World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:7, No:4, 2013

Enhanced Conference Organization Based On Correlation of Web Information and Ontology Based Expertise Search

Hassan Noureddine, Maria Sokhn, Iman Jarkass, Elena Mugellini, and Omar Abou Khaled

***Abstract*—**From the importance of the conference and its constructive role in the studies discussion, there must be a strong organization that allows the exploitation of the discussions in opening new horizons. The vast amount of information scattered across the web, make it difficult to find experts, who can play a prominent role in organizing conferences. In this paper we proposed a new approach of extracting researchers' information from various Web resources and correlating them in order to confirm their correctness. As a validator of this approach, we propose a service that will be useful to set up a conference. Its main objective is to find appropriate experts, as well as the social events for a conference. For this application we us Semantic Web technologies like RDF and ontology to represent the confirmed information, which are linked to another ontology (skills ontology) that are used to present and compute the expertise.

***Keywords*—**Expert finding, Information extraction, Ontologies, Semantic web, Social events.

I. Introduction

TODAY, the Web plays a major role in the interaction between people and communities. Gradually, the world moves all its activities to this global space. From the beginning, the scientific community benefited from the web like other communities, and now uses it mainly to activate the cooperation between researchers and to exchange information between them. So we can find vast quantities of scientific information as researchers, projects, papers... With this enormous amount of data, the automatic or semi-automatic applications become necessity, especially to find appropriate information in a brief time. Therefore, the Web becomes saturated by these applications in all domains.

Despite this, the problem has not been fully solved with the presence of a large amount of conflicted and outdated information. For that in order to get more accurate and ranked results, it was interesting to correlate this information, especially with the existence of information for the same subjects from multiple sources. This is typically useful during setting up a conference, when we need to find information for relevant experts and ranking them depending on their expertise, as well as finding and proposing social events for the conference.

We aim to demonstrate our approach of extracting and correlating information from multiple Web resources within a system that have objective to find appropriate reviewers and propose social events for a conference in a specific domain. In this paper, we present the previous work that address the issues of researcher information extraction, profiling and expert finding, and then we introduce the Framework of our system that provides the mentioned service through exploiting of the correlated information and the semantic Web technologies.

The rest of the paper is structured as follows: in Section II we review the related work on researchers' information extraction and expert finding issues, and discuss the result of these works. In Section III, we propose our system framework design, and describe the scenario of future work. Section IV presents the initial steps in the framework implementation. Conclusion and future works are in Section V.

II. Related Work

Since the main task of our work is extracting and correlating researchers' information in order to obtain a list of ranked experts. Therefore it was necessary to review several researches in this area that have been carried out in the last years.

The expert finding systems have been proposed often within the organizations as a solution to users' problems, who wish to use this expertise knowledge or find a specific expert to perform a certain task. To reach this goal, it is necessary to achieve the task of profiling and social network extraction. From these applications we mention Referral Web [1], Agilience (http://www.agilience.com), BuddyFinder [2], DemonD [3], and SmallBlue [11]. In these applications, the expertise is inferred using keywords extracted from web pages, shared documents, email and instant message transcripts. The social network is also determined from the co-occurrence of names on publications or emails.

Furthermore, there are more directions and efforts towards automating the expert finding process, so the systems went towards improving the presentation of knowledge in their databases and enhancing the expert finding process using the

Hassan Noureddine is with the University of Applied Sciences of Western Switzerland, Fribourg, Switzerland and the EDST, Lebanese University, Beirut, Lebanon (e-mail: Hassan.Noureddine@edu.hefr.ch).

Maria Sokhn is with the Institute of Business Information Systems, University of Applied Sciences Western Switzerland, Sierre, Switzerland. (e-mail: Maria.Sokhn@hevs.ch).

Iman Jarkass is with LSI Department, Institute of Technology IUT, Lebanese University, Saïda, Lebanon. (e-mail: ijarkass@ul.edu.lb).

Elena Mugellini and Omar Abou Khaled are with the ICT Department, University of Applied Sciences of Western Switzerland, Fribourg, Switzerland (e-mail: Elena.Mugellini@hefr.ch).

Semantic Web technologies, as in Semantic Scout project and the Semantic Web based approach to expertise finding at KPMG [16], [17]. Now with the presence of the Semantic Web technologies, the process of expert finding was expanded to a wider scale outside the organizations, on the Web with an enormous quantity of data. We present several works carried out using different Web resources.

