Socialising Around Media (SAM): Dynamic Social and Media Content Syndication for Second Screen[*]

Socialising Around Media (SAM): Sindicación Dinámica de Contenido Social y Multimedia para Segundas Pantallas

David Tomás, Yoan Gutiérrez, Isabel Moreno, Francisco Agulló
University of Alicante, Spain
{dtomas,ygutierrez,imoreno,fagullo}@dlsi.ua.es

Marco Tiemann
University of Reading, United Kingdom
m.tiemann@reading.ac.uk

Juan V. Vidagany
TIE Kinetix, The Netherlands
juanvi.vidagany@tiekinetix.com

Andreas Menychtas
National Technical University of Athens, Greece
ameny@mail.ntua.gr

Resumen: La actual generación de dispositivos móviles ha cambiado la forma en que los usuarios interactúan con los medios digitales, pasado de ser pasivos y unidireccionales a proactivos e interactivos. Los usuarios usan estos dispositivos para comentar y valorar programas televisivos, buscando información relacionada sobre personajes, hechos y celebridades. Este fenómeno se conoce con el nombre de *segunda pantalla*. En este artículo se describe SAM, un proyecto de investigación financiado por la EU y enfocado al desarrollo de una plataforma avanzada de distribución de contenidos digitales basada en segundas pantallas, usando la sindicación de contenidos en el contexto de los medios sociales para proporcionar maneras abiertas y estándares de caracterizar, descubrir y sindicar recursos digitales. Este trabajo resume las principales características y objetivos del proyecto, así como los retos de PLN a afrontar y las tecnologías desarrolladas para tal fin hasta el momento.
Palabras clave: Sindicación de contenidos, segundas pantallas, medios sociales

Abstract: Today's generation of Internet devices has changed how users are interacting with media, from passive and unidirectional users to proactive and interactive. Users can use these devices to comment or rate a TV show and search for related information regarding characters, facts or personalities. This phenomenon is known as *second screen*. This paper describes SAM, an EU-funded research project that focuses on developing an advanced digital media delivery platform based on second screen interaction and content syndication within a social media context, providing open and standardised ways of characterising, discovering and syndicating digital assets. This work provides an overview of the project and its main objectives, focusing on the NLP challenges to be faced and the technologies developed so far.
Keywords: Content syndication, second screen, social media

1 Introduction

The current generation of Internet devices has changed how users are interacting with media, from passive and unidirectional users to proactive and interactive. Users can comment or rate a TV show and search for related information regarding characters, facts or personalities. This phenomenon is known as *second screen*. In today's second screen environment there are no true standards, protocols or commonly used frameworks through which users can discover and access information related to consumed contents. Users have to actively perform searches using web search

David Tomás, Yoan Gutiérrez, Isabel Moreno, Francisco Agulló, Marco Tiemann, Juan V. Vidagany, Andreas Menychtas

engines such as Google to "participate" in TV shows.

This paper describes SAM[1] (Socialising Around Media), an EU-funded project focusing on developing an advanced digital media delivery platform for second screen and content syndication within a social media context, providing open and standardised ways of characterising, discovering and syndicating digital assets (e.g. films, songs, books and metadata).

The potential customers of SAM are both business stakeholders (such as media broadcasters, content asset providers, software companies and digital marketing agencies) and end users. For the former, this platform will provide a number of benefits, including dynamic social and media content syndication, managing of online reputation, listening to customers, tracking real time statistics or monitoring media related social content through second screen. For the latter, SAM will offer a complete solution for people consuming media and TV programs. The platform will integrate context aware information and complex social functionalities providing contextual information about their actual and current interest. This will provide the end users with an augmented experience in which they can discover new information about the subject, and will talk and share their experience with other users that are also interested in the same topic. Users will produce and consume digital assets from different syndicated sources and synchronised devices (e.g. tablets, smartphones and connected TVs) thus creating richer experiences around the original assets.

The key innovation of SAM is that instead of users reaching for the data, it is the data that reaches the users through the syndication approach and their second screen. Core functionalities of SAM are based on Natural Language Processing (NLP) technologies—including sentiment analysis, text summarisation and semantic analysis—which are applied to both formal texts and user generated content in social networks (i.e. user comments).

