HAL
open science

Semantics, Sensors, and the Social Web: The Live Social Semantics Experiments

Martin Szomszor, Ciro Cattuto, Wouter van den Broeck, Alain Barrat, Harith Alani

To cite this version:

Martin Szomszor, Ciro Cattuto, Wouter van den Broeck, Alain Barrat, Harith Alani. Semantics, Sensors, and the Social Web: The Live Social Semantics Experiments. ESWC2010, May 2010, Heraklion, Greece. pp.196-210. hal-00489568

HAL Id: hal-00489568

https://hal.science/hal-00489568

Submitted on 5 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Semantics, Sensors, and the Social Web: The Live Social Semantics experiments

Martin Szomszor[1], Ciro Cattuto[2], Wouter Van den Broeck[2], Alain Barrat[2,3], Harith Alani[4]

[1] City University, UK
[2] Complex Networks and Systems Group
Institute for Scientific Interchange (ISI) Foundation, Turin, Italy
[3] Centre de Physique Théorique (CNRS UMR 6207), Marseille, France
[4] Knowledge Media Institute, The Open University, UK

Abstract. The Live Social Semantics is an innovative application that encourages and guides social networking between researchers at conferences and similar events. The application integrates data and technologies from the semantic web, online social networks, and a face-to-face contact sensing platform. It helps researchers to find like-minded and influential researchers, to identify and meet people in their community of practice, and to capture and later retrace their real-world networking activities at conferences. The application was successfully deployed at two international conferences, attracting more than 300 users in total. This paper describes this application, and discusses and evaluates the results of its two deployments.

1 Introduction

Most conference attendees would agree that networking is a crucial component of their conference activities. It is vital for researchers to network and collaborate, and most of such ties stem from casual meetings and conversations at conferences and similar scientific and social events. Networking at conferences can be very daunting, especially to junior researchers, or to researchers who cross discipline boundaries with limited knowledge and social ties to the new community. Additionally, researchers often lose track of whom they met and where at such event. Such problems become even more evident in medium to large size conferences, where it is easy for individual researchers to get lost in the crowd. However, there are hardly any applications to help researchers to initiate, capture, and preserve their online as well as off line social interactions during conferences.

To this end, we have developed Live Social Semantics (LSS); a novel application that brings together the semantic web, web2.0, and active sensors to encourage and strengthen collaboration and communication between researchers, by supporting their social networking activities during conferences, to find and locate like-minded people, people in their community of practice, and to permanently log, and retrace, their face-to-face (F2F) contacts network at current and past conferences.

LSS integrates (a) the available wealth of linked semantic data, (b) the rich social data from existing major social networking systems, and (c) a physical-presence awareness infrastructure based on active radio-frequency identification (RFID). This

application was deployed at two major international conferences; ESWC 2009 in Crete, and HyperText 2009 in Turin, were it was used by 300 researchers with very promising results.

The next section describes a variety of related works, followed by a full description of LSS in section 3. Results and evaluation of LSS deployments are covered in sections 5 and 6 respectively. Discussion and future work are given in Section 7, followed by conclusions in Section 8.

2 Related Work

The interplay of networking and social contact at a conference gathering was initially investigated in the context of opportunistic networking for mobile devices [12] by using wearable Bluetooth-enabled devices. Subsequent work focused on sensing organisational aspects [6] by using Bluetooth-enabled mobile phones, and on characterising some statistical properties of human mobility and contact [21, 16]. These early experiments could not assess face-to-face (F2F) human contact in a large-scale setting, as they mostly relied on Bluetooth communication. Wu and colleagues used what they call "sociometric badges" to investigate impact of F2F interactions on productivity [22]. These badges used radio frequency to detect physical proximity, infra red to detect F2F body alignments, and voice sensors to detect conversations.

RFID is an increasingly popular technology for location tracking. IBM used RFIDs to track attendees of a conference in Las Vegas in 2007. The devices were used to track session and meal attendance [21]. The information they collected were limited to the name, title and company of attendees. No social or semantic data were collected nor used. Fire Eagle[5] by Yahoo! is a service that detects the geographical location of users (e.g. based on wifi points), and allows them to share it with their online friends.

Recently, the SocioPatterns project[6] investigated patterns of human contact at large-scale social gatherings by deploying a distributed RFID platform that is scalable and attains reliable detection of F2F interactions as a proxy of social contact [3]. The LSS application presented here leveraged that platform to mine real-time social contacts.

To the best of our knowledge, our LSS application is the first where real-world F2F contacts are mashed up in real time with semantic data from online tagging systems. The free nature of tagging generates various vocabulary problems: tags can be too personalised; made of compound words; mix plural and singular terms; they can be meaningless; they can be synonymous, etc. [14, 9, 10]. This total lack of control obstructs its analysis [13]. In our work, we follow the approach of *cleaning* existing tags using a number of term filtering processes, similar in spirit to those used in [11]. LSS constructs semantic models of interests for individuals by merging and processing their tagging activities from multiple online social networking systems (SNS). This process involves dealing with several problems, such as filtering of tags, disambiguating them, associating tags with semantics, and identifying interests.

