PAD: A Semantic Social Network

Samad Paydar, Mohsen Kahani, Fattane Zarrinkalam
Web Technology Lab., dept. of Computer Engineering
Ferdowsi University of Mashhad
Mashhad, Iran
samad.paydar@stu-mail.um.ac.ir, kahani@um.ac.ir, fattane.zarrinkalam@stu-mail.um.ac.ir

Abstract— **In this paper, *PAD* is introduced which is a semantic social network developed at *Ferdowsi University of Mashhad*. It uses the semantic web technologies in different ways: First it publishes parts of the social network data as *Linked Data*. It uses semantic web sources for automatic annotation of the social network entities with resources on the semantic web. Finally it uses the semantic web sources to collect required background or domain knowledge for the purpose of recommending items to the social network users. *PAD* has been designed to provide a medium through which people at *Ferdowsi University of Mashhad* can share their interests and knowledge, and also to provide a native infrastructure for Iranian researchers in different domains like semantic web, social networks, recommendation systems, and knowledge management. In this paper, architecture of *PAD* is described and different aspects in which *PAD* is related to the semantic web are discussed.**

Keywords- Semantic Web; Social Network; Linked Data; Annotation; Semantic Social Network

I. INTRODUCTION

From a social point of view, *social networks* can be considered as the most successful application of the web in the daily life of people. They enable people to easily connect with each other, communicate and share photos, posts and interests.

On the other hand, from a technical point of view, *semantic web* is one of the most essential visions considered for the web. Its mission is to make the web an environment in which not only people, but also machines can understand semantic of the contents published.

The integration of these two successful technologies has been interesting for many researchers and as a result, semantic social networks are emerged which utilize the semantic web technologies to improve different processes associated with social networks [18].

In this paper, *PAD*[1] is introduced which is a semantic social network developed at *Ferdowsi University of Mashhad*. Its aim has been to utilize the semantic web technologies to create a social network through which people at *Ferdowsi University of Mashhad* can share their interests and knowledge.

In addition, another goal that *PAD* is expected to address is to provide a native infrastructure with real data for Iranian researchers in different domains like semantic web, social networks, recommendation systems, and knowledge management.

PAD is implemented and it is in the early days of its deployment[2]. It is interesting to see how it can contribute to knowledge dissemination across its target community. In this paper, architecture of *PAD* is discussed and different aspects in which *PAD* is related to the semantic web are discussed.

The rest of the paper is organized as follows: In Section II the related works are briefly reviewed. Section III describes the architecture of *PAD*, and Section IV discusses different aspects from which *PAD* is related to the semantic web. Finally, section V concludes the paper by mentioning some future works.

II. RELATED WORK

Web was originally emerged as a medium of communicating information between scientists and academic people. Later, as the web based systems evolved, web started to act as a medium for publishing and accessing business processes across the internet. During the last decade, with the emergence of Web 2.0 and its associated technologies, web has become more and more social in the sense that not only computer experts or scientist, but any person can contribute to evolution of the web by simply disseminating his ideas, information, and interests [15], [13].

Social networks, as one of the technologies of Web 2.0, have greatly increased users' involvement on the web. They allow users to communicate with each other based on their commonalities, and easily share whatever they are interested in [3], e.g. photos, video clips, and bookmarks.

Despite interesting success stories attributed to well-known social networks like *Facebook*[3], they also have some limitations [17]. While it is easy to share data on a single social network, it is much challenging to share it across different social networks, because each network has its own data schema and conceptual space [8]. This portability limitation is very important since usually people are simultaneously involved in multiple social networks. If a user decides to migrate from a social network to another, there is no solution available for transferring all his data, including his friends community, from the source to the destination social network [10].

Another issue is that the simplicity of sharing data on social networks leads to a large volume of unstructured

---

[1] PAD is an abbreviation taken from ‹پایگاه اشتراک دانش›

data to be created, which is not readily usable or understandable for machines.

The idea of the semantic web is to uniformly represent data and its semantic in a well-defined format across the web. This semantic augmentation enables machine to better process web data, to the extent that it is usually stated that in the semantic web, machines understand data.

As a result, due to the benefits of the semantic web, researchers have worked on semantic social networks as an integration of the semantic web technologies with social networks to address existing limitations [4], [8], [19].