The most used source in these works is the publications. The VIKEF project [10] uses several collections of papers to construct the profiles of researchers participating in ISWC 2004. DBLPVis [6] uses the publications in DBLP database to search for the relations between different entities. Also AEFS [18] uses the citations of the publications as experts' profiles to rank the experts, and EFS [19] uses the experts' publications as the materials to build their expertise, in addition they use the link structures of Wikipedia to improve the expertise.

We have seen several projects benefit from publications in several ways, and that depend on the application. For this reason we can see also other systems that benefit from multiple sources. In addition to the publications, Flink [4] uses web pages, emails and FOAF profiles [5] to extract the Semantic Web researchers' social networks. Arnetminer [8], [9] uses the home pages to create a semantic-based profile for each researcher and then use them with publications to compute their expertise. And now with the conversion of the existing data on the Web largely in the form of RDF, we find systems that use it as main sources, especially as it presents the relationship between entities. For instance RKBExplorer [7] present unified views of a significant number of heterogeneous data sources (triple stores) regarding a given domain.

A. Discussion

In the mentioned work, ontologies and Semantic Web technologies have proved their efficiency in presenting the researchers domains. We also see the systems operating on the web are benefiting from more sources such as Arnetminer and RKBExplorer obtain more comprehensive and accurate results. Even so, we find incorrect or incomplete results in certain cases due to the conflict and outdated information. This is what motivates us to use all practical sources in order to correlate the extracted information from them. We benefit also from new sources like social networks, as in the algorithm applied for finding experts in Friendfeed [12], and sources have not been used so far like videos and images databases.

On the other hand the expert finding process in most systems is based on the co-occurrence of query keywords in the used sources. Furthermore to improve the result, they extend this process on the expert propagations in their social networks [13]-[15]. Noting that, they mainly use of publications (co-author relationship) within Web is to extract the researchers' social networks. The extracting of new relationships can improve the results, and that what we intend to do in our work using ontology and rules to extract and present these relations. And recently, a new method has emerged in computing the expertise using skills ontology as in [20], which this ontology present the relations between researchers domains and their hierarchy. This ontology was linked to a first one that presents the researchers profiles and their relations in a new method of expert finding and ranking.

III. Framework

In this section, we introduce the preliminaries of our work and describe its architecture overview (Fig. 1).

So far, the studied approaches didn't give optimal results, especially when they take into account a very large base of researchers. But they evolve gradually, trying to overcome the existing problems on the Web, as already described in the discussion section. The proposed approach aim to enhance the two processes of acquiring information and expert finding. The first is achieved by correlating the extracted information from various Web resources and the second by treating this information with relations between researchers and relations between domains. This work is done with the support of the Semantic Web technologies.

A. Scenario

In order to apply these objectives, it is interesting to implement our proposed approach into a system that provides a significant service to help organizers of conferences like proposing a list of ranked experts in a certain domain. In addition, the system is capable to propose social events for making the service more significant.

The proposed scenario to be applied through our system starts by entering information about user's request across the interface, including the scientific domain and the information about the conference location, date, number of participants and halls. After this request the system begins extracting information from heterogeneous sources from the Web. It constructs the researchers' profiles, and then uses these profiles into the new expert finding process. Finally it provides a list of ranked experts. On the other hand the system uses the distributed information along the web to propose the social events depending on the user request. At this stage the organizer can choose their relevant choices in order that the system sends invitations for experts.

B. Architecture

The architecture through which the scenario will be applied is shown below in the Fig. 1. In the right side, the input and output of the system are shown. As mentioned in the previous section, the input is a query including research domain and social events information, and the output is a list of ranked experts with the ability to access their profiles, in addition to a list of the proposed social events.

The system sources are shown in the left side, which are composed of two parts: Web resources for researchers' information and Web pages for social events' information. The Web pages are chosen through a search engine using the query depending on the user request. In regard to the researchers' Web resources, it is composed of several types of sources:

1) Publications: Paper for researcher, which are stored in

several databases (e.g. DBLP, CiteSeer and Google Scholar).

2) Emails: Public collection of emails that show the interaction between researchers.
3) Home pages and projects: Distributed homes pages for researchers on the Web, as well as the projects pages that show information for researchers and projects.
4) Videos and images: Videos and images that show the activities of researchers (e.g. lectures and conferences).
5) Social network activities: Activities and interactions of researchers with their social networks (e.g. Facebook and twitter).
6) Semantic web sources: FOAF profiles and RDF (triples stores) from the Semantic Web.

