An Overview of SMOB 2: Open, Semantic and Distributed Microblogging*

Alexandre Passant and **Uldis Bojars** and **John G. Breslin**[†]
Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland
firstname.lastname@deri.org

Tuukka Hastrup
University of Jyväskylä, Finland
Tuukka.Hastrup@iki.fi

Milan Stankovic[‡] and **Philippe Laublet**
STIH, Université Paris-Sorbonne, Paris, France
firstname.lastname@paris-sorbonne.fr

Abstract

This short paper provides an overview of the architecture and features of SMOB 2 — http://smob.me —, a platform for open, semantic and distributed microblogging combining Social Web principles and state-of-the-art Semantic Web and Linked Data technologies.

Introduction

In a few years, microblogging became a popular trend on the Web, notably with the rise of Twitter. However, in spite of lots of interest from the research community[1], the following issues need to be addressed to exploit this paradigm to its full potential.

Closed-world architecture. Microblogging services are centralised applications that act as close-world walled garden, where information cannot be easily reused across services. Our first requirement is thus to offer a decentralised architecture for microblogging, where everyone can setup her or his own platform. Then, data remains the property of their author and become openly available for any purposes.

Lack of machine-readable meta-data Microblog services also lack open and interoperable meta-data about their posts (creation date, author, recipient, etc.). Twitter has adopted microformats[2] for describing followers (and subscribers) lists, but much more information is require to efficiently use meta-data regarding microblogging. Our second requirement is then to define and use standard models for representing the various microblogging meta-data.

Lack of semantic in microblog posts In addition, microblog posts do not carry any semantics, making their querying and reuse difficult. Once again, Twitter users have adopted #hashtags, but their semantics are not readily machine-processable, thus raising the same ambiguity and heterogeneity problems that tagging practises cause (Mathes 2004). Our third requirement is then to provide meaningful semantics about microblog content.

Information overload Finally, information overload is also an important problem, for several reasons. Figure 1 shows the different factors we identified by running 30-minutes interviews with 10 users of microblogging platform (Stankovic, Passant, and Laublet 2009). Our final requirement is thus to more efficiently direct microblogging updates so that they do not become annoying for their reader(s).

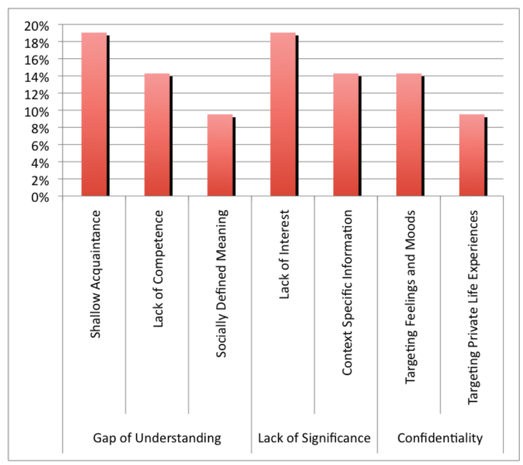

Figure 1: Issues related to the audience of status messages

The SMOB Framework

Based on these requirements, we designed SMOB[3] (available at http://smob.me), a microblogging framework combining existing Social Web paradigms with state-of-the-art Semantic Web and Linked Data (Bizer, Heath, and Berners-Lee 2009) technologies, leading to what is generally known as the Social Semantic Web (Breslin, Passant, and Decker 2009). More specifically, SMOB relies on:

*The work presented in this paper is funded in part by Science Foundation Ireland under grant number SFI/08/CE/I1380 (Líon 2).

†John G. Breslin is also a member of the School of Engineering and Informatics, National University of Ireland, Galway, Ireland.

‡Milan Stankovic is also affiliated with Hypios, Paris, France.

[1] See list at http://www.danah.org/researchBibs/twitter.html

[2] http://microformats.org/

[3] A first version of SMOB has been designed mid-2008. This paper focuses on its v2, a complete re-factoring with a new architecture and improved features.

- *ontologies*, used to define common semantics for representing microblog posts, so that they can be reused by any other service capable of consuming RDF(S)/OWL data;
- *distributed hubs*, spread across the Web and used to publish data, exchanging information (posts and follower / following subscriptions) based on the previous ontologies;
- *interlinking components*, making microblog posts interlinked with other resources from the Web (and in particular those from the Linking Open Data cloud[4]);
- *faceted presence*, so that one can browse status messages corresponding only to his or her current context.