One theme of works on semantic social networks deals with using semantic web technologies, mainly ontologies, to publish social network data in a better representation format. This makes it possible to easily share data across different social networks. For instance, in [10], some well-known ontologies like *FOAF*[4] and *SIOC* [2] are used to provide portability between social media Sites and also to model user information and user-generated content in a machine-readable way.

Since the semantic web technologies are proved to be successful in improving information retrieval tasks and increasing flexibility of data access mechanisms, they can be used to effectively analyze data. As a result, there is another theme of works which uses the semantic web technologies for the purpose of social network analysis [14], [10].

Some other works utilize the semantic web technologies to integrate data of social networks with data from other sources to create a rich content layer for being used in a specific task. For instance, an ontology is used in [12] to integrate information from *DBLP*[5] and *LinkedIn*[6] to identify and extract social relationships between different researchers. As another example, in [5], the authors propose a method to integrate three different networks, i.e. social network, ontology network and concept network, as a single semantic social network. Further, they discuss how it is possible to infer information in a network by utilizing information of the two other networks.

In [9] and [7] also the semantic web technologies are used for enriching contents of a social network through semantic web based annotation.

In [3], an ontology is developed to integrate information of different learning management systems with the purpose of extracting a personalized social network from the corresponding ontology.

Another field of research in semantic social networks, which is rooted at knowledge management disciplines, is dedicated to the concept of emerging knowledge [16]. The focus here is to see a social network as an environment in which some form of knowledge is produced during the time. It is interesting to employ knowledge management technologies and processes to extract, maintain and evolve this knowledge. In [6], the authors have modeled semantic social networks in the context of knowledge management, and have proposed techniques to enable queries on the emerging knowledge of the social network.

Current paper, describes a semantic social network named *PAD*. It is in different ways associated with the semantic web: First it publishes parts of the social network data as Linked Data. Further, it uses semantic web sources to automatically augment social network entities with machine-interpretable annotations. Finally it uses the semantic web sources to collect required background or domain knowledge for the purpose of recommending items to the social network users.

III. PROPOSED ARCHITECTURE

PAD is developed with the goal of creating a social network for academic environments to help users easily connect with each other and share their digital knowledge resources. It provides functionalities similar to all traditional social networks, e.g. adding friends, sharing contents, writing posts and sending messages. In addition, it has a functionality which is much demanding for an academic environment. Actually, *PAD* supports full-text indexing of the files uploaded by the users. As a result, it is possible to search in the social network for all the files which contain a special keyword or expression.

Overview of the architecture of *PAD* is shown in Fig. 1. It can be considered to have two logical layers: application layer and data layer. The application layer includes three main components:

1) A web-based frontend which provides the interface through which end-users interact with *PAD*. This frontend is implemented in PHP and is based on *ELGG*[7] social network engine which provides most required facilities for developing a social network.

2) A Java-based backend which provides core functionalities of the system:

 a) Indexing and searching files uploaded into *PAD*. For this purpose, *Lucene*[8] is used which is de facto standard for full-text indexing. When a new file is uploaded in the frontend, it is added to an index queue, which from an implementation point of view is a table in the database. Then, whenever the administrator runs the backend, it processes the index queue by indexing its files and removing them from the queue. Further, whenever a search is performed in the frontend, a call is made to the backend to perform the search on the *Lucene* indexes and return the results back to the frontend.

 b) Functionalities related to the Semantic web. These functionalities are described in Section IV.

3) *JavaBridge*[9] which enables interaction of the frontend with backend. Actually, it provides a means of calling backend Java code from frontend PHP code, for instance for searching indexes. Since *JavaBridge* is a Java web application, it needs to be deployed on a *Servlet*

---

[4] http://www.foaf-project.org/
[5] http://www.informatik.uni-trier.de/~ley/db/
[6] www.linkedin.com

[7] http://www.elgg.org
[8] http://lucene.apache.org/java
[9] http://php-java-bridge.sourceforge.net

container like *Apache Tomcat* [10]. Further the backend which is in fact a .jar file is deployed in *JavaBridge.*

The data layer includes sources which provide or store data of the social network. There are four essential components in this layer:

1) A *MySQL* database which stores operational data of the system. It is mainly used by the web based frontend for storing entities created or updated in the social network.

2) A set of *Lucene* indexes which store indexes created over the files uploaded in the social network. *PAD* supports multiple indexes to increase indexing performance. The number of indexes is configured by the administrator.