Tag ambiguity is a well recognised problem in folksonomies. Clustering has been investigated as a disambiguation approach, where tags are grouped together based on

[5] `http://fireeagle.yahoo.net/`

[6] `http://www.sociopatterns.org`

their co-occurence, to facilitate distinguishing between their different meanings [4, 23, 17]. While such techniques have demonstrated that the underlying folksonomy structure does contain information that can enable automatic disambiguation, they are too computationally expensive and lack any semantic grounding. Angeletou and colleagues [2] used WordNet to identify ambiguous tags, and compared the WordNet senses for the tag to those of the co-occurring tags, to identify the most similar sense. In our approach, we used DBpedia[7] for disambiguating tags, and to automatically associate them with URIs. Some manually-driven approaches have been proposed for assigning URIs to tags (e.g. [17, 15]. Similarly to [20], we explore a strategy for the automatic selection of URIs using DBpedia concepts [7].

3 Live Social Semantics application

3.1 General Architecture

The system architecture of LSS is shown in figure 1. The diagram is vertically partitioned into two spaces: the virtual world (i.e. data about individuals held on the web), and the real world (i.e. RFID-based contact data). Data in the virtual world is sourced from the following:

- Social networking sites: Tagging and social relation data is collected from Delicious, Flickr, Facebook, and lastFM using the Extractor Daemon process and APIs. This data is then used to reflect the online contact network of individuals. The tagging data is processes by the Profile Builder (center, top of diagram) to infer their interests. The Tagora Sense Repository process is responsible for associating tags to URIs from DBpedia. See section 4 for detail.
- Semantic Web Linked Data: Information on publications, projects, and the Community of Practice (COP) of researchers is retrieved from RKBExplorer[8] [8] and semanticweb.org. This data is used to reflect the contact network of individuals based on their paper co-authorships and project co-memberships.

Real-world data is collected from F2F contacts between individuals which are measured using RFID readings (section 3.2). Such data is fully integrated with the virtual world data in a triple store (center right of figure 1), where all the data is stored. This enriches the visualisation and processing of real-world social contacts with the online social contacts of those individuals.

A focused *Contact* ontology[9] was used to represent real-world social interactions between individuals, recording the total contact time on a daily basis 3.3.

3.2 Real-Time Social Contacts

Real-world interactions of conference attendees are mined using RFID hardware and software infrastructure developed by the SocioPatterns project [3]. Each participant was issued with an RFID badge to monitor his/her F2F contacts with others. The RFID

[7] http://dbpedia.org/
[8] www.rkbexplorer.org
[9] http://tagora.ecs.soton.ac.uk/schemas/LiveSocialSemantics

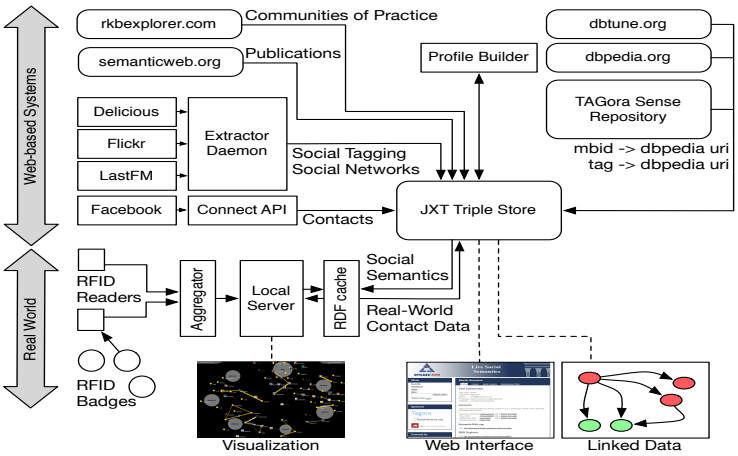

Fig. 1. Live Social Semantics Architecture

badges engage in multi-channel bi-directional radio communication, and by exchanging low-power signals which are shielded by the human body, they can reliably assess the continued F2F proximity of two individuals. A F2F contact is recorded if users face each other for around 10 seconds or more, within a distance of around 1 meter.

We generate a weighted graph to represent the cumulative F2F contacts between the participants. This information is periodically uploaded to the triple store via RDF/HTTP and integrated with the other data layers.

We use the real-world and online social relations to compute simple recommendations. For example, if two attendees are in F2F contact at a given time, the server searches for, and displays, any mutual contacts from the virtual world data, for example if people who are not present at the given time, but are nevertheless connected to the two users in one of the online social networks used by LSS (section 3.3).Details of using RFID in LSS can be found in [5, 1].

