Web Mining Approach for a User-centered Semantic Web

Junichiro Mori[1,2], Yutaka Matsuo[2], Koichi Hashida[2], and Mitsuru Ishizuka[1]

[1] University of Tokyo, Japan
jmori,ishizuka@miv.t.u-tokyo.ac.jp
[2] National Institute of Advanced Industrial Science and Technology (AIST), Japan
y.matsuo@carc.aist.go.jp, hashida.k@aist.go.jp

Abstract. In this paper, we propose a Web mining approach for the Semantic Web. The approach uses a search engine and the traditional web as a source of information to produce semantically rich information. In particular, we assess one community and obtain the social network and related information from the Web. As an example, we extract the social network of an academic society and show that extracted information can be incorporated into FOAF representation and utilized to measure the authoritativeness of a member in terms of social trust or individual trust. To demonstrate our Web mining approach in the real application, we show a researcher mining and retrieval system. Finally, we discuss the manner in which the Web mining approach contributes to availability to users of the Semantic Web.

1 Introduction

The Semantic Web [2] is designed to let users make explicit statements about any resource, and maintain that data themselves in an open and distributed manner. Several standards such as the Resource Description Framework (RDF) [18] and Web Ontology Language (OWL) [19] have been developed to realize the layer cake of the Semantic Web.

From the viewpoint of end users, expressing semantics about people and their relationships has garnered considerable interest. The Friend of a Friend (FOAF) project [4] is an extremely popular ontology of the Semantic Web [6]. It is essentially a vocabulary for describing people and whom they know. The FOAF ontology is not the only one people use to publish social information on the Web. For example, it is reported that more than 360 RDF Schema or OWL classes are defined with the local name "person" [1]. In fact, many vocabularies for user semantics have been developed [20, 5, 12].

Supported by these user-side ontologies, users are gradually coming to adopt Semantic Web technologies both explicitly and implicitly. For example, in Weblogs, which are diary-like sites, users attach a FOAF profile to a Weblog and publish various contents by the RDF site summary (RSS). Some social networking sites that allow users to maintain an online network of friends associates for social or business purposes publish their users' social network data in FOAF format. Approaching the top of the Semantic

[1] http://swoogle.umbc.edu

Web layers, calculation of a “Web of Trust” on a FOAF-based network is also proposed [10].

Users are beginning to accept FOAF and its extensions as something of a standardized ontology for representing user semantics on the Semantic Web. While some users are explicitly authoring their FOAF files, others use FOAF file that systems automatically create using their Web pages. In fact, considering the personal information that the FOAF vocabulary expresses, we find that much information is contained in the traditional Web. For example, imagine a researcher: that researcher’s information might be in an affiliation page, a conference page, an online paper, or even in a Weblog. A method that can process the vast amount of information on the current “non-semantic” Web and can thereafter produce semantic information would facilitate and accelerate the use of the Semantic Web. For example, reusing existing sources of information on the Web would solve semantic annotation problems by helping users to create their metadata.

In this paper, we propose a Web mining approach for the Semantic Web. The approach uses a search engine and the traditional web as an information resource to produce semantically rich information. In particular, we examine one community and extract its social network and related information from the Web. As an example, we infer the social network of an academic society and show that extracted information can be incorporated in FOAF representation. It can then be used to measure the authoritativeness of a member as social trust or individual trust. To demonstrate our Web mining approach in an actual application, we show a researcher mining and retrieval system. Finally, we discuss how the Web mining approach contributes to user aspects in the Semantic Web.

The remainder of this paper is organized as follows: section 2 describes the proposed Web mining method and its application. Section 3 presents discussion of the Web mining approach for user aspects in the Semantic Web. Section 4 shows a comparison of our method with related works. Finally, we conclude this paper in section 5.

2 Web Mining Approach for the Semantic Web

This study specifically addresses one community and obtains the social network and related information from the Web. One reason for focusing on a community is that we believe that a huge “Web of Trust” over the entire Web comprises the superposed local “Webs of Trust” in each community to which a person or an organization belongs to.