3) A semantic repository which stores semantic representation of the entities created in the social network. This repository provides a *SPARQL* endpoint through which it is possible to retrieve stored information using *SPARQL* queries. In Current implementation of *PAD, OpenLink Virtuoso*[11] (open source version) is used as the semantic repository which is an established product in the *Linked Data* domain.

Semantic web sources are used mainly for two purposes: 1) a number of well-known ontologies from the semantic web are used to generate semantic representation of the social network entities, and 2) some semantic web sources, e.g. datasets from the *Linking Open Data (LOD)*[12] cloud are used to annotate social network entities to enrich their semantic. These semantic web aspects of *PAD* are described in more details in the next section.

IV.   SEMANTIC WEB ASPECTS

Since *PAD* is considered a semantic social network, it is from different aspects related to the semantic web. In this section, these aspects are described.

A.   Linked Data publishing

There are different types of entities in the social network: users, categories, files, wall posts, bookmarks, tags, groups. *PAD* publishes some metadata of these entities based on *Linked Data* principles [1]. There are a number of prerequisites for this:

1) It is required to assign a unique identifier of the form URI to each entity. Since *ELGG* framework assigns a *Globally (system-wide) Unique Identifier (GUID)* to each entity, a simple URI schema is used to assign URIs to entities:

http://pad.um.ac.ir/entities/{entity_GUID}

2) It is required to assign an ontology predicate to each attribute of an entity which needs to be included in the semantic representation of that entity. Further a predicate is required to specify *rdf:type* value for that entity. Since structure of each entity, except for entities of

type file, is known at the development time, it has been possible to hardcode corresponding ontology predicates for these entity types. As an example, for wall posts, the predicates are mainly chosen from *SIOC* [2] and *Dublin Core*[13] ontologies, e.g. subject of the post is specified by *dc:subject* and content of the posts are specified by *sioc:content.* There is an issue regarding entities of type file. To semantically organize content, each file uploaded in *PAD* is forced to be uploaded in a specific file category, but file categories along with their attributes are not known at development time and they are defined by the administrator at runtime. As a result, it is the responsibility of the administrator to declare required ontology predicates for the category he defines and also for its attributes. For instance, when creating a category named *'eBook'* it is possible to declare *foaf:Document* or *dcterms:BibliographicResource* predicates to be used for specification of *rdf:type* of the entities of this category, and also assign *dc:title* predicate to the *title* attribute of this category. If no predicate is assigned to an attribute, that attribute is ignored when generating the semantic representation of the entities of the corresponding category. This requires the administrator to be expert at the semantic web domain. Although this might be considered as a critique, but our point of view is that it is a reasonable requirement for administration of a semantic social network.

3) A transformation process is required to generate *Linked Data* representation of the social network entities based on the URIs and predicates specified in the two previous steps. This process is implemented in the Java-based backend. Whenever an entity is created or modified in the frontend, the backend is called through *JavaBridge* to produce the corresponding *Linked Data* representation.

Since producing links to other resources is an important issue in *Linked Data*, it is important to produce these links. Currently, linking logic is hardcoded in the backend implementation. For instance, if the entity is a user, *foaf:knows* predicate is used to create links between that entity and entities corresponding to his friends in the social network. Further his birthplace is linked to the corresponding resources in *DBpedia*[14] and *GeoNames*[15] datasets.

As mentioned before, file categories and their attributes are defined at runtime, which means part of the schema of the social network is defined at runtime. Therefore, linking logic of these entities can only be partially hardcoded in the development time. For instance, it is possible to link each file entity to its owner entity using *dc:creator*, since it is clear at development time that each file has an owner. However, if an attribute named *'publisher'* is defined for a category named *'eBook'*, then for a file in this category, it is not possible to link its publisher attribute to a resource in an external source, since it is not specified how to find the corresponding external resource.