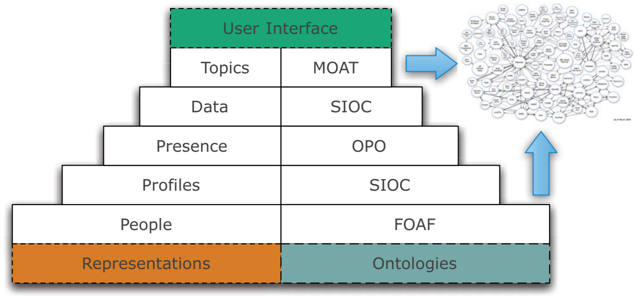

Figure 2: The SMOB ontologies stack

Ontologies for Microblogging

In order to model user profiles, we naturally relied on FOAF — Friend of a Friend (Brickley and Miller 2004)[5] — as it provides a simple way to define people, their main attributes and their social acquaintances. Regarding the modelling of microblog updates themselves, we relied on and extended SIOC — Semantically-Interlinked Online Communities (Breslin et al. 2005)[6] — and its Types module, a standard vocabulary for expressing social data in RDF. Especially, we introduced (i) two new classes in SIOC: `sioct:Microblog` and `sioct:MicroblogPost` as well as (ii) two properties: `sioc:follows` to express follower / following notifications (this property is used for both, thanks to the directed graph model of RDF), and `sioc:addressed_to` to identify who a particular post is intended for. We also relied on the Online Presence Ontology (OPO[7]) (Stankovic 2008) for describing a user's presence as well as their context that can give better insight into their current situation. Finally, we relied on MOAT — Meaning Of A Tag[8] (Passant et al. 2009) — to model semantic tagging capabilities, i.e. linking tags to meaningful resource from the Semantic Web, as we will detail later.

Combined together, these ontologies form a complete stack to represent the various elements involved in microblogging applications (Figure 2) and that can be referred to as a more global ontologies stack for the Social Semantic Web. An example of microblog post modelled with this stack can be found at `http://apassant.net/smob/data/2010-03-03T12:56:35+00:00`.

Distributed Hubs

In order to fulfil our second requirement, we designed an architecture based on distributed microblogging hubs that communicate each other to exchange microblog posts and follower / following notifications. That way, there is no centralised server but rather a set of hubs that contains microblog data and that can be easily replicated and extended.

SMOB hubs can be easily installed on any LAMP environment and hubs communicate each others via HTTP thanks to SPARQL/Update[9] for synchronisation purposes (Figure 3). When a user creates a new post, it is immediately stored in the hub's RDF store, and published in an RDFa-enabled page, see for example `http://apassant.net/smob/post/2010-03-03T12:56:35+00:00`. Notifications about the new post are sent to the followers' stores, that aggregate the newly-created RDF data[10]. While we currently provide direct publishing between hubs, the architecture can be extended to support relays through the use of pubsubhubbub[11], so that SMOB hubs would send only one notice for each update, that would be broadcasted to other hubs.

A similar distributed approach is used with regards to follower / following subscriptions. A simple bookmarklet allows any user to become a new follower of another user when browsing that user's SMOB hub. The subscription is registered in the follower's hub by adding a <local_user> sioc:follows <remote_user> triple, while the same triple is included in the remote store so that both parties are instantaneously aware of this new follower / following relationship.

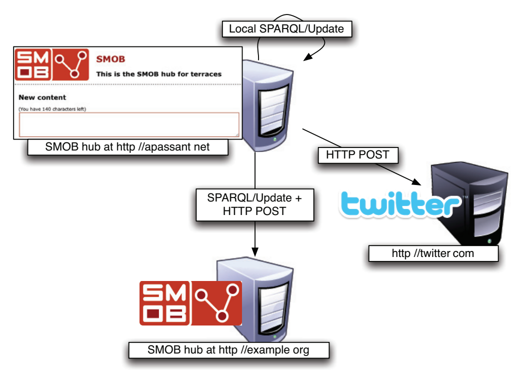

Figure 3: Communication between SMOB hubs using SPARQL/Update

[4] `http://linkeddata.org`
[5] `http://foaf-project.org`
[6] `http://sioc-project.org`
[7] `http://online-presence.net`
[8] `http://moat-project.org`
[9] `http://www.w3.org/TR/2009/WD-sparql11-update-20091022/`
[10] In addition, SMOB allows simple cross posting to Twitter and aggregation of Twitter messages in one's hub.
[11] `http://code.google.com/p/pubsubhubbub/`

In addition, when installing a hub, users can re-use their existing FOAF profile (which can be generated from existing Web 2.0 applications) instead of creating a new account. Thus, their information (name, depiction) is retrieved from these profiles, solving one issue regarding data portability.

Interlinking Components

In order to make microblog interlinked with the rest of the Linking Open Data cloud, we designed interlinking components, in addition to the natural interlinking that is provided by the use of existing FOAF profiles.