Numerous communities exist in the physical world and online. We specifically examine an academic society: Japanese Society of Artificial Intelligence (JSAI). We choose JSAI because of its inherent availability of related information on the Web. Information related to this academic society in computer science is available online to a great degree. Another reason is that we are actually working mainly in JSAI so we can evaluate the extracted information. The following sections show how to automatically obtain JSAI members’ social networks and related information from the Web.

2.1 Social Network Extraction

Before extracting the social network, we choose the participants to the last four annual JSAI conferences as active members of the JSAI community. Each active member of JSAI is represented as a node in a social network. A node is labeled with the name of its corresponding person.

Next, edges between nodes are added using Web information. A simple approach to measure the relevance of two nodes is to use word co-occurrence information. Herein, we define co-occurrence of two words as word appearance in the same Web page. If two words co-occur in many pages, it is assumed that those two have a strong relation. The co-occurrence information is acquired by the number of retrieved documents of a search engine result. For example, assume we are to measure the relevance of two names "Junichiro Mori"(denoted x) and "Yutaka Matsuo" (denoted y). We first address two names $n1$, $n2$ as a query "$n1$ and $n2$" to a search engine and get $|N1 \cap N2|$ documents including those words in the text. Therein, N denotes a Web page set that includes a name n. Additionally, we make another query "$n1$ or $n2$" and obtain $|N1 \cup N2|$ matched documents. The relevance between $n1$ and $n2$ is approximated by the Jaccard coefficient $|N1 \cap N2|/|N1 \cup N2|$. If $n1$ and $n2$ have a strong relation, the retrieved documents might include $n1$'s and $n2$'s homepages, their publication pages, a laboratory's member list page, a conference program page and so on. In that case, $|N1 \cap N2|$ becomes large compared to $|N1 \cup N2|$. However, the Jaccard coefficient generally gives a famous person few edges because the denominator $|N1 \cup N2|$ is very large in comparison to $|N1 \cap N2|$. We can modify denominator $|N1 \cup N2|$ to $min(|N1|, |N2|)$, which places too much weight on a person with few edges. Therefore, the relevance of node $n1$ and $n2$ is represented by the following threshold-based Simpson coefficient:

$$R(n1, n2) = \begin{cases} \frac{|N1 \cap N2|}{min(|N1|, |N2|)} & if |N1| > k \; and \; |N2| > k, \\ 0 & otherwise \end{cases}$$

We set $k = 30$ for JSAI case. If we wish to estimate the co-occurrence more precisely to a person with small hits, we can pursue other alternatives to calculate statistical reliability. If relevance $R(n1, n2)$ of a node pair is larger than the given threshold, an edge is added with its weight equal to the relevance.

In the same manner as with the edge relation extraction, we can extract information of each node by considering the co-occurrence between the name and the term. For example, the search result of a query "Tim Berners-Lee and Semantic Web" returns about 76500 documents while about 9850 documents are returned for the query "Tim Berners-Lee and Software engineering". In this manner, we can infer that "Semantic Web" is more relevant to "Tim Berners-Lee" than "Software engineering". The term set of each node is acquired by retrieving the person's name that represents the node. Among the set, the term that often co-occurs with a person's name is chosen as his or her node keyword [2].

It is more useful to assign each edge a "label" for the relationship between two persons. For example, two nodes have the relation of "colleagues of the same research

[2] As a measure of co-occurrence, we use the Jaccard coefficient.

Table 1. Obtained rules.

Class	Rule[3]
Coauthor	SameLine=yes
Lab	(Number_of_Cooccurrence = more_than_one & Word_Group_in_Title(D)=no & Word_Group_in_First_Five_lines(A, E) = yes) or ...
Proj	(SameLine=no & Word_Group_in_Title(A)=no & Word_Group_in_First_Five_lines(F)=yes) or ...
Conf	(Word_Group_in_Title(A)=no & Word_Group_in_First_Five_lines(B)=no & Word_Group_in_First_Five_lines(D)= yes) or ...