[10] http://tomcat.apache.org
[11] http://virtuoso.openlinksw.com/
[12] http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

[13] http://www.dublincore.org
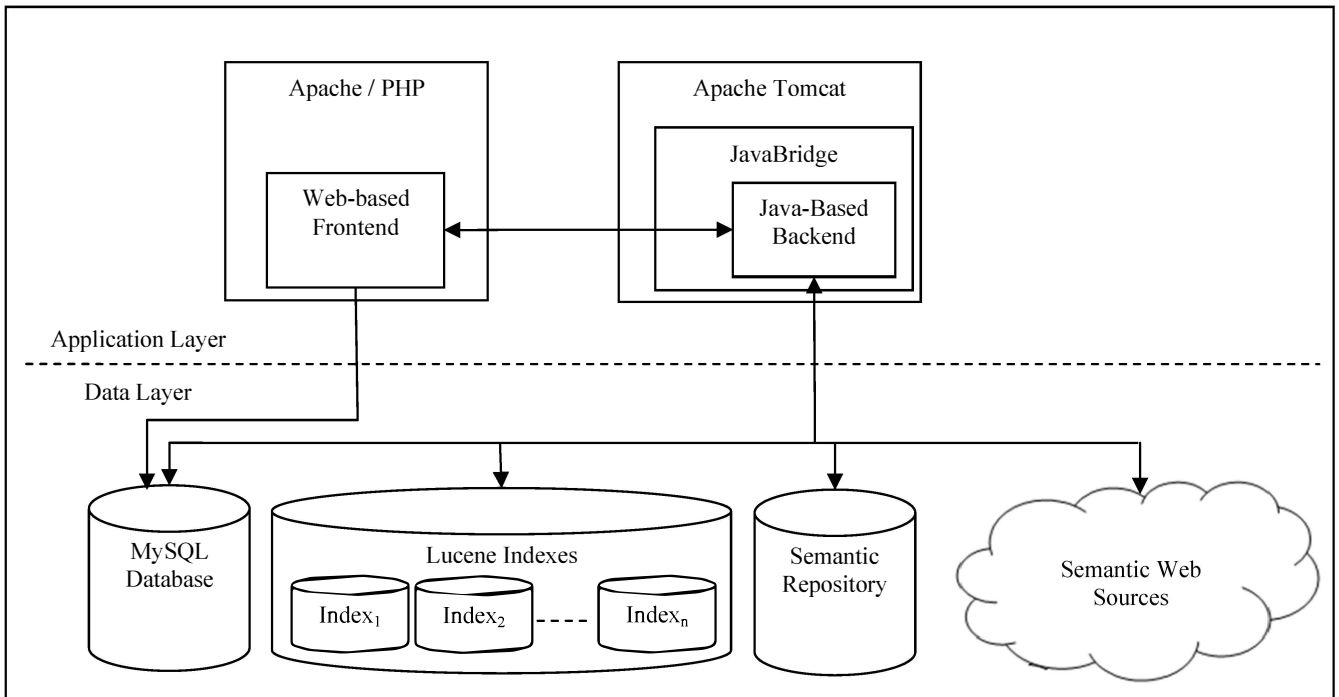[14] http://www.dbpedia.org
[15] http://www.geonames.org

Figure 1. Overview of architecture of PAD

Since each entity, except for *user* entities, has an owner which is a user, different entities owned by a user are interlinked to each other through their common owner. These semantic internal and external links make it possible for software agents to automatically browse entities of the social network and collect information by moving from one entity to another one.

B. Inference Capability

One of the major advantages of the semantic web data model is its inherent support for reasoning over data. In other words, the use of ontologies for specifying schema of data and its semantic provides capability of ontological inference. This benefit has been utilized in *PAD*.

The parser which is responsible for publishing an entity as *Linked Data* uses a local rule base which is in fact a textual file including a set of Jena rules. The parser loads these rules and uses them to infer new triples. For the sake of brevity, details of the process are ignored here, but one of the corresponding rules is shown in Fig. 2. It can be described as *"if entity1 and entity2 are two persons and the web address of their workplaces are the same, then they know each other since they both work for the same organization'*.

This inference capability has reduced cost and complexity of implementing the parser, since it makes it possible to implement some functionality of the parser by simply declaring rules in a textual file, instead of hardcoding the inference logic in the source code of the parser. Further, it improves flexibility of the parser because it is possible to modify its functionality at runtime, simply by modifying the rule base.

