Web Semantics: Science, Services and Agents on the World Wide Web 46-47 (2017) 20–30

Contents lists available at ScienceDirect

Web Semantics: Science, Services and Agents on the World Wide Web

journal homepage: www.elsevier.com/locate/websem

The xLiMe system: Cross-lingual and cross-modal semantic annotation, search and recommendation over live-TV, news and social media streams

Lei Zhang *, Andreas Thalhammer, Achim Rettinger, Michael Färber, Aditya Mogadala, Ronald Denaux

Institute AIFB, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
Expert System Iberia, Madrid, Spain

ARTICLE INFO

Article history:
Received 1 October 2016
Received in revised form 1 March 2017
Accepted 26 March 2017
Available online 5 April 2017

Keywords:
Semantic annotation
Semantic search
Semantic recommendation
Cross-lingual
Cross-modal

ABSTRACT

Modern Web search engines still have many limitations: search terms are not disambiguated, search terms in one query cannot be in different languages, the retrieved media items have to be in the same language as the search terms and search results are not integrated across a live stream of different media channels, including TV, online news and social media. The system described in this paper enables all of this by combining a media stream processing architecture with cross-lingual and cross-modal semantic annotation, search and recommendation. All those components were developed in the xLiMe project.

1. Motivation

The amount of entities in large knowledge graphs (KGs) has been increasing rapidly, enabling new ways of semantic information access, like keyword and semantic queries over entities and concepts mentioned in heterogeneous media items. While *entity search* has become a standard feature, Web search engines are still limited in their semantic processing capabilities: it is not possible to disambiguate search terms manually, search terms in one query cannot be in different languages, the retrieved media items have to be in the same language as the search terms and search results are not collected across a live stream of different media channels.

In this work, we demonstrate a system that intends to break the barriers in between languages and modalities for a seamless semantic access to media streams. We first introduce a real-time processing software architecture and an annotation data model (Section 2) before describing the components for cross-lingual annotation of multilingual text from multiple channels, such as Live-TV, social media and online news (Section 3). The annotated cross-channel media stream allows multilingual and cross-lingual semantic search (Section 4) as well as cross-media recommendation (Section 5). We measure the scalability of the complete system in terms of several metrics (Section 6) before comparing the features of our system to related approaches (Section 7).

This paper provides an overview of components developed in the xLiMe project[1] for the processing of media content across languages, modalities and channels using *explicit semantics*.

2. Cross-lingual and cross-modal processing of semantic media streams

The processing of different multimedia streams is a cost-intensive task. It has been best performed in a distributed manner. Unfortunately, the various sources, their individual particularities, and their distributed processing pose a huge challenge for data integration. As such, we consider three different contributions: (1) multimedia sources; (2) intelligent processing and (3) semantic integration.

The media sources include online news, social media and TV content. All of these sources are multilingual media streams with different and – in the case of social media – changing velocity. The processors include annotation tools for text (i.e., entity linking) and video (i.e., optical character recognition), accompanied by speech-to-text processing (i.e., automatic speech recognition) in the case

* Corresponding author at: Institute AIFB, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany.
E-mail addresses: l.zhang@kit.edu (L. Zhang), thalhammer@kit.edu (A. Thalhammer), rettinger@kit.edu (A. Rettinger), michael.faerber@kit.edu (M. Färber), aditya.mogadala@kit.edu (A. Mogadala), rdenaux@expertsystem.com (R. Denaux).