Word groups
A: publication, paper, presentation, activity, theme, award, authors etc.
B: member, lab, group, laboratory, institute, team, etc.
C: project, committee
D: workshop, conference, seminar, meeting, sponsor, symposium, etc.
E: association, program, national, journal, session, etc.
F: professor, major, graduate student, lecturer, etc.

Table 2. Higher-ranked keywords of the "Mitsuru Ishizuka" node

Yutaka Matsuo, Hiroshi Dohi, Character Agent, Koichi Hashida, Life-like Interface
Naoaki Okazaki, University of Tokyo, Life-like Agent, Hypothetical Reasoning

institute", "professor-student", "members of the same committee", and so on. We discern the relationship by consulting retrieved page contents and applying classification rules. These rules are obtained through a machine-learning approach. We define labels for each edge as follows: *Coauthor* (Coauthors of a technical paper), *Lab* (Members of the same laboratory or research institute), *Proj* (Members of the same project or committee), *Conf* (Participants of the same conference or workshop). Each edge has multiple labels. For example, the relations might be both *Coauthor* and *Lab*. We first retrieve the top five pages returned for the query "$n1$ and $n2$". Then we extract some features from the contents of each page. We apply classification rules to the features and thereby obtain labels of the relation between $n1$ and $n2$. We employ C4.5 [16] to derive classification rules because of their ease of interpretability. Some of the obtained rules are shown in Table 1: For example, if two names cooccur in the same line, they are classified as coauthors. if the number of cooccurrences is more than one, and the title does not include word group D, but the first five line includes word groups A and E, then the relation is classified as members of the same laboratory.

Figure 1 portrays a part of the social network of the JSAI community. A node is labeled as the corresponding participant name (in Japanese), and an edge is labeled as *Coauthor*, *Lab*, *Proj*, or *Conf*. The whole network is shown in Fig. 2. We have more than 1500 people in the community from which we choose about 150 members to illustrate this network. Table 2 shows higher-ranked keywords of the node – "Mitsuru Ishizuka" – a co-author of this paper and current chairperson of JSAI.

Fig. 1. Part of the JSAI social network

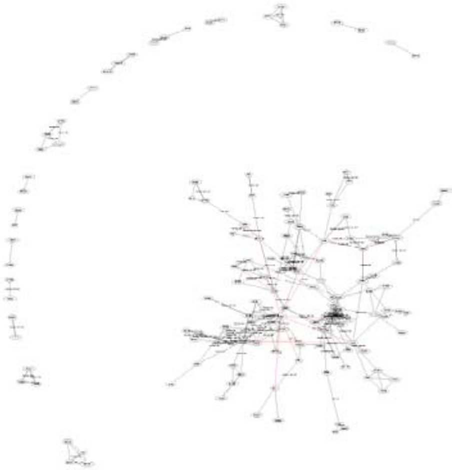

Fig. 2. JSAI social network

2.2 Trust Calculation

Trust on the Social Network Anyone can say anything on the Web. For that reason, lacking trust, we are unable to determine whom to believe. Trust is a necessary condition for users to fully utilize a semantic web.

We focus on the locality of a "Web of Trust". Initially, a local community will develop a small "Web of Trust" within the community. The small "Web of Trust" in a local community is helpful for judging the reliability of a person, an organization, or a piece of information. Some nodes have a high degree of trust edges: they are considered reliable. A newcomer can gain trust by somehow tying himself to a trusted node. The small "Web of Trust" has its *raison d'etre* within the community. Subsequently, small "Webs of Trust" will appear one by one in different communities. These local "Webs of Trust" will be superposed one by one because a person or an organization belongs to several communities at the same time. Finally, they will come to comprise a huge "Web of Trust" that spans the entire Web, encompassing many local trust networks.

The physical world already offers a "Web of Trust", as a kind of social network. I trust one of my friends; consequently, I also trust a person introduced by that friend. I trust a company because one of my companies is dealing with that company. In this way, our social network works well to assess trustworthiness. Such a mechanism is likely to work well on the Semantic Web. Using the social network, we can obtain the authoritativeness of a node. It can be considered as reliability or social trust. On the other hand, the network is used to calculate trust that can be accorded to that person: individual trust.

