Live Social Semantics

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

[1] Intelligence, Agents, Multimedia
School of Electronics and Computer Science
University of Southampton, Southampton, UK
`{h.alani,mns2,gc3}@ecs.soton.ac.uk`
[2] Complex Networks and Systems Group
Institute for Scientific Interchange (ISI) Foundation, Turin, Italy
`ciro.cattuto@isi.it, wouter.vandenbroeck@isi.it`
[3] Centre de Physique Théorique (CNRS UMR 6207), Marseille, France
`alain.barrat@cpt.univ-mrs.fr`

Abstract. Social interactions are one of the key factors to the success of conferences and similar community gatherings. This paper describes a novel application that integrates data from the semantic web, online social networks, and a real-world contact sensing platform. This application was successfully deployed at ESWC09, and actively used by 139 people. Personal profiles of the participants were automatically generated using several Web 2.0 systems and semantic academic data sources, and integrated in real-time with face-to-face contact networks derived from wearable sensors. Integration of all these heterogeneous data layers made it possible to offer various services to conference attendees to enhance their social experience such as visualisation of contact data, and a site to explore and connect with other participants. This paper describes the architecture of the application, the services we provided, and the results we achieved in this deployment.

1 Introduction

Most conference attendees would agree that *networking* is a crucial component of their conference activities. Conferences, and similar events, are indeed often judged not only by their popularity or scientific qualities, but also by the social experiences they provide. Consequently, conference organisers are keen to enhance the social experience by offering activities or technologies that encourage and support social interactions.

We strove to significantly further the state of the art by developing a Semantic Web application that 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). The resulting prototypical application was deployed at the 2009 European Semantic Web Conference (ESWC09).

Making Use of Linked Data. The amount and variety of semantic data available on the web is continuously growing. The Linked Data initiative has been instrumental in this. Data from various conferences (e.g. ESWC, ISWC, WWW) has been consistently collected and published in recent years [12], and can be retrieved from sites such as

`data.semanticweb.org`. This data has been merged with data from several publication databases (e.g. CiteSeer, DBLP) by the RKBExplorer system [5]. From this data we inferred Communities of practice (COP), which offer a first insight into the scientific networks of the participants. We used this to provide awareness of the presence of their COP members at the conference, and of any talks they might be giving there.

Mining folksonomies. The tags that people use on various Web 2.0 sites tend to represent their personal interests [10]. Avid users are often active across several such sites, each of which solicits different aspects of a user's interests. If these are brought together, a far richer understanding of a user's interests can be obtained and subsequently used for superior personalisation, recommendation or awareness services [17]. Users often carry some of their tagging selections and patterns across different folksonomies [18]. Retrieving interests of conference attendees from multiple social sites, interests that might transcend the academic or scientific domain, could lead to more interesting matchmaking services. To this end, we generated Profiles of Interests (POI) for participants to allow people to explore each others' interests.

Meshing online and real-life social networks. Social relationship data from online social networks could provide a useful substrate for constructing social services. However, since such networks generally capture only part of the actual social network, meshing this data with knowledge of real-life social activities would greatly improve this potential. To this end, we deployed a novel active-RFID based sensor platform [1] that is capable of detecting real-life social interactions in terms of sustained face-to-face proximity events. This not only enabled us to provide participants with various novel services, such as logs and summaries of their social interactions, but also to integrate this with information from people's social profiles of interest, scientific communities of practice, and their online social contacts. This meshing not only leads to superior services, but can also facilitate the extension of both networks, online as well as offline [14].

The following Section sheds some light on related work. A full description of the Live Social Semantics application is given in section 3. Section 4 covers various aspects of the results of the application deployment at ESWC. Discussion and future work are given in Section 5, followed by conclusions in Section 6.

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 [9] by using wearable Bluetooth-enabled devices. Subsequent work focused on sensing organisational aspects [4] by using Bluetooth-enabled mobile phones, and on characterising some statistical properties of human mobility and contact [20, 15]. All of these early experiments involved a small number of participant, and could not assess face-to-face human contact in a large-scale setting, as they mostly relied on Bluetooth communication. Recently, the SocioPatterns project[4] investigated patterns of human contact at large-scale social gatherings by deploying a distributed RFID platform that is scalable

---

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

and attains reliable detection of face-to-face interactions as a proxy of social contact [1]. The application presented here leveraged that platform to mine real-time social contacts.

IBM used RFIDs to track attendees of a conference in Las Vegas in 2007. The devices were used to track session and meal attendance [20]. The information they collected were limited to the name, title and company of attendees. No social or semantics 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. To the best of our knowledge, our application is the first where real-world face-to-face contacts are mashed up in real time with semantic data from on-line social networking systems.