3.3 Visualisation

LSS has two visualisations, taken from the SocioPatterns project (detailed in [1]):

- Spatial View: This view provides an overview of the real-time contact graph. It represents the location of RFID-badge wearing participants within range of the RFID readers, as well as their ongoing social contacts. Each participant is represented by a labelled yellow disc or, when available, by the Facebook profile picture. The contacts are represented by yellow edges, whose thickness and opacity reflects the weight of the contact. The edges are decorated, where applicable, with small Facebook, Flickr, Delicious, lastFM or COP icons, marking the occurrence of that relationship in the respective network.
- User-focus view: This view displays the social neighbourhood of a particular user. It shows all participants with whom this user has ongoing contact with (yellow edges for live contacts) or had significant (cumulative) contact with (grey edges for

historical contacts). This view also attempts to *close relevant triangles*, by showing mutual contacts as explained earlier.

4 Semantic Profiles of Interest

Tags usually reflect the interests of their authors. Such interests could range from topics, places, events, people, hobbies, etc. We have developed a tool that processes the public tagging activities of users and automatically generates a list of DBpedia URIs to represent the interests of the taggers [18]. To generate Profiles of Interest (POI) from social tagging, we follow these steps:

1. Collect tagging information: To retrieve tags, their frequencies, and time stamps from social networking systems using APIs and screen scraping.
2. Filter the tag cloud: Produces a compact, clean, set of tags by filtering out non-words, correcting misspellings, split compound tags, etc.
3. Associate tags with semantics: This step links a given tag to a suitable concept URI from DBpedia.org.
4. Tag disambiguation: Selecting the correct meaning URI for a given tag.
5. Calculate interest weights: Each identified concept URI (interest) is given a weight based on frequency and time decay.
6. Create profiles of interest: Generates a ranked list of interests based on the output of the steps above.
7. User verification: Allows users to verify and edit their POI as they see fit.

The above steps are detailed in the following sections.

4.1 Collecting tagging information

The first step in building a POI is collecting social tagging information from folksonomies. In previous work we used Google Social API[10] to find and correlate several social networking accounts of given users [19]. In LSS however, users must explicitly enter their social networking accounts on the LSS website. Therefore users are given full control on deciding which of their accounts will be used and shared.

The data collection process is responsible for harvesting information from a range of social networking sites. In the case of Flickr, Facebook, and Last.fm, APIs are provided that allow us to download a complete history of user tagging activity. However, for Delicious, the API is very limited and so we used custom screen-scraping scripts.

4.2 Filtering the tag cloud

Tags can be misspelled, synonymous and come in a morphologic variety. As a result, important correlations between resources and users are sometimes lost simply because of the syntactic mismatches of the tags they used. To this end, we developed a tag filtering architecture that cleans and condenses user generated tag-clouds [19].

Figure 2 shows the process of tag filtering used in LSS, which consists of the following operations:

[10] `http://code.google.com/apis/socialgraph/`

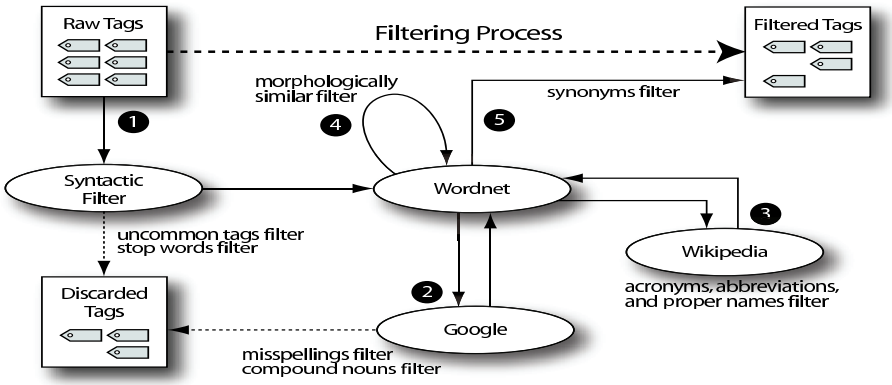

Fig. 2. The tag filtering process

1. Syntactic Filtering: This first step removes tags that are too small (length $= 1$) or too large (length > 25), and converts special characters to their base form (e.g. *Zürich* is converted to *Zurich*). Tags containing numbers are also filtered out, unless the tag is a popular one (e.g. `2009`, `911`, `666`, `7up`, `4x4`). After this process, if a remaining tag has an exact match in WordNet, we pass it directly to the set of *filtered tags* to avoid unnecessary processing.
2. Compound Nouns and Misspellings: We use the Google *did you mean* service to correct misspellings and to resolve compound tags; both commonly found in tags.
3. Wikipedia Correlation: Many popular tags do not appear in WordNet (e.g. tags of famous people, places, companies, acronyms, and contemporary terminology such as `web2.0`, `podcast`, `blog`, etc). In order to verify such tags, and to provide an agreed representation for them, we correlate them to their appropriate Wikipedia page. For example, when searching Wikipedia for the tag `nyc`, the entry for New York City is returned.
4. Morphological Similarity: We use a custom singularisation algorithm, and the stemming functions in the *snowball* library,[11] to reduce morphologically similar tags to a single tag. The shortest term in WordNet is used as the representative term.
5. WordNet Synonyms: The final step in our filtering process is to identify tags that are non-ambiguous synonyms, and merge them.