Social and Individual Trust The Google search engine uses a link structure for ranking Web pages, called PageRank [3]. A page has a high rank if the sum of the its for-

Table 3. Result of Authority Propagation

	Name	Activation	Freq	Comment (in 2004)
1	Toyoaki Nishida	5.53	624	Former Commissioner of JSAI, Prof.
2	Toru Ishida	4.98	574	Former Commissioner of JSAI, Prof.
3	Hideyuki Nakashima	4.52	278	Former Commissioner of JSAI, Prof.
4	Koiti Hashida	4.49	345	Commissioner of JSAI
5	Mitsuru Ishizuka	4.24	377	Commissioner of JSAI, Prof.
6	Hiroshi Okuno	3.89	242	Commissioner of JSAI, Prof.
7	Riichiro Mizoguchi	3.60	404	Commissioner of JSAI, Prof.
8	Seiji Yamada	3.35	168	Associate Prof.
9	Hideaki Takeda	3.22	435	Associate Prof.
10	Takahira Yamaguchi	236	624	Prof.
11	Yukio Ohsawa	2.98	185	Associate Prof.
12	Hozumi Tanaka	2.90	465	Chairperson of JSAI, Prof.
13	Takenobu Tokunaga	2.89	302	Associate Prof.
14	Koichi Furukawa	2.77	141	Former Commissioner of JSAI, Prof.
15	Kawahara Tatsuya	2.74	440	Prof.

Table 4. Result of Authority Propagation from Yutaka Matsuo

	Name	Activation	Freq.	Comment (in 2004)
1	Yutaka Matsuo	230.6	136	Target node
2	Mitsuru Ishizuka	28.7	377	Former supervisor, co-author
3	Yukio Ohasawa	19.5	185	Former project leader, co-author
4	Toyoaki Nishida	14.5	624	Professor of lecture at university
5	Masahiro Matsumura	13.5	82	Former colleague, co-author
6	Seiji Yamada	12.7	168	Acquaintance
7	Yasushi Takama	12.3	16	Former researcher of the former laboratory
8	Toru Ishida	12.1	574	Advisory Board of current research center
9	Takahira Yamaguchi	11.5	236	Acquaintance
10	Hidehiko Tanaka	11.3	842	Professor at university

ward links evenly contribute to the ranks of the pages to which they point. PageRank is a global ranking of all Web pages and is known to perform very well.

We employ here a PageRank-like model to measure authoritativeness of each member [13]. Each node v has an authority value $A_n(v)$ on iteration n. The authority value propagates to neighboring nodes in proportion to the node relevance:

$$A_{n+1}(v) = c \sum_{v^{'} \in Neighbor(v)} \frac{R(v,v^{'})}{Rsum(v)} A_n(v^{'}) + cE(v)$$

$$Rsum(v) = \sum_{v^{''} \in Neighbor(v)} R(v,v^{''})$$

where $Neghbor(v)$ represents a set of nodes, each of which is connected to node v, c is a constant for normalization, and E represents a source of authority value. We set E

as uniform over all nodes for simplicity (but it can be set depending on v). If we set a certain node v_{target} as a source of authority value, the result can be interpreted as showing authority for the node: individual trust. We set the initial authority as follows.

$$E(v) = \begin{cases} 1.0\ if v = v_{target}, \\ 0.0\ otherwise \end{cases}$$

For mathematical details, see [3].

Table 3 shows a result applied to the JSAI community extracted from the Web. Among 1509 people in the community, these people have high authority value $A(v)$ (denoted as Activation) after 1000 iterations. Although the hits (denoted as Freq) are few, some people are ranked highly. Present or former JSAI Commissioners are 9 of 15 people. Others are younger; they are not yet Commissioners, but they are active researchers who are mainly working in JSAI.

The top listed people by this algorithm are authoritative and reliable in the JSAI community. However, authoritative people are not always listed highly by our approach. For example, JSAI currently has 20 commissioners (including a Chairperson and two Vice-chairpersons), but we can extract only 5 current commissioners of the top 15. In other words, our approach seems to have high precision, but low recall. This drawback is attributable to the lack of information online. Especially, elder authorities tend to have produced many publications before the WWW came to daily use.

