Job Matching Information Sharing In EGovernment With P2P Infrastructure

I Wayan Simri Wicaksana*, I Made Wiryana**
Gunadarma University, Jakarta, Indonesia
{iwayan, mwiryana}@staff.gunadarma.ac.id
*Doctoral student at Université de Bourgogne, Dijon, France
**Doctoral student at Bielefeld University, Bielefeld, Germany

Abstract. One of the important public services delivered by government bodies in Indonesia is job matching. Following the change of Indonesia government, the restructuring of government has been performed several times. Public services face problems due to the dynamic of government organization, especially for information sharing. P2P model has some promising features that make possible to handle dynamic environment, such as joint and leave network, no central, distributed, etc. This work is focusing on P2P environment model for successful information sharing in eGovernment with dynamic and inter-organization.

1. Introduction

Internet sites that provide job vacancies or job matching services, such as www.jobsdb.com, or listed in http://www.state.gov/m/dghr/flo/rsrcs/pubs/4510.htm have been very popular lately. Unfortunately, they only present information from themselves. Users cannot extract directly and perform cross-references with other information sources. Consider this following scenario, a company seeks 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.

Job matching is part of public service provided by government bodies in Indonesia. The government has responsibility to support and facilitate by improving it. In this paper, we distinguish between job matching and job searching. In job searching process, employees are more active than job providers, but in job matching process employees and job providers have same level of need (symmetry). Job matching involves job providers, job seekers, and training institution as well as information broker. Job matching service must facilitate information sharing among these parties.

This paper describes a model to find the appropriate information for job matching services from various sources. Peer-To-Peer (P2P) infrastructure will be exploited in this model. In establishing information sharing/interoperability among node, semantic mediation will be employed. The paper will be structured as follows. Part 2 will discuss current solutions and problems in job matching. Next part will introduce of P2P and how P2P can be used as environment for this problem. Last part provides conclusion and further work

2. Background

Formerly, Indonesia government attempted to provide job-matching service by implementing a centralized model. Data which consists of job providers and job seekers from different provinces or districts (*kabupaten*) were gathered into a central database in Ministry of Manpower in Jakarta and 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 structure, job matching services will also be distributed to the district level. The availability data processing facilities and networking infrastructure accelerate the adoption 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 incompatibility data from other institution also make the integration and 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 prefer to use search engine. 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 it does not reflect the quality of sources. From user point of view, it can be said that search engine sites implement more centralized then decentralized model.

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. The cache mechanism should be considered as part of the solution. An efficient protocol which consumes only low bandwidth should be developed.

3. Peer-To-Peer for Job Matching

Peer-to-Peer (P2P) is not a new thing and has been started since 1980 [11], USENET (1979). FidoNet (1984). Napster (2001) is a P2P music file sharing, which is as a trigger of the popularity of P2P. From various definitions of P2P [11], 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 [1].

In performing searching mechanism, a query which consists of a set of terms (words) in P2P environment is different with the centralized model. There are some limitations on P2P environment such as scalability, storage and bandwidth constraints. To respond the query in peer-to-peer or information retrieval some approaches have been introduced [13]. The query result is a list of documents that contain the terms, ranked by a particular scoring mechanism.

In this work, the P2P architecture is used to solve query for job matching. Some approaches from indexing approach [9], Intelligent Search Mechanism [13], pure P2P [12] and super-peer [10] will be combined in this proposed mode. Our proposed model will has general configuration as seen figure 1.

In this model, 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 selected due to the network infrastructure in Indonesia. Broadband connection is still expensive. Therefore, only SuperPeer which should always be connected to the Internet. Peers can connect to Internet whenever they want to supply the data or perform the query.

Peer nodes have data, which can be private or public. For public access the schema of data will be stored in Local Public Component (LPuC). The registry mechanism is used by Peer to join particular SuperPeer. It starts by sending advertising and publishing to a SupePeer. After being accepted by SuperPeer (based on grouping/clustering), Peer sends its schema data and metadata of its data. SuperPeer will cache it and store in Index-Member at SuperPeer.

SuperPeer may be assigned manually or automatically. SuperPeer has Index-Member that stores list of members and the content Index-Member handles dynamic environment with member that can join and leave at any time. Data of members which leaves the system are not deleted immediately. Those Data will be cached for a period of time. Thus, at the next time when Peer registers again, they just need to submit the new schema data and metadata for adjusting Index-Member. Index-Cache keeps track the Peers, which respond a query. Indexing method will be based on statistical replying answer and time or distance of the Peer, which provides answer to the SuperPeer. Index-Related-SuperPeer will maintain information about SuperPeer in same group and closed or related other groups which representation of SuperPeer.

As illustrated in Figure 1a, a company notated as Peer P61 wants to find new employee for technician and testing staff. The testing staff will be sent to training centre to attend training in using the particular testing tools. Group1 which consists of SP5 and SP6 is job provider group. Group2 which consists of SP1, SP2 and SP3 is job searching group. SP3, SP4 , SP5 group to Group3 which is the training institution group. A SuperPeer may be belong to more than one group, for instance SP5 is in Group3 and also Group1.