The output from the processes above is a "cleaned" set of meaningful terms that can now be disambiguated and analysed to identify user interests. Our tag filtering process is now available as an online service.[12]

4.3 Associating tags with semantics

All tagging information is stored by LSS in RDF. We use the TAGora tagging ontology[13] (figure 4.3) which is specifically designed to represent tag assignments (posts) and use (frequency, time, co-occurance with other tags in any post, relation to global

[11] `http://snowball.tartarus.org/`
[12] `http://tagora.ecs.soton.ac.uk/tsr/tag_filtering.html`
[13] `http://tagora.ecs.soton.ac.uk/schemas/tagging`

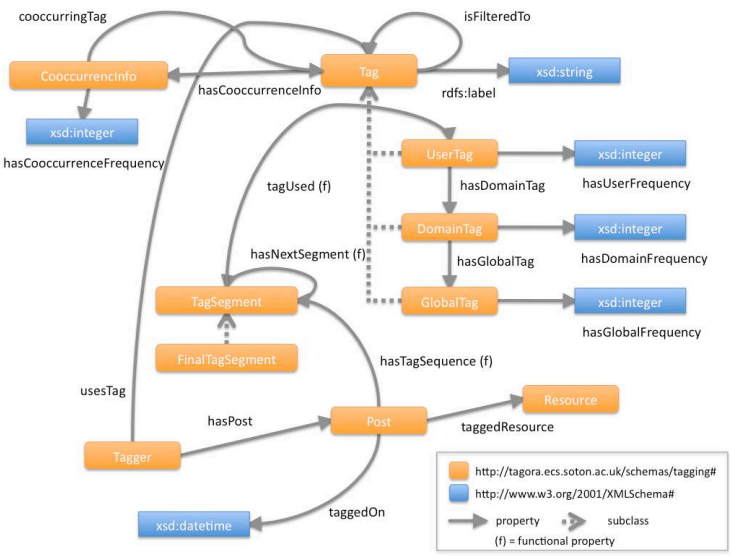

Fig. 3. TAGora Tagging Ontology.

tag, etc). In future work we plan to merge this ontology with SIOC,[14] by extending the latter to include our detailed tagging representations that are necessary for tag disambiguation.

For each user tag, we create a property in the ontology that links it to the Global Tag in the TAGora Sense Repository (TSR).[15] When queried with a tag (e.g. Delicious tag *ontologymapping*), TSR will look for DBpedia URIs and Wordnet Synsets that correspond to the possible meanings of the tag (e.g.`/dbpedia/resource/Semantic_integration`). Figure 4 shows how we generate FOAF files with links to interests inferred from user tags. TSR is available as a linked data endpoint for finding possible tag senses.[16]

4.4 Tag disambiguation

The disambiguation process aims to analyse the context in which a tag has been used to identify the most likely sense among all possible senses for that tag. Tags are considered ambiguous if they are associated with multiple senses (i.e. more than 1 DBpedia resource). For such a tag, its context is captured and represented as a term vector. By context we mean the other tags that were used to annotate a given resource, hence each use of the tag can have different contexts. We construct another vector from term frequencies associated with the possible DBpedia senses. We then measure cosine similarity between these vectors, and if one of the similarity scores is above a the threshold (0.3 in this case), we conclude that this is the correct sense for that tag. If more than

[14] `http://scot-project.org/scot/`
[15] `http://tagora.ecs.soton.ac.uk/`
[16] `http://tagora.ecs.soton.ac.uk/tsr/sense_matching.html`

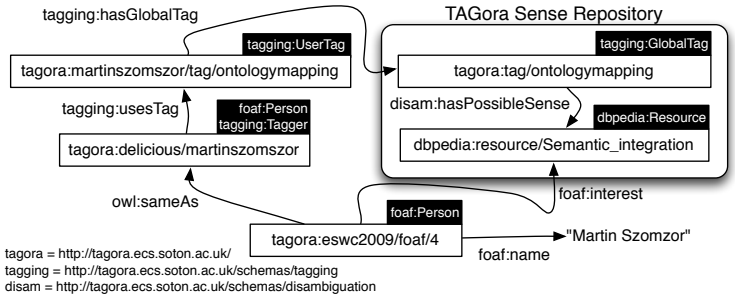

Fig. 4. Linking users to interests inferred from their tagging activities on social networking sites.

one (or zero) senses score above the threshold, we do not associate a meaning to the tag since we cannot reliably choose a correct sense. By iterating through all tags associated with a user (i.e. through Delicious or Flickr), we are able to build a candidate resource list of interests C. Details of our disambiguation algorithm and some initial experiments can be found in [7].

