

A methodology for quality-based mashup of data sources

Muhammad Raza

Digital Ecosystems and Business
Intelligence Institute
Curtin University of Technology,
Perth 6845, Australia
+61-8-92667613

muhammad.raza@postgrad.curtin.edu.au

Farookh khadeer Hussain

Digital Ecosystems and Business
Intelligence Institute
Curtin University of Technology,
Perth 6845, Australia
+61-8-92669268

Farookh.Hussain@cbs.curtin.edu.au

Elizabeth Chang

Digital Ecosystems and Business
Intelligence Institute
Curtin University of Technology,
Perth 6845, Australia
+61-8-92669272

Elizabeth.Chang@cbs.curtin.edu.au

ABSTRACT

The concept of mashup is gaining tremendous popularity and its application can be seen in a large number of domains. Enterprises using and relying upon mashup have improved their mass collaboration and personalization. In order for mashup technology to be widely accepted and widely used, we need a methodology by which can make use of the quality of the input to the mashup process as a governing principle to carry out mashup. This paper reviews the concept of mashup in different domains and proposes a conceptual solution framework for providing quality based mashup process.

Categories and Subject Descriptors

E.m [Miscellaneous]: Trust, H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces – *Web-based interaction*.

General Terms

Algorithms, Management, Measurement, Performance, Reliability and Security.

Keywords

Mashup, Trust, Reputation, Quality, CCCI metrics.

1. INTRODUCTION

Over the past few years, the adaptation of new web-related technologies has been going on at a rapid rate. The emergence of Web 2.0, the application and integration of Web 2.0 based technologies has opened up new and innovative ways of collaboration and research challenges that needs to be address. Among the suite of Web 2.0 technologies is mashup, which is defined as a combination of openly published APIs (Application

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

iiWAS2008, November 24–26, 2008, Linz, Austria.

(c) 2008 ACM 978-1-60558-349-5/08/0011 \$5.00.

Programming Interface), content (eg. data tables), and application components (eg. Web services) to extract information from several different sources, which are combined, and revisualized in other forms (eg., twitter with Google Maps to view images on the maps) [14].

Mashup can be used by people from who are normal internet users (for data mashup) to professionals or application developers. Additionally, the application of mashup can be seen in different domains. A case in point is the VARIAZIONI project, where in semantic content enriching is carried out on a collaborative basis, which would be useful for non-technical users who can easily catalogue the existing digital assets, comprising multimedia master classes audio/video concerts, images, scanned documents, etc and create new contents from the existing contents of other institutions [1]. Mashup architectures extending SOA model are have been used [2] to facilitate service composition in the form of mashup component model and to help developers leverage to create their own composite services. Jin et al [3] have provided some directions for building both simple as well as sophisticated mashup-based applications.

As the number of individuals and organizations utilizing the mashup technology is increasing the crucial issues is to develop a methodology for quality based mashup process. In this paper we present a conceptual overview of a methodology for mashup which is governed or driven by the quality of the underlying data source. In the context of this paper, we refer to mashup of various data sources into a larger information set. However, this method can be equally applicable to services as well.

This paper is divided into four sections. In Section 2 we summarize and discuss the general approaches and applications of mashup. In Section 3 we will present a conceptual solution framework to enable quality based mash up processes. Finally, Section 4 concludes the paper and sets the ground for future work.

2. BACKGROUND

The emerging technologies of Web 2.0 and Semantic Web are providing great facilities to the end users. At the same time they are opening up numerous research issues and problems that need to be addressed. Concepts such as community, usability, organization and presentation are the pillars of Web 2.0. Mashup is one of the most interesting technologies within the framework of Web 2.0 and its usage is growing rapidly. “Crystal Tools” are

used to provide a mechanism for visual mashup of web, image, video, news, blogs and tagging search results, from diverse data sources, into a single integrated display [4].

Hotta et al [6] have proposed the use of web content collaboration and personalization simultaneously. In their method the Client Side Information Management (CSIM) collaborates (they make use of the term collaborate to refer to mashup) web contents without the exchange or personal data. The concept gives a protection on user's personal data. Migration strategy to SOC, based on one of the Web 2.0 challenges, namely the mashup technology is one of the major issues for legacy software users [5][7]. Cetin et al [5], [7] have proposed a roadmap for migration of legacy software to pervasive service-oriented computing. Web information extraction method can generate the virtual

Han et al [8] propose the use of web service functions from web applications at client side. This method called "*World in Web*", integrate the general web applications since integration is limited to the Web sites that provide the open Web service APIs, and currently, most existing Web sites do not provide Web services [8].