The novelty of the user profiling work reported here is in the amalgamation of multiple Web 2.0 user-tagging histories to build up personal semantically-enriched models of interest. This process involves dealing with several problems, such as filtering of tags, disambiguating them, associating tags with semantics, and identifying interests.

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. [11, 6, 7]. This total lack of control obstructs its analysis [10]. 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 [8]. Our filtering process is fully described in [3, 18] and produces a *cleaned* tag cloud for each user.

Tag ambiguity is a well recognised problem, yet still under researched. Clustering of tags was investigated for tag disambiguation [2], where similar tags were grouped together to facilitate distinguishing between their different meanings when searching. Similar clustering techniques were investigated to automatically identify emergent communities that correspond to a tag's different interpretations [21]. 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. The latter has been investigated in [16] where clusters of related tags are grounded to Semantic Web concepts.

The Meaning Of A Tag (MOAT) framework is a system in which users can manually select appropriate semantic web URIs for their tags from existing ontologies [13]. In contrast, the work reported in this paper explores a strategy for the automatic selection of URIs to maintain the essential simplicity of tagging, an approach also followed in [19] where DBPedia[6] concept URIs are automatically suggested for Delicious[7] tags. We make use of our own tagging and disambiguation ontologies since the ones provided by MOAT do not maintain tag ordering - an important feature when automatically determining tag semantics.

3   Live Social Semantics application

At ESWC09, the *semantic web*, the *social web*, and the *physical world* were brought together to create a rich and integrated network of information. Acquiring and integrating

---

[5] http://fireeagle.yahoo.net/
[6] http://dbpedia.org/
[7] http://delicious.com/

these heterogeneous, but overlapping, data sources enabled us to provide a new experience and services to conference attendees. The main goal was to encourage conference participants to network, to find people with similar interests, to locate their current friends, and to make new ones.

The Live Social Semantics application was deployed for 4 days (1-4 June 2009) during the European Semantic Web Conference (ESWC), which was located in Crete. More than 300 people attended the conference, out of which 187 accepted to participate in using our application. Each participant was issued with a uniquely numbered RFID badge. Users were asked to enter their RFID ID number on a website dedicated to this social application. On this website, users were also able to provide their Delicious, Flickr, and lastFM[8] account names, as well as activating a Facebook application that collected their social contacts. Out of the 187 who collected an RFID badges, 139 of them also created accounts in our application site (see Section 4).

3.1 General Architecture

Data from various Web 2.0 sources were imported using APIs or screen scraping, and subsequently converted RDF. The aim was to provide a service endpoint that supports the collection and reasoning over the data. Figure 1 provides a global picture of the Live Social Semantics framework. The vertical axis partitions the diagram according to two spaces: the virtual world (i.e. data about individuals held in the web), and the real world (i.e. RFID contact data). Data in the virtual world is sourced from social networking sites, to obtain social tagging data and contact networks, as well as the Semantic Web (SW), to obtain information about publications, projects, and the individuals COP (via RKBExplorer and semanticweb.org). All data is sourced directly from linked data sites, or converted to a linked data representation via the Extractor Daemon, and stored in a triple store (center, right of diagram). The Profile Builder (center, top of diagram) processes an individual's tagging activities and links them to DBpedia[9] URIs using the TAGora Sense Repository (sec. 3.4). Similarly, their favourite music artists from LastFM are linked to DBpedia URIs using DBTune.[10] In turn, the Profile Builder automatically suggests to users a list of interests that they can edit, and elect to expose to other participants. DBPedia was our choice *lingua franca* for representing participant's interests.

Data from the real world, i.e. that representing the social interactions of the conference participants, is collected and processed by a local server that communicates via RDF / HTTP with the triples store. A custom *Contact* ontology[11] was used to represent social interactions between individuals, recording the total contact time on a daily basis.

3.2 Semantically Interlinked Personal Data

To provide a practical framework that supports the integration of personal data, we employed a technique we refer to as Distinct Separated Identity Management (DSIM). DSIM provides each participant with a foaf:Person URI, that can be linked with other

---

[8] http://last.fm/
[9] http://dbpedia.org
[10] http://dbtune.org/
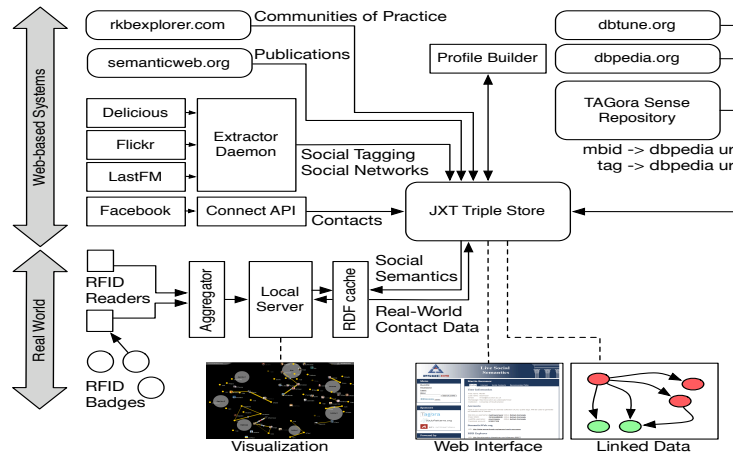[11] http://tagora.ecs.soton.ac.uk/schemas/LiveSocialSemantics