4.5 Calculating interest weights

For each interest (i.e. DPpedia resource) $r \in C$, we calculate a weight $w = f_r * u_r$, where f_r is the total frequency of all tags disambiguated to sense r, and u_r is a time decay factor, where $u_r = \lceil days(r)/90 \rceil$. Hence tags used within the last 3 months are given their total frequency, tags used between 3 and 6 months ago are given $1/2$ their frequency, 6 - 9 months a third, etc. If many tags of a given user are associated with the same interest, then the weight for that interest will increase accordingly. The final list of interests contains only those with a weight above the average weight for that user.

4.6 Creating the Profile of Interest

The lists of interests that comes out of the previous processes are used to generate an RDF Profile of Interest (POI) for each users using the FOAF `interest` property to link the person to the relevant Wikipedia categories. If more than 50 candidate interests have been found, we rank them by weight and suggest the top 50.

4.7 User verification

Once a POI is generated, the user can browse the list of interests and edit as required, by removing or adding new interests as they see fit (figure 5). Users may wish to remove an interest for various reasons, for example if it was incorrectly identified (e.g wrong disambiguation or filtering), if it is not an interest (or a historical one), or if the user chooses not to share it with the community (private or deemed irrelevant). Users can authorise LSS to use their profiles by clicking on a *save* button.

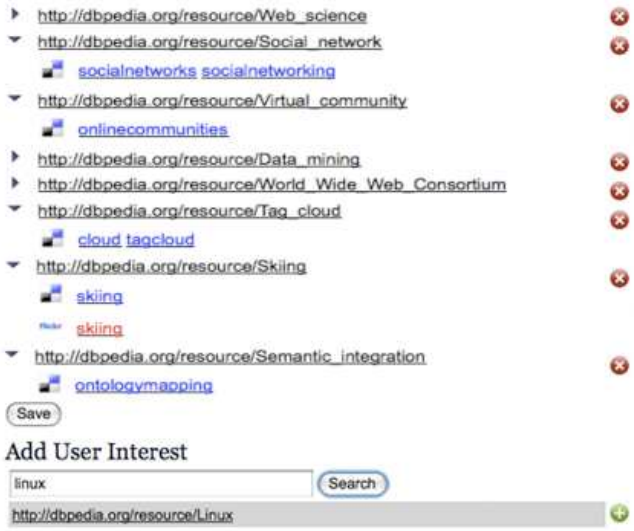

Fig. 5. Users can browse and edit their profiles of interest before authorising their use. The tags and their sources that led to each interest are shown. These profiles are automatically generated from users' public tagging activities.

5 Experiments and Results

LSS application was deployed for a total of 7 days at the following two events:

- European Semantic Web Conference (ESWC) in Crete, 1-4 June 2009: This conference was attended by 305 people, out of which 187 participated in LSS. Out of the 187 who collected an RFID badge, 139 of them also created accounts on our application site.
- HyperText (HT), Turin, June 29-July 1, 2009: Attended by around 150 people. 113 of them collected an RFID, and 88 registered with LSS.

Each participant was issued with a uniquely numbered RFID badge. Users were asked to enter their RFID ID number on a the dedicated LSS website. On this website, users were also able to provide their Delicious, Flickr, and lastFM account names, as well as activating a Facebook application that collected their social contacts. The results reported below focus on user participation and SNS account declaration and POI generation. Results and statistics of RFID use can be found in [5].

Participation results: As mentioned above, out of a total of 455 attendees of ESWC and HT conferences, 300 of them took part in Live Social Semantics (187 at ESWC, and 113 at HT). Out of these 300 users, 227 of them managed to create an account on the application site (139 at ESWC, and 88 at HT). Hence around 24% of the users who collected an RFID badge did not register to submit any information about themselves

(e.g. name, email, social network accounts). F2F contacts of such users were captured, but were not associated with any personal profile.

Declaration of social networking accounts: Users were able to declare on the LSS site their accounts for Delicious, Flickr, lastFM, and Facebook. The numbers of such accounts that our 227 registered users (i.e. users with LSS accounts) declared on the LSS site are shown in table 1.

Experiment \ Account	Facebook	Delicious	lastFM	Flickr	**Total**
ESWC09	78	59	57	52	**246**
HT09	48	28	26	23	**125**
Total	126	87	83	75	**371**

Table 1. Number of social networking accounts entered into LSS in two field experiments.

The number of social networking accounts declared on LSS site by each individual user varied from 0 (i.e did not enter any accounts), to 4 (i.e. entered an account for each of Delicious, Flickr, lastFM, and Facebook). Table 2 shows that about 37% of our 227 registered users did not declare any social networking accounts. It also shows that around 57% of our users declared more than one social networking account.

Experiment \ Number of accounts	0	1	2	3	4	**Total**
ESWC09 users	49	36	28	13	13	**139**
HT09 users	35	18	23	8	13	**97**
Total	84	54	51	21	26	**236**

Table 2. Number of users who entered 0,1,2,3 or 4 social networking accounts into the Live Social Semantics site during experiments at ESWC09 and HT09 conferences.

