

Social Media Meta-API: Leveraging the Content of Social Networks

George Papadakis
Institute of Communication and
Computer Systems - ICCS
9, Iroon. Polytechniou Str.
GR-157 73 Zografou – HELLAS
+30-210-7722132
papadakis@l3s.de

Magdalini Kardara
Institute of Communication and
Computer Systems - ICCS
9, Iroon. Polytechniou Str.
GR-157 73 Zografou – HELLAS
+30-210-7722132
mkardara@sociosproject.eu

Konstantinos Tserpes
Institute of Communication and
Computer Systems - ICCS
9, Iroon. Polytechniou Str.
GR-157 73 Zografou – HELLAS
+30-210-7722132
tserpes@mail.ntua.gr

Athanasios Papaoikonomou
Institute of Communication and
Computer Systems - ICCS
9, Iroon. Polytechniou Str.
GR-157 73 Zografou – HELLAS
+30-210-7722132
tpap@mail.ntua.gr

Emmanuel Sardis
Institute of Communication and
Computer Systems - ICCS
9, Iroon. Polytechniou Str.
GR-157 73 Zografou – HELLAS
+30-210-7722132
sardism@mail.ntua.gr

Fotis Aisopos
Institute of Communication and
Computer Systems - ICCS
9, Iroon. Polytechniou Str.
GR-157 73 Zografou – HELLAS
+30-210-7722132
fotais@mail.ntua.gr

ABSTRACT

Social Network (SN) environments are the ideal future service marketplaces. It is well known and documented that SN users are increasing at a tremendous pace. Taking advantage of these social dynamics as well as the vast volumes, of amateur content generated every second, is a major step towards creating a potentially huge market of services. In this paper, we describe the external web services that SocIoS project is researching and developing, and will support with the Social Media community. Aiming to support the end users of SNs, to enhance their transactions with more automated ways, and with the advantage for better production and performance in their workflows over SNs inputs and content, this work presents the main architecture, functionality, and benefits per external service. Finally, introduces the end user, into the new era of SNs with business applicability and better social transactions over SNs content.

Categories and Subject Descriptors

C.2 [Computer Communication Networks]: Distributed Networks; D.2 [Software Engineering]; D.3 [Programming Languages]; E.1 [Data Structures]: Distributed data structures, Graphs and Networks;

General Terms

Algorithms, Management, Measurement, Documentation, Experimentation, Human Factors, Languages, Theory, Legal Aspects.

Keywords

Social Networks, Service Level Agreement, Internet of Services, Social graph, Event detection, Topic detection, Influence Diffusion, SOA Infrastructure, SN Business Models, Service Provisioning Systems, Evaluation Scenarios.

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.
WWW 2012 Companion, April 16–20, 2012, Lyon, France.
ACM 978-1-4503-1230-1/12/04.

1. INTRODUCTION

This manuscript provides an overall presentation of Social Dynamics and User Created Content as a Service-SocIoS[1] architecture, and main construction modules over Social networks-SNs. The basic topic of this manuscript is to provide the project outputs and status of research and development work, performed by the project coordinator NTUA team [2].

2. SocIoS PROJECT OVERVIEW

The SocIoS platform aims at providing the tools to build applications that leverage on the dynamics and content that "live" in the SNs. These tools are a collection of SOAP web services that make use of SNs metadata in order to organize information residing in SNs, each service for its own purpose. SocIoS platform allows the creative "mix-and-match" of these services to application workflows which are ready to be consumed

In detail, the conceptual elements that constitute SocIoS platform (see Figure 1) are the following:

- *SocIoS API*: An aggregation of methods provided by underlying SNs APIs. SocIoS API will map a standard interface to collections of methods and objects of the SNs APIs. This will allow the developer to initiate multiple instances of the same functionality to various SNs by a single method call.
- *SocIoS Middleware*: The functionality that SocIoS will provide at the level of the SocIoS API, will be leveraged to a web service level so as to make it more usable in the frame of the IoS. Thus, each class in the SocIoS API will be exposed as a web service and SocIoS will provide tools for managing these services. The SocIoS middleware is the service hosting, provisioning and management environment that SocIoS will develop.
- *SocIoS Core Services*: SocIoS, through standard web service technologies (e.g. WSDL) will deliver the SocIoS API functionality as a service. These core services can be used as they are or by combination. They are hosted and provided strictly on the SocIoS middleware.