C. Resource Recommendation

Semantic web sources, for instance the datasets on the *Linking Open Data (LOD)* cloud provide a rich machine interpretable web scale data space. This data space can be used to help users of the social network by providing their required information. For instance, when a user performs a search on the social network content, a list of more specific and more general keywords are provided for him to make his search narrower or broader. This is performed by sending predefined *SPARQL* queries to some *Linked Data* sources like *DBpedia* which provide categorization information, for instance using *SKOS* ontology[16]. As an example, if the user searches *"Semantic Web"*, then some of the terms recommended as broader keywords are *World Wide Web, Knowledge Representation, Data Management, Internet Architecture*, and some of the terms recommended as more specific keywords are *Semantic HTML, Microformats, Triple Stores, Folksonomy, Knowledge bases*. A similar facility is also used in *PAD* for tag recommendation purpose.

```
[rule1: (?entity1  rdf:type  foaf:Person),
    (?entity2  rdf:type  foaf:Person),
    (?entity1  foaf:workplaceHomepage ?homepage),
    (?entity2  foaf:workplaceHomepage ?homepage) ->
    (?entity1  foaf:knows  ?entity2),
    (?entity2  foaf:knows  ?entity1)]
```

Figure 2. A simple Jena rule used by the parser

---

[16] www.w3.org/2008/05/skos

D. *Contribution to the Semantic Web*

PAD contributes to the semantic web data space by exposing parts of its content as *Linked Data*. More specifically, semantic representation of the metadata of each entity is both stored in a repository and also published as *RDFa* inside associated HTML web pages at the presentation layer.

The semantic repository provides a *SPARQL* endpoint which is used to retrieve information of the social network entities through query execution. This has the benefit of making metadata of the entities open to and shared with other applications, even third party applications. Since in the semantic representation, semantic of data is explicitly specified in a machine-interpretable format, the repository is an application-independent and hence shareable view of the social network information. Further, embedding *RDFa* snippets inside HTML pages has the benefit that currently many semantic web search engines, e.g. Watson [17] and Sindice [18], and even general web search engines, e.g. *Google* or *Yahoo*, are interested in gathering *RDFa* from web pages and using them for improving quality of their search service.

V. EXPERIMENTAL RESULTS

The proposed architecture of *PAD* is currently implemented and deployed at *Ferdowsi University of Mashhad*. Since it is still in its early days, it is not possible to fully evaluate different aspects of *PAD*. As time goes by and more users get involved, more potential will be created for assessing the ideas and techniques employed.

Table 1 shows current [19] statistics about some types of entities in *PAD*. Further, Table 2 provides some information about the full-text index of the social network.

As described in Section IV, some information of the entities of *PAD* is published as *Linked Data*. The resulting RDF dataset includes a number of RDF links to external resources on *LOD*. For instance, birthplace of the users are linked to corresponding entities in *DBPedia* and *GeoNames*, and also educational information of the students and professors are linked to *DBPedia* entities. Table 3 provides some information about this process of *Linked Data* publishing.

TABLE I.   SOME STATISTICS ABOUT ENTITIES IN PAD

Entity Type	Number of Entities
User	2239
Group	31
File	725155
File Downloads	42429
Message	35932
Basic Search	26846
Advanced Search	2445

[17] http://watson.kmi.open.ac.uk/WatsonWUI/
[18] http://sindice.com/
[19] Statistics correspond to October 25, 2013

TABLE II.   SOME STATISTICS ABOUT FULL-TEXT INDEX OF PAD

Number of indexed files	718433
Total size of indexed files (TB)	7.12
Average indexing speed (MB/S)	3.18
Average search time (S)	0.78

TABLE III.   SOME STATISTICS ABOUT LINKED DATA PUBLISHING

Number of triples related to users	8921
Number of triples related to files	4351128
Number of triples inferred (through reasoning)	1025
Number of external links	1138

VI. CONCLUSION

In this paper, *PAD* is introduced which is a semantic social network developed at *Ferdowsi University of Mashhad*. It uses the semantic web technologies in different ways: First it publishes parts of the social network data as *Linked Data*. Further, it uses semantic web sources for automatic annotation of the social network entities with resources on the semantic web. Finally it uses the semantic web sources to collect required background or domain knowledge for the purpose of recommending items to the social network users.

PAD is designed to provide a medium through which people at *Ferdowsi University of Mashhad* can share their interests and knowledge, and also to provide a native infrastructure for Iranian researchers in different domains.

In this paper, architecture of *PAD* is introduced and different aspects in which it is related to the semantic web are discussed.

Since *PAD* is in its early days of deployment, our future works include a wide range of research and development tasks, for instance:

- Providing a content recommendation widget for *PAD* based on collaborative filtering techniques.