Semantic Profiles-of-Interest results We analysed 72 POIs that were verified and activated by our users (section 4.7). Table 3 shows the total number of interests that were automatically generated, and those that were removed manually by users during both field experiments. A total of 2114 DBPedia concepts were proposed, out of which 449 were removed by users (21%). Although a facility was included on the website for users to add new interests, only 19 new concepts were added.

	Global	Delicious	Flickr	lastFM
Concepts Generated	2114	1615	456	43
Concepts Removed	449 (21%)	307 (19%)	133 (29%)	9 (21%)

Table 3. Table shows the number of interests generated from tags taken from Delicious, Flickr, or lastFM, and how many were removed by users. These statistics are based on 72 POIs verified and saved by their owners.

6 Evaluation

Table 3 above showed that 29% of interests suggested from Flickr tags were removed by users, in comparison to 19% and 21% for Delicious and lastFM respectively. This suggests that Delicious and lastFM are perhaps more reliable sources of user interests than Flickr. Inspection of the concepts removed shows that Flickr was likely to suggest concepts referring to years, names, or to places that users visited in the past, but are no longer interested in.

To evaluate the accuracy of our interest suggestions, we examined the interests that our users removed from their profiles during the HT09 experiment.[17] Users may choose to delete an interest because it is simply inaccurate (i.e. wrong DBpedia match), it does not reflect an actual interest (i.e. a very general concept), or it is something they prefer to keep private. Users seem to have different perceptions of what an interest is, or which ones are worthy of sharing in this context. Some users were very conservative and only kept a few of the interests that our system generated for them, while others kept almost all their proposed interests. In future LSS implementations we intend to instantly capture user incentives and rational behind any interest removal. Understanding these drivers will help us to better design and tune the POI generation process. However, for this evaluation, we will focus of finding out how many of the removed interests were based on tags that our POI process matched to irrelevant DBpedia URIs.

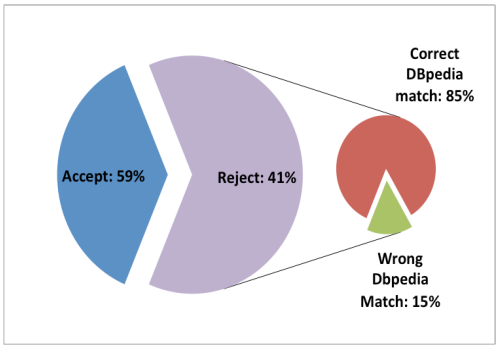

Fig. 6. 11 users edited their POIs in HT09. They accepted 59% of the interests that our system generated, and rejected 41%, out of which 15% were matched by our system to incorrect DBpedia URIs (6% of all suggested interests).

Although 36 of our users at HT09 have activated their POIs (by saving them - section 4.7), only 11 of them removed any interests. The other 25 users might have been totally satisfied with their original POIs, or perhaps they saved their profiles without reviewing them. To be on the safe side, in this evaluation we focus on the POIs that were clearly *scrutinised* and *corrected* by their owners. On average, those 11 HT09 users kept 63%, 57%, and 49% of the interests that the system suggested based of their

[17] Profiles of interests from ESWC 2009 were later anonymised and hence could not be included in this detailed evaluation since it requires an examination of users' original tags and tagged resources.

Delicious, Flickr, and lastFM tagging activities. Several users removed lastFM-based interests, although those interests referred to the music bands that these users listened to the most.

Figure 6 shows the percentages of tags that these 11 users removed from their automatically generated POIs. Although these users removed 41% of their POIs, only 15% (30 tags out of 203 removed ones) of these removed tags were given the wrong semantics (i.e. matched to the wrong DBpedia URIs).For example, for most users the tag "km" was wrongly matched to the concept "Kilometre" in DBpedia, instead of "Knowledge Management". With a closer look at those 15% of tags, we find that 2% of them originated from Flickr, and the rest came from Delicious. This is hardly surprising since Delicious tags tend to be more diverse than those from Flickr. Majority of Flickr tags referred to known geographical places that has dedicated DBpedia URIs.

In addition to the above, we have also evaluated the sharability of SNS accounts by our users. As mentioned in section 5, a total of 84 registered users did not enter any social networking accounts on the LSS site. To understand the drivers behind this, we ran a survey where we asked each of these users to pick their main reason out of the 5 options shown in table 4. We received 36 responses to our survey so far, and the main reasons the users picked are listed in table 4. It is clear that no having any SNS account is the most common reason for not declaring any. LinkedIn[18] and xing[19] were mentioned by several users as alternative SNS accounts, which LSS does not cater for yet. Although several users mentioned privacy concerns, only 8% of the users selected this as their primary reason.

Option	**Reason**	**No. Users**	**%**
a	don't have those accounts (or rarely use them)	16	44%
b	use different networking sites	10	28%
c	don't like to share them	3	8%
d	didn't get a chance to share them (eg no computer, slow internet)	6	17%
e	other	1	3%
	Total	36	100

Table 4. Reasons why some users didn't enter any social network accounts to our application site

Table 7 shows a comparison of the answers to our survey (table 4) from the ESWC and HT experiments. It is interesting to see that answer *b* was very common for ESWC attendees, who often blamed the unreliable internet connection at the venue for their inactive participation, whereas this was not an issue for HT.