Vancea et al [10] explore the application of mashup technology in databases. Due to the powerful features of the database systems, the authors propose a web mashup process that is supported by integration and synchronization at the database level. This is an interesting aspect because traditionally the source of the contents used for mashup process was from feeds or API-based web services, which have limitations such as working with complex data structures and different APIs for each web server that needs different implementations for mashup users. The authors proposed a generic proxy mechanism for data integration and synchronization to support database-driven web mashup architecture [10]. Nakano et al [11] have proposed a method for creating web services from web applications. They proposed an effective method of creating wrappers that makes Web applications usable as Web services. This method extracts important segments, such as search results of a hotel search application, from an HTML document generated by the web application and generates extraction rules for the wrappers.

2.1 Mashup in Enterprises

Leveraging service orchestration and integration features in BPEL (Business Process Execution Language) to implement an enterprise composite service is one of the popular approaches in today's SOA industry. As there are a lot of constraints in BPEL, Service Component Architecture (SCA) can be the replacement that aims to provide a more open, flexible and scalable framework to create a composite service in SOA [12]. The authors [12] propose the implementation of an enterprise mashup composite service in SOA through a use case study and User Profile Use Case. In their proposed method a user profile use case was to create a universal interface to access information including person info, travel info, project info, property info, community info etc.

Another work in enterprise service mashup [13] introduces the main components of this new paradigm, and discusses the design principles of this architecture, upcoming intermediaries and mass collaboration, lightweight composition as well as perpetual beta development model. This architecture is able to provide

individual and heterogeneous enterprise applications in a shorter time.

Mashup has a lot of importance, benefits and possibility in enterprises besides its effectiveness in personalization and support for user choice [14]. Makki et al [14] have shown the benefits of such enterprise mashups in terms of personalization, organization, ease of navigation, and interpretation value for users and presentation, although the aspects of insecurity and instability on a webpage associated with data mashups makes it difficult to implement mashups in enterprises.

2.2 Mashup in specialized domains

Mashups are successfully used in some specialized domains where the benefits can be seen obviously. Support for e-learning has made Web 2.0 most attractive modern web technology for the users, replacing the previously used technologies and the success and failures of different approaches used in e-learning [15]. The concepts of ownership, blogs, simulation etc are seen as a success factor in e-learning. These factors can be easily seen in Web 2.0 especially in mashup. Data integration is another major issue in Bioinformatics domain [16]. The two web-technologies i.e. *semantic web* and *Web 2.0's* mashup, have the potential to address the important issue of data integration in bioinformatics. Mashup and Web 2.0 are also being used in online travel information systems [17]. Wang et al [17] have proposed an ontology and Bayesian Network based methodology for tourism. The efficiency of their proposed method is presented in the form of an intelligent recommendation system. Their system works by integrating heterogeneous online travel information through ontology and personalized recommendation system. Bayesian network technique method is used for modeling and determining users' preferences in order to provide recommendations to the users. Wang et al [17] have proposed a system that recommends tourist attractions to a user by taking into account the travel behavior both of the user.

Yan et al [18] have proposed the integration of GIS and hydrological models in order to predict the stream flow in the Oak Ridge Moraine, Southern Ontario, Canada. They have integrated Google maps and IHACRES (Identification of unit Hydrographs and Component flows from Rainfall, Evaporation and Stream flow data). Their proposed method has lots of advantages like, platform independence no need of any GIS software installation etc. The proposed method is cost effective, as Google Maps can provide crucial spatial information including high-resolution images and road networks. Additionally, it is less time consuming as developers only need to process part of spatial data, such as basin boundary, stream network and temporal data. Wood et al [19] have introduced the notion of Geovisualization mashup, where in freely available functionality and data are loosely but flexibly combined using de-facto exchange standards. Using this approach, novel combinations of interaction and visual encoding can be developed including spatial 'tag clouds', 'tag maps', 'data dials' and multi-scale density surfaces.