The remainder of this article is organised as follows: next section provides information about the members of the SAM consortium; Section 3 describes the three core technolo-

gies supporting the development of this project; Section 4 summarises the main goals of the project; Section 5 describes the challenges faced in SAM by NLP technologies; finally, Section 6 describes the current development status of the project.

2 The Consortium

SAM involves nine partners from a pan-European consortium, presenting a well-balanced combination of research institutions, software developers and user companies (including a media broadcaster, a content provider and a smart TV manufacturer):

- *Research institutions*: National Technical University of Athens (Greece), University of Reading (United Kingdom) and University of Alicante (Spain).
- *Technology SMEs*: TIE Kinetix, coordinator (The Netherlands), Ascora GmbH (Germany) and Talkamatic (Sweden).
- *User companies*: Deutsche Welle (Germany), Bibliographic Data Services (United Kingdom) and TP Vision (Belgium).

3 The Three Pillars of SAM

The SAM platform has been designed around three pillars that highlight the main research and business directions of this project: content syndication, second screen and social media. By combining these pillars together, SAM will implement the distribution of media assets to the end users through their devices, including linked content and related information to enrich the user experience.

Content Syndication Technologies like Really Simple Syndication (RSS) have their place in the syndication world, but when an organisation needs to push more enriched and target-adapted information to partner websites or social networking sites (from product details to full microsites with rich media), approaches such as RSS are not powerful enough. Content syndication solutions today have evolved beyond RSS, allowing vendors, distributors and publishers to issue, control and track rich content experiences on third-party websites. In SAM, content syndication techniques will allow content providers to prepare their digital assets and associate them to specific media and usage context, offering mechanisms for these enriched assets to

---

[1]http://www.socialisingaroundmedia.com.

be delivered in the expected format and to be consumed by the users in a specific context.

Second Screen This concept refers to any electronic device (broadly a mobile device, such as a tablet or smartphone) that allows users to retrieve additional information about the content they are watching on the first screen (usually a TV set). The SAM platform includes a multi-device representation layer that provides syndicated information in the appropriate format to be consumed by different types of devices. This generic approach is used in order to access the asset-related syndicated information while it is being consumed, commented, or interacted with by the users, creating a second screen experience.

Social Media These technologies are changing the way in which users interact and communicate with each other, expressing their feelings, opinions and thoughts about almost anything—including products, personalities and TV shows. In social media, users not only share comments or articles, they also exchange different types of digital assets such as videos, photographs and documents. In recent years, the user activity in social networks has significantly increased, making it into a key area of interest for media business and advertisers. Decision makers try to find ways in which commercial products can make profitable use of applications such as YouTube, Facebook and Twitter. In SAM, social interaction around digital media items will provide the context in which the syndicated content will be consumed. SAM will incorporate complex context extraction mechanisms based on NLP technologies, such as sentiment and semantic analysis, for the creation of dynamic social communities based on users' actions and their context (e.g. assets consumed, demographic profile and preferences).

4   Objectives

The main objective of the SAM project is the development of an advanced digital media delivery platform that provides an open environment for defining, characterising, discovering, socially consuming, syndicating and interacting with media assets. This main goal can be decomposed into several research and technological objectives. One of these objectives is the dynamic creation of social communities based on user-consumed media, user behaviour and interests expressed while inter-

acting with the system and with other users. NLP technologies are crucial in this context to provide advanced techniques for data extraction, analysis and characterisation. Sentiment analysis and text summarisation will help to achieve the objective of enabling companies, especially SMEs, to analyse crowd-sourced reactions to the assets they publish. Another objective is the use and definition of open and standardised formats for the description of media assets, along with a framework for their configuration and use that could be exploited by third party software companies to easily build second screen social orientated applications.

Besides these research and technological objectives, the project also aims to accomplish different goals related to the exploitation and sustainability of the platform. These objectives include addressing issues paramount for commercial exploitation, such as content curation, user privacy and brand integrity. Another goal in this category is the provision of business models and exploitation plans for further development of the project results and future commercialisation activities.

5   NLP Challenges

Some of the core functionalities of SAM will be supported by NLP technologies, ranging from entity linking to sentiment analysis. This section describes the main challenges that will be faced in this project by means of human language technologies.