Fig. 1. Live Social Semantics Architecture

Semantic Web URIs that expose different metadata about the individual, whether that be an external linked data source such as data.semanticweb.org, or internally created data such as the contact data derived from RFID badges. This means that individuals contact data is stored in a separate graph to that of their Facebook friends, Delicious tags, COP, etc. The advantage of this approach is that it closely approximates a distributed linked data scenario (i.e queries must be expanded and run over multiple SPARQL endpoints, contact networks must be flattened), as well as allowing different processes to update the data model asynchronously. This asynchronous nature proved particularly useful when managing the real-time contact data since it enabled separate systems to simply push/pull data whenever needed, whether that be the local RFID server updating the central Triple Store with participants contact data, or the inclusion of Social Semantics (i.e. social networking contacts and profiles) in the visualisation client.

3.3 Real-Time Social Contacts

In order to mine the real-world interactions of conference attendees, we deployed the hardware and software infrastructure developed by the SocioPatterns project [1]. The name badges of those attendees who volunteered to become users of the application were equipped with active RFID badges.[12] The RFID 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 face-to-face proximity of two individuals.

We assume continued face-to-face proximity, within a distance of approximately 1 meter, to be a good proxy for a social interaction between individuals. Contacts are not recorded if people are facing the same direction (e.g. listening to a speaker), unless they turn and face each other for around 10 seconds or more. This kind of resolution is a result of the particular distributed sensing technology we use here, which pushes the state of the art of RFID platforms.

---

[12] Each RFID is equipped with a common button-size battery cell that can last for one week on average.

The real-world proximity relations are relayed from RFID badges to RFID readers installed in the conference venue. The readers encapsulate the RFID packets into UDP packets and forward them over a local Ethernet network to a central server. There, the UDP packets from RFID badges are aggregated and fed to a post-processing server that builds and maintains a real-time graph representation of the proximity relations among the tagged attendees. This instantaneous contact graph is represented as a time-dependent adjacency matrix A_{ij}^t, such that $A_{ij}^t = 1$ if individuals i and j are in contact at discrete time t, and $A_{ij}^t = 0$ otherwise. The adjacency matrix was updated every 5 seconds.

The post-processing server also maintains a weighted graph representation of their cumulative proximity relations of the tagged attendees over time. The (normalized) adjacency matrix of the cumulative graph during the time interval $[t_1, t_n]$ is defined as $C_{ij}(t_1, t_n) = (1/n) \sum_{k=0}^{n} A_{ij}^{t_k}$. The matrix element $C_{ij}(t_1, t_n) \in [0, 1]$ is the fraction of application time that individuals i and j spent together. Periodically, the cumulated proximity graph is thresholded, and those relations for which $C_{ij} > C_0$ are represented as a set of RDF triples describing the cumulated real-world proximity of attendees, and periodically uploaded to the triple store via RDF / HTTP.

The real-world proximity relations of the instantaneous proximity graph are mashed up by the server with the web-based attendee relations that it periodically pulls from the triple store. This allows the visualisation clients to display real-world relations in the context of their on-line counterparts.

Moreover, the post-processing server uses the real-world and web-based relations to compute simple recommendation schemes. For example, if two attendees are in contact at a given time, the server provides access to those attendees who may not be present at the same time, but are nevertheless connected to the two users in one of the web-based social networks covered in the application. The visualisation clients (specifically, the user-centerer views) can then use this information to enhance the presented information and support browsing of the social network. More precisely, when creating a personalised view for attendee u_0, the system considers the set $\mathcal{V}$ of attendees who are currently in contact with u_0, or who have been (significantly) in contact with u_0, i.e.,

$$\mathcal{V}(u_0) = \{v \in \mathcal{U} \mid A_{u_0,v} \neq 0 \vee C_{u_0,v} > C_0\} ,$$

where $\mathcal{U}$ is the set of all attendees. Subsequently, the system considers the neighbors of attendees $\mathcal{V}$ along the web-based social networks obtained from the triple store, and builds a new set of attendees $\mathcal{W}$ such that members of $\mathcal{W}$ are connected (along web-based systems) to both a member of $\mathcal{V}$ *and* to the focused attendee u_0, closing triangles that have one edge grounded in (current or cumulated) physical proximity, and two edges grounded in on-line relationships. The instantaneous contact graph, the cumulated contact graph, and the web-based graphs are then restricted to the set $u_0 \cup \mathcal{V} \cup \mathcal{W}$ and sent to the visualisation clients, that lay them out for the final user.

