SEMANTIC MATCHMAKING FOR JOB MARKET IN PEER-TO-PEER ENVIRONMENT

I Wayan Simri Wicaksana *, and Suryo Guritno †
* Faculty of Computer Science, University of Gunadarma
Jl. Margonda Raya 100, 16424, Depok, Indonesia
† Faculty of Mathematic and Science, University of Gadjah Mada
Skip Utara, BLS 21, Yogyakarta, Indonesia
email : iwayan@staff.gunadarma.ac.id, suryoguritno@ugm.ac.id

ABSTRACT

One of the important public service for developing country is job matching service. Public services face problems due to the dynamic of government organization, large number and diversity of sources for information sharing. P2P model has some features that make possible to handle dynamic environment, such as joint and leave network, no central, and distributed. Semantic and web semantic have introduced a solution to handle heterogeneity in semantic level, especially in Internet era. This work is focusing on emerging advantages of semantic web and P2P environment model for successful information sharing with case for job matching.

Keywords : job market, matchmaking, P2P, semantic, semantic web.

1 INTRODUCTION

Internet sites that provide job vacancies or job matching services, such as www.jobsdb.com, http: // www.state.gov/m/dghr/flo/rsrcs/pubs/ 4510.htm have been very popular lately. Unfortunately, they provide only information from themselves. Users cannot directly extract and perform cross-references with other information sources. Consider a scenario, one company requires employees with various requirements. The company has to visit many sites, and performs the data collection manually as well as "copy and paste" operation to produce the final report. This effort is time consuming and human intensive.

Matchmaking is the problem of matching offers and request such as supply and demand in a marketplace, where both partners are peers in the transaction. Currently, process of matchmaking utilize sources from Internet, however sources of Internet are more dynamic, heterogeny, open, and large. Semantic web has proposed some approach to overcome problem of diversity in structure and semantic by consider semantic distance based on ontology.

Peer-to-Peer (P2P) allow information sharing within a community of users with a common interest based on reference ontology. P2P can retrieve information based on semantic description and data source discovery. Emerging of semantic web and P2P can reduce problem of diversity semantic in discovery sources, writing request and matchmaking.

This paper describes a model to find the appropriate information by using P2P environment and semantic web in case of job matching services from various sources. In establishing information sharing/interoperability among node, semantic similarity based on ontology will be employed. The paper will be structured as follows. Part 2 will discuss current solutions and problems in job matching and related works. Next part will introduce of architecture and discovery sources of P2P, and semantic web for running example. Last part provides conclusion and further work

2 BACKGROUND

2.1 Job Matching in Indonesia

Formerly, Indonesia government attempted to provide job matching service by implementing centralized model. Data which consists of job providers and job seekers from different provinces or districts (kabupaten) were collected into a central database in Ministry of Manpower in Jakarta. Then, the data is centrally processed. This centralized model was implemented due to government structure and also lack of the data processing facilities in regions. Computers and network connections were still rare and slow. This approach introduced many problems such as processing delay and reliability.

Following the change in Indonesia government from centralized to decentralized model , the job matching services will also be distributed to the district level. The availability data processing facilities and networking infrastructure accelerate the acceptance of decentralized model. These technologies make the developing data or information sources easier to be done. However, data between districts are very diverse. Thus, it creates some difficulties in performing job matching across regions or districts. Moreover, the data from other institution also has diversity model and concept, so the interoperability becomes more difficult to achieve.

Internet has been helping in connecting many information sources for sharing purposes. However, it is still not easy to find particular information according to particular request. Usually, to find information provider

in Internet, we use search engine which based on keyword. Some popular search engines such as Google, Yahoo, Altavista, and Vivisimo are not designed to search very specific topic in Internet. They try to cover all of the world and all topics. It is very difficult to judge the quality of sources. Most search engines rank the information but not reflecting the quality of sources. It can be said that search engine site, from the user point of view implement more centralized then decentralized model [1].

Although, the Internet and computer facilities have been available in district levels in Indonesia, Internet connection still has low bandwidth and high latency. Therefore, the searching mechanisms which rely on the high bandwidth and stable connection (such as using centralized search engine) cannot be easily adopted. A search mechanism that can be performed for various information sources that connect temporarily with the system, should be employed. An efficient protocol which consumes only low bandwidth should be chosen.

2.2 Related Works

Refer to survey from [2, 3], earliest matchmakers, based on KQML. The matching process is carried out some stages from classical Information Retrieval (IR) analyst of text to semantic match via Θ-subsumption. The approach were deployed in SIMS and InfoSleuth project. Others matchmakers approaches are: *matchmaking framework* which operates on service description, *approximate matches* which proposed in the absence of exact matches, and *symmetric matchmaking* which use OO and implement in P2P.