Assume a company joins to SuperPeer SP6, and it sends its schema and metadata. 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, it will be sent to SP6. Firstly, SP6 will try to find in Index-Member and Index-Cache in SP6. Assume, the answer is known to be available at Peer P21 for employee information and Peer P31 for training information. Unfortunately, both Peers are not belonging to SP6, and there is no cache information in SP6 which has the answer of this query. Therefore, the Index-Related-SuperPeer will be searched. Assume there is data about job searching group with SP2 and training group with SP3. Some process will be performed with SP6. In this illustration, the search of job information will traverse from SP5 to SP2 and then to Peer P21. The search of training information flows from SP2 to Peer P21. The answers are stored into cache in SP6. These answers can be retrieved by more than one peer. This mechanism will reduce the traffic, bandwidth usage and speed up searching process.

However, in developing this architecture some questions still have to be solved. How if there is no super-peer? As suggested by Löser [10], if there is no super-peer, peers can be a super-peer for themselves. Is it possible for one peer to join more than 2 super-peers? Ideally, referring to the characteristic of P2P, it is possible. What is the optimal searching and register of peer? What is

optimal mechanism in writing and update: index member, index caching and index related super-peer at super-peer? How to optimize clustering or grouping super-peer?

4. Conclusions and Further Work

P2P is a promising model 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.

Government can take main initiative by supporting super-peer and clustering / grouping in some region in Indonesia. This initiative can be part of eGovernment services. By establishing the SuperPeer, other parties such as private companies, universities, and other non-government organizations can join and share information for job matching more easily. User will be able to search information easier to get appropriate answer from appropriate peer.