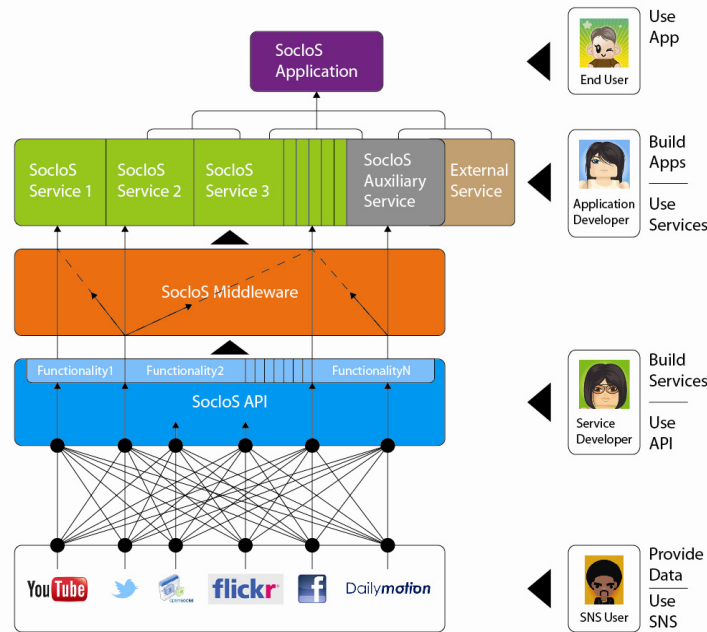


Figure 1. SocIoS Architecture.

- SocIoS Application: A combined set of SocIoS added value services to create business-oriented applications that utilize UCC and SG information from SNSs. SocIoS applications may even combine services external to SocIoS middleware.
- SocIoS Auxiliary Services: It is expected that there are some auxiliary services that can be built by combining SocIoS Services that will be useful for supporting the larger workflows of the SocIoS applications. These services are not part of the core set of functionality described under SocIoS API. They are expected to be more complex and to combine the core set (SocIoS Services) in order to be implemented, still, they will be hosted in the SOA middleware. They are called SocIoS Auxiliary Services and they will come bundled with the whole SocIoS platform when it is to be installed. Such auxiliary services are the reputation and recommendation services
- SocIoS Platform: The whole set of SocIoS items described above. The software bundle of SocIoS.

In the following sections we present the latest achievements and developed functionalities that SocIoS project provides.

3. SocIoS OUTPUTS

3.1 SocIoS API

SocIoS created its own SNSs-like API that captures the data specifications and methods that summarizes the core objects of the SNSs API layer which in turn can -at least- support the particular scenarios in which SocIoS is working. The definition of an Object Model in SocIoS to assist the implementation and integration and to add value to the exploitation potential of the project, becomes a point of reference that enables the extension of

its functionality, and becomes a compatibility point to other solutions or even new SNSs APIs. SocIoS team worked on the definition of this object model, through the: the identification of objects that appear consistently in the SNSs APIs and the identification and definition of the objects that capture in the most complete way an abstract SNSs API domain.

We studied the most prominent SNSs APIs (according to criteria relevant to the SNSs API popularity, maturity, openness, scope and necessarily, whether they are or not fit for SocIoS purposes). This analysis led us to a good understanding of what are the conceptually common objects that the SNSs APIs are using. Finally, in order to test our conclusions and ensure that there is semantic equivalence between these objects, we implemented a showcase scenario in which the notion of MediaItem was tested under various SNSs[1],[4],[5],[6],[7].

3.2 SocIoS Core Services

The SocIoS Core Services are SOAP web services that provide a basic management layer to entities that "live" in SNSs. They support a specific API, that is, the SocIoS Core API (based on an adaptation of the OpenSocial data specification), thus, implementing data transformations between primitive SNSs types and OpenSocial ones as well as delivering a single interface that masks various (sometimes even complex) underlying SNSs API methods. In terms of implementation, the SocIoS Core Services are comprised of a central component acting as the single point of reference for the layers above, and a set of adaptors that implement the SNSs API calls and wrap the data types to OpenSocial ones.

In order for this mechanism to operate, there is the need for the adaptors to have access to SNSs users' data. For that purpose, each adaptor needs to run a registered application in its associated SNS using an assigned API key. In turn, the application needs to

execute API calls on behalf of a registered SNS user, therefore, the adaptors need to be authenticated for that purpose.

The SocIoS Core Services receive and return SocIoS entities, most notably the four basic objects that SocIoS adapts from OpenSocial : Person, Activity, MediaItem, and Message.

3.3 SocIoS Auxiliary Services

Some auxiliary services that are developed in the frame of SocIoS project are presented in the following subsections.

3.3.1 Event Detection

Journalists seek to exploit the high potential of the user-generated content (UGC) that is continuously posted on-line. The reason is that it allows for significantly faster detection of events and that it offers direct information from eye-witnesses. However, its extreme volumes pose a serious challenge: Web 2.0 technologies and Social Networks do not facilitate the navigation through the content relevant to specific events. The common solution to this problem is to group the streaming UGC into clusters pertaining to a particular event and to assign a descriptive label to each cluster. Traditional streaming algorithms are incapable of processing the voluminous UGC, due to its sparse, noisy and multilingual textual content that is not restricted to a standard vocabulary, but abounds in constantly evolving neologisms.

