Introduction to FREME: Data meets Language meets Business

Felix Sasaki[1], Tatiana Gornostay[2], Milan Dojchinovski[3], Michele Osella[4], Erik Mannens[5], Giannis Stoitsis[6], Phil Ritchie[7], Kevin Koidl[8]

[1] DFKI, felix.sasaki@dfki.de; [2] Tilde, tatiana.gornostay@tilde.lv; [3] InfAI, milan.dojchinovski@fit.cvut.cz; [4] ISMB, osella@ismb.it; [5] iMinds, erik.mannens@ugent.be; [6] Agro-Know, stoitsis@agroknow.gr; [7] VistaTEC, philr@vistatec.ie; [8] Wripl, kevin@wripl.com

Abstract. This short paper introduces the FREME project, a new Horizon 2020 innovation action. FREME aims at building an open framework of e-Services for multilingual and semantic enrichment of digital content, based on a reusable set of open Application Programme Interfaces and Graphical User Interfaces to FREME enrichment services. As a result, FREME will improve the existing processes of digital content management.

1 Introduction

FREME is a Horizon 2020 innovation action that started in February 2015 with the duration of two years. Partners are DFKI (as coordinator), Tilde, iMinds, Agro-Know, Wripl, VistaTEC, InfAI and ISMB. FREME aims at building an open framework of e-Services for multilingual and semantic enrichment of digital content.

2 Background: The Concept of FREME

The growing amount of digital content across languages, sectors and domains leads to both challenges and business opportunities for many industries. Linked data (LD) and language technology (LT) solutions exist, providing e.g. machine translation, entity recognition, or multilingual linked data sets. These solutions face several issues, e.g.: a plethora of content formats to process; adaptability and "silo solution" dependency; and usability in an industry application scenario: the lack of adequate tooling for a given or new group of user types (authors, translators, data wranglers or scientists etc.) in selected business scenarios.

The FREME framework addresses these issues by providing a **reusable set of open Application Programme Interfaces and Graphical User Interfaces** to FREME **enrichment services**. In this way, FREME will improve the existing **processes of digital content management**. The improvement will go through the whole content value chain: content creation (or authoring), content translation/localization, publishing and access to content including cross-language sharing and personalized content recommendations. Thus, FREME aims at opening new opportunities for all sectors that are involved in digital content management.

The FREME framework will develop and integrate the following **e-Services** based on existing and mature technologies:

- e-Internationalization based on Internationalization Tag Set (ITS) 2.0;
- e-Link based on Natural Language Processing Interchange Format (NIF) and various (linked open) data sets;
- e-Entity: based on entity recognition and existing linked entity datasets;
- e-Terminology: based on cloud terminology services for terminology management and terminology annotation;
- e-Translation: based on cloud machine translation services for building custom machine translation systems;
- e-Publishing: based on cloud content authoring environment and its export for publishing in the EPUB 3 format.

The innovation, robustness and usability of the FREME framework of e-Services will be shaped by the following four FREME real world **business cases**:

1. Authoring and publishing multilingually and semantically enriched eBooks;
2. Integrating semantic enrichment into translation and localization processes;
3. Enhancing cross-language access & sharing for agricultural and food data;
4. FREME-empowered personalized content recommendations.

3 Networking Expectations

The FREME presentation will show prototype implementations of the e-Services in action. We will discuss applications and the benefit gained in data value-chains created via FREME, across languages, domains and sectors. We also will explore research potentials, taking current technology developments in the data and language communities into account. The FREME project is unique in that it gathers key players from both communities.

Via the networking session FREME wants to connect to others that are interested in multilingual data sets and data processing. This will be of interest for

- data set providers: what data sets could benefit from FREME as input to enrichment, or for being enriched via FREME e-Services?
- data sets users: what use cases can benefit from multilingual semantic enrichment?
- providers of multilingual data set related technologies: what FREME technology components are useful for your research and development? Complementary, what components could improve FREME, e.g., in terms of language or domain coverage?

The FREME focus of domains, data sets and languages is developed in an agile manner: FREME prototyping and business case needs are interacting constantly. The current state will be presented during the networking session. A special – but not exclusive – focus is on data sets that are part of the linguistic linked data (LLD) cloud. These are crucial for multilingual content analytics tasks. FREME is closely cooperating with the FP7 LIDER and FALCON projects, e.g., deploying LIDER best practices for creating and accessing LLD.