Table 4 shows a result obtained by setting v_{target} as node "Yutaka Matsuo". The familiar persons for him, e.g., a supervisor, a project leader, colleagues, and co-authors are ranked highly. This ranking is useful as a proxy for individual trust. For example, if a person is judged as very familiar to me, then she can automatically have permission to access my work libraries. Otherwise, she must ask my permission.

2.3 Application

To demonstrate our Web mining approach in the real application, we develop a researcher mining and retrieval system called Polyphonet (Fig. 3). The system is an example of an end-user application that integrates Web mining into the Semantic Web. The system is intended to provide a search function based on the relation of researchers and promote efficient collaboration. For example, a user can find what research topic a researcher is doing or whom she is working with. Social networks is used for finding path to other researchers or recommending related researchers. If the researcher is not found in the system, a user can register his name. Subsequently, the system automatically extracts information from the Web using the proposed Web mining method.

Extracted users' information is easily incorporated in the RDF representation [11]. For example, the network ties and the interest associations are represented in RDF using the `foaf:knows` and `foaf:interst` properties. Similarly, the relation become `foaf:Persons` with the appropriate relations. Some extensions of the FOAF model are necessary for expressing the relation labels. Figure 4 shows a FOAF file that was generated based on extracted information. Each researcher can have metadata included in the system. because extracted information is stored as a FOAF file.

Trust gives an authoritativeness of a person which is useful when finding an important researcher in the field. If we trace the node which has high individual trust from

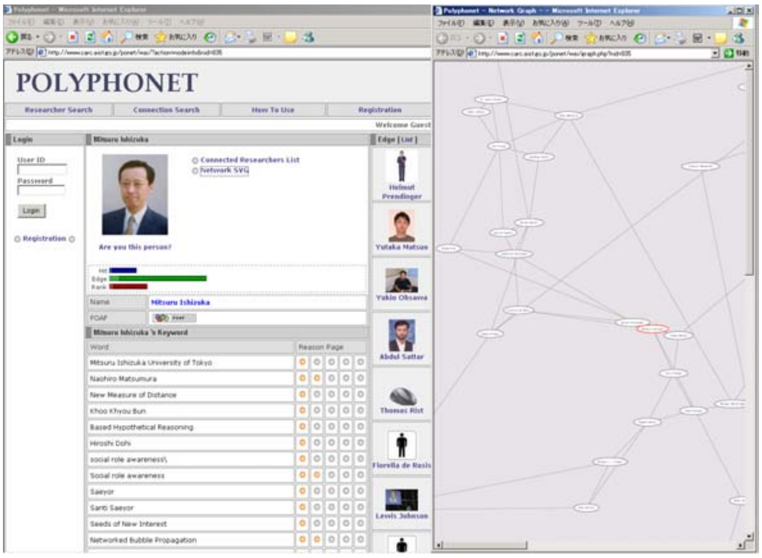

Fig. 3. Polyphonet: a researcher mining and retrieval system

antecedent node, we can find the circle of trust which comprises the small “Web of Trust” in a community.

3 Discussion

Hereafter, we address some of the workshop issues and discuss how our approach contributes to user aspects in the Semantic Web.

– Which baseline technologies are used and how are they combined?
– What aspects of end user activity does the technique affect?
– Can you describe convincing use-case scenarios demonstrating the power and usefulness of this approach?