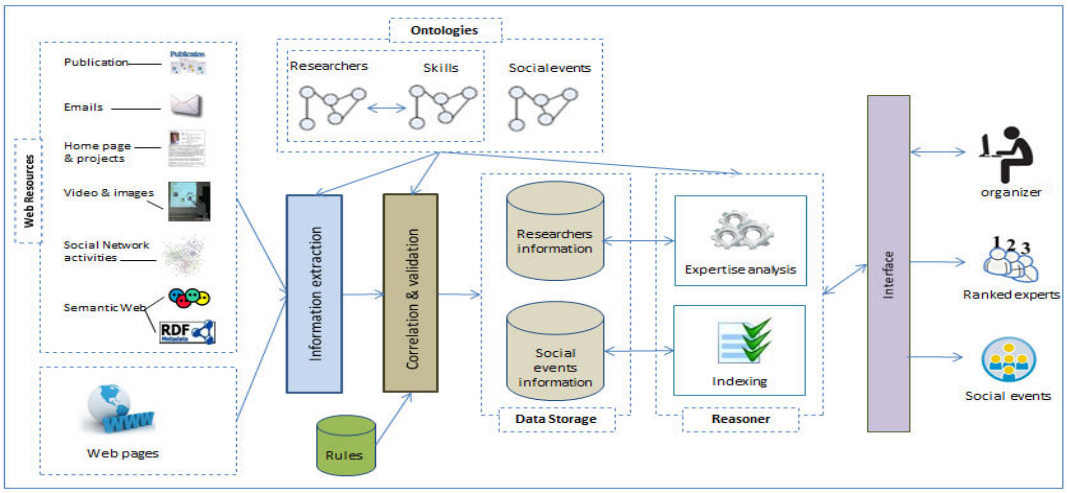

Fig. 1 Our framework global view

In the middle, the first block applies the process of information extraction from the defined sources, according to concepts and properties presented by the three ontologies on the top. Ontology for social events, it presents the relevant properties of desired social events and it is independent from two other linked ontologies, one for the researchers' information, which includes profiles' information and different relations with other researchers (researches' propagation), and the other includes scientific domains depending on their relations and hierarchy (skills ontology). Now the principal publications databases are integrated into RKBExplorer databases in form of triples (RDF), which makes it machine-readable and easier to use. This also applies on several FOAF profiles databases that are used as semantic web sources. The information from textual sources such as emails and home pages are extracted using gate library as detailed in the next section. The extraction can be applied on the textual parts in the social networks sites, in addition to advantages taken from the graph information of these sites. Finally, the extraction from videos and images databases (e.g. flicker and video lectures) is divided on two parts, extraction from metadata and extraction from images content.

The second block applies the correlation between extracted information according to several rules that indicates the priority of each source. In addition to other rules that determine the likelihood of any information through specific characteristics and rules for comparison between repetitive information from multiple sources. And then the validation comes to validate the relevancy of the correlated information to the ontologies' components (concepts, properties and relations). The validation process depends on several rules to associate the information to their relevant components.

In the next block, the system saves separately the confirmed information by correlation and validation as shown in Fig. 1. In this case the researchers' information includes experts' local information, publications, social networks (relations), as well as videos and images for their scientific activities. On the other hand, social networks' information includes hotels, halls and touristic sites.

The final block in the process is the reasoner. It computes researchers' expertise, taking into account the expert propagation from researchers' ontology and the domains' classification from skills ontology as a new method. On the other hand, it use the social events' information to index them depending to the user request and finally present the results through the interface.

IV. Experimental Setup

After this general view of the Framework, we start implementing the issue of ontology based information extraction by applying relevant methods of extraction from each source separately to evaluate the results and give each source its appropriate degree or coefficient.

The beginning was in the selection of three heterogeneous sources in order to extract different information according to a simple ontology. The information that we want to extract are: Personal data (name, address…), graduate certificates, workplace, hobbies, relations with others, photos and events performed by the user as a conference. It is possible later to add other information to be extracted using ontology, and that is the aim of this preliminary stage. From which we will enrich the ontology by inferring new constraints (concepts,

properties and relations), and also improving the process by inferring new rules for correlation and comparison, in addition to the determination of the relationship between extracted entities.