Travel service is another domain where the concept of mashup can benefit the users [20]. Various travel services are integrated to form a mashup of these services. Algorithms are developed in order to find the shortest path for traveling from one location to another with the cheapest prices. The shortest path algorithm is

used to find the cheapest price via the combination of flight, train, and bus.

2.3 Technology Integration

Recently there has been discussion on the two important web technologies i.e. *Web 2.0* and *Semantic Web* as being complementary to each other rather than competitors [17, 21]. Each field can and must draw from the other's strengths. It is envisaged that the future Web applications will retain the Web 2.0 focus on community and usability, while drawing on the Semantic Web infrastructure to facilitate mashup-like information sharing [21]. Semantic Web technologies must integrate with Web 2.0 services in order for both to leverage each other's strengths. Wang et al [17] propose to create a semantic mashup. In their proposed method Semantic Web technologies can support information integration.

Computer Telephony Integration (CTI) allows a computer application to access features from a telephone device or telephony network. These are commonly proprietary-based, platform dependent and usually designed and implemented by telephony experts [22]. Venkatchalam et al [22] propose a web application mashup that uses a high-level communication service, which are highly efficient and cost effective, and also present experiences from the prototype implementation. A communication-enabled kiosk as a representative example is used for the application mashup [22].

In their proposed method, Yelmo et al [9] combine web and network services. This can be viewed as a creative step towards complex application development or mashup. These user-centric service creation environments have the ability to improve the service offering with profitable, value-added services faster and cheaper than ever before [9].

2.4 Content ranking and recommendations

Researchers have proposed the use of ranking algorithms before the mashup process in order to ensure the quality of the resultant content. Alba et al [23] have proposed the use voting theory as means to enable quality based mashup process in the music domain. Their proposed method examines a Top 10 artist chart based on user comments and listening behavior from several Web communities [23]. Two major voting schemes are employed namely the majority-based and position-based schemes to combine rankings in a fair and effective manner.

Chen et al [24] examine the effectiveness of combining opinions from multiple sources. Their method proposes is based on information recommendation and is supported by Ubiquitous Personal Study (UPS). UPS is a framework of personalized virtual study and is successfully implemented with Web 2.0. Wang et al [17] propose an intelligent recommendation system for tourism domain that is leverages on an ontology and Bayesian Network based methodology.

The difficulty in finding and suggesting the most relevant service candidates for new mashups is increasing due to the increasing number of web services. Blake et al [25] propose a novel approach for enhanced syntactical matching for service mashups. They propose to use an overarching matching algorithm called

Tendency-Based Syntactic Matching-Levenshtein Distance and Letter Pairing (TSM-LP). Their methods can also be used for prediction service mashups in the future.

Accountability related issues in mashup services are more complex when compared to non-repudiation in an eCommerce transaction. As the service is mashed up from a number of sources the content presented may not be delivered by the content originator, and the source content may be altered or extended during the mashup process. In order to enable accountability of mashup of services there should be disclosure, trust, and un-deniability [26].

From the above comprehensive review of the literature it can be observed that the concept mashup is gaining tremendous popularity and acceptance in various domains. However, it is very important to note that in the existing literature there is no such mechanism that can be used to ensure that the content of the mash up process is reliable and trustworthy. The quality of the output mashup process would be directly related to the inputs to the mashup process. In the following section we will propose a conceptual solution framework, which can solve this issue.

3. CONCEPTUAL FRAMEWORK

Let us assume that a given end-user (say user A) intends to create a mashup of data sources on a given coherent topic. Throughout the discussion assume that the user intends to create a mashup from a finite number of data sources. However, in reality our proposed methods could be used to form a mashup of services in order to make a composite service.

We assume the existence of a semantic service search engine (like the one proposed by Dong et al [28]) or a semantic information search engine which would be able to find out those data sources/services specified by the end-user. Using such a search engine the user would be able to find out various information (or services) that can be combined with each other in a mash up process. The crucial question that arises at this stage is to have a methodology by which the Quality of this information (or service) can be decided before the end user initiates the mashup process. In other words we intend to propose a methodology where in the end user can make an informed decision of the components that (s)he is using for the mashup process based on the quality of those components.

