. Dr. Asunción Gómez-Pérez

Tel:+34 913367417

E-mail: asun@fi.upm.es

Organização: 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

Cronograma do projeto

Duração: 24 meses

Data de inicio: primeiro de novembro de 2013

Data de finalização: 31 de outubro de 2015

Chefe do projeto: Susan Fraser

Projeto 610782, CSA

Para solicitar mais informacoes de como participar:

- Página web do projeto: <http://www.lider-project.eu/>
- Portal da comunidade: <http://www.multilingualweb.eu/>
- Grupo de trabalho W3C sobre “Linked Data for Language Technology”: <http://www.w3.org/community/ld4lt/>
- Lista de email lider-community@listas.fi.upm.es para boletins informativos.
- Escreva-se a través de <http://www.lider-project.eu/get-involved>.
- Atividades nas redes sociais mediante a hashtag #LiderEU. Twitter: <https://twitter.com/multilingweb>