Practically, Gmail, Facebook and Flickr were the three chosen sources to apply the mentioned process. First we download the personal data from the user profiles. With this step we get the information introduced at three different times and servers, allowing us to compare and validate it based on several rules. There is other information, such as relations with others or events performed by user may not exist on the structured data despite its great importance. That prompts to infer and extract them from unstructured data like unstructured text from home pages or emails. For this task, we implement an algorithm of extraction from Gmail's text messages using gate library (Fig. 2).

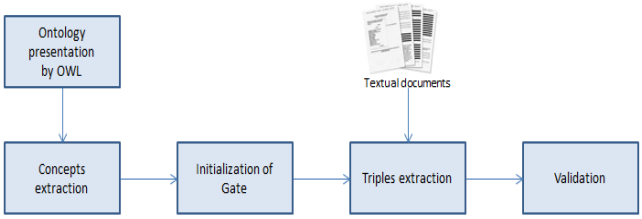

Fig. 2 Extraction algorithm from textual documents

The first block presents researchers' ontology using protégé tool. In the second block, the Java library « Jena 2.0 » is used to extract the different classes of ontology for using them in the extraction. The initialization of the tool "gate" is done in the third block. Gate is a popular tool for information extraction used by scientific communities. It uses linguistic rules and relations to extract scientific information in form of triples (Subject, predicate, object). Therefore in the fourth block, it take a textual document as input (Gmail's text message in this case), to begin the process of triples extraction through several modules. Starting with Tokenizer, passing by Sentence Splitter, Part Of speech Tagging, Gazetter, Named-entity recognition, Coreference resolution, Dependency graph and ending with Triples Extraction. After this step, we obtain number of triples, which only part of them related to the ontology domain. In the last block, each triple t = [Sub, Predi, Object] undergo a validation through an algorithm (Fig. 3), which is composed of several rules applied progressively. If there is a rule consistent with the triple then it is considered relevant to the ontology, else it is discarded.

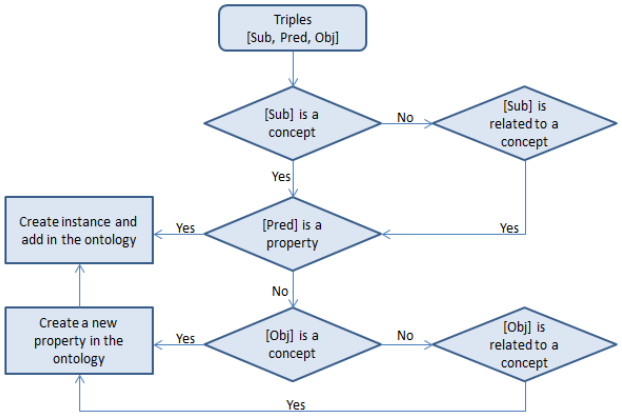

Fig. 3 Validation algorithm

We test this algorithm using corpus taken from several homes pages and emails messages. This method provides repeated information that has already been extracted from other sources, and this enriches the correlation process. On the other hand it provides more comprehensive information, but it does not provide all the desired information. Furthermore, we continue with extracting other kind of documents such as photos, in order to extract semantic information from them.

After the extraction process, the information must be saved on relevant database that achieve quick and accurate search. Orient Db database was used for this task as NoSQL database. It supports the "Graph databases" and "document databases" in order to save files and triples, and then deploy them to a Linked Data System.

Finally, an Interface was performed for presenting the extracted information from different sources (Fig. 4).

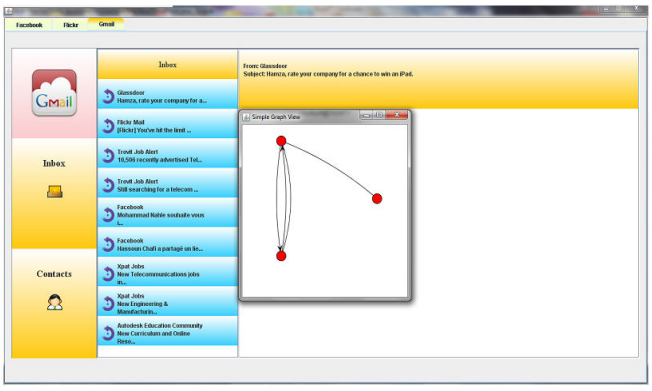

Fig. 4 Example of our system interface

V. Conclusion

In this paper, we have proposed a new approach of ontology based information extraction from various types of Web resources in order to correlate them and confirm their correctness. The scientific domain is a relevant area for demonstrating this approach, with the enormous quantities of discarded information along the web. Therefore we have described the framework of expert finding system through which we aim to validate this proposed approach, and

continue with expert and social events ranking processes as a service to help conferences' organizers.

