



LIDER – *Verlinkte Daten als Anwendungsmöglichkeit von multimedialer und multilingualer Inhaltsanalytik für Unternehmen in Europa*

Hintergrund

Die Absicht von LIDER ist es, die Herausforderungen des explosionsartigen Wachstums von Umfang, Vielfalt und Geschwindigkeit in der Entstehung von Webinhalten zu thematisieren, da diese neue Herangehensweisen im Bereich der Inhaltsanalytik verlangen. Schlüsselemente dabei sind das Verstehen und Interpretieren von Sprache im Kontext, sowohl in unstrukturierten Textinhalten als auch in sprachlichen Inhalten, die im Kontext verschiedener Medienströme auftreten.

LIDER wird einen Leitfaden für die Erstellung eines neuen Linked Open Data (LOD) basierten Ökosystems von offenen, verlinkten und semantisch interoperablen Ressourcen sowohl aus dem Bereich der Sprachen ("Linguistic Linked Data" Representationen von Korpora, Wörterbüchern, lexikalischen and syntaktische Metadaten, etc.) als auch der Medien (Bild, Video, Metadaten ect.) erstellen. LIDER wird untersuchen, wie diese Ressourcen zur technologischen Entwicklung im Bereich der Inhaltsanalytik im multilingualen Web innerhalb von Unternehmen beitragen. Dies bezieht sich ebenfalls auf multilinguale Inhalte, die in verschiedenen Medien vorhanden sind.

Hauptergebnisse im Jahr 1 von LIDER

➤ **Kontaktaufnahme mit Communities, Erstellung von Unternehmens-Anwendungsfällen.** LIDER hat vier Roadmapping-Workshops ausgerichtet, um Anregungen für Unternehmens-Anwendungsfälle und Anforderungen für linguistische verlinkte Daten zusammenzutragen. Die Workshops richteten sich an vier Gruppen: die Daten-Community, die Web-Community im Allgemeinen, die Lokalisierungs-Community und die Analytik-Community. Die Rückmeldungen wurden durch eine dazu

eingeschichtete W3C Community-Gruppe, LD4LT ("Linguistic Linked Data for Language Technology"), gesammelt. Des Weiteren wurden sie in den Workshop-Berichten dokumentiert und mit in den LIDER Fahrplan einbezogen. Zusätzlich war LIDER bei einer großen Anzahl von Forschungs- und Industrieereignissen anwesend und hat Tutorien sowie Präsentationen über linguistische verlinkte Daten und Inhaltsanalytik gehalten.

➤ **Richtlinien und Leitfäden.** LIDER hat eine weitere W3C Community-Gruppe, BPMLOD ("Best Practices for Multilingual Linked Open Data"), eingerichtet, um Leitfäden zur Erstellung von multilingualen verlinkten Datenquellen zusammenzutragen. Konversionen in ein verlinktes Datenformat für bereits existierende Sprachressourcen wurden entwickelt. Dies betrifft zuallererst die Metadaten der Sprachressourcen. LIDER hat eng mit den META-NET- und CLARIN-Communities an den Konversionen zusammengearbeitet, um eine breite Akzeptanz und Verteilung von bereits existierenden Metadaten schemata sicherzustellen. Anschließend hat LIDER Richtlinien für diverse Typen von Ressourcen, wie beispielsweise bilinguale oder multilinguale Wörterbücher, verfasst und mit Konversionen für weitere Ressourcen, wie terminologischen (TBX) Daten oder Korpusdaten begonnen.

➤ **Referenzarchitektur und Roadmapping.** LIDER hat einen ersten Entwurf für eine Referenzarchitektur für linguistische verlinkte Daten für den Bereich der Inhaltsanalytik entwickelt. Diese Architektur definiert ein allgemeines Modell zur Erstellung von Diensten, die sensibilisiert für linguistische verlinkte Daten sind sowie einiger Muster zur Anfertigung von Anwendungen für Inhaltsanalytik. Zudem gibt sie einen Überblick über

mlider

Werkzeuge und Initiativen, die relevant für die Architektur sind.

LIDER hat außerdem eine erste Version eines Fahrplans für linguistische verlinkte Daten für Inhaltsanalytik entworfen. Der Fahrplan interpretiert Beiträge aus diversen Quellen: LIDER-Aktivitäten wie Umfragen oder Roadmapping-Workshops, diverse Berichte, Förderungseinrichtungen und verschiedenen Forschungsgruppen. Der Fahrplan ist in Bezug auf drei Hauptanwendungsgebiete verfasst: globales Kundenengagement, öffentlicher Sektor/Zivilgesellschaft und der linguistische verlinkte Datenlebenszyklus sowie die damit verbundene linguistische verlinkte Datenwertkette. Für jeden dieser Bereiche stellt der Fahrplan einige Anwendungsfälle dar, die Prognosen, Zeitrahmen und relevante Handlungspersonen umfassen.

Weiteres Vorgehen

LIDER wird die Ergebnisse vom ersten Jahr auswerten und mit der Community im Allgemeinen diskutieren. Dies betrifft Konversionen von Sprachressourcen und deren Metadaten, die linguistische verlinkte Referenzarchitektur für Inhaltsanalytik und ebenso den Fahrplan für linguistische verlinkte Daten für Inhaltsanalytik. Darüber hinaus wird LIDER Kontakt zu neuen Gruppen aufnehmen, die im ersten Jahr noch nicht

gezielt angesprochen wurden, vor allem im Bereich der Informationen aus dem öffentlichen Sektor.

Koordinatorin

Kontaktperson: Prof. Dr. Asunción Gómez-Pérez
Tel:+34 913367417

E-mail: asun@fi.upm.es

Organisation: Universidad Politécnica de Madrid (UPM)

Projektteilnehmer

- Universidad Politécnica de Madrid (UPM)
- Trinity College Dublin (TCD)
- Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI)
- INSIGHT National Center for Data Analytics, National University of Ireland, Galway
- Institut für Angewandte Informatik (InfAI), Universität Leipzig
- Universität Bielefeld
- Sapienza University of Rome
- W3C / Geie ERCIM

Projektaufbau

Dauer: 24 Monate

Beginn: 1. November 2013

Ende: 31. Oktober 2015

Projektbeauftragte: Susan Fraser

Projekt 610782, CSA

Wie man sich informieren und zu LIDER beitragen kann:

- Projekt-Webseite: <http://www.lider-project.eu>
- Community-Portal: <http://www.multilingualweb.eu>
- W3C Community-Gruppe "Linked Data for Language Technology": <https://www.w3.org/community/ld4lt>
- Mailing-Liste: lider-community@listas.fi.upm.es, um anonym Feedback und Informationen zu sammeln. Zur Einschreibung <http://www.lider-project.eu/get-involved> verwenden
- Soziale Medienpräsenz via Hashtag #LiderEU e.g. auf Twitter: <https://twitter.com/multilingweb>