Proximity data from RFID devices were taken in the conference area only, covering conference sessions and coffee breaks, but excluding breakfasts, lunches, and evenings.

3.4 Profiles of Interest

In previous work [17], we devised an architecture to automatically generate a list of DBpedia URIs to represent interests a person might have by processing their social
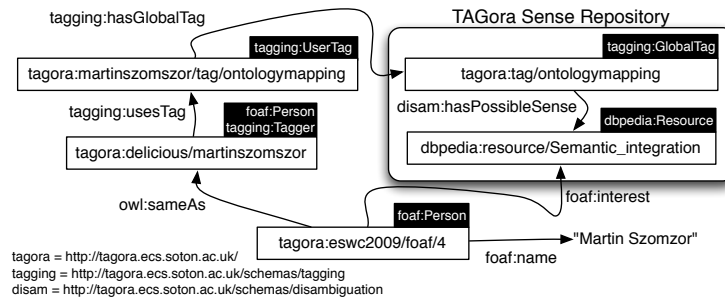
Fig. 2. Linking participants to their interests. The boxes in the diagram represent linked data URIs that provide metadata about various aspects of a participant's social networking data. They link participants to each other through the contact data exposed in various social networking sites, as well as associating them with interests that have been mined from their tagging activity.

tagging activity. Under the assumption that the tags used most often by an individual correspond to the topics, places, events and people they are interested in, we sought to provide a novel dimension to the social interaction at the conference by providing people with a basis to expose their interests, both professional and personal, and see those of others at the conference. Central to this idea is that these profiles can be built automatically, only requiring a short verification phase from the user.

Within the Live Social Semantics architecture, any social tagging information from Delicious and Flickr is collected and converted to an RDF representation (according to the TAGora tagging ontology[13]) by the Extractor Daemon (Figure 1). For each URI that represents a user's tag (for example a Delicious tag *ontologymapping*), a property is created that links it to the Global Tag in the TAGora Sense Repository (TSR).[14] When queried with a tag, the TSR will attempt to find DBpedia.org URIs and Wordnet Synsets that correspond to the possible meanings of the tag. This linked data resource provides information about the possible senses of a tag with mappings to DBpedia resources. Figure 2 contains example URIs that show how the FOAF file produced by our system for Martin (`http://tagora/eswc2009/foaf/4`) is linked to the interest *Semantic integration* via the delicious tag *ontologymapping*.

Profile Building Algorithm To build a Profile of Interests (POI), we first check to see if the user has a LastFM account. Using DBTune, a linked data site providing metadata about music, we can map the MusicBrainz[15] ID associated to their top artists in LastFM to a resource in DBpedia. The top 5 artists with a DBpedia mapping are added to the user's POI. The second phase of the profile generation procedure is to map the user's tags to DBpedia resources that represent their topics of interest. This is achieved with the following steps:

1. **Disambiguate Tags** When tags are associated to multiple senses (i.e. more than 1 DBpedia resource), we compare the similarity (using a cosine measure) of the

---

[13] `http://tagora.ecs.soton.ac.uk/schemas/tagging`

[14] `http://tagora.ecs.soton.ac.uk/`

[15] `http://musicbrainz.org/`

user's cooccurrence vector for that tag (i.e. all other tags that occur in the same post, and their frequencies) against the term frequencies associated with the possible DBpedia senses. If one of the similarity scores is above a threshold value, (0.3 in this case), we conclude that this is the correct sense for that tag. If more than one (or zero) senses score above the threshold, we do not associate a meaning to the tag. By iterating through all tags associated to a user (i.e. through Delicious or Flickr), we are able to build a candidate resource list C.

2. **Calculate Interest Weights** For each 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 a time decay factor. This factor $u_r = \lceil days(r)/90 \rceil$. 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.

3. **Create Interest List** If more than 50 candidate resources have been found, we rank them by weight and suggest the top 50. Since users are required to edit and verify this list, we believe it important to keep the number of suggestions to a reasonable amount.

Such semantic POIs could be used to find users with similar interests.

3.5 Visualisation

Two kinds of real-time visualisations were provided. The first, the *spatial view*, was publicly displayed on large screens in the main lobby area. The second, the *user focus view*, was accessible by means of a web browser on the conference LAN, and is linked to from each user's account page on the application site. Both are dynamic visualisations driven by regular updates received through a TCP socket connection with the local post-processing server.