As shown, the last section has described the implementation of the framework in its initial steps, in which we extracted several researchers' information from multiple sources according to a determined researchers' ontology. The obtained results demonstrate the importance of the correlation process in confirming the correctness, especially in the presence of repeated information from multiple sources. This confirmed information provide robust platform for expert finding process. Therefore the future steps consist of inferring and presenting the final form of ontologies and rules for completing the extraction process from all defined sources.

REFERENCES

[1] H. KAUNTZ and B. SELMAN. Referral web: Combining social networks and collaborative filtering. Communications of the ACM, 403, 63-65.
[2] J. ZHU and M. EISENSTADT. Buddyfinder-corder: Leveraging social networks for matchmaking by opportunistic discovery. ISWC2005, Galway Ireland.
[3] C. Delalonde and E. Soulier. Collaborative Information Retrieval in R&D distributed teams. 13th International Conference on Concurrent Engineering, Sophia-Antipolis.
[4] P. Mika. Flink: Semantic Web technology for the extraction and analysis of social networks. Journal of Web Semantics (2005).
[5] D. Brickley and L. Miller: FOAF Vocabulary Specification. FOAF Project, http://xmlns.com/foaf/0.1/. (2004).
[6] Remo Lemma. Ebon: Visualizing the DBLP Database (June 17, 2010).
[7] Glaser, H., Millard, I.C.: RKBPlatform: Opening up Services in the Web of Data. In: International Semantic Web Conference (2009)
[8] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su. ArnetMiner: Extraction and Mining of Academic Social Networks. Proc. of 14th Intl. Conf. on Knowledge Discovery and Data Mining (SIGKDD 2008). Henderson, Nevada, 2008, pp.990-998.
[9] Juanzi LI, Jie TANG, Jing ZHANG, Qiong LUO, Yunhao LIU, Mingcai HONG. Arnetminer: expertise oriented search using social networks. Frontiers of Computer Science in China, 2008: 94~105.
[10] The VIKEF Consortium, VIKEF Technology Catalogue (January 2007).
[11] C-Y. Lin, N. Cao, S. X. Liu, S. Papadimitriou, J. Sun and X. Yan. SmallBlue: Social Network Analysis for Expertise Search and Collective Intelligence. IEEE International Conference on Data Engineering.
[12] A. Kardan, A. Omidvar and F. Farahmandnia. Expert Finding on Social Network with Link Analysis Approach. 19th Iranian Conference on Electrical Engineering (ICEE), 2011 Page(s): 1- 5.
[13] J. Tang, J. Zhang, D. Zhang, L. Yao, C. Zhu and J. Li .ArnetMiner: An Expertise Oriented Search System for Web Community.
[14] J. Zhang, J. Tang, and J. Li. Expert Finding in a Social Network. In: DASFAA 2007. LNCS, vol. 4443, pp. 1066–1069. Springer, Heidelberg (2007).
[15] E. Smirnova. A Model for Expert Finding in Social Networks. Proceeding SIGIR '11 Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval Pages 1191-1192 New York, NY, USA ©2011. C.
[16] Baldassarre, E. Daga, A. Gangemi, A. Gliozzo, A. Salvati, and Gianluca Troiani. Semantic Scout: Making Sense of Organizational Knowledge. EKAW2010.
[17] E. A. Jansen. A Semantic Web based approach to expertise finding at KPMG.(2010)
[18] C. C. Chou, K. H. Yang, and H. M. Lee. AEFS: Authoritative Expert Finding System Based on a Language Model and Social Network Analysis.(2007)
[19] K. H. Yang, C. Y. Chen, H. M. Lee, and J.M. Ho. EFS: Expert Finding System based on Wikipedia Link Pattern Analysis. 2008 IEEE International Conference on Systems, Man and Cybernetics (SMC 2008).
[20] R. Punnarut and G. Sriharee. A Researcher Expertise Search System using Ontology-Based Data Mining. Seventh Asia-Pacific Conference on Conceptual Modelling (APCCM 2010), Brisbane, Australia, January 2010.