Semantic web is to enhance current web by providing better understanding machine-machine and machine-human. One of the important factor is semantic similarity which relates to computing similarity between concepts of resources. There are some approaches to calculate semantic similarity, according to survey [4], the approach can utilize string analysis, latent semantic, WordNet, internal and external structure.

Peer-to-Peer (P2P) has been started since 1980, USENET (1979). FidoNet (1984). Napster (2001) is a P2P music file sharing, which is as a trigger of the popularity of P2P. As stated by Milojicik [5] P2P is not a new thing. From various definitions of P2P [6], P2P has main characteristics as self-organizing, symmetric communication, decentralized, autonomy, joint and leave at any time, anonymity. P2P implementation should be considered based on these characteristics. Based on degree of centralization of P2P model can be classified as pure, partially and hybrid P2P [6].

3 EMERGING SEMANTIC WEB AND P2P FOR MATCHMAKING

A query consists of a set of search terms (words). In performing searching mechanism, the query process in P2P environment is different with the centralized model. To respond the query in peer-to-peer or information retrieval some approaches have been introduced [7]. The query result is a list of documents that contain the terms, ranked by some scoring mechanism. The query will based on matchmaking process. Matchmaking process can be done by manual or automatic, to bring in automation level some method is needed.

The general process of matchmaking of job market are:

- Provide Super Peer(s).
- Create cluster of community of users with common interest.
- Publish Meta-Source of peer to Super-Peer(s).
- Communicate between Super Peer(s) to create Meta-SuperPeer
- Discovery sources will based on Meta-Source(s) and Meta-SuperPeer(s)
- After finding source(s), a query is sent directly to source(s) without SuperPeer
- Handling matching for the query will based on semantic similarity.

3.1 Peer Architecture

The P2P architecture is used to solve discovery resources in dynamic environment [8]. First popular P2P is Napster as music file sharing services [9]. The other advantage, peer community will share information in common interest which called clustering to reduce semantic diversity [10, 11, 12]. Combination of P2P and semantic web can bring an approach to current issues. Some approaches from Intelligent Search Mechanism [7], pure P2P [5] and super-peer [13] will be combined in this proposed mode. Our proposed architecture will has general configuration as seen figure 1 b and c.

In this architecture, there are two different types of nodes, Peer and SuperPeer. SuperPeer is a peer with more computing power and higher bandwidth. SuperPeer has more reliable connection than Peer. This model is chosen due to the network infrastructure in Indonesia, such as: broadband connection is still expensive. Therefore, only SuperPeer which should always be connected to the Internet. Peers can connect to Internet when they want to supply the data or perform the query. Peer nodes have data to be shared. For public access the schema of data will be stored in Export Schema/Ontology. The registry mechanism is used by Peer to join particular SuperPeer. It starts by sending advertising and publishing to a SuperPeer, after being accepted by SuperPeer (based on grouping/clustering), Peer sends its schema data and metadata of its data. SuperPeer will store in Meta-Sources at SuperPeer.

SuperPeer will have Meta-Sources which save list of members and the content. There is a mechanism to handle dynamic environment with member that can join and leave at any time cause of dirty or clean log-off. For member that leaves the system, their data are not deleted immediately. The data will be cached for a period of time. Thus, for next time when Peer registers again, they just need to submit the new schema data and metadata for adjusting Meta-Source. Meta-SuperPeer will maintain information about SuperPeer in same cluster and closed or related other clusters which representation of SuperPeer.

3.2 Process of Discovery Sources

As illustrated in Figure 1a, a company which is notated as Peer P61 wants to find new employee for technician and testing staff. The testing staff will be sent to training center to attend training in using the particular testing tools. Cluster 1 which consists of SP5 and SP6 is common interest of community in job provider. Cluster 2 which consists of SP1, SP2 and SP3 is community of job searching. Cluster 3 which interest in training institution, the members are SP3, SP4 , SP5. A SuperPeer or a peer can be member more than one cluster, for instance SP5 is in cluster 3 and cluster1.