As a current practise in microblogging is the use of `#hashtags` to represent topics, locations, etc. we extended this paradigm to semantic hashtags, where hashtags are not simple tags anymore, but URIs of existing resources from the Semantic Web — solving the ambiguity and heterogeneity issues of tagging. In order to define and assist users in creating these mappings, SMOB hubs feature a set of wrappers that automatically suggest relevant URIs of existing resources for each `#hashtag` used in microblog updates, as well as for `L:xxx` content (location patterns) by querying existing services such as Sindice (the Semantic Web index[12]) or DBpedia (the RDF export of Wikipedia[13]), whilst also letting people write their own wrappers. The mappings are then modelled using MOAT and exposed in the microblog posts as RDFa for further reuse.

While such manual interlinking may sound complex, we recently demonstrated the usefulness of MOAT in a corporate context (Passant et al. 2009), showing (i) how users can benefit from it when searching for information, and (ii) that in spite of an additional effort, end-users were willing to do it. In addition, we should mention that while we rely only on Sindice and DBpedia, new wrappers can be added, for instance to enable interlinking in corporate microblogging, using internal knowledge bases.

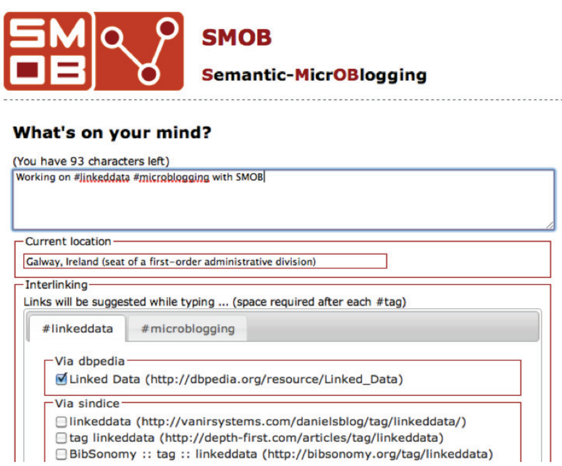

Figure 4: Publishing and interlinking content with SMOB

Thanks to these interlinking components, microblogging content becomes more discoverable. For example, using Sindice (that is pinged each time a new content is created), one can identify all microblog posts about a particular topic or location. Moreover, new browsing interfaces can be deployed such as real-time geolocation mash-ups, as seen in Figure 5 (using GeoNames[14] data).

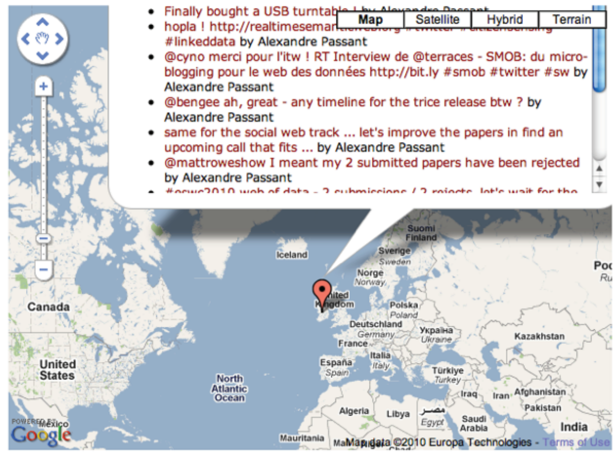

Figure 5: Real-time geolocation using GeoNames

Faceted Presence

Finally, we defined models for *faceted presence*, in order to model the intended audience of status messages as dynamic information, by defining "Sharing Spaces", modelled by extending the Online Presence Ontology. A Sharing Space, in our specification[15], is a group of people (or agents) with whom particular information can be shared. It is defined with a list of properties to allow the representation of common attributes that bind members of the Sharing Space together (*e.g.* common interests, current location, etc.). In order to express the semantics of those attributes, we relied on concepts from widely-used vocabularies and data from the Linking Open Data cloud. Moreover, we use SPARQL and its CONSTRUCT pattern to define Sharing Spaces dynamically, and provide user-interface to let users define them. Figure 6 demonstrates the SPARQL pattern used to define a Sharing Space of all the people interested in the Semantic Web, that are currently in Paris (we rely on GeoNames to provide the URI of Paris, and on DBpedia to provide the URI for the Semantic Web)[16]. By identifying people who are intended to receive a status message, the notion of Sharing Space can help software systems to deliver status messages to specific people (members of the Sharing Space) and thus deal with information noise and even ensure confidential status message exchange.