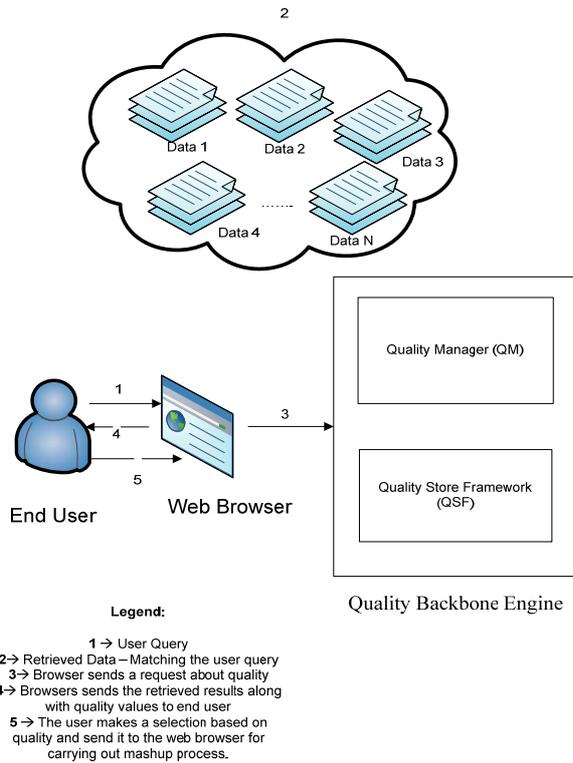


Figure 1: Conceptual Framework of quality based mashup process

For simplicity purposes let us assume that the user there are n data sources, as shown in Figure 1, that match the user search query. The user has to select m of these data sources (where $m \subseteq n$) based on the quality of the underlying m data sources.

In order to address this, we propose that existence of Quality Backbone Engine (QBE). For simplicity purposes we assume that the QBE would be a third party service that would store all the quality related information about web content. The browser on getting a mashup request from the end user would send a request to the Quality Backbone Engine (QBE) in order to make judgments about quality of data sources as shown in Figure 1. For simplicity purposes, QBE comprises of two components, namely Quality Manager and Quality Store Framework. The role of the Quality Manager lies at the core of the functioning of our proposed methodology. The task of the Quality Manager (QM) is to compute the quality of a givens source content based on the feedback provided by the user. QM makes use of a mathematical function to adjust the quality of a given source based on the feedback provided by the user. The Quality Manager makes use of the Quality Store Framework (QSF) to get quality related information. In other words the QSF would act like a repository of the quality information of data sources.

During the first step the user sends a request to the browser (for mashup content only). The syntax of the proposed request is as follows:

$$\text{Request (SType, QLevel, PType)}$$

Where,

SType is a variable that denotes the type of source that the end-user is requesting for the mashup process. The Source Type (SType) variable can have values in the form the type of the information or service the users want like, courier service, car rental, hotel, airline ticketing etc (and);

QType is a variable that denotes the desired quality of the information sources/services that are being returned by the QSF (and);

PType is a variable denoting *Preference Type*. This variable can have the following four values, namely. The value of PType is used to determine the value of QType.

The end-user may choose the parameter PType as EXACT. In this case those sources would be selected which have a quality value equal to the value specified in QLevel. Mathematically, this may be expressed as follows:

$$\begin{aligned} \text{If PType} &= \text{EXACT then} \\ \text{Quality} &= \text{QLevel} \end{aligned}$$

Alternatively, the end-user may choose the parameter PType as BEST. In this case the sources with the highest quality are selected. Mathematically, this can be expressed as follows:

$$\begin{aligned} \text{If PType} &= \text{BEST then} \\ \text{Quality} &= 5 \end{aligned}$$

Moreover, the end user may choose the parameter PType as AVAILABLE. In this scene value they would receive all the available sources of the specified irrespective of their quality level.

$$\begin{aligned} \text{If PType} &= \text{AVAILABLE then} \\ \text{Quality} &\geq 0 \end{aligned}$$

Finally, the end user may choose the parameter PType as SUITABLE. In this case they would receive all those sources whose quality value is higher than the QLevel provided by the user.

$$\begin{aligned} \text{If PType} &= \text{SUITABLE then} \\ \text{Quality} &\geq \text{QLevel} \end{aligned}$$

The Quality Manager would then pass on the retrieved results back to the browser along with their quality levels. The end user can then choose from the sources in a customized fashion based on the quality assessments. Subsequently the mashup process is carried out.

