

Survey Outline

This document specifies the approach taken in LIDER to the survey of industry to determine:

- To enumerate baseline awareness of the potential of:
 - Linked data in general
 - Linked data in applications of NLP
 - NLP for content analytics
- To enumerate baseline of maturity of above as input for WP3
- To elicit use cases or pain points that could be addressed by the use of linked data in application of NLP for content analytics
- To elicit requirements and constraints that may impact applications of linked data in NLP applications for content analytics
- Implicit goals:
 - Improve awareness of potential of linked data for NLP applications and the expertise in Europe
 - Identify potential partners for H2020

Survey Content

Introduction

LIDER is assembling an industrial community to gather information on the key data requirements in the application of natural language technology to content analytics in different sectors. The outcome will be: guidelines, best practices and a reference architecture targeted at high priority industry use cases and addressing knowledge gaps in the market. It will also identify a roadmap for future research development and innovation activities to address any major gaps in applying open data techniques to content analytics.

The analysis of content exchanged on the web is a key function in more and more industries. This requires the integration analysis of the metadata associated with digital content and the natural language contained in the content. Natural language technologies such as machine translation, sentiment analysis, term extraction etc. are increasingly being deployed in content processing applications. The effectiveness of such NLP technology is in turn dependent on the relevance of the training data used to the content in the application domain. Accessing such training data and other relevant lexical and semantic data via the web can therefore make the accurate use of NLP in different content analysis scenarios easier and cheaper.

Survey Questions

Name and email address (this will be treated as confidential by the LIDER consortium and all survey results will be presented anonymously)

Organisational profile

- large company
- SME
- public sector
- non profit
- sole trader
- other

Industry Sector: Please identify the industry sectors in which you are most active (scale: core business activity, secondary activity, planned business activity, not relevant)

NLP Use: Which of the following NLP technologies do you employ or plan to employ (scale: in use, trialed, considering, not relevant)

- terminology extraction
- named entity recognition/entity linking
- word sense disambiguation
- machine translation
 - statistical machine translation
 - rule based machine translation
- sentiment analysis
- trend analysis
- event detection
- information extraction
- other (please specify)

Data Use: what types of data are relevant to your use of the NLP technologies you employ (scale: critical, desirable, not relevant, not used)

Priorities: What aspect of this data are most important to your use of it with NLP (scale: critical, desirable, not relevant, not used)

- provenance of data

- quality of value
- quality of links
- license terms under which data is available
- the format under which the data is available
- the use of open formats for data
- the persistence of the data source
- the cost of the data
- other

Use Cases: The development of guidelines & best practice, linked data reference architecture and roadmapping for future R&D, standardisation and innovation actions depends of a deep understanding of current industrial business cases where data is used or shared for nlp based applications. Your input on business cases is invaluable in this analysis. You may therefore provide such information below:

- actors involved
- benefits to actors
- steps or scenarios

Pain points and constraints: Alternatively it is useful to understand current pain points in content processing that could be potentially addressed by NLP and linguistic data, even if the business case is not yet apparent.

Further Engagement: The LIDER project will be organizing a number of workshops to further refine our understanding of use cases as well as communicating updates on a regular basis and provide. Please tick the box below if you would like to register to the lider industrial community.

H2020: The EU funding of the H2020 programme provides a number of opportunities for collaboration between industry and researchers for research, development and innovation, with deadline for proposals related to content processing with deadlines in April and October 2014. If you are interested in exploring partnerships for H2020 you can enter information below and it will be shared in a closed user group:

- your interests on a partnership, what outcome you would like to achieve
- the specific industry problem you would like to address
- what you might be able to contribute to a research and innovation partnership
- the types of knowledge, skills or capabilities you would seek in potential partners