[12] `http://sindice.com`

[13] `http://dbpedia.org`

[14] `http://geonames.org`

[15] `http://online-presence.net/opo/specs/2009/OPO-20090501/`

[16] We shall mention that we could have used RIF — Rule Interchange Format — to define these Sharing Spaces but that we rely on SPARQL since there are no mature RIF implementation in PHP thus far.

```
CONSTRUCT {
  <http://example.org/ns#SWPeopleInParis> rdf:type opo
      :SharingSpace;
    foaf:member ?person.
} WHERE {
  ?person foaf:topic_interest <http://dbpedia.org/
      resource/Semantic_Web>.
  ?person opo:declaresOnlinePresence ?presence .
  ?presence opo:currentLocation <http://sws.geonames.
      org/2988507/>.
}
```

Figure 6: Sharing Space using SPARQL CONSTRUCT

Related Work

Apart the first release of SMOB in mid-2008, it being the first semantic microblogging platform, various systems recently emerged, including StatusNet[17] (formerly Laconica), an open-source platform that powers Identi.ca[18], and smesher[19], a semantic microblogging client with local storage, that integrates with Twitter and Identi.ca and that features structure identification and a dashboard for custom filters. StatusNet also uses the OpenMicroBlogging protocol[20] for client-server communication, and we recently discussed how our proposal and ontologies can be simply mapped to it[21]. However, they do not provide interlinking with the Linked Data Cloud and focus mainly on representing metadata about the containers or structure using semantics, while smesher also permits the generation of new RDF statements.

Various systems also implement semantic capabilities on the top of existing systems, such as semantictweet[22] (exporting Twitter user profiles in FOAF), the Chisimba Twitterizer[23] (providing microblog data using open formats), HyperTweeter[24] (providing user-driven structuring of Twitter hashtags), Twitris[25] (extracting semantics from Twitter messages), as well as rtsw[26] (providing real-time RDF streams from Twitter data).

Finally, in order to extend current microblogging systems, various syntax extensions have been proposed, including MicroTurtle[27], the Star Priority Notation[28], microsyntax[29], nanoformats[30], Twitter Data[31], Twitterformats[32] and TwitLogic[33], this one translating tweets into RDF. However, many of these applications and syntaxes rely on the Twitter infrastructure, and do not offer the distributed and open architecture that we provide with SMOB.

Conclusion and Future works

In this paper, we overviewed the features and architecture of the new release of SMOB, a system providing an open, distributed and semantic alternative to existing microblogging systems. In particular, we showed how it fulfils four main requirements to enhance the microblogging experience in terms of data portability, ownership and discovery while at the same time making microblogging an integral component of the ongoing Social Semantic Web and Linking Open Data initiative. Thanks to the system, and relying only on existing Web standards, microblog posts then becomes more discoverable, meaningful and mashable with other kind of content from the Web.

References

Bizer, C.; Heath, T.; and Berners-Lee, T. 2009. Linked Data — The Story So Far. *International Journal on Semantic Web and Information Systems (IJSWIS)* 5(3):1–22.

Breslin, J. G.; Harth, A.; Bojars, U.; and Decker, S. 2005. Towards Semantically-Interlinked Online Communities. In *Proceedings of the Second European Semantic Web Conference, ESWC 2005, May 29–June 1, 2005*.

Breslin, J. G.; Passant, A.; and Decker, S. 2009. *The Social Semantic Web*. Springer.

Brickley, D., and Miller, L. 2004. FOAF Vocabulary Specification. Namespace Document 2 Sept 2004, FOAF Project. http://xmlns.com/foaf/0.1/.

Mathes, A. 2004. Folksonomies: Cooperative Classification and Communication Through Shared Metadata.

Passant, A.; Laublet, P.; Breslin, J. G.; and Decker, S. 2009. A URI is Worth a Thousand Tags: From Tagging to Linked Data with MOAT. *International Journal on Semantic Web and Information Systems (IJSWIS)* 5(3):71–94.

Stankovic, M.; Passant, A.; and Laublet, P. 2009. Status Messages for the Right Audience with an Ontology-Based Approach. In *The First International Workshop On Collaborative Social Networks - CollaborateSN 2009 - at the Fifth International Conference on Collaborative Computing*.

Stankovic, M. 2008. Modeling online presence. In *Proceedings of the First Social Data on the Web Workshop*, CEUR Workshop Proceedings.

[17] http://status.net/

[18] http://identi.ca

[19] http://smesher.org/

[20] http://openmicroblogging.org/

[21] See http://apassant.net/blog/2010/02/09/proposal-semantic-omb and http://status.net/wiki/Semantic_OMB

[22] http://semantictweet.com

[23] http://trac.uwc.ac.za/trac/chisimba/browser/modules/trunk/twitterizer

[24] http://semantictwitter.appspot.com/

[25] http://twitris.knoesis.org/

[26] http://realtimesemanticweb.org

[27] http://buzzword.org.uk/2009/microturtle/

[28] http://civilities.net/Star_Priority_Notation

[29] http://www.microsyntax.org/

[30] http://microformats.org/wiki/microblogging-nanoformats

[31] http://twitterdata.org/

[32] http://twitterformats.org/

[33] http://wiki.github.com/joshsh/twitlogic/syntax-conventions