Ontology Exploitation One of the challenges in SAM is to define an ontology to represent the properties of the assets and the relations existing between them and with other external sources (such as Wikipedia). All the assets in SAM will be stored as instances of this ontology. In order to facilitate the process of importing external data into the SAM platform, ontology matching technologies will provide a mapping between the structure of the incoming data and the ontology defined. This task requires the implementation of CRUD operations—create, read, update and delete—over ontology instances.

Data Characterisation This task comprises different problems related to the semantic analysis of text. One of the issues to be solved here is entity linking, the task of identifying entities mentioned in text and

David Tomás, Yoan Gutiérrez, Isabel Moreno, Francisco Agulló, Marco Tiemann, Juan V. Vidagany, Andreas Menychtas

connecting them to instances in a knowledge base (Mihalcea and Csomai, 2007). In SAM, both assets' content and user comments will be analysed in order to identify mentions to entities existing in Wikipedia and assets already stored in SAM. The aim of this analysis is to give content providers the possibility to enrich their assets by linking them to additional internal (assets) and external (Wikipedia) sources of information, and also to identify comments on specific assets, made by the end users, for business intelligence purposes.

Another task related to this subject is asset discovery. In this case, the system will recommend assets to the users based on their context (e.g. assets consumed and user preferences). This task requires context analysis and ontology exploitation to identify suitable instances stored in the platform.

Social Mining Sentiment analysis and text summarisation techniques will be applied to social media in SAM. The purpose is to obtain suitable information from user-generated content to help the business intelligence module of the platform to create advanced reports for content providers. Sentiment analysis will extract intensity, polarity and emotions from users opinions over the assets consumed. Moreover, aspect-based sentiment analysis (Pontiki et al., 2014) will be applied to identify opinions about specific features of an asset (e.g. an actor in a film).

SAM will bring added value to this research area by using these techniques in a novel domain: the creation of dynamic social communities based on the feelings, opinions and interests expressed by the users.

Besides that, novel text summarisation techniques will be applied not only to identify the most relevant comments posted by users regarding an asset or subject, but also to detect the most salient parts of these comments. As it was the case of sentiment analysis, this information will be used by the business intelligence module to supply content providers with advanced reports on user opinions regarding their assets.

6   Current Status of the Project

The project started on September 2013 and will run for 37 months, finishing by the end of October 2016. In its first year, most of the efforts were focused on defining and specifying the SAM concept and the project vision, and carrying out the first stages of the software development life cycle: requirement analysis, global architecture definition, functional specification and technical specification.

SAM is currently in its second year, and at this stage the development of most of the modules in the platform has been started. Regarding NLP functionalities in the project, the following tasks were carried out:

- *Entity linking*: given an input text, the system identifies the occurrence of Wikipedia entities in its content. This module is based on OpenNLP[2] and DBpedia Lookup.[3]

- *Ontology definition and exploitation*: the SAM ontology has been defined using Europeana[4] as a basis to represent the properties of the assets, and including additional concepts and attributes to store specific information related to the SAM platform (such as information on the assets' owners).

- *Sentiment analysis*: a first approach has been developed to polarity and intensity detection on user generated content based on machine learning techniques.

- *Ontology matching*: a module has been defined to match incoming data structures with the SAM ontology. The mapping is based on Levenshtein distance between concept/label names, combined with a measure based on the density of the graph obtained from the structures.

News, updates and additional information on the progress of the project can be found in the official web page and in the wiki page of SAM.[5]

References

Mihalcea, R. and A. Csomai. 2007. Wikify!: Linking documents to encyclopedic knowledge. In *CIKM '07*, pages 233–242, New York, NY, USA. ACM.

Pontiki, M., H. Papageorgiou, D. Galanis, I. Androutsopoulos, J. Pavlopoulos, and S. Manandhar. 2014. Semeval-2014 task 4: Aspect based sentiment analysis. In *SemEval 2014*, pages 27–35.

---

[2]`https://opennlp.apache.org/`.
[3]`http://wiki.dbpedia.org/Lookup`.
[4]`http://www.europeana.eu/`.
[5]`http://wiki.socialisingaroundmedia.com/`.