Users are coming to accept FOAF and its extensions as something of a standardized ontology for representing user semantics on the Semantic Web. It has been a popular ontology of the Semantic Web. In other words, users are actively disseminating their social information on the Semantic Web. Our approach is to support those user-side trends by reusing the current Web as a source to produce such users’ information. We employ various Web mining techniques such as a search engine, statistical word co-occurrence information and machine learning. Our approach assists users in extracting relevant information from the Web and integrating it with the Semantic Web technologies. Furthermore, it encourages users to publish their information on the Semantic Web. In the proposed researcher mining and retrieval system, novice users can naturally approach the Semantic Web technologies such as Ontology and “Web of Trust” because those technologies are included in the system.

```
<rdf:RDF
xmlns:rdf=”http://www.w3.org/1999/02/22-rdf-syntax-ns#”
xmlns:foaf=”http://xmlns.com/foaf/0.1”
xmlns:acsn=”http://www.carc.aist.go.jp/ y.matsuo/acsn/0.1”>
<foaf:Person>
<foaf:mbox rdf:resource=”ishizuka@miv.t.u-tokyo.ac.jp”/>
<foaf:name>Mitsuru Ishizuka</foaf:name>
<foaf:interest rdfs:label=”Character agent”
rdf:resource=”http://www.miv.t.u-tokyo.ac.jp”/>
<foaf:currentProject rdfs:label =”Life-like interface”
rdf:resource=”http://www.miv.t.u-tokyo.ac.jp”/>
<foaf:workplaceHomepage rdfs:label=”University of Tokyo”
rdf:resource=”http://www.miv.t.u-tokyo.ac.jp”/>
<acsn:Coauthor>
<foaf:Person>
<foaf:mbox rdf:resource=”y.matsuo@aist.go.jp”/>
<foaf:name>Yutaka Matsuo</foaf:name>
</foaf:Person>
</acsn:Coauthor>
</foaf:Person>
```

Fig. 4. An example of a FOAF file tha tis based on extracted information from the Web.

– What is its potential to improve/simplify users' tasks?

There is often discussion about how metadata annotation is facilitated and accelerated. Consequently, users often find it difficult to collect and describe their information according to the Semantic Web standards. Reusing the existing sources of information on the Web would be a solution of the semantic annotation problem by minimizing the associated effort and helping users create their metadata.

In the Semantic Web, it is important to know whom to believe so that users can determine whether or not the source of information is reliable and credible. However, users often find it difficult to determine whom to believe in the distributed and heterogeneous environment of the Semantic Web. Our community-based approach would provide important clues for a Web of Trust on such a Semantic Web. Based on such a trust network, the system can help users determine the veracity of trustworthy persons, resources, and information.

– Why do we need the Semantic Web for this?

In the process of reusing the current Web as a source of information to produce semantically rich information, the Semantic Web provides a rich framework to describe semantics of the extracted information. In addition to the FOAF ontology and its extensions that we are currently using, we are extracting myriad community information in different contexts from the Web and converting it into semantic information. Our future work will explore the kinds of service that can be provided using semantically rich information as a resource.

4 Related works

The emerging field of social network mining provides methods for discovering social interactions and networks from legacy sources such as e-mail archives [1, 17], schedule data, Web citation information [15], and FOAF files [6]. It would be useful to incorporate such other information sources to obtain a more accurate social network, but such resources involve particular concerns of privacy: people do not want e-mail data to be analyzed.

Kautz and Selman developed a social network extraction system from the Web, called *referral web* [9]. This pioneering work particularly emphasizes co-occurrence of names on Web pages using a search engine. Mika pursued a similar approach [14] to extract a social network of a community. He also proposed a method to determine whether or not a certain person is associated with a certain interest. Both studies employ the Jaccard coefficient as a co-occurrence index. Although the fundamental idea resembles that of our approach, we further develop the mining algorithm. We use an overlap coefficient rather than a Jaccard coefficient based on experimental evaluation. We apply text processing and machine learning to determine the class of relation. Whereas Mika gives a list of interests, we can capture the various aspects of personal information from different Web pages. Furthermore, our method demonstrates the applicability of calculating the trust of each person.

Golbeck proposed an algorithm for generating locally-calculated trust ratings from a FOAF-based social network [10]. In a peer to peer context, Kamvar developed the EigenTrust system [8], which computes global trust values for peers. Although both approaches calculate trust on the network, we extract a social network of a community from the Web, which realizes more end-users and real-world oriented design for a "Web of Trust". Many research issues require investigation to realize a "Web of Trust" on the Semantic Web.