Assume a company joins to SuperPeer SP6, and it sends its schema and metadata (Meta-Sources). Basically, searching mechanism in this architecture is performed by finding Peers that can provide appropriate answer for the query. Whenever this company performs a query, firstly it will discovery sources based on information at SP6. SP6 will try to find in his meta-sources. Assume, the source is known to be available at Peer P21 for employee information and Peer P31 for training information. Unfortunately, both Peers are not member of SP6, and there is no cache information in SP6 which has the answer of sources of the query. Therefore, the Meta-SuperPeer of SP6 will be searched. Assume there is a data about job searching cluster in SP2 and training cluster in SP3. The answers are stored into cache in SP6. The purpose is the answer of sources can be retrieved by other query of peer. After finding the sources peer, the query can be sent directly to source peers without through super peer. This mechanism will reduce the traffic, bandwidth usage, speed up for searching and single point failure.

3.3 Semantic Web and Matchmaking

If there is a query from a company to find an employee with request has skill in hardware and testing. For example, there are some sources provide information for the vacant employee, source-1 and 2 with different concept as table 1.

Table 1: Different Concepts of Request and Sources

Request	***Peer-1***	***Peer-2***
employee	resources	person
skill	education	background
hardware	hardware	PC
testing	quality control	check

If the query send to source 1 and 2 with query **?** employee(where skill=hardware, skill=testing)**?** , source-1 and 2 can not answer because there is no employee concept. Currently manual approach did to solve the problem by rewriting a query or matchmaking [14, 15]. For human is easy task to match employee is similar or not to resources or person. However for machine, it is very hard task to handle it. In our approach, semantic similarity or distance is implemented for matchmaking. The first step a linguistic preprocessing is used, such as expand abbreviations and replace acronyms. Next, the concept are matched by determining relation between them in WordNet relation. "WordNet is a semantic lexicon for the English language. It groups English words into sets of synonyms called synsets, provides short, general definitions, and records the various semantic relations between these synonym sets" [http: // en.wikipedia.org/wiki/Wordnet].

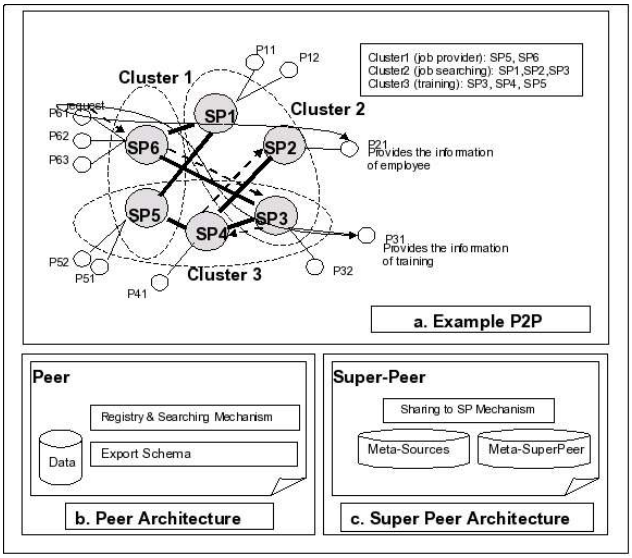

Figure 1. Example P2P Environment

Refer to Wu-Palmer (WUP) equation, semantic similarity of concepts can be calculated based on WordNet as:

$$similar_{WUP} = max[\frac{(2.depth(LCS(a,b)))}{(length(a,b)+2depth(LCS(a,b)))}] \quad (1)$$

where *length*(*a*,*b*) is number of path between a and b; *depth* (*LCS*(*a*,*b*)) is number of path from common concept of a and b to root.

The result of calculation matchmaking based on WordNet by using WUP equation can be seen at table 2. By using a threshold value τ, we can consider which concepts is match or not. Let say, we consider a threshold value is 0.6, the result implementation τ of can be seen at table 2. Refer to result table 2, the above query can be respond by source-2 because concept of employee is similar to concept person, however source-1 can not answer because concept of employee is not similar to concept of resources. In detail, source-2 can match if concept skill is similar to background, etc. In other word, matchmaking the original query can be read as local query as follow:

- original_query= **?**employee(where skill=hardware, skill=testing)**?**
- source-2_query= **?**person(where background=PC, background=check)**?**

Currently in our approach has not yet discussed to differ between concept of class, property and value. The above example consist of class (e.g. employee, person, resources), property (e.g. skill, background, education) and value (e.g. hardware, testing, PC). In our in going research, we evaluate and develop better approach for more specific concept in class, property and value.