At the end of the process the user may provide feedback about the quality of the data sources. These values are then used to populate the Quality Source Framework.

4. CONCLUSION AND FUTURE WORK

In this paper we carried out a review of the existing literature on mashup. We discussed and pointed out that mashup has different applications in different domains. Additionally, we discussed that in the current literature there is no methodology by which the quality of the inputs to the mashup process can be determined. Once the quality of the inputs to the mashup process has been determined the user can subsequently choose those inputs that are of high quality, thereby resulting in a higher quality output.

We proposed a conceptual solution that can address this issue. We intend to continue working on this issue along many directions. One direction would be to develop the precise mathematical measures and metrics for determining and accessing quality. Another direction would be to find out how the QBE can be best organized and hosted and how the quality information in QBE can be organized and stored. Another additional direction that we intend to work on is to develop a fuzzy-mechanism for search results retrieval during the user query process.

5. REFERENCES

- [1] C. A. Iglesias, M. Sánchez, and F. Spadoni, "VARIAZIONI: Collaborative Authoring of Localized Cultural Heritage Contents over the Next Generation of Mashup Web Services" Proceedings of the Third International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution (AXMEDIS'07), 2007, pp. 225-230.
- [2] X. Liu, Y. Hui, W. Sun, and H. Liang, "Towards Service Composition Based on Mashup," in IEEE Congress on Services (Services 2007), IEEE SCW 2007 Salt Lake City, UT, 2007, pp. 332-339.
- [3] H.-J. Jin and H.-C. Lee, "Web services development methodology using the mashup technology," in International Conference on Smart Manufacturing Application, 2008. ICSMA 2008. Gyeonggi-do: IEEE, 2008, pp. 559-562.
- [4] A. Spoerri, "Visual Mashup of Text and Media Search Results," in 11th International Conference on Information Visualisation, IV 2007 Zurich, Switzerland, 2007, pp. 216-221.
- [5] S. Cetin, N. I. Altintas, H. Oguztuzun, A. H. Dogru, O. Tufekci, and S. Suloglu, "Legacy Migration to Service-Oriented Computing with Mashups," in International Conference on Software Engineering Advances, 2007, pp. 21-21.
- [6] H. Hotta, T. Nozawa, and M. Hagiwara, "A Design of Client Side Information Management Method for Web Services Collaboration," in Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Workshops IEEE Computer Society, 2007, pp. 95-98.
- [7] S. Cetin, N. I. Altintas, H. Oguztuzun, A. H. Dogru, O. Tufekci, and S. Suloglu, "A Mashup-Based Strategy for Migration to Service-Oriented Computing," in Pervasive Services, IEEE International Conference 2007, pp. 169-172.
- [8] H. Han and T. Tokuda, "A Method for Integration of Web Applications Based on Information Extraction," in Eighth International Conference on Web Engineering ICWE 2008 Yorktown Heights, New York, USA: IEEE, 2008, pp. 198-195.
- [9] J. C. Yelmo, J. M. del Alamo, R. Trapero, P. Falcarin, J. Yu, B. Carro, and C. Baladron, "A user-centric service creation approach for Next Generation Networks," in Innovations in NGN: Future Network and Services, 2008. K-INGN 2008. First ITU-T Kaleidoscope Academic Conference Geneva, Switzerland: IEEE, 2008, pp. 211-218.
- [10] A. Vancea, M. Grossniklaus, and M. C. Norrie, "Database-Driven Web Mashups," in 2008 Eighth International Conference on Web Engineering, 2008, pp. 162-174.
- [11] Y. Nakano, Y. Yamato, M. Takemoto, and H. Sunaga, "Method of creating web services from web applications," in IEEE International Conference on Service-Oriented Computing and Applications 2007, pp. 65-71.
- [12] F. Yang, "Enterprise Mashup Composite Service in SOA – User Profile Use Case," in 2008 IEEE Congress on Services - Part I, 2008, pp. 97-98.
- [13] V. Hoyer, K. Stanoesvka-Slabeva, T. Janner, and C. Schroth, "Enterprise Mashups: Design Principles towards the Long Tail of User Needs," in 2008 IEEE International Conference on Services Computing Vol. 2, 2008, pp. 601-602.
- [14] S. K. Makki and J. Sangtani, "Data Mashups & Their Applications in Enterprises," in Internet and Web Applications and Services, 2008. ICIW '08. Third International Conference Athens: IEEE, 2008, pp. 445-450.
- [15] M. Eisenstadt, "Does Elearning Have To Be So Awful? (Time to Mashup or Shutup)," in Advanced Learning Technologies, 2007. ICALT 2007. Seventh IEEE International Conference Niigata: IEEE, 2007, pp. 6-10.
- [16] C. Goble and R. Stevens, "State of the nation in data integration for bioinformatics," J Biomed Inform, 2008.
- [17] W. Wang, G. Zeng, D. Zhang, Y. Huang, Y. Qiu, and X. Wang, "An Intelligent Ontology and Bayesian Network Based Semantic Mashup for Tourism," in 2008 IEEE Congress on Services - Part I: IEEE, 2008, pp. 128-135.
- [18] Y. Yuan and Q. Chenga, "Integrating Web-GIS and Hydrological Model: a Case Study with Google Maps and IHACRES in the Oak Ridges Moraine area, Southern Ontario, Canada," in IEEE International Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. Barcelona,: IEEE, 2007, pp. 4574-4577.
- [19] J. Wood, J. Dykes, A. Slingsby, and K. Clarke, "Interactive Visual Exploration of a Large Spatio-temporal Dataset: Reflections on a Geovisualization Mashup," in IEEE Transactions on Visualization and Computer Graphics, . vol. 3 Research Triangle Park, NC, USA: IEEE, 2007, pp. 1176-1183.
- [20] S. Navabpour, L. S. Ghorai, A. A. Malayeri, J. Chen, and J. Lu, "An Intelligent Traveling Service Based on SOA," in