- Proposing techniques and methods for improving automatic annotation of contents in *PAD* with resources across *Linked Data* cloud.

- Assessment of the capabilities of *PAD* in answering information needs of the research community at *Ferdowsi University of Mashhad*.

- Proposing automatic personalization techniques based on users' activities and interests.

- Using social network mining techniques to detect hidden association rules inside users' activity logs.

REFERENCES

[1] T. Berners-Lee, "Linked Data," 2006, http://www.w3.org/DesignIssues/LinkedData.html .

[2] J. G. Breslin, A. Harth, U. Bojars, and S. Decker, "Towards Semantically-Interlinked Online Communities," In The 2nd European Semantic Web Conference (ESWC '05), Heraklion, Greece, Proceedings, pp. 500-514, May 2005.

[3] M.P. Cuéllar, M. Delgado, and M.C. Pegalajar, "Improving learning management through semantic web and social networks in e-learning environments," J. Expert Systems with Applications, vol. 38, pp. 4181–4189, 2011.

[4] J. Breslin and S. Decker, "The future of social networks on the internet: The need for semantics," IEEE Internet Computing, vol.11 No.6, pp.86–90, 2007.

[5] J. J. Jung and J. Euzenat, "Towards semantic social networks," In ESWC'07: Proceedings of the fourth European conference on the semantic web, Berlin, Heidelberg: Springer-Verlag, pp. 267-280, 2007.

[6] M. Schatten, "Knowledge management in semantic social networks," Computational & Mathematical Organization Theory, pp.1-31, 2012.

[7] C. R. Pablo, S. L. Abraham, and N. R. David, "Moveek: A semantic social network," WILE 2012 (Fifth Workshop on Intelligent Learning Environments), pp.1-10, 2012.

[8] L. Razmerita, M. Jusevicius, and R. Firantas, "New Generation of Social Networks Based on Semantic Web Technologies: the Importance of Social Data Portability," In Workshop on Adaptation and Personalization for Web 2.0, UMAP'09, pp. 22-26, 2009.

[9] L. Reeve and H. Han, "Semantic Annotation for Semantic Social Networks Using Community Resources," AIS SIGSEMIS Bulletin, Vol. 2, Issue (3&4), pp. 52-56, 2005.

[10] G. Erétéo, F. L. Gandon, O. Corby, and M. Buffa, "Semantic social network analysis," In Proceedings of CoRR, 2009.

[11] U. Bojars, A. Passant, J.G. Breslin, and S. Decker, "Social Networks and Data Portability Using Semantic Web Technologies," The 2nd Workshop on Social Aspects of the Web (SAW 2008) at the 11th International Conference on Business Information Systems (BIS 2008), Innsbruck, Austria, pp.1-15, 2008.

[12] J. J. Jung, "Integrating Social Networks for Context Fusion in Mofile Service Platforms," Journal of Universal Computer Science, vol. 16, no. 15, pp. 2099-2110, 2010.

[13] I. Cantador and P. Castells, "Multilayered semantic social network modeling by ontology-based user profiles clustering: Application to collaborative filtering," In S. Staab & V. Svtek (Eds.), EKAW, Lecture notes in computer science, Vol. 4248, pp. 334–349, 2006.

[14] P. Mika, "Flink: Semantic Web technology for the extraction and analysis of social networks," Web Semantics: Science, Services and Agents on the World Wide Web, vol. 3, pp. 211-223, 2005.

[15] M. ALLEN, "What was Web 2.0? Versions as the dominant mode of internet history," New Media & Society, vol. 15 no. 2, pp.260-275, 2013.

[16] G. V. Krogh, "How does social software change knowledge management? Toward a strategic research agenda," The Journal of Strategic Information Systems, vol.21 no.2, pp.154-164, 2012.

[17] S. Bortoli, T. Palpanas, and P. Bouquet, "Decentralised social network management," International Journal of Web Based Communities, vol.7 no.3, pp-276-297, 2011.

[18] S. Kinsella, J. G. Breslin, A. Passant, and S. Decker, "Applications of Semantic Web Methodologies and Techniques to Social Networks and Social Websites," Lecture Notes in Computer Science, vol.5224, pp.171-199, 2008.

[19] Y. Wang, "A User-Centric Approach for Social Data Integration and Recommendation," 3rd International Conference on Human-Centric Computing (HumanCom), pp.1-8, 2010.