5 Conclusions

This paper presents an advanced Web mining approach to extract users' social networks and their related information from the current Web for the Semantic Web. In particular, we focus on an academic community and then argue the manner in which local trust networks will finally constitute a huge "Web of Trust". We show that the social relation is utilized to measure the authoritativeness of a member as social trust or individual trust. As an actual application that integrates Web mining with the Semantic Web, we presented a researcher mining and retrieval system.

We target researchers because of their associated information has relatively high availability on the Web, but our approach is not limited to that domain by any means. More and more information related to ordinary people online makes our approach feasible in various domains. More possibilities for using a search engine and mining the "non semantic" Web will arise in the future. For example, an ontology can be constructed using a search engine. We believe that merging the vast amount of information on the current Web and producing semantic information might help users fully utilize a Semantic Web and contribute to its further diffusion.

References

1. L. Adamic and E. Adar. Friends and Neighbors on the Web, http://www.parc.xerox.com/istl/groups/iea/ papers/web10/, 2001.
2. T. Berners-Lee, J. Hender and O. Lassila. The Semantic Web. Scientific American, 2001.
3. S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine, *In Proc. 7th WWW conf*, 1998.
4. D. Brickley and L. Miller. FOAF: the 'friend of a friend' vocabulary. http://xmlns.com/foaf/0.1/, 2004.
5. I. Davis and E. Vitiello Jr. RELATIONSHIP: A vocabulary for describing relationships between people, http://vocab.org/relationship/, 2004.
6. L. Ding, L. Zhou, T. Finin and A. Joshi. How the Semantic Web Is Being Used: An Analysis of FOAF Documents. *In Proc. of the 38th Ann. Hawaii International Conference on System Sciences*, 2005.
7. L. C. Freeman. Centrality in social networks: Conceptual clarification, *Social Networks*, Vol.1, pp.215–239, 1979.
8. S. D. Kamvar, S. T. Mario and H. Garcia-Molina. The EigenTrust Algorithm for Reputation Management in P2P Networks, *In Proc. WWW 2003*, 2003.
9. H. Kautz, B. Selman and M. Shah. The Hidden Web. *AI Magazine*, Vol. 18, No. 2, pp.27–36, 1997.
10. J. Golbeck, J. Hendler and B. Parsia. Trust networks on the semantic web, *In Proc. WWW 2003*, 2003.
11. J. Mori, Y. Matsuo, M. Ishizuka, and B. Faltings. Keyword Extraction from the Web for FOAF Metadata. *In Proceedings of the 1st Workshop on Friend of a Friend, Social Networking and Semantic Web*, 2004.
12. Y. Matsuo, M. Hamasaki, J. Mori, H. Takeda, and K. Hasida. Ontological Consideration on Human Relationship Vocabulary for FOAF. *In Proceedings of the 1st Workshop on Friend of a Friend, Social Networking and Semantic Web*, 2004.
13. Y. Matsuo, H. Tomobe, K. Hasida, and M. Ishizuka. Finding Social Network for Trust Calculation, *In Proc. 16th European Conf. on Artificial Intelligence (ECAI2004)*, 2004.
14. P. Mika, Bootstrapping the FOAF-Web: an experiment in social networking network mining, *Proc. of 1st Workshop on Friend of a Friend, Social Networking and the Semantic Web*, 2004.
15. T. Miki, S. Nomura and T. Ishida. Semantic Web link analysis to discover social relationship in academic communities. *In Proc. 2005 International Symposium on Applications and the Internet (SAINT)*, 2005.
16. J. R. Quinlan, *C4,5: Programs for Machine Learning*, Morgan Kaufmann, California, 1993.
17. J. Tyler, D. Wikinson and B. Huberman. Email as spectroscopy: automated discovery of a community structure within organizations. *Communities and Technologies*, Kluwer, B.V. pp.81–96, 2003.
18. Resource Description Framework(RDF) Schema Specification. In *W3C Recommendation*, 2000.
19. OWL Web Ontology Language. In *W3C Recommendation*, 2004.
20. Representing vCard Objects in RDF/XML. http://www.w3.org/TR/2001/NOTE-vcard-rdf-20010222/