xLiMe
FP7-ICT-2013-10
ICT-2013.4.1
Contract no.: 611346
www.xlime.eu

Deliverable D7.3.2

Intermediate Prototype and Validation Report MONITOR

Editor:	Eduardo Torres Schumann, VICO
Author(s):	Eduardo Torres Schumann, VICO
Deliverable Nature:	Prototype (P)
Dissemination Level:	Consortium (CO)
Contractual Delivery Date:	M24 – 31 October 2015
Actual Delivery Date:	M24 – 31 October 2015
Suggested Readers:	All project partners, users of (social) media monitoring systems.
Version:	1.0
Keywords:	VICO; Use Case; Monitoring

Disclaimer

Full Project Title:	xLiMe – crossLingual crossMedia knowledge extraction
Short Project Title:	xLiMe
Number and Title of Work package:	WP7 Use Cases and Evaluation
Document Title:	D7.3.2 - Intermediate Prototype and Validation Report MONITOR
Editor:	Eduardo Torres Schumann, VICO
Work package Leader:	Ronald Denaux, ESI

Executive Summary

This document describes the development of the Intermediary Prototype for the Use Case MONITOR performed by the XLiMe partner VICO Research & Consulting GmbH during Year Two of the project.

While for the Early Prototype developed during Year One the focus was on the integration of data delivered by xLiMe partners into VICO's framework for social media monitoring, the further development of the prototype aimed at supporting two specific, particularly laborious processes related to the Use Case by using technology delivered by xLiMe partners. The first process supported by the Intermediary Prototype arises at the beginning of the implementation of a monitoring project and constitutes the modelling of Research Objects to be monitored by writing queries. The second one is related to service activities involved in the Use Case such as customer reporting and represents the process of categorizing and classifying data corresponding to research objects already being monitored, with the purpose of obtaining valuable insights.

Modelling Research Objects for monitoring was achieved in the Intermediary Prototype by using queries that select relevant data by referring to annotations made on the data as it is delivered by the consortium partners. This enables to get rid of the keyword queries used before, whose definition is very time consuming, and at the same time, to retrieve relevant data in a cross-lingual and cross-media way. Experiments were conducted using a specific methodology, trying to evaluate the effort required for modelling the queries based on annotations and the quality of the corresponding data results and comparing these aspects to the previous stage represented by the keyword queries. As outcome, modelling Research Objects representing single entities is in general easier than for the keyword pendants, whereas queries representing broad topics are still difficult to define even on the basis of semantic annotations. The quality of results delivered by semantic queries vary significantly along the different sources (news articles, social media and TV transcripts), but improvements in processing social media and TV transcript text from WP2 presumably will resolve this.

Easing the process of interpreting monitoring results and writing customer reports was achieved in the Intermediary Prototype by implementing a special component, the Topic Annotator, based on the EventRegistry API developed within WP4. Events mentioned in monitoring results are detected using the API and clustered together, each cluster representing a topic. The data is then annotated and categorized in terms of these topics mentioned in the corresponding text and displayed in the fronted, making easy to visualize the evolution of different detected topics over time. An in-depth evaluation of the automatically categorized results that compares them to available manually categorized data is planned, but preliminary results also indicate that a significant facilitation of this process can be achieved.

Although the described improvements are not in a stage ready for start of production, we conclude that they constitute a gain in efficiency that can help opening up new business opportunities.