Spatial view This view provides an overview of the real-time contact graph. It represents the RFID-badge wearing participants within range of the RFID readers, as well as ongoing social contacts (see section 3.3). 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. This approach constitutes a projection of said networks onto the real-time contact network.

The SocioPatterns project is primarily concerned with the real-time detection of the contact topology. The precise localisation of the participants in the physical space is of lesser concern. However, a coarse-grained localisation of the participants with respect to the RFID readers is possible. This enabled us to not only represent the contact topology, but also give an indication of which area the participants are in. To this end, the RFID readers were represented by labelled grey shapes, equiangularly laid out on a circumcentric oval, and the participants' shapes are positioned near or in between the readers' marks they are close to. This approach adds spatial structure to the contact graph representation.
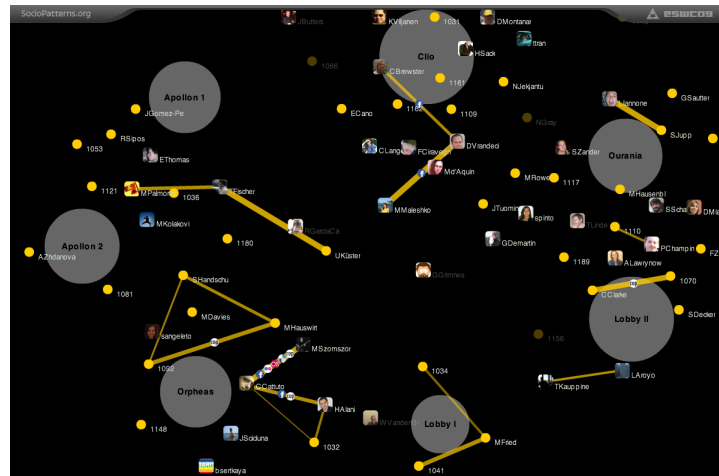
Fig. 3. Screenshot of the spatial view grabbed during a session.

User-focus view This view displays the social neighbourhood of the focussed upon participant. It represents all participants with whom this user has ongoing contact or had significant (cumulative) contact with so far. All physical interactions between these participants are shown as edges, the current ones in yellow, the historical in grey.

This view furthermore attempts to *close relevant triangles*, by which we mean that all participants that are in some way linked to both the focus participant and any of the initially included participants (i.e. those with whom the focus participant has or had contact), are also included, as well as the concerned links, decorated with the relevant icons like in the spatial view. The objective was to provide the users, after focussing upon themselves, with an overview of that subsection of their social neighbourhood that is relevant for their networking activities at that moment.

3.6 Privacy

Permission was sought from all participants for collecting and using their data. A form was prepared which explained what the data is, how it was going to be used, and for how long. Users were shown how the RFID badges are used, and the geographical limits of where their face-to-face contacts can be detected (conference building). When creating an account on the application site, each user was given the option of destroying their data after the end of the event.

As explained in Section 3.1, a POI was generated for each user who declared an account in any of the tagging systems we supported (Delicious, Flickr, and lastFM). To ensure that the users are happy with those interests to be viewed by others, each user was asked to verify and edit their list of interests. These profiles only become visible to other users once their owners activate them.

As an extra security, all data from the RFID devices were encrypted to ensure that could only be processed by our systems. All the data gathered by this application were stored in private triplestores, only accessible to the developers of this application.
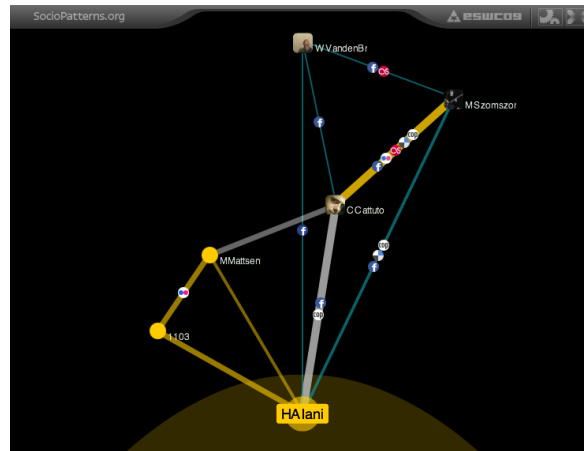
Fig. 4. User-focus visualisation in which *HAlani* has the focus. He has ongoing contacts with *MMattsen* and an anonymous user with badge id *1103*, as indicated by the yellow edges. These two users are also in contact, and they are Flickr friends as indicated by the yellow edge and the Flickr icon that decorates it. There has been significant contact between HAlani and *CCattuto*, as indicated by the thick grey line. They are also Facebook friends and share a COP. Both WVandenBr and MSzomszor were included in order to close relevant triangles. The cyan coloured edges indicate that the users are (only) linked in one or more of the social or COP networks.