- Congress on Services - Part I, 2008. SERVICES '08. IEEE, 2008, pp. 191 – 198
- [21] A. Ankolekar, M. Krotzsch, T. Tran, and D. Vrandeic, "The two cultures: Mashing up Web 2.0 and the Semantic Web," in 16th international conference on World Wide Web Banff, Alberta, Canada 2007, pp. 825 - 834.
- [22] L. Venkatachalam, K. K. Dhara, V. Krishnaswamy, and M. Vernick, "Communication enabled web applications The evolution of computer telephony integration," in Communication Systems Software and Middleware and Workshops, 2008. COMSWARE 2008. 3rd International Conference IEEE, 2008, pp. 275-278.
- [23] A. Alba, V. Bhagwan, J. Grace, D. Gruhl, K. Haas, M. Nagarajan, J. Pieper, C. Robson, and N. Sahoo, "Applications of Voting Theory to Information Mashups," in 2008 IEEE International Conference on Semantic Computing: IEEE, 2008, pp. 10-17.
- [24] H. Chen, N. Ikeuchi, and Q. Jin, "Implementation of Ubiquitous Personal Study Using Web 2.0 Mash-up and OSS Technologies," in 22nd International Conference on Advanced Information Networking and Applications - Workshops, 2008. AINAW 2008. Okinawa: IEEE, 2008, pp. 1573-1578.
- [25] M. B. Blake and M. F. Nowlan, "Predicting Service Mashup Candidates Using Enhanced Syntactical Message Management," in 2008 IEEE International Conference on Services Computing. vol. 1, 2008, pp. 229-236.
- [26] J. Zou and C. J. Pavlovski, "Towards Accountable Enterprise Mashup Services," in IEEE International Conference on e-Business Engineering, ICEBE 2007 Hong Kong, China, 2007, pp. 205 - 212.
- [27] E. Chang, T. Dillon, and F. Hussain, Trust and Reputation for Service-Oriented Environments: Technologies for building Business Intelligence and consumer confidence vol. 1: John Wiley & Sons, 2006.
- [28] H. Dong, F.K. Hussain, E. Chang, "A Transport Service Ontology-based Focused Crawler, Proceedings of the 4th International Conference on Semantics, Knowledge and Grid (SKG 2008), Beijing, China, December 3-5 2008, Accepted for publication.