7 Discussion and Future Work

The first phase of LSS development, which lead to the prototype tested at ESWC09 and HT09, was focused on architectural design and technology integration to demonstrate a novel proof-of-concept application. The second phase of development will focus on scalability, extendibility, and services.

[18] `http://www.linkedin.com`

[19] `http://www.xing.com`

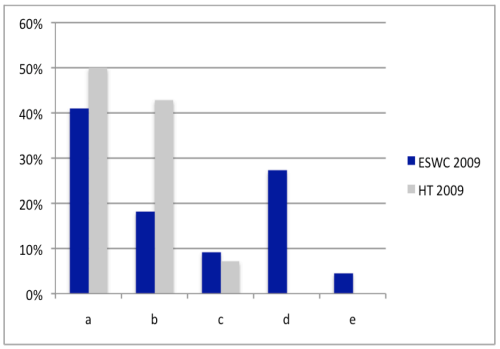

Fig. 7. Comparison of users answers to our participation survey for the two experiments.

LSS has so far only supported 4 currently popular SNSs; Delicious, Flickr, lastFM, and facebook. We plan to extend LSS to allow users to submit their FOAF files, and to support other networking sites (e.g. LinkedIn and xing). We also plan to develop a open plug-in architecture to allow external parties to develop connection to other networking systems to LSS.

The generating of profiles of interests from social tagging systems produced promising results. These POIs highlighted various general interests of users that usually cannot be inferred from their publications or project descriptions (e.g. "skiing", "iphone", "sewing", "The Beatles"). However, many users did not take the crucial step of verifying and editing these profiles. This might be due to a misunderstanding of the purpose or value for taking this step. We hope that more users will be encouraged to edit their profiles once we provide additional services that use these profiles, for example, to highlight people with similar interests.

Users who saved their profiles during both experiments removed approximately 21% of the interests the system suggested. This goes up to 41% if we exclude users who saved their profiles without any modifications. Although there could be many reasons for why users choose to remove an interest, our investigation showed that only 15% of the removed interests were totally inaccurate tag-to-concept associations (section 6). The other 85% were proper associations, but did not necessarily represent an interest. This is a clear indication that we need to develop more sophisticated methods for determining what constitutes an interest and what does not. One promising approach is to tap into our users' collective intelligence to improve our POI generation process, for example by filtering out the interests that most users tend to reject ("Tutorial", "API") or those that are too common or too general (e.g. "web 2.0", "semantic web").

Next step for interest identification will be to model user's interests in semantic hierarchies which will enable us to represent interests at different levels of granularity. For example, if someone is interested in "Visualbasic", "Perl", and "C++", then one can infer that this person is interested in "Programming languages". The hierarchy can show how general the user interest is, so one user may use the tag "music" very often, while another might tag with "jazz" or "Hip hop", which are more specific concepts than "music". People tag with different levels of specificity, and this usually reflects their level of expertise in the subject [9].

Extractions of POIs has so far been limited to users' online tagging activities. However, many of the participants have authored papers which can be used to determine their research interests, and some of these interests are already available on `semanticweb.org` in the form of paper keywords. Acquiring such interests can be added to the system and used to improve recommendations on talks or sessions to attend, or people to meet. Also, information from social networking accounts can be used to avoid recommending existing friends.

Many users expressed their interest in retrieving their data after the conferences. The next version of LSS will give users permanent access to their LSS accounts, to enable them to revisit their logs of face-to-face contacts, to modify or regenerate their POIs, and to access all the services LSS provides. This will not only enable them to access their activity log, but it will also allow them to carry their accounts across conferences where this application is deployed.

More services will be provided in future LSS deployments, such as a 'search for person', 'I want to meet', and 'find people with similar interests'. Data from RFIDs can be used to identify 'best attended session or talk'. Social contacts from social networking systems and COPs could be used to find out who has made new contacts, especially if we can compare data over several LSS deployments.

8 Conclusions

The Live Social Semantics application is pioneering the full integration of active RFIDs with semantics and social networking systems. The paper described and evaluated the generation of Profiles of Interests for individuals by analysing their public tagging activities on Flickr, Delicious, and lastFm. The paper reported results from deploying this application at two international conferences; ESWC 2009 and HyperText 2009, during which 300 people too part in LSS. 227 users shared 371 SNS account on LSS site. A POI was generated for each of these users, and saved by 72 of them. Overall, 21% of the interests suggested by our system were removed by users. When analysing logs of 11 HT09 users who clearly edited their POIs, we found that 15% of the interests they rejected were due to incorrect semantic association. Further research is required to better understand user's rational for removing/keeping interests, and for using their collective intelligence to improve our POI generation processes.