4 Results

In this section we will report on various results of the application launch at ESWC09; numbers of participants and their shared networking accounts, interest profile generation, RFID usage, and privacy outcomes.

4.1 Participation

Out of the 305 conference attendees, 187 of them took part in Live Social Semantics. Out of these 187 users, 139 of them managed to create an account on the application site, Hence about 26% of the users who collected an RFID badge did not submit any information about themselves (e.g. name, email, social network accounts). Face-to-face contacts of such users were captured, but were not associated with any personal profiles.

4.2 Social networking accounts

The application site allowed users to declare their accounts on Delicious, Flickr, lastFM, and Facebook. Table 1 shows how many social networking accounts were entered into our system by all our 139 registered participants.

Account	Facebook	Delicious	lastFM	Flickr	Total
Quantity	78	59	57	52	246

Table 1. Number of social networking accounts entered by users into the application site.

Table 2 shows that about 35% of our registered users did not declare any social networking accounts (49 users). It also shows that over 61% of our 139 users had more than one social networking account.

Number of Social Networking Accounts	0	1	2	3	4	Total
Number of Users	49	36	28	13	13	**139**

Table 2. Number of users who entered 0,1,2,3 or 4 social networking accounts into the Live Social Semantics site.

After the conference, we emailed all 49 users who did register on our site, but did not enter any social networking accounts. Aim was to understand the reasons behind that. Table 3 lists the 22 responses we received so far. Out of those 22 participants, 9 (41%) of them simply did not have *any* social networking accounts, and only 1 of these 9 indicated that s/he have an almost empty Facebook account. Four participants (18%) indicated that they use *other* networking accounts, (LinkedIn was named twice). Only 2 (9%) of the 22 replies we received cited privacy reasons for not sharing their social networking accounts. Six replies (27%) picked answer *d*, and four of them blamed the slow internet connection at the conference venue. One participant (5%) picked *e* for being "too busy' during the conference'.

Option	Reason	No. Users	%
a	don't have those accounts (or rarely use them)	9	41%
b	use different networking sites	4	18%
c	don't like to share them	2	9%
d	didn't get a chance to share them (eg no computer, slow internet)	6	27%
e	other	1	5%
	Total	22	100%

Table 3. Reasons why some users didn't enter any social network accounts to our application site

4.3 Social Profiles-of-Interest Results

Out of the 90 people who entered at least one social networking account (Table 2), 59 of them entered at least one account from Delicious, Flickr, or lastFM (remaining 31 only entered Facebook accounts, which we do not use when generating POIs). Although our profile building framework had the potential to utilise all three of these accounts, the linked data site DBTune was offline for the duration of the conference, and hence, we were unable to associate a user's favourite lastFM artists to a DBPedia concept. 41 individuals viewed and saved their POI, of which 31 had a non-empty profile generated. Empty profiles were generated for a number of users who registered that had a very small or empty tag-cloud. Table 4 summarises the results in terms of the number of concepts automatically generated, the number that were removed manually by users, the number that were added manually, and the size of the final profile they saved.

A total of 1210 DBPedia concepts were proposed (an average of 39 per person across the 31 non-empty profiles), out of which 247 were deleted. While it would be useful to know exactly why users deleted a concept, whether it be simply inaccurate (i.e. incorrect disambiguation), it didn't reflect an actual interest (i.e. a very general concept), or it was something they wished to keep private, we considered it too much of a burden to ask users this question when editing their profiles. The total number of

	Global	Delicious	Flickr
Concepts Generated	1210	922	288
Concepts Removed	247	156	91
Concepts Added	19		
Concepts Saved	982	766	197

Table 4. Statistics of the profile generation, editing, and saving.

concepts deleted was 20% of those suggested. Although a facility was included on the website for users to add their own interests, few did - only 19 new concepts were added. When comparing the results from Delicious and Flickr, we see that 17% of concepts proposed from Delicious Tags were deleted, and 32% respectively for Flickr tags. This suggests that the accuracy of topics harvested from Delicious tags was more accurate than those from Flickr. Inspection of the concepts removed shows that Flickr was likely to suggest concepts referring to years and names.

4.4 RFID results

Data from RFID badges were taken for a continuous interval of about 80 hours, fully covering the three days of the ESWC09 main track. During that interval, one snapshot of the instantaneous contact graph was recorded every 5 seconds, for a total of $57,240$ snapshots covering the approximate location and the proximity relations of 174 RFID devices[16].