To overcome these challenges, SocIoS introduces a novel event detection algorithm that is based on three major steps: first, it identifies terms with a substantial increase in their frequency and considers them indicative of new events. Second, it boot-straps the creation of event-related clusters by grouping together the documents containing the identified event terms. Co-occurring terms belong to the same event and produce an independent cluster. To capture the aggregate textual patterns of each cluster, we employ the language-neutral, noise-tolerant method of n-gram graphs, which identifies distinguishing sequences of character n-grams. At the third stage, the cluster graphs are used to associate the remaining documents into one of the identified events so as to build a comprehensive summary of them.

3.3.2 Topic Detection

At the core of Event Detection lies the task of topic classification: how to assign documents to the event clusters created by the highly frequent terms. This task is typically tackled through patterns of co-occurring words or characters that are extracted from traditional document representation models (i.e., Term Vector Model and Character N-Grams). However, not all documents are of the same quality; for example, the curated content of news articles usually entails lower levels of noise than the user-generated content of the blog posts and the other Social Media.

To identify the most appropriate model for the documents processed by SocIoS, we conducted a thorough experimental study comprising large collections of documents from three different domains: curated news articles, semi-curated blog posts and raw microblog messages (Figure 2).

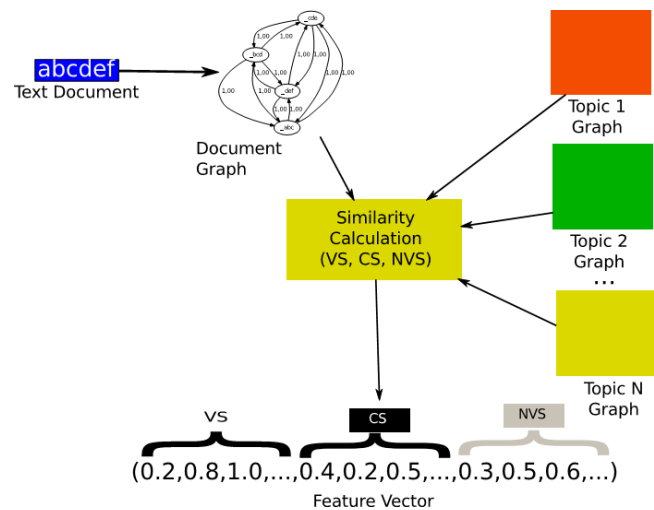


Figure 2. Topic detection with N-Gram graphs.

Despite the different classification settings required for each category, we found out that character n-gram graphs exhibit the highest accuracy across all domains. The reason is that they go beyond the established bag-of-words models, representing each document as a graph; individual graphs can be combined into a class graph and graph similarities are then employed to position and classify documents into the vector space. Accuracy is increased due to the contextual information that is encapsulated in the edges of the n-gram graphs, while efficiency is boosted by reducing the feature space to a limited set of dimensions that depend on the number of classes, rather than the size of the vocabulary.

3.3.3 Influence Diffusion inside Topic Communities

Users of Social Media typically gather into communities on the basis of some common interest. Their interactions inside these on-line communities follow several, interesting patterns. For example, their members differ in the level of influence they exert to the rest of the group: some of them are actively involved, affecting a large part of the community with their actions, while the majority comprises plain participants (e.g., information consumers). Identifying users of the former category lies on the focus of interest of many recent works, as they can be employed in a variety of applications, like targeted marketing.

SocIoS offers novel tools for detecting influence patterns in the context of Social Media topic communities. These are spontaneous user groups that are created on-the-fly by people posting messages about a particular topic, without necessarily being connected with each other through friendship links on the social graph. In this way, SocIoS can accommodate heterogeneous Social Networks, which rely on different definitions of friendship. SocIoS encompasses a set of influence criteria that are able to map users into a numeric space quantifying their level of influence (see Figure 3) in order to facilitate the identification of the most influential community members. Large scale experiments with real-world datasets have estimated the effect of top-k influencers on the rest of the community and the time it takes to follow their actions and behavior.

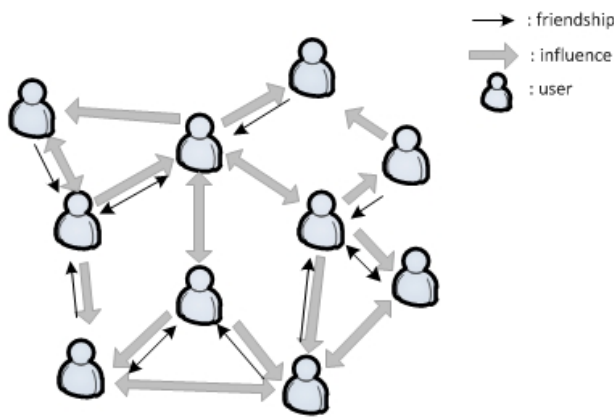


Figure 3. Influence diffusion.