Table 2: Semantic Similarity based on WordNet and WUP

Concepts	***WordNet & WUP***	τ=0.6
Source-1		
employee-resources	0.267	0
skill-education	0.750	1
hardware-hardware	1.000	1
testing-quality control	0.889	1
Source-2		
Employee-person	0.833	1
Skill-background	0.750	1
Hardware-PC	0.75	1
Testing-check	0.880	1

4 CONCLUSION AND FUTURE WORKS

In our approach, emerging semantic web and P2P can cons tribute to handle discover sources and matchmaking query for semantic diversity. P2P is promising for enhancing the information sharing of job matching services. P2P will not replace search engine soon but as a complement in information sharing services. P2P has some characteristic, which is suitable for the conditions in Indonesia, such as autonomy, decentralization, constraint in storage and bandwidth.

By establishing the SuperPeer(s), other parties such as private companies, universities, and other non-government organizations can join and share information for job matching more easily. User(s) will be able to search information easier to get appropriate answer from appropriate peer/source.

The further work will focus in some optimizing mechanisms in order to provide correct answer of the user query. To judge which mechanisms that will be chosen is based on time of searching, correctness and completeness of answer. Issue of single point failure of super peer is our concern in next step by consider replica super peer, combination flooding mechanism or copying to peer members.

In order to fusion various information from various source with different information structure, our next research will also look at efficient mechanism for schema/ontology mapping consider matchmaking with class, property and value.

REFERENCE

[1] A. Cali, C. Diego, G. Giuseppe De, and L. Maurizio, *On The Role of Integrity Constraints in Data Integration*, 2002, accessed: August 2005 http: // www.dis.uniroma1.it/pub/calvanes/ cali-calv-degi-lenz-IEEE-BDE-2002.pdf.

[2] T.D. Noia, E. D. Sciascio, F.M. Donini, and M. Mongiello. Semantic Matchmaking in a P2P Electronic Marketplace, *SAC2003*, 2003, 582-586.

[3] L. Li, and I. Horrocs, A Software Framework for Matchmaking Based on Semantic Web Technology, WWW2003, 2003, 331- 339.

[4] IW.S. Wicaksana, *A Peer-to-Peer Based Semantic Agreement Approach for Spatial Information Interoperability*, Thesis Doctoral, University Gunadarma, Jakarta, 2006.

[5] D.S. Milojicic, K. Vana, L. Rajan, N. Kiran, P. Jim, R. Bruno, R. Sami, and X. Zhichen, *Peer-To-Peer Computing*, 2002, accessed: May 2005 http: // www.hpl.hp.com/techreports/2002/ HPL-2002-57.pdf.

[6] A. Loser, S. Wolf, and W. Nejdl, *Information Integration in Schema-Based Peer-To-Peer Networks*, 2003, accessed: May 2005, http: //cis.cs.tu-berlin.de/ aloeser/ publications/ Caise03_submission _Loeser_Nejdl_Wolpers_Siberski.PDF.

[7] M. Panti, P. Loris, and S. Luca, *A Pure P2P Approach To Information Integration*, 2001, accessed: March 2005, http: zSzzSzlsirpeople.epfl. chzSzabererzSzcitationszSztr_2002_02_19.pdf/ a-pure-p-p.pdf.

[8] A. Crespo, and G-M. Hector, *Routing Indices For Peer-To-Peer Systems*, 2001, accessed: March 2005, http: //www-db.stanford.edu/ crespo/ publications/crespoa_ri.pdf.

[9] A-T. Stephanos, *A Survey of Peer-To-Peer File Sharing Technologies*, White Paper, Athens University of Economics and Business, Greece, 2002, accessed: May 2006, http: //www.eltrun.aueb.gr/whitepapers/ p2p_2002.pdf.

[10]D. Barkai, *An Introduction to Peer-To-Peer Computing*, 2001, accessed: May 2005, http: // www.intel.com/ update/departments/initech/it02012.pdf

[11]F. M. Cuenca-Acuna, and N. Thu , *Text-Based Content Search and Retrieval in ad hoc P2P Communities*, 2002, accessed: May 2005, http: // www.elet.polimi.it/p2p/ papers/22.pdf.

[12]G. Flammia, Peer-To-Peer Is Not for Everyone, *IEEE Intelligent Systems* , 2001, 78-79.

[13]J. Li, L.B. Thau, H. Joseph M, and K.M. Frans, *On The Feasibility of Peer-To-Peer Web Indexing and Search*, 2002, accessed: May 2005 http: // www.pdos.lcs.mit.edu/ rtm/ papers/search_feasibility.ps.

[14]A. Doan, D. Pedro, and H. Alon, *Learning to Match the Schema of Data Sources: A Multistrategy Approach*, 2002, accessed: May 2006, http: // citeseer.ist.psu.edu/ doan03learning.htm.

[15]A, Kramer J Zisman, *Toward Interoperability in Heterogeneous Database Systems*, Imperial College Research Report, London, 1995.