The first column of Table 5 reports the fraction of possible pair-wise contacts that involve face-to-face proximity for a time interval longer than a given threshold. As expected, the cumulative contact graph is dominated by contacts of short duration, and the introduction of a threshold on contact significance makes the graph more and more sparse as the threshold increases. Table 5 also reports standard network metrics for the cumulative contact graphs over the entire conference duration, for different values of the contact duration threshold. It is apparent that the heterogeneity of the contact graph makes it impossible to choose a single threshold for the significance of social contacts.

threshold	edge fraction	avg. degree	clustering	# conn. comp.	avg. comp. size	isolated nodes
1 min	17.1%	14.9	0.36	1	173.0	0.5%
2 min	11.4%	10.2	0.34	2	84.5	2.9%
5 min	5.5%	5.4	0.20	1	153.0	12.1%
15 min	1.7%	2.6	0.21	7	14.6	41.4%
30 min	0.5%	1.4	0.08	18	3.1	68.4%
60 min	0.1%	1.3	0.17	6	2.3	92.0%

Table 5. Properties of the cumulative contact graph, as a function of the contact duration threshold. The edge fraction is the percentage of possible edges that are present. The average degree is the average number of contacts to distinct attendees. Clustering is the average node clustering of the graph. The number of connected components and their average size is also reported, together with the fraction of isolated nodes.

Table 6 reports the number of distinct attendees met by the 5 most social attendees of the conference, for different values of the contact duration threshold. As the RFID

---

[16] Out of the 187 RFIDs we gave out, 13 were used during workshops only

contact data were taken during the conferences sessions and coffee breaks (and not during lunchtime, for example), only few contacts of long duration were observed.

threshold	#1	#2	#2	#4	#5
1 min	61	59	40	35	27
2 min	49	44	24	21	18
5 min	22	17	15	11	10
15 min	7	7	5	4	4
30 min	2	2	2	2	2
60 min	2	2	1	1	1

Table 6. Number of individuals met by the 5 most social attendees at the conference, as a function of the contact duration threshold.

4.5 Privacy results

Naturally, privacy is always a concern in such contexts, where personal data is being collected and processed in various ways. As explain in section 3.6, we took various measures to secure the data and protect privacy, even though most of the data we were gathering was actually in the public domain (e.g. shared tags).

Some participants asked for the data to be kept without any anonymisation, to be stored for reuse in coming events, and even to be published so they can link to their profiles and contacts logs from their websites and blogs. On the other hand, some participants were only prepared to take part if the data is anonymised. Table 7 shows the two options given to the users in the Terms & Conditions form (section 3.6) when they come to register on the application website. The table shows that 61% of the participants were happy for their data to be kept, while 39% requested the destruction of their data.

Option	No. Users	%
I agree for my data to be used for research purposes after the end of this experiment if properly anonymised	85	61%
I do not agree for my data to be kept after the end of this experiment	54	39%
Total	139	100%

Table 7. Numbers of participants who chose for their anonymised data to be kept, or destroyed.

The numbers in section 4.2 indicate that the majority were happy to share their social networking accounts. However, we cannot extend this observation to the other 48 who collected and RFID but never registered any information into our site.

5 Discussion and Future Work

The deployment of the application at ESWC2009 was the first where all components were put together and a good number of participants got to use it. We observed quite a few technical and sociological issues, which we discuss in the following.

There are many social networking sites, but in this first version we only supported four currently popular ones. We are working on a open plug-in architecture that allows external parties to develop the functionality needed to connect to, and crawl data from, other networking systems. We also plan to let users submit their FOAF files.

The number of available social networking sites on the web is always on the increase, and the popularity of such sites is never constant. In our application, only four of such networking systems were taken into account. Although the ones we selected are currently amongst the most popular ones, several users wished to add other accounts, such as FOAF files and LinkedIn. One approach to increase extendibility and increase coverage is to use an open architecture to allow external parties to develop and plug applications and services to connect to, and crawl data from, other networking systems, or sources such as FOAF files.

We had devised a privacy and data retention policy in which we pledged to anonymise the resulting data set, and allowed participants to request the complete removal of their data after the end of the event. This approach introduced a number of inconsistencies and ambiguities. As highlighted in section 4.5, many users expressed their interest in acquiring their data after the conference. However, the anonymisation actually precludes that. Other issues, such as whether a participant holds the right to access information on recorded contacts with participants that choose to have their data fully removed, points to the need to reconsider our privacy and data retention policies for future deployments. We believe it would be important to allow people to retain all their data, including user accounts, profiles, contact logs, etc. This will not only enable them to access their activity log, but it also allow them to carry their accounts across conferences where this application is deployed.