3.3.4 Sentiment Analysis

The abundance of subjective Social Media content turns it ideal for mining the aggregate opinion of its users towards products or political debates. As mentioned above, however, it poses serious challenges to the applicability of traditional sentiment analysis techniques, due to its inherent intricacies: sparse, multilingual and noisy content expressed in terms of a continuously evolving vocabulary.

To deal with these challenges, SocIoS introduces a novel tool that relies on two orthogonal, but complementary sources of evidence: content-based features captured by n-gram graphs and context-based features captured by the metric of polarity ratio (e.g., the aggregate polarity of the authors posting about a specific theme, of their friends as well as of the relevant links and hashtags). Both types of features are language-neutral and noise-tolerant, thus guaranteeing high effectiveness and robustness in the settings SocIoS is operating. In addition to high effectiveness, this tool is highly efficient, as well, handling large volumes of data, due to a series of efficiency optimization techniques it employs: low extraction cost for its features, restricted search space (i.e., limited number of features) as well as discretization techniques that ensure high accuracy even when combined with fast classification algorithms, like Naive Bayes.

4. SocIoS APPLICATIONS

In the frame of the evaluation of the technologies that are developed by SocIoS, but also with the parallel purpose to create an integrated solution for covering real business needs, SocIoS work on scenarios that involve all the actors in the supply chain and span through different business models. As such, the scenarios cover the business needs of: a) *service providers*: anyone that wishes to use the SocIoS platform to increase service utilization; b) *resource providers*: anyone that wishes to lease their resources for hosting SocIoS services; and, c) *brokers*: any service provider that wishes to play the role of a SocIoS service broker hosting the SocIoS SOA infrastructure. These scenarios are the following:

- *Journalistic Use Case Scenario*, where a journalist wants to monitor the activity of a group of SNS users who are related to a certain topic or a location. In detail she wants to get all the information, taking advantage of the sharing of content from a vast amount of information sources (possibly original user created content) and the viral dissemination effect of the SNS

(get access to news in a timely fashion). Monitoring the content-related activities of a group of people and use the system to effectively sort these activities based on whether they are related to a real-life important event or not.

- *Casting Extras*, where, a TV commercial producer wants to cast a number of extras for a new commercial. The producer wants to send out the description and get back proposals. Taking advantage of fact that in SNS users are already providing information and photographs about themselves. He only worries about the reliability of the source of the information, but SocIoS has the answer to that.
- *Location Scouting*, where a TV commercial producer wants to locate a place where they can have the shooting. The producer wants to send out the description and get back photographs that depict a place that looks as close as possible to the target location. He wants to take advantage of the "wisdom of the crowd" that will help him answer his question, quickly and at a low cost.

The above scenarios using the framework of SNS offers an ideal environment to tackle problems related to the content and collaboration described above and to extend the SocIoS research:

Third Party Services support toolset, tools for managing UCC, Dynamic Services, SOA infrastructure, and SocIoS API.

5. CONCLUSIONS

This manuscript provides a poster screenshot of research and development performed under SocIoS project. Aiming to introduce and provide to the end user of Social Networks another aspects of the social dynamics that could be transformed into services over the web through business applications and SNS auxiliary services is our article target.

6. ACKNOWLEDGMENTS

This work has been supported by the SocIoS project and has been partly funded by the European Commission's 7th Framework Programme through theme ICT-2009.1.2: Internet of Services, Software and Virtualisation under contract no.257774.

7. REFERENCES

- [1] SocIoS project home page, <http://www.sociosproject.eu>
- [2] Institute of Communication and Computer Systems/National Technical University of Athens, ICCS/NTUA, <http://www.ntua.gr>
- [3] SocIoS, showcase demo <http://www.sociosproject.eu/Showcases/VeryInitialShowcase/tabid/382/language/en-US/Default.aspx>
- [4] GNIP API, <http://gnip.com/>
- [5] HootSuite Social Media dashboard for Teams http://www.google.gr/url?sa=t&ret=j&q=hootsuite&source=web&cd=1&ved=0CCgQFjAA&url=http%3A%2F%2Fhootsuite.com%2F&ei=YaX1TsaNJYab8gOsu5G5AQ&usg=AFQjCNHDlysbFj2h6Mef_LO7XqdiIDU-7A
- [6] Stroodle, Social Media Aggregator, <http://www.stroodle.it/>
- [7] OpenSocial API, <http://code.google.com/apis/opensocial/>