http://dx.doi.org/10.1016/j.websem.2017.03.002
1570-8268/

[1] http://www.xlime.eu

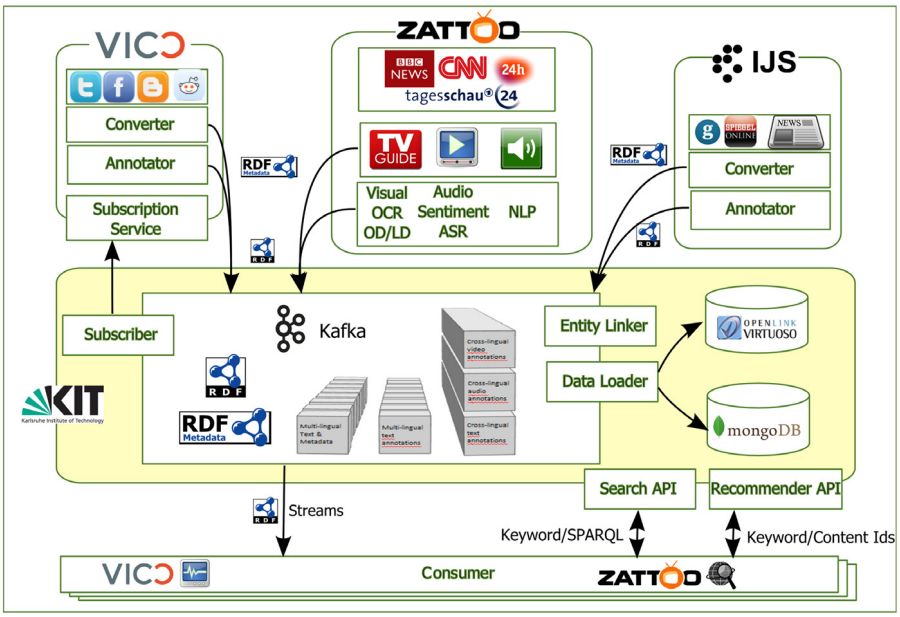

Fig. 1. The xLiMe architecture.

of the audio streams of the TV content. Most of this processing involves a high demand in computational power and sophisticated machine learning models. The semantic integration of different media streams poses the challenge to identify a common model that suits the diversity of the data sources and the output of the processing engines. Further, we combine the processed data with additional background knowledge from knowledge bases.

2.1. System architecture

The architecture of the xLiMe project is divided into multiple components (see Fig. 1). For practical reasons, the multimedia source and initial processing infrastructure is respectively attached to the institutions that provide the respective data. Raw data as well as (intermediately or fully) processed data is directly sent to an Apache™ Kafka message broker that enables a multitude of different topics (communication channels). The partners that provide data processing capabilities provide meta and provenance data in accordance to the xLiMe data model (that will be introduced in the next section). As also the raw data is pushed to the message broker, every partner that has processing capabilities and tools can provide enhanced or alternative services. Along with the message broker, a triple store (i.e., Virtuoso) and a NoSQL database (i.e., MongoDB) provide further data integration and query capabilities. Like this, individual hooks are subscribed to specific Kafka topics and constantly load data into the respective store.

The xLiMe triple store is individually queryable and enables restrictions and aggregates on multiple modalities, languages, and sources. The same accounts for the NoSQL database. This enables the flexible operation of services that build on live streaming data in combination with additional background knowledge. Based on the integrated data in the triple store and NoSQL database, xLiMe components enable us to ask complex questions using SPARQL queries and to search for different media channels using keyword queries.

2.2. Data model

The xLiMe data model is defined as an RDF vocabulary and tailored specifically to the different modalities: text, audio, and video. It extends other vocabularies such as Dublin Core,[2] SIOC,[3] and KDO.[4] Its main scheme is depicted in Fig. 2. Similarly to the Web Annotation Model,[5] it enables to relate text and (parts of) video or audio streams to real world entities. In the xLiMe project we refrained from using the Web Annotation Model in order to reduce the amount of unnecessary blank nodes and thus, at query time, joins. The predicates that define the start and stop positions can be used in a flexible manner and may define character positions, in the case of text, or milliseconds/frame numbers in case of audio/video. In order to describe rectangular fragments of videos or images, there is a specific class that defines the visual position. In any case, the recognized entity should relate to a resource in the knowledge base.

Another particularity of the xLiMe data model – that is not depicted in Fig. 2– is extensive use of named graphs where we use the W3C provenance data model[6] in order to provide meta data for the respective processing of one or more media items.

3. Cross-lingual semantic annotation

In this section, we present *X-LiSA* [1], an infrastructure for *cross-lingual semantic annotation*, which supports interfaces for annotating media data with resources in knowledge bases. It helps to bridge the ambiguity of unstructured data and its formal semantics as well as to transform such data in different languages into a unified representation.

3.1. System architecture

The architecture of *X-LiSA* is shown in Fig. 3, where *cross-lingual groundings extraction* is performed offline to generate the indexes used by the online *cross-lingual semantic annotation*.

Cross-lingual groundings extraction. For matching words and phrases in different languages against entities in knowledge bases,

2 http://dublincore.org/documents/dces/
3 http://www.w3.org/Submission/sioc-spec/
4 http://render-project.eu/resources/kdo/
5 http://www.w3.org/TR/annotation-model/
6 https://www.w3.org/TR/prov-dm/