The visualisation displays of live contacts were popular points of attraction during the conference. They gave accurate reflections of real-time social interactions during the conference. People were often coming to those displays, searching for their colleagues, session chairs, organisers, etc. We plan to extend these visualisations to highlight conference organisers, session chairs, authors. Furthermore, we plan to introduce support for Twitter, both as another source of on-line social links and as a way of providing additional conference-related content in the visualisations. For example, the node corresponding to a person in the visualisation could be highlighted when s/he sends a message on Twitter that relates to the conference.

The results in table 3 show that 27% of our users could not log into our system to enter further data (social networking accounts, edit POI, etc) because of bandwidth issues at the conference venue. To avoid rushing everyone to enter their data while at the conference, we plan to make the application site available well ahead of the starting date of conferences where this application will be deployed next.

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. We furthermore believe it will be advantageous to organise the interests URIs into hierarchies, to support inference and fuzzy matching.

The use of Flickr tags to identify interests seemed to be less accurate than when using Delicious tags. This indicates the need for alternative approaches when dealing with Flickr tags.

In the application deployment reported here, recommendation of attendees based on physical proximity and on links in social networking systems was performed by means of the simple scheme of section 3.3. One could consider ranking schemes for suggested attendees, to make the recommendation more useful and serendipitous. Specifically, a representation of the context of the conference and of the attendees' interests can help in ranking suggested social connections in terms of their predictability based on the conference context.

More services will be provided in future application runs, 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 application launches.

6   Conclusions

The Live Social Semantics application was a demonstration of how semantics from several different sources can be harnessed and used to enhance the real-world interactions of people at a social gathering. In particular, the combination of semantic data from social media with the real-world encounters of attendees provides a new way of connecting to people, both in the real world and on-line.

Exposing real-world encounters in digital form facilitates mining interesting and serendipitous social connections, and greatly facilitates the process of establishing new on-line connections to encountered people. On the other hand, connections in social networking systems such as Facebook can be used to stimulate real-world encounters on the basis of shared acquaintances and interests. All of these opportunities were explored by the participants of ESWC09, and their reactions, observations, and responses provided valuable input on the future evolution of the platform.

In general, this application goes in the direction of making the co-evolution of real-world social networks and on-line social networks more transparent to users, more lightweight, and more usable. The Live Social Semantics application also provided a first opportunity to expose the semantics of social encounters, and investigate recommendation schemes in bodies of data that mix links from social media with links from real-world encounters.

Overall, this application has great potential for further growth over future deployments at conferences and similar event in a wide variety of domains.

Acknowledgement

References

1. 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.

2. 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.

3. I. Cantador, M. Szomszor, H. Alani, M. Fernändez, and P. Castells. Enriching ontological user profiles with tagging history for multi-domain recommendations. In *Proc. Workshop on Collective Semantics: Collective Intelligence and the Semantic Web (CISWeb 2008), in 5th ESWC, Tenerife, Spain*, 2008.

4. N. Eagle and A. (Sandy) Pentland. Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.*, 10(4):255–268, 2006.

5. 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.

6. S. A. Golder and B. A. Huberman. Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32:198–208, 2006.

7. M. Guy and E. Tonkin. Tidying up tags? *D-Lib Magazine*, 12(1), 2006.

8. 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.

9. 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.

10. 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.

11. A. Mathes. Folksonomies - cooperative classification and communication through shared metadata. *Computer Mediated Communication - LIS590CMC*, December 2004.

12. K. Mller, T. Heath, S. Handschuh, and J. Domingue. Recipes for semantic web dog food - the eswc and iswc metadata projects. In *Recipes for Semantic Web Dog Food - The ESWC and ISWC Metadata Projects*, Busan, Korea, 2007.

13. 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.

14. A. Passant, I. Mulvany, P. Mika, N. Maisonneauve, A. Löser, C. Cattuto, C. Bizer, C. Bauckhage, and H. Alani. Mining for social serendipity. In *Dagstuhl Seminar on Social Web Communities*, 2008.

15. 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.

16. L. Specia and E. Motta. Integrating folksonomies with the semantic web. In *Proc. 4th. European Semantic Web Conf. (ESWC), Innsbruck, Austria*, 2007.

17. 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.

18. M. Szomszor, I. Cantador, and H. Alani. Correlating user profiles from multiple folksonomies. In *Proc. Int. Conf. Hypertext (HT08), Pittsburgh, PA, USA*, 2008.

19. M. Tesconi, F. Ronzano, A. Marchetti, and S. Minutoli. Semantigy del.icio.us: automatically turn your tags into senses. In *Social Data on the Web, Workshop at the 7th ISWC*, 2008.

20. P. Thibodeau. IBM uses RFID to track conference attendees, 2007. http://pcworld.about.com/od/businesscenter/IBM-uses-RFID-to-track-confere.htm.

21. 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.