References

1. H. Alani, M. Szomszor, C. Cattuto, W. V. den Broeck, G. Correndo, and A. Barrat. Live social semantics. In *8th International Semantic Web Conference (ISWC)*, near Washington, DC, USA, 2009.
2. S. Angeletou, M. Sabou, and E. Motta. Semantically enriching folksonomies with flor. In *Workshop on Collective Intelligence & the Semantic Web (CISWeb), ESWC*, Tenerife, Spain, 2008.
3. A. Barrat, C. Cattuto, V. Colizza, J.-F. Pinton, W. V. den Broeck, and A. Vespignani. High resolution dynamical mapping of social interactions with active rfid, 2008. http://arxiv.org/abs/0811.4170.
4. G. Begelman, P. Keller, and F. Smadja. Automated tag clustering: Improving search and exploration in the tag space. In *Proc. 17th Int. World Wide Web Conf., Edinburgh, UK*, 2006.

5. W. V. den Broeck, C. Cattuto, A. Barrat, M. Szomszor, G. Correndo, and H. Alani. The live social semantics application: a platform for integrating face-to-face proximity with on-line social networking. In *Workshop on Communication, Collaboration and Social Networking in Pervasive Computing Environments (PerCol 2010), IEEE Int. Conf. on Pervasive Computing and Communications (PerCom)*, Mannheim, Germany, 2010.
6. N. Eagle and A. (Sandy) Pentland. Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.*, 10(4):255–268, 2006.
7. A. Garca-Silva, M. Szomszor, H. Alani, and O. Corcho. Preliminary results in tag disambiguation using dbpedia. In *Knowledge Capture (K-Cap'09) - Workshop on Collective Knowledge Capturing and Representation - CKCaR'09*, CA, USA, 2009.
8. H. Glaser, I. Millard, and A. Jaffri. Rkbexplorer.com:a knowledge driven infrastructure for linked data providers. In *Proc. European Semantic Web Conference*, Tenerife, Spain, 2008.
9. S. A. Golder and B. A. Huberman. Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32:198–208, 2006.
10. M. Guy and E. Tonkin. Tidying up tags? *D-Lib Magazine*, 12(1), 2006.
11. C. Hayes, P. Avesani, and S. Veeramachaneni. An analysis of the use of tags in a log recommender system. In *Int. Joint Conf. Artificial Intelligence (IJCAI), Hyderabad, India*, 2007.
12. P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot. Pocket switched networks and human mobility in conference environments. In *WDTN '05: Proc. 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, New York, NY, USA, 2005. ACM.
13. X. Li, L. Guo, and Y. E. Zhao. Tag-based social interest discovery. In *Proc. 19th Int. World Wide Web Conf (WWW), Beijing, China*, 2008.
14. A. Mathes. Folksonomies - cooperative classification and communication through shared metadata. *Computer Mediated Communication - LIS590CMC*, December 2004.
15. A. Passant and P. Laublet. Meaning of a tag: A collaborative approach to bridge the gap between tagging and linked data. In *Workshop on Linked Data on the Web (LDOW), Int. Word Wide Web Conference, Beijing, China*, 2008.
16. A. Scherrer, P. Borgnat, E. Fleury, J.-L. Guillaume, and C. Robardet. Description and simulation of dynamic mobility networks. *Comput. Netw.*, 52(15):2842–2858, 2008.
17. L. Specia and E. Motta. Integrating folksonomies with the semantic web. In *Proc. 4th European Semantic Web Conf. (ESWC)*, Innsbruck, Austria, 2007.
18. M. Szomszor, H. Alani, I. Cantador, K. O'Hara, and N. Shadbolt. Semantic modelling of user interests based on cross-folksonomy analysis. In *Proc. 7th Int. Semantic Web Conf. (ISWC)*, Karlsruhe, Germany, 2008.
19. M. Szomszor, I. Cantador, and H. Alani. Correlating user profiles from multiple folksonomies. In *Proc. Int. Conf. Hypertext (HT08), Pittsburgh, PA, USA*, 2008.
20. M. Tesconi, F. Ronzano, A. Marchetti, and S. Minutoli. Semantify del.icio.us: automatically turn your tags into senses. In *Social Data on the Web, Workshop at the 7th ISWC*, 2008.
21. P. Thibodeau. IBM uses RFID to track conference attendees, 2007. http://pcworld.about.com/od/businesscenter/IBM-uses-RFID-to-track-confere.htm.
22. L. Wu, B. Waber, S. Aral, E. Brynjolfsson, and S. Pentland. Mining face-to-face interaction networks using sociometric badges: Evidence predicting productivity in it configuration. In *The 2008 Winter Conference on Business Intelligence*, University of Utah, 2008.
23. C.-M. A. Yeung, N. Gibbins, and N. Shadbolt. Tag meaning disambiguation through analysis of tripartite structure of folksonomies. In *Web Intelligence and Intelligent Agent Technology Workshops, 2007 IEEE/WIC/ACM International Conferences on*, pages 3–6, 2007.