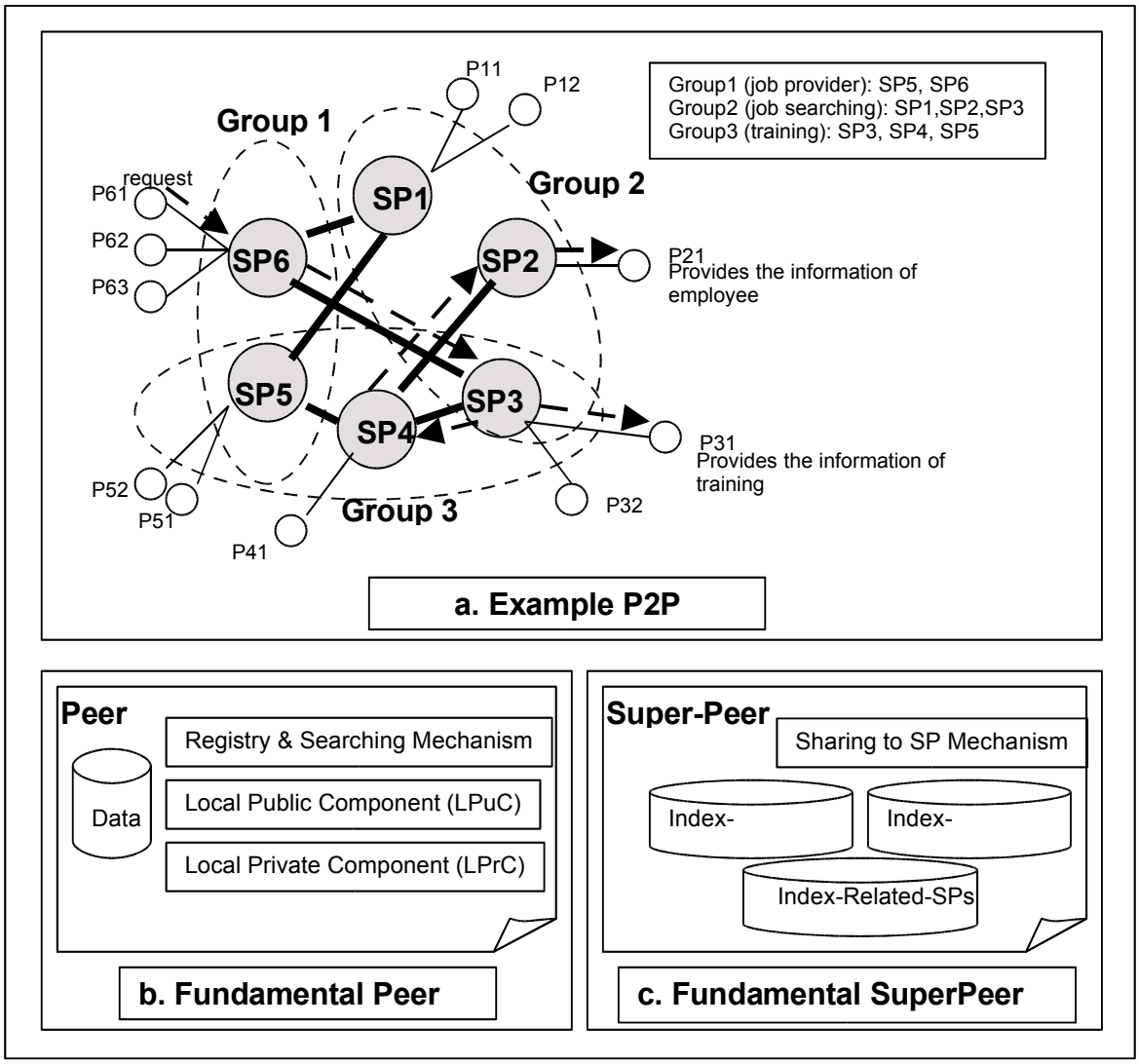

Figure 1: P2P Model, Peer and Super-Peer Architecture

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. However, there is still main problem in fusioning answers from some peers, because they are difference in syntactic, structural and semantic as well as context. Each peer which acts as the information source has different model and level of authority. Which peers that can be trusted?

In order to fusion various information from various source with different information structure, our next research will also look at efficient mechanism for schema mapping and ontology sharing. Currently, schema mapping and ontology sharing are based on semantic mediation and hybrid ontology. However, in P2P environment, the method to map and share between peers should be performed on the 'fly'. It is allowed peers to join and leave at anytime.

References

[1] ANDROUTSELLIS-THEOTOKIS Stephanos, "A Survey of Peer-To-Peer File Sharing Technologies", White Paper, Athens University of Economics and Business, Greece, 2002, <http://www.eltrun.aueb.gr/whitepapers/p2p_2002.pdf>

[2] BARKAI David, "An Introduction to Peer-To-Peer Computing", 2001, <http://www.intel.com/update/departments/initech/it02012.pdf>

[3] BAWA Mayank, Manku Gurmeet Singh, Raghavan Prabhakar, "SETS: Search Enhanced by Topic Segmentation," <2003, http://www-db.stanford.edu/~bawa/Pub/sets.pdf>

[4] CALI Andrea, Calvanes Diego, Giacomo Giuseppe De, Lenzerini Maurizio, "On The Role of Integrity Constraints in Data Integration", 2002, <http://www.dis.uniroma1.it/pub/calvanes/cali-calv-degi-lenz-IEEE-BDE-2002.pdf>

[5] CRESPO Arturo, Garcia-Molina Hector, "Routing Indices For Peer-To-Peer Systems", 2001, <http://www-db.stanford.edu/~crespo/publications/crespoa_ri.pdf>

[6] CUENCA-ACUNA Francisco Matias, Nguyen Thu D, "Text-Based Content Search and Retrieval in ad hoc P2P Communities", 2002, <http://www.elet.polimi.it/p2p/papers/22.pdf>

[7] DOAN AnHai, Domingos Pedro, Halevy Alon, "Learning to Match the Schema of Data Sources: A Multistrategy Approach", 2002, <http://citeseer.ist.psu.edu/doan03learning.htm>

[8] FLAMMIA Givanni, "Peer-To-Peer Is Not for Everyone", IEEE Intelligent Systems (2001) page 78-79, 2001, <http://www.computer.org/intelligent/homepage/2001/03Flammia/Flammia1.htm>

[9] LI Jinyang, Loo Boon Thau, Hellerstein Joseph M, Kaashoek M Frans, "On The Feasibility of Peer-To-Peer Web Indexing and Search", 2002, <http://www.pdos.lcs.mit.edu/~rtm/papers/search_feasibility.ps>

[10] LÖSER Alexander, Siberski Wolf, Wolpers, Wolfgang Nejdl, "Information Integration in Schema-Based Peer-To-Peer Networks", 2003, <http://cis.cs.tu-berlin.de/~aloeser/publications/Caise03_submission_Loeser_Nejdl_Wolpers_Siberski.PDF>

[11] MILOJICIC Dejan S, Kalogeraki Vana, Lukose Rajan, Nagaraja Kiran, Pruyne Jim, Richard Bruno, Rollins Sami, Xu Zhichen, "Peer-To-Peer Computing", 2002, <http://www.hpl.hp.com/techreports/2002/HPL-2002-57.pdf>

[12] PANTI Maurizio, Penserini Loris, Spalazzi Luca, "A Pure P2P Approach To Information Integration", 2001, <http://citeseer.ist.psu.edu/cache/papers/cs/26827/http:zSzzSzlsirpeople.epfl.chzSzabererzSzcitationszSztr_2002_02_19.pdf/a-pure-p-p.pdf>

[13] YAZTI Demetrios Zeinalipour, "Information Retrieval in Peer-To-Peer Systems", Master Thesis, University of California, Riverside, 2003, <http://www.cs.ucr.edu/~csyiazti/papers/msc/html/>

[14] ZISMAN A, Kramer J, "Toward Interoperability in Heterogeneous Database Systems", Imperial College Research Report, London, 1995, <http://citeseer.ist.psu.edu/cache/papers/cs/10387/http:zSzzSzwww.doc.ic.ac.ukzSzdeptechrepzSzDTR95-11.pdf/zisman95toward.pdf>